The Journal of

Economic Perspectives

Winter 2017

The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents   *Volume 31 • Number 1 • Winter 2017*

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Is China Socialist?

Barry Naughton

I t has been 40 years since Deng Xiaoping broke dramatically with Maoist ideology and the Maoist variant of socialism. Since then, China has been transformed. Spectacular growth, powered by the expansion of markets, has made urban China into a mainly middle-class society and lifted hundreds of millions of rural residents out of poverty. Throughout these enormous changes, China has always officially claimed to be socialist. Since 1992, China has called itself a "socialist market economy." Does the "socialist" label make sense when applied to China today?

There is no generally accepted definition of "socialism," and there seems little point in arguing over whether a complex reality coincides with a simple and arbitrarily defined label. Instead, the strategy of this piece is to advance four general characteristics that are plausibly related to a broad range of conceptions of socialism: that is, we will be talking about descriptive characteristics of socialism, rather than "models." Whether or not the reader accepts that these features are related to a coherent ideal of "socialism," raising the question of socialism in this way can help to gain a fresh perspective on the current reality of the Chinese economy.

In this spirit, a plausibly socialist system would be judged on the following four criteria: capacity, intention, redistribution, and responsiveness. First, a socialist government controls a sufficient share of the economy's resources that it has the *capacity* to shape economic outcomes. One traditional definition of socialism includes "public ownership of the means of production," but "capacity" is here

■ *Barry Naughton is the Sokwanlok Chair of Chinese International Affairs, School of Global Policy & Strategy, University of California–San Diego, La Jolla, California. His email address is bnaughton@ucsd.edu.*

broadened to include the ability to control assets and income streams, through taxation and regulatory authority. Second, a socialist government has the *intention* of shaping the economy to get outcomes that are different from what a noninterventionist market would produce. Third, because a socialist government typically justifies itself as benefitting those citizens who are less well off, it is natural to look for evidence of whether such policies are succeeding in the outcomes involving growth, social security, and pro-poor *redistribution.* Fourth, a socialist government should have some mechanism through which the broader population can influence the government's economic and social policy, so that policy shows at least some partial *responsiveness* to the changing preferences of the population.

In thinking about China's economy through the prism of these four descriptive characteristics, we must be sensitive to how the definition of socialism in China has evolved and thoroughly changed. Forty years ago, in 1978, China was unquestionably a socialist economy of the familiar and well-studied "command economy" variant, even though it was more decentralized and more loosely planned than its Soviet progenitor. Twenty years ago—that is, by the late 1990s—China had completely discarded this type of socialism and was moving decisively to a market economy. At that time, the question "Is China Socialist?" seemed meaningless to most people. China had shrugged off its old model of socialism, and obviously was never going back. China had officially recognized that no economy that excluded the market could hope to deliver satisfactory outcomes. Moreover, powerful trends at this time were limiting the scope of what China's government could achieve. Government tax revenues relative to GDP had declined dramatically, substantially limiting government capacity. Social service provision had collapsed in most rural areas; inequality soared and a new wealthy class emerged; and de facto privatization enriched a group of people. At the time, it appeared that China's economic success had been achieved at the cost of discarding socialist values. In the mid-1990s, the important question seemed to be: Would China continue to be a kind of "Wild West Capitalism," in which almost anything might be for sale, or would it converge with the developed market economies, with improved regulation and rule of law?

China today is quite different both from the command economy of 40 years ago, and from the "Wild West Capitalism" of 20 years ago. The government in China has much more influence over the economy than in virtually any other middle-income or developed economy. State firms and state banks remain prominent. Government five-year plans command attention, both domestically and internationally. The Communist Party remains in power. What, then, is the relationship between China's government and the economy? And, thus, to what extent is it reasonable to think of China as a form of socialism?

Does China's Government Have the Resources to Shape the Economy?

In the 1980s and 1990s, as China's government surrendered much of its direct control over the economy, it gradually lost effective control over a sufficient share

Figure 1
Public Revenue in China
(% of GDP)
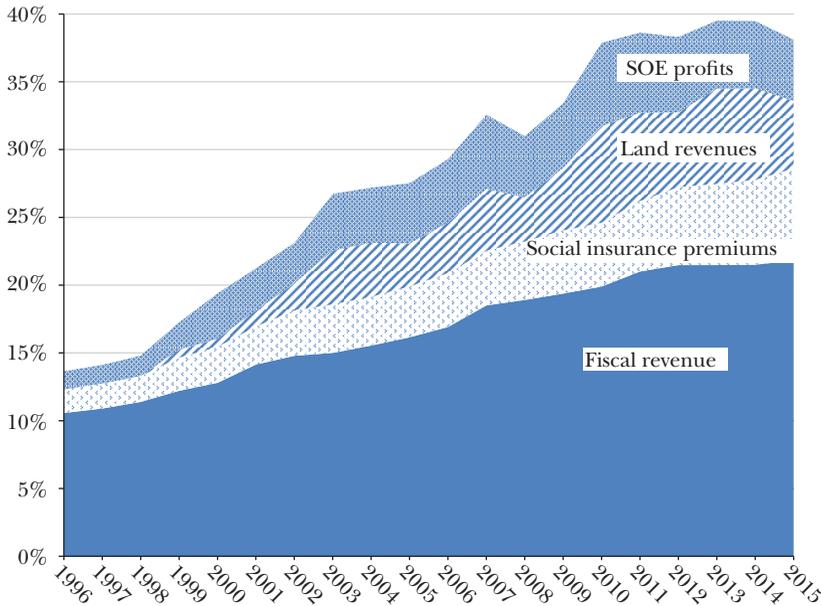
Source: Chinese government sources; for details, see Data Appendix available with this paper at http://e-jep.org.
Note: SOE is "state-owned enterprises."

of national income to carry out a socialist program. It faced, instead, a crisis of state capacity. However, in the last 20 years this situation has changed dramatically. Today, China is a predominantly market economy in which government is unusually large, powerful, and intrusive. Moreover, government ownership remains substantial and concentrated in strategic, large-scale, and capital-intensive sectors of the economy.

Control over National Income

China's government controls an unusually large proportion of national income flows, and these income streams have grown dramatically as a share of GDP since the mid-1990s. Figure 1 shows an expanded concept of government revenues that includes four main components: 1) budgetary revenues (not including social security); 2) social insurance premiums; 3) land revenues; and 4) net income from state-owned enterprises. This expanded concept provides an index of government control of resources, which tripled as a share of GDP between 1996 and 2013. This expanded concept is admittedly an *ad hoc* definition, but it gives a much better indication of the overall size of government than any of its components by itself. The

first two components can be placed in both international and temporal context, while the final two can be traced over time in China, but lack clear international comparisons.

Generally, fiscal revenues as a share of GDP increase as an economy becomes richer (Besley and Persson 2014). According to the dataset for 1975–2006 assembled by Baunsgaard and Keen (2010), low-income countries on average raise 14.5 percent of GDP in taxes; middle-income countries, 18.7 percent; and high-income countries, 32.5 percent. Thus, China was well *below* the low-income average in 1996 with 11 percent of GDP in fiscal revenues; and subsequently increased to well *above* the middle-income average at 21.8 percent of GDP in 2015. The figures of Baunsgaard and Keen (2010) do not distinguish clearly between ordinary taxes and social security contributions, but OECD data permit us to do so for a set of developed countries (OECD 2014). In 2012, 30 OECD countries collected on average 22.4 percent of GDP in taxes, plus 10 percent of GDP in social security contributions.[1] China's fiscal revenue of 21.8 percent in 2015 excluding social security is thus already equal to the OECD average (and also above the US level of 18.9 percent).

This pattern is surprising not only because of China's overall level of development, but also because China relies very little on personal income tax, the adoption of which is highly correlated with an increasing share of taxation in the economy (Besley and Persson 2014). Personal income tax was only 1.3 percent of GDP in China in 2015, and property taxes are also small. Instead, taxes on goods and services (especially the value-added tax at 4.6 percent of GDP) and enterprise income taxes (4 percent of GDP) are the largest categories, along with a dizzying array of transaction taxes (on land transactions and stock transactions, plus business and luxury taxes).

Social insurance contributions in China were 6.8 percent of GDP in 2015. These are not included in the Chinese figures for budgetary revenue (with some exceptions; budgetary contributions to the social security fund have been deducted from budgetary revenues to avoid double counting). While lower than the level for OECD countries, this is high for a middle-income country. Population aging is a key driver of the size of social security, and while China has a long-term aging problem, China *today* is still young, with an extremely low elderly dependency rate. Its social security revenues are consistently above social security outlays, so that its social security fund has an annual surplus (1 percent of GDP in 2015) and a relatively large and growing balance. Compared to OECD countries as a group, China's public revenues are relatively unencumbered, in the sense that they are not pre-committed to paying for existing services or social security outlays. Adding ordinary budgetary revenues and social insurance contributions together gives 28.6 percent of GDP, which is clearly higher than China's expected level as a middle-income economy.

The next two categories of government revenues are possibly unique to China. While not easily placed in a comparative context, their control clearly pushes the

---

[1]A Data Appendix available with this paper at http://e-jep.org offers additional discussion.

Chinese government further above its expected position, and they represent important dimensions of control of the economy's resources. All land in China is publicly owned: urban land is owned by the state, and housing is "privately owned" only in the sense that households hold long-term leases, typically of 50–70 years. Commercial developers pay local governments to lease development rights: those fees go into a separate account, not included in the formal budget, but under the control of local officials. State-owned enterprises also generate large profits. State-owned enterprises pay profit taxes and a changing—but relatively small—proportion of after-tax profits to their governmental owners (these are included in budgetary revenues). The majority of profits remain in the firm, where it is subject to government influence but not direct government allocation. The financial system in particular is overwhelmingly dominated by the state-owned banks, which generate substantial profits. These profits and, more importantly, the pattern of lending are subject to governmental influence. These funds together reached a peak in 2010, when land revenues accounted for 7.1 percent of GDP and after-tax profits of state-owned financial and nonfinancial enterprises were 6.1 percent of GDP. Since 2010, both have drifted down with the cooling of China's economy, and they totaled 9.5 percent of GDP in 2015.

Summing all four components, the Chinese government had direct or indirect control of 38 percent of GDP in 2015. Overall, the Chinese government is large, well-resourced, and potentially highly intrusive. This is both a distinctive feature of the Chinese economy in comparative context, and a dramatic change from the China of 20 years ago.

Assets: Does China's Government Own the Means of Production?

In 1978, virtually all of China's productive assets were owned by the state or by rural agricultural collectives. Since that time, a gradual process of liberalization, entry, and *de facto* privatization has fundamentally transformed the composition of ownership in goods-producing sectors, which are now predominantly private. Lardy (2014) describes this transformation as the dominant driver of China's economic performance. Crucially, the diversification and privatization of productive assets is a movement in the *opposite* direction from the increased government control of income streams described in the previous section. Since the 1990s, government has gotten bigger even as it has released its control over many productive assets.

Today's pattern of state ownership has been shaped by an evolutionary process of entry and competition, which is particularly clear in the industrial sector. The state share of industry declined as private and foreign-invested firms entered and grew rapidly. "Bottom-up" transformation was then complemented by a major downsizing of the state sector in the late 1990s, in which uncompetitive and loss-making state-owned enterprises were closed down or sold off. As a result, when small-scale firms and family artisans are included, workers in state-controlled firms make up only 12 percent of the total industrial labor force (according to China's 2013 Economic Census). In this sense, state ownership of industry in China is probably less than it

was in France or Italy before the 1980s wave of privatizations, and China's industry and agriculture have both become predominantly private and market-driven.

However, the evolutionary processes at work require several important qualifications to this generalization. First, the Chinese government has never overtly accepted privatization. The official position—written into China's Constitution—is that while multiple ownership systems coexist, state ownership is the "leading force." Second, barriers to entry—either legally created or driven by economies of scale and capital market inefficiencies—meant that substantial government ownership was retained in capital-intensive sectors: resources, utilities, and several heavy industrial sectors including steel, nonferrous metals, and transportation equipment. Government ownership is tiny (but not zero) in food products, textiles, and garments. Because state ownership is concentrated in capital-intensive sectors, when we look at the large-scale formal sector for which annual data are regularly available (in this case, data for 2014), state ownership of industrial assets (39 percent of the total) is substantial; much more so than the state share of revenues (24 percent), or workers (18 percent).

State ownership is actually somewhat more prevalent in the service sector than in industry. The government controls at least 85 percent of banking sector assets; the entire telecommunications and transport network; and essentially all education and scientific and technological services. In addition, the Communist Party owns and controls virtually all public media. Small-scale and labor-intensive services, such as retail and restaurants, are of course predominantly private, but government ownership and control is evident in the more capital-intensive and the more human-capital-intensive sectors.

As part of the marketization process, government relaxed its control over most productive assets. Today, the vast majority of Chinese workers are self-employed or work for firms. However, government has maintained substantial control of a number of upstream sectors and large intermediate good and machinery producers. Some sectors, notably oil and gas, are structured to generate monopoly rents. Moreover, government has maintained control over all land and almost all financial institutions, as described above. In an increasingly marketized environment, it is increasingly easy to "monetize" that control, generating the large revenue streams described in the previous section. Thus, while the Chinese government owns a relatively small share of overall productive assets, the assets that it owns often give it a monopoly position (land, natural resources, transport, and communication), or are strategically positioned upstream in the production system (intermediates and production equipment).

There have been several serious efforts to draw up a national balance sheet for China in recent years. Based on the extensive results of Li, Zhang, and Chang (2015; see also Ma, Zhang, and Li 2012), we can derive a highly simplified government balance sheet. Table 1 shows that government assets in China are about three times GDP, and net assets (after accounting for debt) are almost one-and-a-half times GDP. The four primary government asset holdings are land, assets in nonprofit "public service units," state-owned enterprises,

Table 1
Simplified Sovereign Wealth Balance, China 2013

	Trillion RMB	*Percent of GDP*		*Trillion RMB*	*Percent of GDP*
Government assets	**194.6**	**306%**	**Liabilities**	**104.8**	**165%**
of which:			of which:		
Land	62	97%	Central government debt	11.8	19%
Public service units	11.8	19%	Local government debt	8.6	14%
Nonfinancial SOEs	96.4	151%	Policy bank bonds	8.9	14%
Financial SOEs	20.3	32%	Local government funding vehicle debt	15.5	24%
			Debt or nonfinancial SOEs	51.6	81%
All other assets	4.1	6%			
			All other liabilities	8.4	13%
Memo: GDP	63.65		**Government net equity**	**89.8**	**141%**

Source: Li, Zhang, and Chang (2015, pp. 29).
Note: "SOEs" means state-owned enterprises.

and state-run banks and other financial enterprises. Public service units are government-owned nonprofit entities: they provide services for payment, and dominate the health care, environmental services, education, and cultural services sectors.

The value of government assets in China, relative to GDP, is much higher than in other countries. The US government has nonfinancial assets worth about 34 percent of GDP (12 percent of which is land; while 22 percent is made up of structures, equipment, and intellectual property assets) (Larson 2015; Board of Governors 2016). Government debt in China is substantial, although not as large as in developed countries. The right-hand column of the table shows government debt in four main categories summing to 71 percent of GDP; in addition, state-owned enterprises have debt equal to 81 percent of GDP. These numbers are not small, but they are much less than gross assets. Moreover, outstanding debt is owed mainly to the banks, which are themselves predominantly government owned. Thus, China's government maintains significant net wealth and control of a broad swath of society's assets.

The Chinese government clearly does not own the "means of production," and in fact by relaxing its restrictions on ownership, the Chinese government has been able to reap the benefits of competition and create a much more efficient and productive economy. However, the government still has a strong ownership position in the economy overall. Moreover, in a market economy, it is relatively easy for the state to transform its ownership of urban land, resources, and other assets into robust income streams. This has produced a qualitative change over the past 20 years. Twenty years ago, in the late 1990s, the Chinese government faced a crisis of state capacity, with budgetary resources below the low-income average. Since then, with the rationalization of the tax system and the ability to monetize its ownership

of land and other assets, the government has steadily increased the share of income under its control. Today, the Chinese government is wealthy and clearly has the capacity to intervene strongly in the economy.

Does the Chinese Government Self-Consciously Steer the Economy?

The Chinese government consistently emphasizes how it steers the economy because it sees a strong economic performance as a core part of its political legitimacy. Beyond the tools of funding direct government programs and ownership, the Chinese system has two distinctive mechanisms through which it attempts to foster development: 1) a set of bureaucratic incentives that reward officials for growth (of GDP and revenue); and 2) planning that is centered on the national-level Five-Year Plans but includes a broad array of sectoral, regional, and project plans that propose a certain trajectory for growth.

The Pro-Growth Incentive System for Bureaucrats

China's system of incentives for local bureaucrats to encourage growth is extremely unusual, and seems to exist only in China. It is a blunt but powerful instrument, encouraging growth and indirectly promoting investment and high-profile development projects. Ever since the success of township and village enterprises in the early 1980s, Chinese policymakers have seen value in linking the career incentives of local officials with the economic performance of that locality. Beginning in the 1980s, formal "target responsibility systems" established targets—or success indicators—for bureaucrats at all levels and gave them explicit weight in an evaluation function. These targets were predominantly economic, with GDP growth and increase in fiscal revenues the most important (Whiting 2001; Chan and Gao 2008). Good performance in meeting these targets brings cash awards and an increased chance of promotion. An extensive political economy literature has developed on the operation and impact of these incentives (among others, see Li and Zhou 2005; Shi, Adolph, and Liu 2012; Lü and Landry 2014). It needs to be stressed how unusual it is for such high-powered incentives to be introduced into government bureaucracies. Normally, high-powered incentives are avoided for publicly accountable officials, because incentivizing for one or two main targets draws effort away from other objectives (Acemoglu, Kremer, and Mian 2008; Naughton 2008). Such an incentive system can only make sense when a single objective—such as economic development—is seen as an overwhelming priority. This has been true of the Chinese leadership in the last few decades, and may reflect a broader social consensus.

The power of these formal incentives is enhanced by two other characteristics of the Chinese system. First, China's bureaucratic system is a vast pyramid, with norms that establish term limits and rotation of power at the top. All bureaucrats want to move up, but the number of slots for promotion is limited. Thus, incentives for performance become even stronger as officials approach ages associated with

"up or out" career intersections. Second, development incentives are quite compatible with the ordinary deal-making incentives of local officials. Land development and support for businesses can be ways to enrich cronies and relatives, and may also contribute to growth and investment (if the deal-making is handled with restraint). When attitudes toward corruption are relatively permissive, as has usually been the case, the personal and systemic incentives in favor of deal-making reinforce each other and drive local government behavior.

These targets are one way in which China has created a growth and investment machine. Knight (2014) argues that China is a "developmental state" because he finds this priority for developmental objectives built into the system at all levels, with officials given resources and instruments to carry out these goals. As one example, China since 2009 has been investing 48 percent of its GDP, a figure unmatched in large economies (ever), and some portion of China's extraordinary savings and investment effort is clearly attributable to the energy officials put into raising funds for local development projects. These incentives create a pro-growth environment that systematically encourages investment and growth and is diffused throughout the entire bureaucratic apparatus. They also affect and distort the allocation of resources; for example, urban land is one of the most important resources that a local government official controls. The official's discretion in the use of land almost always comes down in favor of industrial or commercial development. Su and Tao (2015, pp. 13–14) show that Chinese municipal governments allocate about three times as much land to industry proportionately as other countries, and the price of land for industrial uses is one-sixth that of land for residential use. Such policies clearly have distortionary consequences, not least of which is restricting the supply and pushing up the price of residential land.

Under extraordinary circumstances, government officials can coordinate control of many different kinds of assets. For example, within a few months of the global financial crisis in 2008–2009, China mobilized an investment effort equal to over 10 percent of GDP. Crucial aspects of this investment surge included the ability to mobilize local governments to initiate infrastructure projects; state-owned banks to loan without restraint to those projects; and state-owned enterprises to undertake the business and construction work. Policymakers issued consistent commands down through separate government and Party hierarchies, and the result was a remarkably large and prompt stimulus effect (Naughton 2009). Of course, such cases are very rare.

Planning

Forty years ago, China was a command economy, with central plans that purported to steer the economy but that actually didn't work. This type of planning was largely abandoned over the next 20 years. By the late 1990s, China essentially had no long-range plans or industrial policies. Zhu Rongji, the premier at the time, was extremely suspicious of grandiose and ineffective government programs. However, since the turn of the 21st century, long-range plans and industrial policies have made a strong comeback (Chen and Naughton 2016). China today operates with

scores of national plans, and literally hundreds of local government plans. At the center of this activity are the successive five-year plans, including the just completed 13th Five-Year Plan (2016–2020). These plans combine three elements:

1. A vision statement. Each of the recent five-year plans has had a theme and a set of overall motifs. These vision statements are consensus documents that communicate some of the hopes and objectives of policymakers, but they do not necessarily involve operational targets. They perform a coordination function, and allow local actors to get on the same page with central policymakers. In this sense, they resemble the "Indicative Planning" that the French pioneered in the 1950s. This function of the plan has changed relatively little over the last four five-year plans.

2. A handful of binding targets. Each of the last three plans has included a few environmental targets that, unlike other targets in the plan, are "compulsory." Beginning with a 20 percent energy conservation target in the 11th Five-Year Plan (2006–2010), these have expanded to include several targets for emissions reduction, carbon emission, and water and energy conservation. The 13th Five-Year Plan (2016–2020) also includes a binding target of bringing all 55 million of China's people below the poverty line out of poverty by 2020. The government then organizes a special data collection and enforcement program for these "compulsory" targets.

3. A panoply of associated sectoral and local plans. The national five-year plan is simply the beginning of a broader nationwide planning process. As soon as the overall plan is completed, local governments and ministries write their own plans that incorporate national objectives. Over the past 20 years, these follow-on plans have proliferated enormously. In the early 2000s, the State Council approved between zero and two industrial policies per year, whereas during 2016 at least 50 sectoral plans associated with the 13th Five-Year Plan have been given official status. There are plans for specific high-technology sectors (like semiconductors or electric vehicles) as well as portmanteau plans for technology development and "strategic emerging industries," as well as some more loosely integrated concept plans such as "Internet Plus," and one on industrial upgrading ("Made in China 2025"). The number and density of these associated plans is a recent and distinctive feature of China's system.

The different elements of China's planning process need to be considered separately, especially in a comparative context. The combination of a vision statement and a few concrete targets is a fairly standard, common-sense approach to planning. The Chinese plan incorporates and summarizes many existing programs, so it may not be particularly consistent, or announce much new (Kennedy and Johnson 2016). The addition of a vision statement gives it coherence and a more attractive message. The introduction of "binding" targets sends an additional strong measure to decentralized actors that they must take environmental objectives seriously (Heilmann and Melton 2013). But described in this way, China's planning process is not terribly different from how it is done in, for example, the state of California, where binding targets are established through legislation and regulation. In China, the target is enforced through administrative action, but decentralized actors are left free to choose specific steps.

The proliferation of associated sectoral and local plans, on the other hand, pushes decentralized actors to specific interventions that have no parallel in other market-oriented economies. The plans in China have teeth: while they rarely specify output targets, they specify desired outcomes very precisely and are intended to guide the allocation of resources. They influence local governments, state financing institutions, and other bodies. They convey government priorities and encourage local actors to align their investments with national targets, reducing risk for investments. The increase in government resources documented in the previous sector has been accompanied by a notable increase in targeted interventions. Financing guided into targeted sectors can be very large: for example, a high-priority sector like integrated circuits can expect access to over a hundred billion dollars in investment funding over the 13th Five-Year Plan.

These policies target multiple sectors, employ multiple instruments, and incentivize multiple actors. An additional layer of regional and land-use plans is layered onto the sectoral plans. This complexity means it is quite difficult to determine the net incidence of planning, much less its effectiveness. Resources pour into the highest priority sectors, which brings in more investors and creates unpredictable competition down the road. With so many plans and industrial policies, it is in some ways difficult to say whether China has any plan or a coherent industrial policy. As one example, the powerful sectoral incentives created may not align very well with the overall "vision" of planners. The current 13th Five-Year Plan envisions a transition to service-led development, but the sector-specific interventions channel scarce resources into high-tech industry, potentially diverting finances and high-skilled labor away from services.

Most of the current industrial policies are relatively new, and we rarely have good information about costs or a good measure of the benefits. What features of the Chinese economy today might *not* be in evidence if there had been no "planning"? So far, nobody has suggested a convincing example. Anecdotal evidence suggests very little correspondence between specific initiatives envisaged in successive plans and any of the subsequent outcomes. On the one hand, cases of apparently successful development—such as China's solar panel industry—have been achieved at the cost of excessive, duplicative, and failed investment. On the other hand, some successes that one would assume are due to planning turn out not to be. For example, the world's largest high-speed rail network is in China, but that network didn't appear in the five-year plans until 2016, by which time 19,000 kilometers had already been built.

We might conclude that there is little evidence so far that planning in itself is effective in shaping development; or that it is a cost-effective instrument. However, planning is part of a complex system of national institutions that overall have generally fostered development and growth.

Interactions Between the Two Types of Steerage

What is the relationship between the highly incentivized bureaucratic environment and the official plans, and how will the relationship change as China's economy

evolves? There are some indications that during China's high-speed growth era (approximately 1980–2010), the two were complements. The bureaucratic incentive system was a blunt instrument that fostered investment across the board, so the plans did not have to take on this function and indeed sometimes could lean against the wind. The five-year plans generally target "soft" GDP growth rates well below the economy's potential. For example, the 11th Five-Year Plan (2006–2010) called for more sustainable, balanced, environmentally friendly economic growth, and greater attention to rural development. It projected annual GDP growth at 7.5 percent, which was below China's potential growth rate. The "message," therefore, was that local governments should take the environment and long-term sustainability into consideration, and not merely push for the highest attainable GDP growth rate.

Was that message received? Apparently not, since China's GDP growth rate actually accelerated to over 10 percent in the three years beginning in 2005 and peaked at 14.2 percent in 2007. Moreover, in many respects, China's growth became less balanced and more high-investment driven over the plan period. Nevertheless, the plan may have played some positive role in moderating the worst aspects of an investment-driven growth model. In fact, the tension between bureaucratic incentives—which have consistently rewarded growth since the mid-1980s—and the more specific objectives of the planning process like environmental goals has been a long-term feature of the Chinese system that worked out reasonably well during the high growth era.

Since about 2010, China's "miracle growth" era has come to an end, due to rapid changes in labor force conditions, incomes, and external markets. In the wake of this change, the complementarity between bureaucratic incentives and the plan appears to be declining. The Xi Jinping administration has since 2012 reduced the focus put on economic growth in official success indicators, squeezing in new indicators such as managing local government debt, reducing poverty, and improving the environment (*New Capital Post* 2014; Wei 2016). At the same time, the most recent 13th Five-Year Plan (2016–2020) GDP target of "more than 6.5%" growth is close to, or perhaps even above, growth potential. The 13th Five-Year Plan seems to have become an instrument to maintain high-speed growth, even as the bureaucratic incentive system becomes less single-minded. It is a paradox of contemporary China that administrative interventions are increasing in size and multiplying in form just as growth potential is slowing and the benefits to a less-interventionist stance would seem to be increasingly evident.

Does China Intervene to Help the Less Well Off?

China's economy has grown faster for longer than any other economy in history. Growth of per capita GDP at the rate of 7–8 percent per year for almost 40 years has meant 2014 GDP per capita is 20 times what is was in 1978. There is no serious doubt that the benefits of this extraordinary growth have been broad-based. According to the widely used Human Development Index, which includes data on income, health, and education, China has improved from the "low human development"

Figure 2
Evolution of China's Gini Coefficient
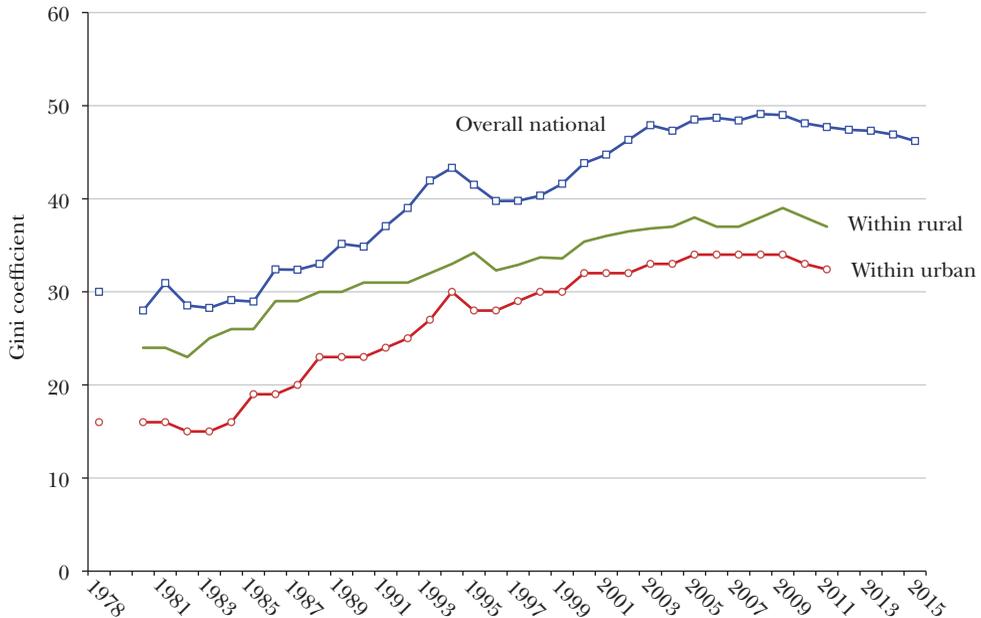
Source: Author's calculations based on official household survey data.

level of .42 in 1980 to the "high human development" level of .73 in 2014. This extraordinary improvement in overall living standards must be the starting point for any discussion of well-being and redistribution in China.

In contrast, China's efforts to improve distribution and increase the supply of public goods have met with only limited success. During the same period that incomes grew explosively, China's income inequality substantially worsened. From being one of the most equal large-population societies in the world in the early 1980s, China has become a relatively unequal society. It follows that improvement of living standards for those at the bottom of the distribution would have been even more rapid if distribution had not deteriorated. Figure 2 shows China's income Gini coefficient according to official household survey data. The Gini peaked at .49 in 2008–2009, and since then has drifted down slightly. The national Gini is much higher than both the within-urban and within-rural Gini because of the large gap between urban and rural incomes in China. Average urban incomes were less than twice rural incomes in 1983, at the minimum, and increased to 3.3 times rural incomes in 2009. The *hukou* household registration system divides all Chinese households into two categories, those with and without an urban household residence permit. Lack of an urban household residence permit continues to limit

access to income-earning opportunities and consigns rural residents to a second-class package of social benefits. As China's explosive growth increased incomes in the cities, rural incomes grew as well, but at a substantially slower rate though 2008–2009. In addition, entrepreneurship and diversification of income sources within urban and rural areas also increased overall inequality (Li and Sicular 2014; Zhou and Song 2016). The moderate decline in inequality since 2009 is almost certainly due to structural changes in China's labor markets. Since the mid-2000s, wages for unskilled and migrant workers have increased relatively rapidly, and more rapidly than those of skilled urban workers. Reflecting these changes, the urban/rural household income ratio has narrowed to less than 3.

Poverty Alleviation

China's record in poverty alleviation since the start of the reform period has been excellent. According to World Bank data, the reduction in the absolute number of those living in poverty in China between 1981 and 2010 accounted for 95 percent of the total reduction worldwide of those in poverty (using the traditional $1 per day—now $1.90 per day at 2011 purchasing power parities). Some of the World Bank's data are based on backward reconstruction that may not correctly capture changes in China's prices and rural distribution system in the early 1980s. However, the most conservative calculation, which would start with the years 1989–1992 when there were on average 260 million rural people below the poverty line, still finds a reduction of more than half to 114 million a decade later in 2001. By 2015, this total had again dropped by more than half, down to 56 million as of 2015 (Wei 2016), notwithstanding a Chinese domestic poverty line that is now slightly above the World Bank's standard $1.90 benchmark set in 2011. There is no question that China's overall record in poverty alleviation has been outstanding.

How much of this poverty reduction is due to government policy? China has had an active policy of targeting poor counties since 1986. About one-quarter of China's counties are designated "poor," which provides them with fiscal transfers for local economic development and infrastructure, food for work, and subsidized loans. The strategy is very much one of fostering local economic growth, especially nonagricultural production. As a method of poverty alleviation, the program has been criticized: geographic targeting misses poor people outside designated regions and benefits wealthier people within targeted regions; there is evidence that a portion of budgetary transfers inflate administrative expenses rather than helping poor people; and agricultural growth may be more fundamental to the success of these areas than industrial investments (Montalvo and Ravallion 2010). Clearly, agricultural and rural growth are the main drivers of poverty reduction in China, not government antipoverty programs. However, China's antipoverty approach is utterly consistent with the pro-growth bureaucratic incentive system. The antipoverty program at a minimum ensures that bureaucrats at the local level have some resources to feed into their competitive growth effort (Ang 2016). There is some evidence that counties near the poverty cutoff grew faster if they were designated "poor" but the effect is small (Meng 2013).

Redistribution in China

China is certainly not a welfare state. China has virtually nothing in common with countries like Cuba or Brazil—otherwise so different—which have made substantial efforts to transfer money directly to the poor and to invest in social services. In China, transfer payments of all kinds are low and are also restricted in important ways. However, even in this respect, there are important differences with the situation 20 years ago. In the mid-1990s, China was doing almost nothing to redistribute income. The social safety net in the countryside, tied to the agricultural collectives, had collapsed. The only social insurance system that remained intact was the one protecting urban workers in state firms, and that system was threatened by the dramatic downsizing of the state sector.

Since the 1990s, the Chinese government has sought to restore the urban social compact and spread a social safety net across the entire population. The existing urban system for those with permanent residence permits has been updated and adapted to a market-based economy, and it was not too difficult to update this existing, rather generous system. Systems for rural residents and rural-to-urban migrants, by contrast, had to be recreated from scratch. Over the past decade, the Chinese government has created new low-cost and low-benefit programs that cover the rural and migrant populations—for example, for medical insurance and pensions. The current extraordinarily complex system is—theoretically, at least—the beginning of a protracted transition toward an integrated national social security system. Today, there are four different medical insurance and pension systems, which between them cover virtually all Chinese citizens. However, the actual flow of resources through these combined systems is still modest. The government's direct contribution to social welfare benefits is still strikingly small. In 2014, budgetary outlays for education by China's government were 3.6 percent of GDP; for health, 1.6 percent of GDP; and for public housing, 0.8 percent of GDP. These are all strikingly low by comparative international standards.

In a similar spirit, China has universalized its system of direct income transfers. After rapid expansion, today about 20 million urban dwellers and 50 million rural residents receive direct minimum income payments. Actual direct relief payments to these 70 million recipients are set by local governments and vary substantially across regions. However, average payments are very low. Even a wealthy province like Zhejiang gives average payments of about US $50 per month, and in most places, the payments are far lower.

What are the proximate causes of China's relatively low redistributive effort? In practical terms, China has not established an institutional framework that makes broader redistribution feasible. As described earlier, personal income tax is still only 1.3 percent of GDP; moreover, income tax is levied only on wage income and some types of interest income, so almost all entrepreneurial and investment incomes escape coverage. As a result, the scope for redistribution through progressive taxation is small. According to the large household survey carried out by Li et al. (2014), the net redistributive impact of taxation on income distribution was essentially zero.

On a deeper level, the hesitation of Chinese policymakers to push for a greater degree of redistribution is driven by two structural factors. First, the deep institutional divide between rural and urban residents continues. Even though hundreds of millions of farmers have already left the countryside and settled in the cities, China has not found a way to extend the benefits of urban citizenship to migrants or rural residents. In theory, all employees, including rural migrant workers, should be covered by the social insurance system, to which both employers and employees should contribute. The proportion of migrant workers covered by social insurance increased until 2013, but only to 15.7 percent for pension plans and 17.6 percent for medical insurance, and the coverage ratio seems to have stagnated since (NBS 2014). Simply extending urban benefits to the rural and migrant population would be prohibitively expensive and administratively difficult, so China has instead settled for a provisional system that extends a much cheaper and less-comprehensive set of benefits to rural dwellers. As a result, the new systems replicate the old urban-rural divide. Urban dwellers with permits have inherited the benefits of the old command economy system. Social services for rural residents have improved significantly, but they remain far below those enjoyed by urban residents.

Second, although China's population today is young, it stands on the brink of an extraordinarily rapid aging process. The current urban system already faces funding challenges in some localities because retirement ages are low and benefits generous (pensions have been raised annually for the past 15 years). However, this generosity extends to only a relatively small proportion of the total population. Contribution rates of current workers are high, which raises labor costs overall, because policymakers worry that setting contributions lower will create funding shortfalls as population aging accelerates. Population aging is not just an urban phenomenon: rural out-migrants are disproportionately young, and the population they leave behind obviously has a higher share of elders. Policymakers struggle to find a solution that covers the rural left-behind aged and also integrates younger workers as contributors into a national system. These factors lead policymakers to take an extremely cautious attitude toward expansion of retirement and other social welfare provisions.

China has taken real and meaningful steps toward re-establishing at least a rudimentary welfare system that covers its entire population. The change from 20 years ago has been dramatic in terms of the creation of an institutional framework to cover the entire population. After a decade of effort, the level of benefits provided is still quite low and the safety net has some large holes. Moreover, the system has failed to tackle the fundamental inequalities of the old society, so that it redistributes income primarily among the comparatively well-off urban population. Overall, the limited scope of the effort is striking, especially when set against China's massive physical asset investment effort.

Public Goods

For the past 20 years, as government resources have grown, China has engaged in a massive physical infrastructure effort that has transformed China's built environment and dramatically expanded the stock of physical infrastructure. Any of a

number of quantitative indicators could be selected, but perhaps the most striking single fact is simply that China is now knit together with a complete grid of high-speed limited-access highways, and is well on the way to completion of a similar grid of high-speed rail. This rail system is not only far and away the largest high-speed rail network in the world, but it also seems poised to ultimately create a high-speed transportation network that goes beyond China's borders and creates a new reality of integrated transport across the Eurasian land mass. These regional infrastructure programs have benefits in promoting overall growth, but they are certainly far less cost-effective as means of redistributing income to the less well-off than direct social programs that Chinese policymakers have decided they cannot afford.

An alternative method of helping those who are less well-off is to provide public goods, like environmental protection, rather than direct income transfers. In sharp contrast to its aggressive and activist record in physical infrastructure, China's record with respect to the environment is fair to poor. Although strong conscious-ness of environmental issues spread in China since about 2000, it took a decade before serious measures began to reduce the emission of air and water pollutants and create conditions for an improved environment. While China has steadily stepped up its investment into environmental protection, it has done no more than follow a standard "environmental Kuznets curve"—the observed pattern in which environmental damage increases in the early stages of economic development but then is reduced in later stages (Dasgupta, Laplante, Wang, and Wheeler 2002). So far, China has experienced substantial environmental deterioration that shows clear signs of flattening out, but only limited indications of significant improvement.

How Responsive Is China's Government to Popular Demands?

Socialist systems, such as those in the former Soviet Union, have a long tradi-tion of being nondemocratic but claiming to speak for the interests of the masses or the working class. China certainly fits comfortably into this tradition. China is an autocratic government, with all political processes controlled by the Communist Party. One peculiar characteristic of China's Communist Party is how frankly elitist it is. In 2002, the Party redefined its claim to legitimacy and focused it on educated elites and knowledge workers, rather than its traditional working class constitu-ency. In the "Three Represents," proposed by Jiang Zemin and formally adopted by a Party Congress in that year, the Communist Party declared that it represented "advanced productive forces, the orientation of China's advanced culture and the fundamental interests of the overwhelming majority of the Chinese people." In this formulation, although the Party is charged with representing the interests of the "overwhelming majority," the Party itself is made up of "advanced" elites. The Party recruits heavily from elite schools and emphasizes that recruitment of Party members is highly selective. Among today's Party members there are three times are many managers and technicians as there are ordinary workers, and nearly all government employees (and teachers) are Party members.

The Chinese system has become partially institutionalized in a way that sets it apart from other Communist systems. The Communist Party has developed a system of term limits, leadership rotation, regular promotion, and credentialing for its officials. There are channels for the 83 million Party members to express their views. These mechanisms give far more weight to the views of urban dwellers than to rural residents. In policy formulation, the Party and government go to great lengths to solicit opinions broadly and gradually form and enforce a policy consensus (Chen 2011; Wang and Fan 2013). Moreover, the Party focuses single-mindedly on public opinion, primarily to shape it through media control, but also secondarily to understand policy preferences and to absorb and blunt criticism. In China, as in other East Asian societies after World War II, there has almost certainly been a profound social and political consensus in favor of growth. A major achievement of the Chinese system has been its ability to fulfill the deep popular desire for development.

Despite the quasi-institutionalization of the Chinese system, it remains wholly without external accountability. The enormous increase in the flow of revenues through the government since the late 1990s, the return to profitability of state-owned enterprises, and the enormous growth in the value of government assets clearly creates a dynamic that is different from the system in the 1990s. With power so concentrated, with little transparency and few checks or balances, it is inevitable that insider control will be pervasive and corruption a major problem. Indeed, in that sense, the major alternative label for China is "state capitalism" (Naughton and Tsai 2015). The term "state capitalism" was invented by Lenin to describe a state that pursues its own interests in the marketplace. Bremmer (2009, p. 41) calls state capitalism "a system in which the state functions as the leading economic actor and uses markets primarily for political gain." While this captures some of the Chinese reality, Bremmer (along with Lenin) assumes the state is a unitary actor, and for that reason underemphasizes the capture of rents and resources by individuals working under the state umbrella. Is the Chinese system responsive to a unified vision of development (albeit developed by autocrats), or is it responsive to a diverse collection of interest groups? For example, the land revenues and profits of state-owned enterprises described in an earlier section are not under the complete control of the state, but rather are partially controlled by the interest groups that grow up around them.

A number of the policy shortcomings described in the previous section are consistent with an interest group interpretation of the policy process: the failure to redistribute income from urban to rural households; preference for expensive physical capital investments over modest social expenditures; difficulties in shifting from an investment-driven to a consumption-driven economy; sluggishness in addressing issues of environmental deterioration; and others. One possibility is that the very revival in the fortunes of the government described earlier led the Chinese system to evolve a network of entrenched interest groups. Redistributive policies cannot be carried through without fundamental reforms of the fiscal, financial, and decision-making systems, which interest groups have so far been able to stall and deflect. In this view, the system ends up reflecting the interests of insiders: for example, of managers of state-owned enterprises. In a broader sense, the interests are those of a

larger group of Communist Party officials, politicians and technocrats, and even the urban population—at least those with urban residence permits—as a whole.

The response of President Xi Jinping to this state of affairs has been to attempt to make China "more socialist." Xi has of course pursued a high-profile anticorruption campaign since he came to power in 2012. The anticorruption campaign revealed that many state enterprises, particularly in the petroleum sector, had been largely captured by individual interests. This could be an argument for improved transparency, competition, and public oversight (and even privatization). Instead, Xi Jinping has shown himself inclined to a revival of socialist ideals and increased loyalty to the Communist Party. He has increased Party discipline and increased the rhetorical commitment to the "China dream" and "socialist core values" (which on the surface closely resemble universal human values, except that the Chinese Communist Party interprets all the practical details). All these actions can be interpreted as an acknowledgement that internal interest groups—rather than the public interest—have been driving specific economic outcomes. What is particularly striking is that the response has been to tack in an increasingly political direction, essentially gambling that socialism still has strong ideological appeal for ordinary Chinese.

Conclusion

Today, the question "Is China Socialist?" can reasonably be asked and left open. Sixty years ago, everyone knew that Mao was leading a socialist China; twenty years ago, everyone thought that political leaders like Jiang Zemin and Zhu Rongji were leading China away from socialism. China today clearly fulfills our first two criteria: the government has the capacity and intention to shape economic outcomes. On the third and fourth of our criteria—redistribution and responsiveness—China scores less highly than on the first two.

The objective of China's state intervention has clearly shifted from growth at any price to a more complex set of goals that includes redistribution and social and economic security. China under its current President (and General Secretary of the Communist Party) Xi Jinping is moving toward a more explicit embrace of socialism and a stronger commitment to socialist goals, as exemplified by the ambitious five-year-plan target of eliminating absolute poverty by 2020. Thus, it seems broadly fair to view China as moving towards a version of "socialism," albeit a very particular flavor of socialism that is authoritarian and top-down, but with a market economy based primarily on private ownership.

It is also possible to take a more skeptical view of China's trajectory. China's system of incentivized hierarchy—the authoritarian growth machine—was effective in mobilizing resources and maximizing growth during a "miracle growth" phase, when demographic, structural, and international factors all came together to raise growth rates. It also gave the Chinese government unprecedented control of resources and incentives, which it used predominantly to drive an enormous physical investment effort. The positive achievements are remarkable: the world's

best record of growth, tremendous success in alleviating poverty, and a national physical infrastructure built at unprecedented speed that is quickly approaching developed country standards. However, this "growth miracle" phase is now ending. Fundamental demographic changes, completion of many infrastructure programs, and a much-reduced distance to the global technological frontier are combining to lower China's potential growth rate in a dramatic manner. China has less *need* for growth-before-all-else, but this also means that the incentivization of the hierarchy, so fundamental to the past growth model, is no longer central to China's most important goals. The Chinese government has only belatedly begun to introduce a new set of instruments to achieve other objectives, and so far there is little evidence that China has developed a new way to steer the economy in a "socialist" direction while retaining some of the benefits of the developmental state.

It is in this context that we must view China's striking policy shortcomings: its weak institutional effort to redistribute income reflected in the emergence of a highly unequal society; the consistent underfunding of social expenditures; and a problem of environmental deterioration that is evident in air, water, soil, and contribution to global warming. China's policymakers have taken notice and increased their rhetorical commitment to addressing such issues, but the hoped-for amelioration of these conditions has barely begun.

In my opinion, China cannot be considered a socialist country until it makes much greater progress fulfilling its own declared policy objectives of universal social security, modest income redistribution, and amelioration of environmental problems. In turn, reaching these objectives will almost certainly require much more robust programs of economic reform. When the predominant objective of policy was economic growth, it was not particularly important to whom policy was responsive, since all groups shared an interest in growth. Today, as the government tries to redistribute and provide more public goods, policy must reflect the interests and more diverse preferences of a broader population. So far, China has not found a way to do this. China's relatively weak performance in achieving broadly redistributive policies, social fairness, and improved public goods provision appears to reflect the limits of responsiveness and the power of entrenched economic and political interests. Given socialism's authoritarian history, some would argue that responsiveness is not a necessary condition of socialism. However, if the lack of responsiveness prevents the system from devising and implementing efficient redistributive and public goods solutions, then a country cannot achieve the core objectives of a socialist system.

The Chinese leadership today, headed by Xi Jinping, has launched a broad campaign against corruption that implicitly acknowledges these problems. At the same time, Xi seems to favor an effort to make the system more socialist, by stressing collective goals and top-down direction. Thus, even those who judge that the Chinese system today is not socialist might consider that the socialist ideal is still influential, and the system may continue to evolve in the direction of stronger "socialist" and redistributive institutions. As that happens, the mix of attributes will change, and a "Chinese model" of socialism may begin to emerge.

References

Acemoglu, Daren, Michael Kremer, and Atif Mian. 2008. "Incentives in Markets, Firms and Governments." *Journal of Law, Economics and Organizations* 24(2): 273–306.

Ang, Yuen Yuen. 2016. *How China Escaped the Poverty Trap.* Ithaca, NY: Cornell University Press.

Baunsgaard, Thomas, and Michael Keen. 2010. "Tax Revenue and (or?) Trade Liberalization." *Journal of Public Economics* 94(9–10): 563–77.

Besley, Timothy, and Torsten Persson. 2014. "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives* 28(4): 99–120.

Board of Governors, Federal Reserve Board. 2016. Table L106 of *Financial Accounts of the United States: Second Quarter 2016*, p. 83 of the Federal Reserve Statistical Release, https://www.federalreserve.gov/releases/z1/current/z1.pdf.

Bremmer, Ian. 2009. "State Capitalism Comes of Age: The End of the Free Market." *Foreign Affairs,* May/June, pp. 40–55. https://www.foreignaffairs.com/articles/united-states/2009-05-01/state-capitalism-comes-age.

Chan, Hon S., and Jie Gao, ed. 2008. "Performance Measurement in Chinese Local Governments" [with translated documents]. *Chinese Law and Government* 41(2–3): 4–111.

Chen, Ling. 2011. *System, Elites and Consensus: In Search of a New Explanatory Framework for China's Policy Process* [in Chinese]. Beijing: Tsinghua University Press.

Chen, Ling, and Barry Naughton. 2016. "An Institutionalized Policy-making Mechanism: China's Return to Techno-Industrial Policy." *Research Policy* 45(10): 2138–52

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler. 2002. "Confronting the Environmental Kuznets Curve." *Journal of Economic Perspectives* 16(1): 147–68.

Heilmann, Sebastian, and Oliver Melton. 2013. "The Reinvention of Development Planning in China, 1993–2012." *Modern China* 39(6): 580–628

Kennedy, Scott, and Christopher K. Johnson. 2016. *Perfecting China, Inc.: The 13th Five-Year Plan.* Report of the CSIS Freeman Chair in China Studies, May. Washington, DC: Center for Strategic and International Studies.

Knight, John B. 2014. "China as a Developmental State." *World Economy* 37(10): 1335–47.

Lardy, Nicholas R. 2014. *Markets over Mao: The Rise of Private Business in China.* Washington, DC: Institute for International Economics.

Larson, William. 2015. "New Estimates of Value of Land of the United States." Bureau of Economic Analysis, April 3. http://www.bea.gov/papers/pdf/new-estimates-of-value-of-land-of-the-united-states-larson.pdf.

Li, Gan, Zhichao Yin, Nan Jia, Shoo Xu, Shang Ma, and Lu Zheng. 2014. *Data You Need to Know about China: Research Report of China Household Finance Survey 2012.* Berlin: Springer

Li, Hongbin, and Li-An Zhou. 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China." *Journal of Public Economics* 89(9): 1743–62.

Li, Shi, and Terry Sicular. 2014. "The Distribution of Household Income in China: Inequality, Poverty and Policies." *China Quarterly* 217: 1–41.

Li, Yang, Xiaojing Zhang, and Xin Chang. 2015. *Zhongguo Guojia Zichan Fuzhaibiao 2015—Gangan Tiaozheng yu Fengxian Guanli.* Beijing: Zhongguo Shehuikexue.

Lü, Xiaobuo, and Pierre Landry. 2014. "Show Me the Money: Interjurisdiction Political Competition and Fiscal Extraction in China." *American Political Science Review* 108(3): 706–22.

Ma, Jun, Xiaorong Zhang, and Zhiguo Li. 2012. *Zhongguo Guojia Zichan Fuzhibiao Yanjiu [A Study of China's National Balance Sheet].* Beijing: Shehuikexue Wenxian.

Meng, Lingsheng. 2013. "Evaluating China's Poverty Alleviation Program: A Regression Discontinuity Approach." *Journal of Public Economics* 101: 1–11.

Montalvo, Jose G., and Martin Ravallion. 2010. "The Pattern of Growth and Poverty Reduction in China." *Journal of Comparative Economics* 38(1): 2–16.

Naughton, Barry. 2008. "Market Economy, Hierarchy and Single Party Rule." In *Market and Socialism Reconsidered (with Particular Reference to China and Vietnam),* edited by Janos Kornai and Yingyi Qian, 135–61. London: Macmillan, for the International Economic Association.

Naughton, Barry. 2009. "Understanding the Chinese Stimulus Package." *China Leadership Monitor,* Issue 28. http://www.hoover.org/research/understanding-chinese-stimulus-package.

Naughton, Barry, and Kellee S. Tsai, eds. 2015. *State Capitalism, Institutional Adaptation, and the Chinese Miracle.* New York: Cambridge University Press.

NBS [National Bureau of Statistics]. "National Survey Report on Migrant Workers 2013" [in Chinese]. May 12, 2014. http://www.stats.gov.cn/tjsj/zxfb/201405/t20140512_551585.html.

New Capital Post. 2014. "State Council Announces Deepening of Fiscal Reform; Local Government

Debt to Be Included as Political Success Indicator" [In Chinese]. October 9. Accessed April 12, 2016, at http://finance.china.com.cn/money/bond/zqzx/20141009/2714779.shtml.

OECD. 2014. "Table 5: Tax Revenue of Main Headings as % of GDP, 2012." of *Revenue Statistics 2014.* http://www.oecd-ilibrary.org/taxation/revenue-statistics-2014_rev_stats-2014-en-fr.

Shih, Victor, Christopher Adolph, and Mingxing Liu. 2012. "Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China." *American Political Science Review* 106(1): 166–87.

Su, Fubing, and Ran Tao. 2015. "The China Model Withering? Institutional Roots of China's Local Developmentalism." *Urban Studies,* published online before print, July 7. DOI: 10.1177/0042098015593461.

Wang, Shaoguang, and Peng Fan. 2013. *The Chinese Model of Consensus Decision-Making: A Case Study of Healthcare Reform* [in Chinese]. Beijing: Renmin University Press.

Wei, Wei. 2016. "Rural Poverty Declined by 14.42 million in 2015" [in Chinese]. From State Council Poverty Alleviation and Development Leadership Group, February 29, 2016, accessed May 24, 2016 at http://www.cpad.gov.cn/art/2016/2/29/art_50_45702.html.

Whiting, Susan H. 2001. *Power and Wealth in Rural China: The Political Economy of Institutional Change.* New York: Cambridge University Press.

Zhou, Yixiao, and Ligang Song. 2016. "Income Inequality in China: Causes and Policy Responses." *China Economic Journal* 9(2): 186–208.

Human Capital and China's Future Growth

Hongbin Li, Prashant Loyalka, Scott Rozelle, and Binzhen Wu

China's real GDP per capita has increased at a rate of nearly 9 percent annually since the start of its economic reforms in 1978, the fastest rate of growth that any large country has sustained for such a long period of time. Output is equal to the number of workers multiplied by productivity per worker. Thus, China's dramatic growth can be broken down into the increase of the size of working-age labor force as a proportion of the population (or, in other words, the decrease of the dependency ratio) and improving labor productivity. In turn, labor productivity can be broken down into factors such as rising human capital and the reallocation of labor to more efficient sectors.

However, each of these determinants of output growth in China are undergoing substantial change. The size of China's working-age population peaked in 2014 and has now started to decline. Labor reallocation from rural to urban areas has decelerated; less than 10 percent of young, able-bodied working-age individuals

■ *Hongbin Li is the co-director of the China Enterprise Survey Center at the Institute for Quality Development Strategy, Wuhan University, Hubei Province, China, and Visiting Scholar at Stanford Center for International Development (SCID), Stanford Institute for Economic Policy Research (SIEPR), Stanford University, Stanford, California. Prashant Loyalka is Center Research Fellow at the Freeman Spogli Institute for International Studies and Assistant Professor (Research) at the Graduate School of Education, and Scott Rozelle is Helen F. Farnsworth Senior Fellow at the Freeman Spogli Institute for International Studies, both at Stanford University, Stanford, California. Binzhen Wu is Associate Professor of Economics, Tsinghua University, Beijing, China. Hongbin Li is the corresponding author at lihongbinsem@gmail.com.*

in the labor force are still living and working on farms (Li, Liu, Luo, Zhang, and Rozelle 2010; Li, Li, Wu, and Xiong 2012). The potential for reallocation of labor from state-owned enterprises to the private sector is limited because the majority of China's workers are already in the private sector, and many of the remaining state-owned enterprises actually behave much like private firms in terms of their labor management (Li, Shi, and Wang 2016).

In this paper, we will consider the sources and prospects for economic growth in China with a focus on human capital. We begin with an overview of the role that labor has played in China's economic success. We then turn to describing China's *hukou* policy, which divides China's labor force into two distinct segments, one composed of rural workers and the other of urban workers. Because of the *hukou* policy and its implications, any study of China's labor force needs to analyze rural and urban workers separately. For the rural labor force, we focus on the challenges of raising human capital by both increasing basic educational attainment rates as well as the quality of education. For the urban labor force, we focus on the issues of further expanding enrollment in college education as well as improving the quality of college education. In the next part of the paper, we use a regression model to show the typical relationship between human capital and output in economies around the world and demonstrate how that relationship has evolved since 1980. In the context of this relationship between human capital and output, we also show that China has made substantial strides both in the education level of its population and in the way that education is being rewarded in its labor markets (and is producing higher levels of output). However, as we look ahead, our results imply that China may find it impossible to maintain what appears to be its desired growth rate of 7 percent in the next 20 years. From our paper's focus on human capital, we believe that a growth rate of 3 percent over the next two decades seems more plausible. Finally, we conclude with some policy recommendations, rooted in the belief that China continues to have substantial room to improve the human capital of its labor force.

Labor and China's Economic Success Since 1978

China experienced a high fertility rate of about six births per woman in the 1960s and 1970s. By 2010, China's total fertility rate had dropped to 1.4 per woman. The reasons include higher wages and education for women, the economic reforms that started in 1978, as well as the family planning efforts enacted in the 1970s and the one-child policy instituted in 1979. Those earlier higher birthrates caused China's working-age population (age 15–64) to rise from 625 million in the early 1980s to one billion in 2014. The drop-off in birth rates starting in 1980 meant that China's working-age population, as a proportion of the total population, rose from 62 percent in the early 1980s to 75 percent in 2010. This *demographic dividend* has raised per capita GDP because a higher share of the nation's population has been in the labor force.

After the early 1980s, labor productivity in China also rose dramatically, a pattern reflecting both improvements in human capital and the reallocation of labor to more

Figure 1
The Share of Employment in the Private Sector (1978–2014)
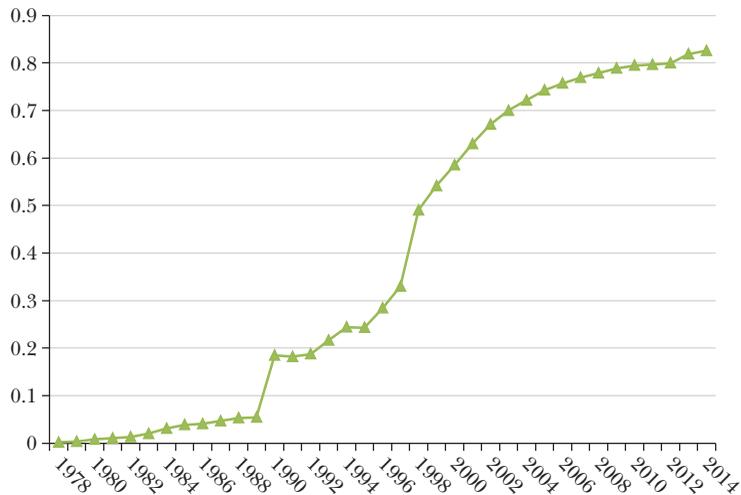
productive sectors of the economy. China started to restore education in 1976 at the end of the Cultural Revolution, a movement that had literally shut down colleges and high schools for most of its duration (1966–1976). Our calculation based on the Chinese Population Census in 1980 and 2010 shows that education has risen rapidly since then, with the average years of schooling for China's adult labor force (ages 25–64) rising from 4.3 years in 1980 to 9.6 years in 2015. The share of the labor force having at least senior-level high school education increased from 6.1 percent in 1980 to 28.8 percent in 2015. The proportion of China's labor force having a college education increased from only 1.1 percent in 1980 to 12.5 percent in 2015. The rapid rise of the level of schooling during the past 30 years was a result of both higher demand, driven in no small part due to rising returns to education (Li, Li, Wu, and Xiong 2012), and higher supply, driven by aggressive government policies that engineered a rising supply of slots in public schools (Li, Ma, Meng, Qiao, and Shi 2016).

Labor productivity also has risen due to movement among sectors of the economy. Similar to the development processes in almost all other countries that have experienced successful periods of rapid growth, over the past 30-plus years a large share of China's labor force has shifted sharply from agriculture to the more-productive industrial and service sectors. From 1978 to 2014, the share of workers in the agricultural sector fell from 71 to 30 percent. During this same period, the share of workers in industry rose from 18 to 30 percent and the share in services rose from 12 to 40 percent (according to various years of the *China Statistical Yearbook*).

Another shift is from employment by government in the state sector to employment by the private sector. Figure 1 shows the share of employment in the private sector from 1978 to 2014. We see that the share of employment in the private sector

was literally zero in 1978. During the pre-reform era, the central planning agency set wages. Workers were not allowed to freely move between firms or between cities/regions. As Figure 1 shows, the most dramatic reforms in terms of moving employment to the private sector happened in the mid-1990s, when millions of workers were laid off from state-owned enterprises. The state made it clear that they were not responsible for employment decisions and the workers needed to search for employment in the newly emerging labor markets. By 2014, the share of employment in the private sector had risen to over 83 percent. Today, wages in China are mostly set by market forces. Li, Shi, and Wang (2016) show that, after controlling for human capital attributes, China's state-owned enterprises pay the same wage as private firms.

Hukou and the Duality of China's Labor Force

One of the most important institutions for people in China over the past several decades, including those in the labor force, is the household registration system, termed the *hukou* system in Chinese. China's leaders created the *hukou* system in the 1950s as a way of managing labor and the movement of population when China was being managed as a planned economy. An individual's *hukou* is inherited from his/her parents, which not surprisingly is also the birthplace for most individuals. An individual's *hukou* has two dimensions: location (of residency) and sector of the economy (rural or urban). During the planned system (1950–1980), a person could only live and work in the location specified in the *hukou* card (county or city): rural people could only live in rural areas and work in agriculture; urban people could only live and work in cities. After the economic reforms were launched, the *hukou* system was relaxed in the early 1990s. People were allowed to move around and to work and live in non-*hukou* localities. The sector-based restriction also was relaxed. Rural individuals could get urban-based (industrial/service sector) jobs rather than being restricted to only farming.

Despite the relaxation of the *hukou* system along of some dimensions, nearly all administrative activities, such as land distribution, housing, the issuance of identity cards, school enrollment, medical insurance, and social security were—and still mostly are—based on an individual's *hukou* status. For example, only children with a *hukou* in a specific jurisdiction are entitled to go to a public school in that particular jurisdiction. This means, of course, that the quality of schooling depends on the local public finance (among other things) of the location of an individual's *hukou*. While there are calls for reform, we believe that the *hukou* system will remain fundamentally intact for at least 20 more years (the timespan for our prediction of education and income later). *Hukou* reform has been and almost certainly will continue to be difficult in China, given entrenched interests of those who have benefited from the system—largely the holders of urban *hukou* status.

The *hukou*-created urban–rural divide has created an economy characterized by dual labor markets, one rural, one urban. People with urban *hukou* live permanently in cities. By 2014, 36 percent of the Chinese population, or 497 million

people in China, had an urban *hukou*. Based on the 2010 population census, of this total 73 percent (or 364 million individuals) were in the working-age bracket of 15–64 years. The 2010 population census shows that the majority (61 percent) of the urban workers have jobs in the service sector; about one-fifth (20 percent) work in industry; and the remaining one-fifth (19 percent) work in agriculture (author's calculation based on Census 2010 and the Annual Monitoring Survey on Rural Workers, 2010, 2011). Almost all residents with an urban *hukou*, of working age or not, are eligible for social service benefits provided by their city's government, including relatively high levels of social security and health insurance, subsided housing, and access to well-funded public schools.

In contrast, 64 percent of individuals living in China (871 million individuals) have a rural *hukou*. Of those with a rural *hukou*, 76 percent (659 million individuals) are of working-age (aged 15–64). But today, after the launch of the economic reforms in the late 1970s and 1980s, not all workers with a rural *hukou* live and work in rural areas. Nearly one-third (31 percent) of workers with a rural *hukou* had migrated to cities to work in 2014. Almost all work in either industry (57 percent) or the service sector (43 percent) (Annual Monitoring Survey on Rural Workers 2014). About 25 percent of individuals with rural *hukou* live in rural areas but work in local industries or the service sector (Population Census 2010). The rest of those rural individuals that are of working age (about 44 percent) are either still in school or working, mostly in agriculture or in the household. As shown in Li, Liu, Luo, Zhang, and Rozelle (2010), of those working-age rural individuals that are working in farming (or are at home doing household work) and are not working off the farm (in either industry or the service sector) almost all are in the age cohorts that are 40 and older. Hence, given the age structure of those that are not in the industrial or service sector yet, there really isn't much scope for China to increase the quantity of labor into the nonfarm industrial/service economy through additional rural-to-urban migration.

The provision of social services for rural workers is a challenge, given the fact that this population is larger and poorer and the process of establishing such programs began later. Rural residents today in China are gradually gaining access to local government-provided social services, including the nine-year compulsory schooling, health insurance, welfare payments, and social security. However, in most rural jurisdictions, the levels of support provided by the local governments remain low.

Rural workers who have migrated to urban areas face challenges, too. Migrant workers are only rarely covered by the social security net or the urban insurance scheme that are enjoyed by urban residents. In 2014, only 16 percent of rural migrant workers working in cities were covered by social security, 18 percent had health insurance in the cities, and 10 percent had unemployment insurance (Annual Monitoring Survey on Rural Workers 2014).[1] Children of migrants are not guaranteed slots in

---

[1] Rural migrants can choose to participate in health insurance and social security programs for rural residents. The New Cooperation Medical Scheme for rural residents was launched in 2003 and covered 99 percent of the rural residents in 2014 (NHFP 2015). The rural social security program started in 2009, and covered about 50 percent of rural residents in 2011 (*Chinese Statistical Yearbook* 2012).

public schools in the urban areas, and as a result, migrant children are typically not found in a city's higher-quality schools. The denial of access to local public services, particularly the local education system, makes it difficult for rural migrants to live permanently and raise their families in cities. The absence of access to social services also helps to perpetuate dual labor markets and dual education systems.

Lagging Human Capital Development for Rural Residents

Despite significant increases in educational attainment since 1978, the overall education level of rural Chinese workers is still extremely low. When examining the stock of human capital in the labor force in 2015, only 11.3 percent of adult workers in the 25–64 age bracket from rural areas had attained at least high school education. In contrast, 44.1 percent of individuals from urban areas had high school education (based on the authors' calculations using data from the Census 2010 and the Annual Monitoring Survey on Rural Workers 2014). The flow of rural students through high school (and college) is also low. School-age children from rural areas will ultimately comprise most of China's future labor force (Khor et al. forthcoming). However, less than 40 percent of children from poor rural areas were attending high school between 2007 and 2012 (Khor et al. forthcoming; Shi et al. 2015).

The low rate of high school attendance can be attributed to high and rising costs, both direct out-of-pocket costs and opportunity costs. Primary and middle schools are free in China. However, whereas most developed countries, and many developing countries, do not charge tuition for high school, China charges a minimum of $160 (in US dollars) per year, which is 27 percent of rural per capita income (Liu et al. 2009). Indeed, academic high school tuition rates in China are among the highest in the world (Liu et al. 2009). College tuition fees also are burdensome for students from poor rural areas, and students from those areas often do not qualify for need-based financial aid (Li, Meng, Shi, and Wu 2013; Loyalka et al. 2013). Even more importantly, rural students face high and increasing opportunity costs from staying in school, even through junior high school. China's labor shortage has rapidly increased the wages of low-skilled workers (Li, Li, Wu, and Xiong 2012), which is a major reason for students dropping out from junior high school (Yi et al. 2012).

Competition is another reason why rural youth leave school at such young ages (for a full summary, see Shi et al. 2015). China's government sets quotas for the total number of students who can be admitted to high school and to college. High-stakes exams are used to determine grant access to each level of schooling. Only about 40 percent of all rural students pass the high school entrance exam and are thus allowed to enter academic high school (Song, Loyalka, and Wei 2013). Only 10 percent of rural students from poor rural areas pass the college entrance exam and enter college (Li, Loyolka, Rozelle, Wu, and Xie 2015).

The problems of youth in rural areas go beyond low rates of matriculation and poor performance on high-stakes exams. In competitive systems, even outside of China, dropout rates are known to be high (Battin-Pearson et al. 2000). In addition,

and in no small part connected to the high pressure in China's schools, high levels of anxiety and depression have been shown to plague youth in China (Wang et al. 2015). Mental health problems are associated with lower performance and dropping out (Shi et al. 2015).

An underlying problem is that children from rural areas are more likely to suffer from learning impairments, which makes advancing in China's school hierarchy even more difficult. Nearly 60 percent of China's elementary school children (ages 6 to 12) have at least one health or nutrition problem that can seriously affect early cognitive skill development (Zhou et al. 2015).[2] The situation is most severe for the approximately 22 percent of all children and youth (ages 0 to 17) in China (61 million total) who are left behind in rural areas by their migrating parents (Shi et al. 2015). These left-behind children may be hindered in developing noncognitive and social skills due to the absence of their parents (Zhou et al. 2015). Approximately 39 percent of the infants and toddlers (ages 0 to 3) that are born and raised in rural Chinese villages exhibit cognitive or psychomotor delay. The problem is mainly the absence of modern parenting/stimulation; only about one-third of caregivers read, talk, sing, or play with children in ways that have been shown to be important for cognitive development (Yue et al. 2016). Given the science showing that the first 1,000 days of life play an important role in determining an individual's life-long cognitive capabilities (Heckman 2006; Currie and Almond 2011), the rates of impairment that researchers have been observing mean that in the coming years, as these children reach adulthood, more than one-third of China's adult population—over 400 million people—will tend to have low levels of cognitive ability.

Besides the fact that the ability of students to learn is compromised due to health and nutrition (as discussed above), the quality of rural schools also is poor—especially relative to urban schools. Spending on facilities and programs is much lower on a per capita basis (Wang and Zhao 2014). Also, research has shown that it is difficult to keep high-quality teachers in rural schools (Wei 2016).

Perhaps the most serious school quality issue, however, is that for the case of migrant students. Because of the *hukou* system, the children of rural migrant workers cannot readily attend good-quality public schools in urban areas. As such, approximately 29 million rural children attend low-quality migrant schools in cities. In these private migrant schools, teachers systematically have lower levels of formal education and have less experience than those in rural public schools (Wang et al. 2016). Class sizes are larger. The commitment to quality education of administrators is generally lower. Research shows that migrant children that attend migrant schools perform even worse than children left-behind in the rural areas by their migrant parents (Wang et al. 2016).

---

[2]According to the World Food Program (WFP) data, other countries may have a similar problem (if not worse) than what exists in China. For example, the shares of preschool-aged children that have anemia in selected countries are: Brazil, 55 percent; Mexico, 25 percent; Thailand, 25 percent; and Turkey, 33 percent. At this age, according to the WFP data, only 20 percent of China's preschool-aged children are anemic.

A number of relatively inexpensive micro-interventions have proven useful in raising educational performance in studies done with rural populations in China. Many of these studies focus on improving the health and nutrition of elementary school students and enhancing early childhood education opportunities. Providing nutritious lunches for rural children (Luo et al. 2012), eyeglasses for the nearsighted (Ma et al. 2014), and deworming medicines (Liu et al. 2015) can improve student scores. Providing parenting, nutritional training, and other services to the caregivers of infants and toddlers could be done for almost no additional fiscal outlay if the responsibilities of the half-million or so family planning officers were changed from restricting the quantity of children (which really is not needed today in the aftermath of China's fertility decline and the movement from a one-child to a two-child policy) to running parenting classes and passing out nutrition supplements for children.

Why has China's government done so little to improve the education of rural children? Why has China, a nation that under Chairman Mao was known as a society that expressed concern about the needs of the rural population (Oi 1989), regressed to a point that rural residents now get access to such paltry government benefits? We believe that much of the neglect of rural education is due to the nature of the incentive system set up by China's growth-centered government as well as the decentralization of almost all basic social services. Studies in the literature provide evidence for these hypotheses.

First, the Communist Party has used GDP growth as a measure of local government performance and has been promoting officials governing jurisdictions that experienced high rates of growth (Li and Zhou 2005; Chen, Li, and Zhou 2005). This system provides strong short-term incentives to local leaders to invest in manufacturing and infrastructures, which can show up in the GDP number. Unfortunately, local officials often have few incentives to make long-term investments in human capital such as education and health because the terms of most government officials are less than five years. The incentive is further distorted by the externalities associated with human capital investments. In rural China today, a student who receives an academic high school education and continues on to college will almost certainly never return to their original county to live and work. While China as a whole benefits from higher levels of education, local leaders see little gain to their county from making such an investment.

Second, another reform launched in the 1980s decentralized China's fiscal system (discussed in Jin, Qian, and Weingast 2005). After the reform, China's tax system was heavily dependent on industrial taxes. Because of this, localities that built factories and those that became urbanized had access to a larger share of locally generated tax revenues to support local services, including education and health. It has been documented that, while some areas developed rapidly and built high-quality schools and other social services, most localities did not (Wong 1997; Huang, Rozelle, and Wang 2006).

It is in such a system that each of China's nearly 3,000 county-level governmental jurisdictions is in charge of delivering education and health to its own population. This responsibility is the same regardless of a jurisdiction's fiscal capacity, regardless of the

Figure 2
College Admissions in China
(in millions)
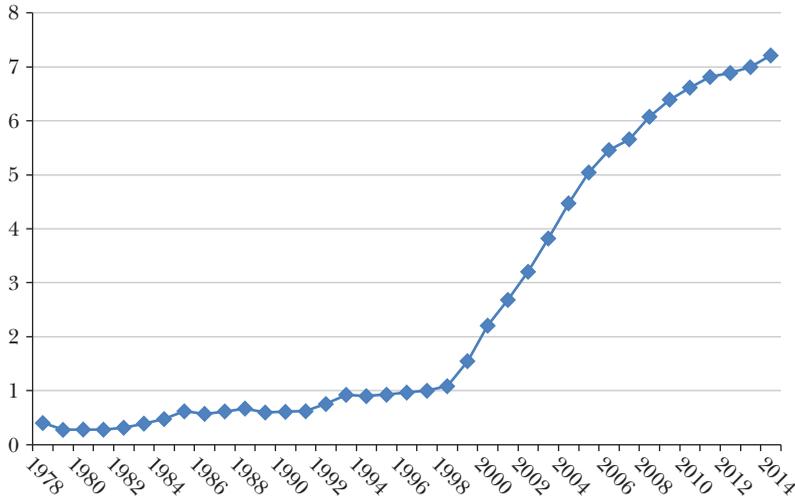
need, and regardless of the incentives (or absence of incentives) to provide quality services. Hence, the low and relatively lagging (compared to the pre-reform era) levels of investment into rural services can almost certainly be traced to the same set of fiscal and government reforms that were responsible for a substantial part of China's early growth.

Too Many Urban-Area College Students?

In contrast to the lagging human capital development for rural workers (and their children), China's urban labor force is relatively well-educated. In 2015, the average years of schooling of the urban labor force was 10.7 years. Of this group, 44 percent had at least a senior high school education and 21 percent had a college education (according to Census 2010 data). Importantly, rates of the urban labor force having college and high school levels of education compare favorably to the average levels of the G20 major economies (61 percent and 29 percent, respectively).

China's urban areas have seen a fast increase in the number of college students since 1999. Figure 2 shows the number of college admissions increased from 0.4 million in 1978 to 1 million in 1998, which implies a modest annual growth rate of 5.1 percent. Then in 1999 alone, China's central government increased the college admission quota by 43 percent. From 1999 to 2005, China's college admissions quadrupled. Between 1999 and 2009, the annual growth rate in college admissions was 18 percent. During this time, the proportion of high school graduates going to

college out of the whole nation increased from 43 percent in 1998 to 78 percent in 2009. During the same years, the gross enrollment rate in college of the 18–22 year-old age cohorts increased from 10 percent in 1998 to 24 percent in 2009. As the majority of college graduates will stay in the urban areas, the dramatic college expansion means a rapid improvement of human capital of the urban labor force, especially of the younger workers.

This rapid rate of expansion has created a perception that there is an over-supply of recent college graduates in China's labor market. Numerous media outlets have stoked this concern by reporting about the inability of recent college graduates to find gainful employment.[3] On the surface, the concern appears to be justified, as the premium to college education, when comparing college grad-uates with high-school graduates, has declined for young workers in recent years from about 21 percent in 2000 to 15 percent in 2009 (Li, Liang, and Wu 2016). In the months after college graduation, many college graduates are still searching for employment. For example, in 2013, 6.4 million college students graduated from college; in June (the time of graduation) about 30 percent of the graduates had not found jobs, and 10 percent still had not found jobs by December.[4] The starting sala-ries of many fresh college graduates are often reported to be on par with the wages of unskilled workers (Li, Li, Wu, and Xiong 2012; Li, Meng, Shi, and Wu 2012).

In contrast to this common perception of too many college students, we believe that college expansion is a great policy achievement of China. If we assume that the demand for human capital is fixed in the short-run, then given the unprecedented increase in the supply of college graduates since 1999, as documented in Figure 2, it is not surprising that the return to college for young college graduates would decline for a time. However, in the long run, human capital investment can lead to investment in physical capital and skill-biased technological changes (Acemoglu 1998; Hanushek and Woessmann 2015), which ultimately will increase the produc-tivity of and return to human capital. In addition, regions and cities in developed nations that experience arguably exogenous shocks to the supply of human capital ultimately also experience increases in the productivity of skilled labor due to human capital spillovers (Acemoglu and Autor 2011). There is no obvious reason to expect that China's case would be different in this respect.

Moreover, college expansion could well be a result of rising demand for human capital. Our analysis of data from China shows that the return to college education for the labor force as a whole has continued to rise despite the fast expansion of China's colleges (Li, Liang, and Wu 2016). In particular, the return for those with 5–20 years of work experience has risen from around 34 percent in 2000 to 41 percent in 2009. A possible reason is the rising demand for skilled workers driven by the influx of

---

[3]For some short essays on this subject, see the "Room for Debate" feature at the *New York Times* website, "Educated and Fearing the Future in China," for March 7, 2010 (at http://roomfordebate.blogs.nytimes.com/2010/03/07/educated-and-fearing-the-future-in-china/).
[4]Chinese speakers can see the article titled as "The Employment Rate of College Graduates in 2015" (http://www.cnrencai.com/qiuzhiguide/203276.html) [in Chinese].

foreign direct investment and expansion of trade starting from the early 1990s. The high return to college education for experienced workers implies a high lifetime return (the 10-year lifespan return to college education for the year 2000 graduate cohort is as high as 42 percent), which explains why urban students flood into colleges in spite of the seemingly low short-term return. Moreover, a substantial increase in college enrollment also appeared in other fast-growing economies such as Korea, Taiwan, and Thailand, which do not have a rigid central-planned quota system for college enrollment, suggesting that demand-side factors played a role as well. From this perspective, the theoretical and empirical findings of Li, Liang, and Wu (2016) support policies that seek to continue the expansion of college enrollments.

Of course, it is plausible that the quality of college graduates has declined, either because the average quality of college students has declined with the fast expansion or because it has been difficult for China's higher education system to maintain the same quality through the process of rapid expansion. Some troubling evidence suggests that China's college graduates are not learning much during their university years. Loyalka et al. (2016) assess nationally representative (random) samples of engineering and computer students in four-year undergraduate programs in China and compare them with similar students in the United States and Russia. Students were tested in both academic skills (using vertically scaled, major-specific math and physics tests) and higher-order thinking skills (using critical thinking and quantitative literacy tests). They find that although students in China score much higher on standardized tests at the start of their freshman year of college compared to their counterparts in the two other countries, they, on average, experience zero or negative growth in cognitive skills after two years of college. In contrast to students in China, students in Russia and the United States make cognitive skill gains over the first two years of college.[5]

Taking these various findings together, China needs both to expand the quantity of its college-educated labor force and improve the quality of college education. Despite the rapid expansion since 1999, only 12.5 percent of China's labor force attended college by 2015. As shown in Figure 3, this rate is lower than that of many other developing economies: for example, Mexico (16 percent), South Africa (15 percent), Philippines (27 percent), and Malaysia (16 percent). The gap in high school attainment rate is also large between China (28.8 percent in 2015) and these countries: for example, Mexico (36 percent); South Africa (42 percent); Philippines (58 percent); and Malaysia (51 percent). The rates of college and high school attendance/attainment in high-income OECD countries are, of course, even higher: Korea, 45 and 86 percent, respectively; Japan, 50 and 100 percent; Germany, 28 and 87 percent; and the United States, 45 and 90 percent.

---

[5] The math and physics tests were constructed by a team of educational researchers. Detailed information about the construction and validity of these two tests can be found in Kardanova et al. (2016). The critical thinking and quantitative literacy tests were created and validated by the Educational Testing Service (see Liu, Mao, Frankel, and Xu 2016; Roohr, Graf, and Liu 2014). The fact that students in the United States make substantive gains in the first two years of college is also collaborated by a handful of studies; for a comprehensive review, see Mayhew et al. (2016).

Figure 3
Education Attainments of the Labor Force in Selected Countries

A: Proportion of the Labor Force Having College Education in 2015
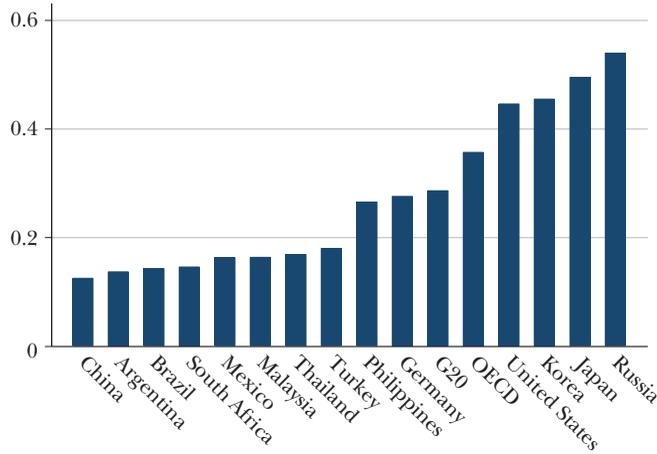
B: Proportion of the Labor Force Having High School Education in 2015
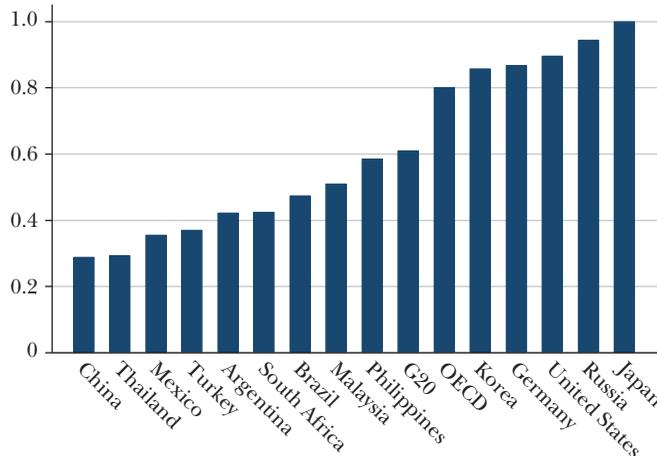
Source and Note: The numbers for China are from Population Census 1982, 1990, 2000, and 2010 and the 1 percent population sample survey in 1995 and 2005. The numbers for other countries are from "Education at a Glance" by OECD, 2016, and UNESCO Institute of Statistics (UIS). The numbers for Malaysia, Thailand, and the Philippines refer to the proportion among the population above age 25, coming from UIS. The numbers in 2015 are not available for the following countries, and we use their information available in the latest year prior to 2010: Argentina (2003), Brazil (2014), Indonesia (2013), Malaysia (2010), the Philippines (2013), Russia (2013), South Africa (2014), and Thailand (2013).

Human Capital and Income: A Regression Approach

To understand how economic development (or income per capita) is corre-
lated with China's human capital and labor market development, we compare
China's experience to the rest of the world. Specifically, we use an ordinary least
squares regression to examine the simple cross-country correlation between per
capita income and education in five specific years: 1980, 1990, 2000, 2010, and
2014. Our dependent variable is the log per capita GDP. The only explanatory vari-
able is the years of schooling for those that are aged 25 or above in a country (no
data are available on the education of the labor force aged 25–64 or 16–64 for
most non-OECD countries). In the analysis, we use the GDP levels of countries as
weights, which means that we weight the human capital–income relationship for
larger economies more than that of smaller ones. The data come from UNESCO
Institute of Statistics (UIS) and Education Statistics in the World Bank (EdStats).[6]
The numbers of countries for the five years we have chosen are: 112 in 1980; 124 in
1990; 144 in 2000; 155 in 2010; and 153 in 2014. To see China's position relative to
the rest of the world, we exclude China for all regressions, and then examine the
position of China relative to the regression line in each of the five years.

In Figure 4, we report the data points for the two variables, per capita income
and education, and overlay the fitted linear line that is based on estimated coef-
ficients reported in Table 1. The figure shows a clear positive correlation between
income and education level of the sample countries for all five years of data. As can
be seen on Figure 4, most countries are either on the line or close to the line. The
R-squared of these regressions are also high, especially for later years: for example,
0.74 for 2014. The regression coefficient on education and the constant term are
also quite stable since 1990. Looking at column 5 of Table 1, the coefficient of
0.259 means that a rise in the years of schooling by 0.1 years is associated with a
2.6 percent rise in income.

In interpreting this analysis, it is useful to spell out what this kind of regres-
sion can and cannot tell us. These simple correlations are not part of a growth
decomposition exercise. A growth decomposition would be built upon production
function analysis, and then use data on initial income, labor, education, and capital
investment (and perhaps other factors) to fit income growth. After estimating the
growth regression, analysts attribute any unexplained residual to changes in *tech-
nology*, where technology is to be understood broadly. A growth decomposition also
requires lagging the right-hand side variables, whereas we use concurrent years of
schooling in our regressions. Moreover, the regression presented here is, of course,
not a measure of the causal returns to education, because other factors are not
being held constant. Indeed, most countries with higher levels of human capital will

---

[6]When the relevant information is not available in a particular year, we use the information from
neighboring years. Specifically, we use 1978–1982 for 1980 (when the data for 1980 are not available);
1988–1992 for 1990; 1998–2002 for 2000; 2008–2012 for 2010; and 2010–2013 for 2014. The numbers for
China are derived from its population census (1982, 1990, 2000, and 2010).

Figure 4

Income per Capita and Education: Scatter Plots and Fitted Lines, 1980, 1990, 2000, 2010, and 2014

A: 1980

B: 1990

C: 2000

D: 2010

E: 2014
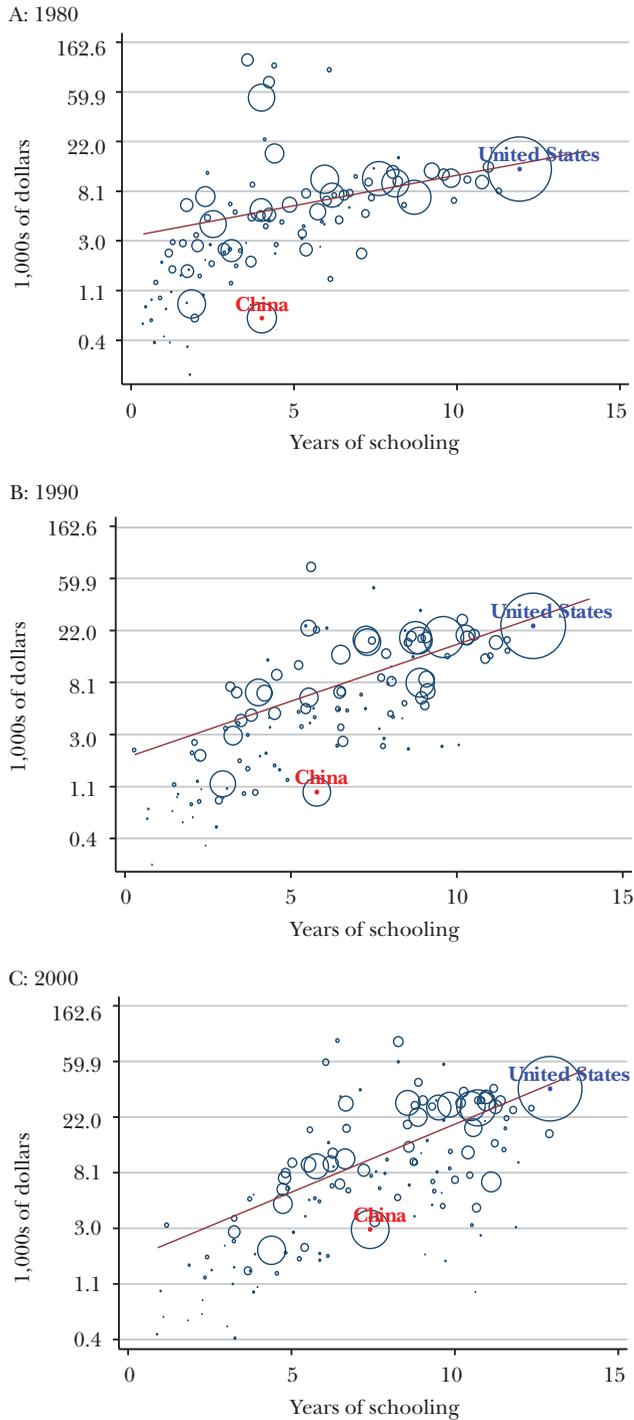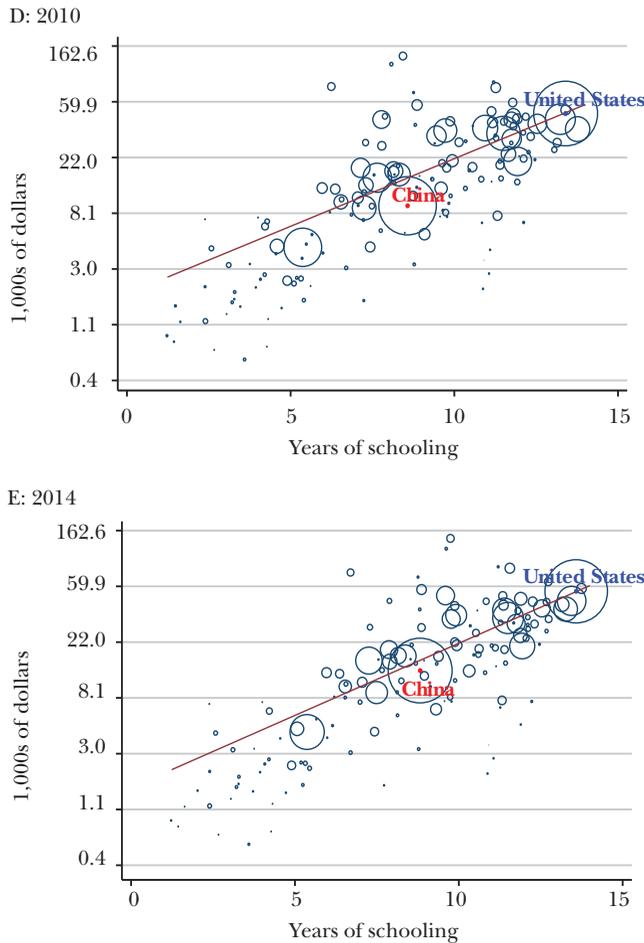
Source: The numbers for China are derived from Population Census (1982, 1990, 2000, and 2010). The data for other countries come from UNESCO Institute of Statistics (UIS) and Education Statistics in the World Bank (EdStats).
Notes: Education is measured by years of schooling of all adults above age 25. The fitted lines use GDP as the weight. We exclude China for the fitted lines. The vertical axis is labeled with actual dollar amounts (in thousands), but the regressions are fitted with log income. Income per capita is measured using the current (each of the five years) PPP dollars.

also tend to have higher levels of physical capital, more advanced technologies, and better market institutions.

Instead, the regressions in this paper's analysis are reduced-form correlations between years of schooling and per capita income for a set of countries globally. Given the years of schooling, the fitted line predicts the income level of a country, assuming that the country has the world average level of all other growth-augmenting factors such as physical capital, technology, institutions, the size of the labor force,

Table 1

Ordinary Least Squares Regressions Estimating the Effect of Education on Income for a Cross-Section of Countries

	Year				
	1980 *(1)*	*1990* *(2)*	*2000* *(3)*	*2010* *(4)*	*2014* *(5)*
Years of Schooling	0.123*** (0.0405)	0.219*** (0.0324)	0.245*** (0.0298)	0.243*** (0.0241)	0.259*** (0.0208)
Constant	8.093*** (0.401)	7.547*** (0.313)	7.425*** (0.324)	7.552*** (0.287)	7.390*** (0.255)
Observations	112	124	144	155	153
R^2	0.225	0.610	0.634	0.685	0.740

Notes: Robust standard errors in parentheses. The numbers for China are derived from Population Census (1982, 1990, 2000, and 2010). The data for other countries come from UNESCO Institute of Statistics (UIS) and Education Statistics in the World Bank (EdStats). The regressions use GDP as the weight. We exclude China for these regressions.
*,**, and ***** indicate significance at 10, 5, and 1 percent levels, respectively.

and the quality of education. The residuals of the regressions—as shown in the figure as the distance between the dots and the fitted line—reflect the gap in these growth-augmenting factors between this country and the world average.

Given that the regression line is more or less stable over time, one can think of income growth as having two components: one represented by the upward slope of the regression line in any given year (the estimated coefficient on education in Table 1), and the other by the distance a country is above or below the regression line in a given year (the residual from the regression). Moving along the regression line means that higher per capita income is positively associated with higher levels of human capital, which should be understood as being combined with the mixture of physical capital investment, technological aptitude, and improved institutions that on average accompanies higher human capital. Being below the regression line means that given its years of schooling, the country is performing below the global average for that level of human capital (and other inputs to production typically associated with that level of human capital). The possible reasons for the poor performance may include a number of factors, including the fact that the country has a relatively low share of its population in the labor force or a low quality of education. It also may be that the country is not taking advantage of its human capital: for example, perhaps the nation is failing to make corresponding investments in physical capital or technology similar to those made by other countries at that level of schooling. Another reason could be that the types of institutions or governance are limiting the ability of the market to reallocate labor to more efficient sectors.

China's position relative to the rest of the world in the five years shows the evolution of its labor market in these terms. As shown in Figure 4, China was far below the regression line in 1980, with per capita income only about 10 percent

of the level predicted by the global averages shown by the regression line, which suggests that China's human capital was not being fully utilized at that time. The inefficiency in the 1980s likely is due both to the fact that China's economy was being operated as a planned economy and that there was considerable labor redundancy in agriculture.

As the economic reforms unfolded during and after the 1980s, China steadily moved closer to the regression line. By 2014, after more than three decades of reform, China's income is about 81 percent of the level of the average global relationship. This implies that China's per capita income rose closer to the line by 6.3 percent per year from its 1980 base. Interestingly, this rate of movement represents about two-thirds of China's annual growth in national income growth of 9.7 percent a year between 1980 and 2014. Also impressively, China moved *along* the regression line by increasing its human capital level significantly in the years from 1980 to 2014. The years of schooling for adults aged 25 or above in China rose from 4.0 in 1980 to 8.8 in 2014, an annual increase of 2.3 percent in the sample period.[7] According to the regression line, this rise of human capital would typically be associated with a rise of 3.8 percent per year during the 34-year period, which is about 40 percent of China's actual growth rate.

The narrative of the regression analysis in this section is that China's gains in human capital, along with its gains in the accompanying inputs to production and its improvements in efficiency, all matter. Importantly, the gains from moving toward global averages (shown by China's increasing closeness to the regression line) seem to be larger than the gains from improvements in human capital and the other inputs to production. From 1980 to 2004, China's per capita income increased from $599 (in US dollars) to $13,206. About 40 percent of the gain in per capita income is what would typically be associated with the higher levels of education that occurred in China, using the coefficients from the regression line for 2014. About 65 percent of the gain in per capita income can be accounted for by China's movement up towards the regression line (of 1980).[8] In this framework, the shift of the regression line from 1980 to 2014 (that is, the change of the intercept and the differing coefficient on education) accounts for negative 5 percent of China's actual growth of per capita income.

[7]This rate is similar to the world average (4.8 years in 1980 to 10.8 years in 2014, increasing by 2.4 percent per year), and about 1 percentage slower than India's levels of 1.9 years in 1980 to 5.4 years in 2010 (no data available in 2014), increasing by 3.5 percent per year, or Brazil's levels of 2.5 years to 7.7 years, increasing by 3.4 percent per year.

[8]Of course, we have also considered a number of alternative specifications, but the same narrative emerges. For example, a specification that controls for the size of population in log form and the size of the labor force as a proportion of the population provides results similar to those in Table 1. We also experimented with other measures of the human capital stock, for example, using the education level of the labor force (age 25–64) as a measure of human capital stock, which is only available for a sample of OECD countries and China after 2005. These results are also similar. See Table A.1 in the online appendix for results from alternative specifications.

How Fast Will China's Economy Grow in the Next 20 Years?

Here we speculate on the development of China's human capital in the next few decades and the implications for China's rate of economic growth. In particular, we find that forecasts that China can maintain a 7 percent annual growth rate going forward are not plausible.

In the context of the framework laid out in the previous section, the forces in China for moving toward the line have changed. First, due to the fall of fertility since the early 1980s, the size of the 16–59 working-age population (this age bracket is China's official definition of its labor force) peaked in 2011 and has declined since then. China's demographic dividend, from a larger-than-usual share of its population in prime working age, has come to an end. Second, China's labor reallocation from rural to urban areas has decelerated. Most individuals in the rural labor force already are working off the farm, and most of them are already working in cities. The annual growth rate of rural-to-urban migration has declined from over 11 percent in the 15 years before 2000 to only 3 percent since then.[9] The growth of migration almost certainly will fall further given that rural-based surveys are finding that less than 10 percent of young able-bodied rural individuals of working age (ages 20 to 35) are now living (and working on farms) in rural areas (Li, Liu, Luo, Zhang, and Rozelle 2010; Li, Li, Wu, and Xiong 2012). Third, China's reallocation from state-owned firms to the private sector has slowed as well. Most of China's workers have already moved to the private sector. Moreover, a large share of the remaining state-owned firms behaves similarly to private firms in terms of labor usage (Li, Shi, and Wang 2016).[10]

Given this summary, in the rest of this section we carry out an illustrative exercise that offers predictions about China's economic growth in the next 20 years. First, consider the remaining potential for *towards-the-line* growth. It is true that there is still some scope for enhancing the quality of education. Recall, for example, the discussion above of how China still has some room for raising the quality of both rural pre-tertiary education and college education. If such investments can be made, they should help each measured unit of education be translated into higher income levels. But overall, the analysis here suggests the gains for China's economy from *toward-the-line* growth will be smaller in the future than in the past. Given the schooling level of nearly 9.0 years (for adults 25 or above) in 2015, the predicted per capita income would be $16,664 (converted to US dollars at the purchasing power parity exchange rate). If China could raise its income by 18 percent from the

[9]The number of migrants prior to 2009 comes mainly from the survey on rural labor that was conducted by China's Ministry of Human Resources and Social Security and the National Bureau of Statistics. The numbers post-2009 come from the annual Nationwide Migrant Worker Monitoring Survey Report conducted by the National Bureau of Statistics.

[10]Li, Shi, and Wang (2016) show that state-owned enterprises compete with private and foreign firms in the labor market, and they pay the same wage as other firms, controlling for the human capital of workers. Moreover, the education premium of workers within state-owned enterprises is similar to that in other types of firms.

2015 level of $14,117,[11] then China will hit the regression line. If China could do this before 2035, then this will add 0.8 percent to the average annual growth rate of income for the next 20 years.

What about the prospects for *moving-along-the-line* growth? If China can adopt an aggressive scenario for education expansion, we can assume that the number of individuals enrolled in college expands at 5 percent per year and that the enrollment rate for high-school for rural students increases at 11 percent a year (which would mean that the high school enrollment rate in rural areas would reach 100 percent by 2020). Under the aggressive scenario, we also assume that the high-school enrollment rate for urban students stays at the current level of 100 percent. In this best-case scenario, 26 percent of China's adults will have a college degree and 42 percent will have at least a high school education by 2035. Average years of schooling for adults would reach 10.7 years with an increase of 1.7 years in 20 years. As the fitted line in Panel E of Figure 4 (and Column 5 of Table 1) shows, every 0.1 years of increase in schooling is associated in these global patterns with a rise in per capita income of 2.6 percent. Of course, this positive correlation is based on China experiencing the ongoing improvements in physical capital investment, technology, and market-supporting institutions that are also positively associated with the projected growth in human capital. According to this scenario, China's *moving-along-the-line* growth would boost the nation's annual per capita GDP rate by 2.2 percent for the next 20 years.

Adding the toward-the-line and along-the-line growth numbers, we project an annual economic growth rate for China's per capita GDP of 3.0 percent in the next 20 years. At this growth rate, China's per capita income would reach the level $25,497 in 2035, similar to the level of Greece today.[12] This forecast is of course meant only to be illustrative of some basic calculations. But it makes the point that from the perspective of human capital development, and using what we view as optimistic assumptions about the expansion of education in China, it seems almost impossible for China to grow at an annual rate of 7 percent for the next 20 years, a rate that was in the government's economic plan.

For a different perspective on why China is unlikely to experience a 7 percent annual rate of growth moving forward, consider a comparison with the US economy. At 7 percent annual growth, China's per capita income would reach the level of $54,682 (in purchasing power exchange rate terms) by 2035, which is almost exactly the per capita income level of the US economy in 2014 ($54,629). In 2014, about 44 percent of the US labor force had at least a college education (and many more have attended college, although not graduated) and 89 percent of the labor force had at least a high school diploma. Even given the optimistic predictions above, China's education levels will be far below these US levels in 2035. Thus, the unlikely

[11] The income level in 2015 is USD 14,117, which is the income level in 2014 (USD 13,206) multiplied by 1.069 (the officially announced growth rate of 6.9 percent), and the income gap between the predicted level and China's current level is then $16,664/$14,117 = 1.18.
[12] If we use the years of schooling for the labor force (age 25–64) as the measure to predict, then the growth rate will be 4.0 percent and reach $30,932 in 2035, or the level of Spain today.

hope for 7 percent annual growth in China over the next 20 years would mean that China would need to have a relationship between human capital and per capita income that is considerably higher than the typical global experience would suggest is plausible.

Policy Suggestions and Conclusions

When China was a low-income country chasing middle-income status, the nation required a labor force that was numerate, literate, and highly disciplined. China delivered. It got almost all children into primary and middle school, taught them the basic math and language skills, and trained them to be disciplined and productive citizens. Such an education was adequate for building a labor force that could supply large volumes of low-wage, low-skill labor for the types of manufacturing that drove China's growth in the 1980s, 1990s, and early 2000s. However, China's labor force does not yet have the skills or level of human capital that is needed in a high-wage, high-skill, and innovation-based economy. What could the Chinese government do today to further raise its human capital up to the standard of high-income countries in the next two to three decades?

First, we propose that China's government should decentralize college education. Rather than setting an enrollment quota for each college, the government should allow at least some colleges to choose their own levels of enrollments. In the current system, college education is centrally controlled by the Ministry of Education and its local branches, and the Chinese government almost fully funds tertiary education. As a result, college officials and (even) professors have an incentive to care more about what they are asked to do by government officials than what is being demanded by the market. Once enrollment quotas are filled and a college is funded, the incentives for college administrators and faculty to provide high quality are weak. If college management and funding were decentralized, we believe that this would make China's colleges more market-oriented and improve the quality of education.

One recent example of China's intention to loosen its central control over higher education is the founding of a new college in Shenzhen in 2011, Southern University of Science and Technology of China (SUSTC), which is widely considered as "China's first independent university." Unlike other Chinese universities, which admit students only based on the scores of the centralized college entrance exams, SUSTC relies on multiple admission criteria (similar to the approach used by most US universities), including face-to-face interviews, high school grades, and abilities shown by other tests as well as the entrance exams scores. SUSTC also has an independently designed curriculum that emphasizes general education, employs 90 percent of the faculty from overseas, and offers many courses that are taught in English. Although the university is financed by the Shenzhen government, it has considerable freedom in making its own management decisions.

Second, for rural education, we propose the opposite policy prescription: specifically, we believe the central government should centralize funding of pre-tertiary,

twelve-year rural education. Centralization can address the budgetary problem of low-income counties. Perhaps more important, it could also overcome the incentive problems faced by local government officials. For China, centralized funding can also help to overcome the *hukou* hurdle and provide better-quality education for migrant children. City governments could be incentivized (or required) by the central government to provide education for migrant children whose parents are working in the city.

We believe that China can afford free high school education for all rural children. The tuition cost of an average rural high school student is about $300 in US dollars. Thus, the total cost of providing a high-school education for 24 million rural students (the size of three birth cohorts) would be $7.2 billion a year. This amount is less than 2 percent of the total outlay for education in China in 2015 ($388 billion). For comparison, China invested $141 billion on infrastructure in 2014. It is time for China's government to place less emphasis on heavy investment in physical capital for public purposes, where marginal additions have a relatively low return, and place greater emphasis on investment in human capital, which we believe will be a more important driver of China's future growth.

References

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics* 113(4): 1055–89.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics* vol. 4, Part B, edited by David Card and Orley Ashenfelter, 1043–1171. Elsevier.

Battin-Pearson, Sara, Michael D. Newcomb, Robert D. Abbott, Karl G. Hill, Richard F. Catalano, and J. David Hawkins. 2000. "Predictors of Early High School Dropout: A Test of Five Theories." *Journal of Educational Psychology* 92(3): 568–82.

Chen, Ye, Hongbin Li, and Li-An Zhou. 2005. "Relative Performance Evaluation and the Turnover of Provincial Leaders in China." *Economics Letters* 88(3): 421–25.

Currie, Janet, and Douglas Almond. 2011. "Human Capital Development before Age Five." In *Handbook of Labor Economics* vol. 4, Part B, edited by David Card and Orley Ashenfelter, 1315–1486. Elsevier.

Hanushek, Eric A., and Ludger Woessmann. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth.* Cambridge, MA: MIT Press.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312(5782): 1900–1902.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24(3): 411–82.

Huang, Jikun, Scott Rozelle, and Honglin Wang. 2006. "Fostering or Striping Rural China:

Modernizing Agriculture and Rural to Urban Capital Flows." *Developing Economies* XLIV-1(March): 1–26.

Jin, Hehui, Yingyi Qian, and Barry R. Weingast. 2005. "Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style." *Journal of Public Economics* 89(9–10): 1719–42.

Kardanova, Elena, Prashant Loyalka, Igor Chirikov, Lydia Liu, Guirong Li, Huan Wang, Ekaterina Enchikova, Henry Shi, and Natalie Johnson. 2016. "Developing Instruments to Assess and Compare the Quality of Engineering Education: The Case of China and Russia." *Assessment & Evaluation in Higher Education* 41(5): 770–86.

Khor, Nina, Lihua Pang, Chengfang Liu, Linxiu Zhang, Fang Chang, Di Mo, Prashant Loyalka, and Scott Rozelle. Forthcoming. "China's Looming Human Capital Crisis: Upper Secondary Educational Attainment Rates and the Middle Income Trap." *China Quarterly*.

Li, Hongbin, Lei Li, Binzhen Wu, and Yanyan Xiong. 2012. "The End of Cheap Chinese Labor." *Journal of Economic Perspectives* 26(4): 57–74.

Li, Hongbin, James Liang, and Binzhen Wu. 2016. "Labor Market Experience and Returns to Education in Fast Growing Economies." Unpublished paper, Tsinghua University.

Li, Hongbin, Prashant Loyalka, Scott Rozelle, Binzhen Wu, and Jieyu Xie. 2015. "Unequal Access to College in China: How Far Have Poor, Rural Students Been Left Behind?" *China Quarterly* 221: 185–207.

Li, Hongbin, Yueyuan Ma, Lingsheng Meng, Xue Qiao, and Xinzheng Shi. 2016. "Skill Complementarities and Returns to Higher Education: Evidence from the College Enrollment Expansion in China." Unpublished paper, Tsinghua University.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu. 2012. "Does Attending Elite Colleges Pay in China." *Journal of Comparative Economics* 40(1): 78–88.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu. 2013. "Poverty in China's Colleges and the Targeting of Financial Aid." *China Quarterly* 216: 970–992.

Li, Hongbin, Xinzheng Shi, and Rui Wang. 2016. "Illusory SOE Wage Premium: Evidence from Matched Data in China." Unpublished paper, Tsinghua University.

Li, Hongbin, and Li-An Zhou. 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China." *Journal of Public Economics* 89(9–10): 1743–62.

Li, Xiaofei, Chengfang Liu, Renfu Luo, Linxiu Zhang, and Scott Rozelle. 2010. "The Challenges Facing Young Workers during Rural Labor Transition." *China Agricultural Economic Review* 2(2): 185–99.

Liu, Chengfang, Renfu Luo, Hongmei Yi, Linxiu Zhang, Shaoping Li, Yunli Bai, Alexis Medina, Scott Rozelle, Scott Smith, Guofei Wang, and Jujun Wang. 2015. "Soil-Transmitted Helminths in Southwestern China: A Cross-Sectional Study of Links to Cognitive Ability, Nutrition, and School Performance among Children." *PLOS Neglected Tropical Diseases* 9(6): e0003877.

Liu, Chengfang, Linxiu Zhang, Renfu Luo, Scott Rozelle, Brian Sharbono, and Yaojiang Shi. 2009. "Development Challenges, Tuition Barriers, and High School Education in China." *Asia Pacific Journal of Education* 29(4): 503–520.

Liu, Ou Lydia, Liyang Mao, Lois Frankel, and Jun Xu. 2016. "Assessing Critical Thinking in Higher Education: The HEIghten™ Approach and Preliminary Validity Evidence." *Assessment & Evaluation in Higher Education* 41(5): 677–94.

Loyalka, Prashant, Elena Kardonova, Lydia Liu, Guirong Li, Huan Wang, Natalie Johnson, and Henry Shi. 2016. "Assessing Skill Levels and Gains in Engineering Programs across the US, China, and Russia." Unpublished paper.

Loyalka, Prashant, Yingquan Song, Jianguo Wei, Weiping Zhong, and Scott Rozelle. 2013. "Information, College Decisions and Financial Aid: Evidence from a Cluster-Randomized Controlled Trial in China." *Economics of Education Review* 36: 26–40.

Luo, Renfu, Yaojiang Shi, Linxiu Zhang, Chengfang Liu, Scott Rozelle, Brian Sharbono, Ai Yue, Qiran Zhao, and Reynaldo Martorell. 2012. "Nutrition and Educational Performance in Rural China's Elementary Schools: Results of a Randomized Control Trial in Shaanxi Province." *Economic Development and Cultural Change* 60(4): 735–72.

Ma, Xiaochen, Zhongqiang Zhou, Hongmei Yi, Xiaopeng Pang, Yaojiang Shi, Qianyun Chen, Mirjam E. Meltzer, Saskia le Cessie, Mingguang He, Scott Rozelle, Yizhi Liu, and Nathan Congdon. 2014. "Effect of Providing Free Glasses on Children's Educational Outcomes in China: Cluster Randomized Controlled Trial." *BMJ* 349: g5740.

Mayhew, Matthew J., Alyssa N. Rockenbach, Nicholas A. Bowman, Tricia A. D. Seifert, and Gregory C. Wolniak. 2016. *How College Affects Students*, Vol. 3: *21st Century Evidence that Higher Education Works*. With Ernest T. Pascarella and Patrick T. Terenzini. Kindle Edition. Wiley.

NHFP (National Health and Family Planning Commission of the People's Republic of China). 2015. "Statistical Bulletin on the Development of Health and Family Planning in China in 2015" [in Chinese]. Available at http://www.moh.gov.cn/guihuaxxs/s10742/201511/191ab1d8c5f240e8b2f5c81524e80f19.shtml.

Oi, Jean C. 1989. *State and Peasant in*

Contemporary China: The Political Economy of Village Government. University of California Press.

Roohr, Katrina Crotts, Edith Aurora Graf, and Ou Lydia Liu. 2014. "Assessing Quantitative Literacy in Higher Education: An Overview of Existing Research and Assessments with Recommendations for Next-Generation Assessment." *ETS Research Report* 2: 1–26.

Rozelle, Scott, Jikun Huang, and Xiaobing Wang. 2014. "The Food Security Roots of the Middle Income Trap." Chap. 3 in *The Evolving Sphere of Food Security,* edited by Rosamond L. Naylor. Oxford University Press.

Shi, Yaojiang, Linxiu Zhang, Yue Ma, Hongmei Yi, Chengfang Liu, Natalie Johnson, James Chu, Prashant Loyalka, and Scott Rozelle. 2015. "Dropout in Rural China's Secondary Schools: A Mixed-Methods Study." *China Quarterly* 224: 1048–69.

Song, Yingquan, Prashant Loyalka, and Jianguo Wei. 2013. "Determinants of Tracking Intentions, and Actual Education Choices among Junior High School Students in Rural China: Evidence from Forty-one Impoverished Western Counties." *Chinese Education & Society* 46(4): 30–42.

Wang, Huan, Chu Yang, Fei He, Yaojiang Shi, Qinghe Qu, Scott Rozelle, and James Chu. 2015. "Mental Health and Dropout Behavior: A Cross-sectional Study of Junior High Students in Northwest Rural China." *International Journal of Educational Development* 41: 1–12.

Wang, Wen, and Zhirong Zhao. 2014. "Spatial Decomposition of Funding Inequality in China's Basic Education: A Four-Level Theil Index Analysis." *Public Finance and Management* 14(4): 416–36.

Wang, Xiaobing, Renfu Luo, Linxiu Zhang, and Scott Rozelle. 2016. "The Education Gap of China's Migrant Children and Rural Counterparts." Working paper, Rural Education Action Program, Freeman Spogli Institute, Stanford University.

Wei, Yi. 2016. "Teacher Mobility in Rural China: Evidence from Northwest China." Doctoral dissertation, Michigan State University.

Wong, Christine P. W. 1997. *Financing Local Government in the People's Republic of China.* Oxford University Press.

Yi, Hongmei, Linxiu Zhang, Renfu Luo, Yaojiang Shi, Di Mo, Xinxin Chen, Carl Brinton, and Scott Rozelle. 2012. "Dropping Out: Why Are Students Leaving Junior High in China's Poor Rural Areas?" *International Journal of Educational Development* 32(4): 555–63.

Yue, Ai, Lauren Marsh, Huan Zhou, Alexis Medina, Renfu Luo, Yaojiang Shi, Linxiu Zhang, Kaleigh Kenny, and Scott Rozelle. 2016. "Nutritional Deficiencies, the Absence of Information and Caregiver Failures: A Qualitative Analysis of Infant Feeding Practices in Rural China." *PLoS ONE* 11(4): e0153385.

Zhou, Chengchao, Sean Sylvia, Linxiu Zhang, Renfu Luo, Hongmei Yi, Chengfang Liu, Yaojiang Shi, Prashant Loyalka, James Chu, Alexis Medina, and Scott Rozelle. 2015. "China's Left-Behind Children: Impact of Parental Migration on Health, Nutrition, and Educational Outcomes." *Health Affairs* 34(11): 1964–71.

From "Made in China" to "Innovated in China": Necessity, Prospect, and Challenges

Shang-Jin Wei, Zhuan Xie, and Xiaobo Zhang

From 1980–2015, China's economy grew at an average annual rate of 8.7 percent. During these 35 years, real per capita income increased by a cumulative rate of 1,759 percent, from $714 in 1980 to $13,277 in 2015 (based on the IMF's World Economic Outlook data, expressed in 2011 international purchasing power parity dollars). Apart from Equatorial Guinea, a country of less than a million people that literally struck oil, no other economy grew as much during the same period. China's growth performance is clearly spectacular and exceptional.

But China's economy has reached a crossroads. The annual growth rate has slowed to about 6–7 percent since 2014 and will likely moderate further. Part of the reason for the slowdown could be cyclical, a result of a relatively weak world economy. But a major part of the reason is structural and fundamental. China's economic growth of the previous three and a half decades was based on several key factors: a sequence of market-oriented institutional reforms, including openness to international trade and direct investment, combined with low wages and a favorable demographic structure. Chinese wages are now higher than a majority of non-OECD economies. For example, China's wages are almost three times as high as India, an economy with almost the same-sized labor force. The Chinese working-age cohort has been shrinking since 2012.

■ *Shang-Jin Wei is N. T. Wang Professor, Columbia University, New York City, New York, and China Center for Economic Research. Zhuan Xie is Research Fellow, State Administration of Foreign Exchange (SAFE) of China, Beijing, China; Xiaobo Zhang is Distinguished Professor, Peking University, Beijing, China, and Senior Research Fellow, International Food Policy Research Institute (IFPRI).*

The first section of this paper will review what factors have propelled China's economic growth in the past, and explain why they are unlikely to provide the same kind of boost going forward.

Future growth in China has to come mostly from the growth of labor productivity. Since China's investment-to-GDP ratio was already a remarkable 43.3 percent in 2015, it is hard to expect a high growth rate of productivity from continued physical investment. Indeed, Bai and Zhang (2014) estimated that the returns to investment have shown signs of decline since 2008. Some productivity increase could come from reducing resource misallocation (Hsieh and Klenow 2009), which could be accomplished by further reforms in the factor and product markets, including reforms of state-owned enterprises. However, the pace of reform in the future is unlikely to be as aggressive as in the past, partly because interest groups across China now have more means to block reforms than in the past and partly because the low-hanging fruit in the area of institutional reforms has been picked. Thus, productivity growth from this source also faces a limit.

Since the onset of the global financial crisis in 2008, the external demand for Chinese products has weakened, and wages in China have meanwhile increased faster than in almost all other major economies. A growth model based on exploiting the use of cheaper labor is no longer viable. While a strict family planning policy implemented since the early 1980s once produced an unnaturally low birth rate and therefore an unusually favorable dependence ratio for China, the same force has now produced relatively few people entering the labor force today relative to the new retirees, hence yielding an unusually unfavorable dependence ratio.

Facing rising labor costs and weak external demand, China's firms have to make a tough choice: in, out, up, or down. "In" is to move factories to inland areas where the wage is lower than coastal China. Given the pace of convergence within the country and the cost of logistics facing firms inland, this is at best a temporary strategy. "Out" means engaging in outbound direct investment, combining Chinese know-how with low wages in other countries. "Up" means innovation and upgrading, so that the firms no longer need to depend on low-paying unskilled labor. "Down" means closing the business; it is an option for individual firms, but not for the country as a whole. While a portfolio of these strategies will be employed by firms, a decisive factor for China's economic future is whether its firms can innovate and upgrade and how fast they can do so. In the next section, we focus on innovation and quality upgrading, and ask the question: Is China investing enough and wisely in research and development, and can it transition to a more innovative economy?

We study three questions in particular. First, how strong is China's national investment in research and development (R&D)? We do so by comparing the Chinese trajectory in recent years with international experiences.

Second, what is the growth of innovation by Chinese firms? To answer this question, we make use of data on patents from China State Intellectual Property Office (SIPO), the United States Patent and Trademark Office (USPTO), and World Intellectual Property Office (WIPO). We use the data on patents to compare China's rate of innovation as compared to other BRICS (that is, Brazil, Russia, India, and

South Africa) economies and high-income economies (such as the United States, Germany, Japan, and the Republic of Korea). We will use patent applications and patents granted by firms both at home and in the United States as proxies for innovative activities. China's performance on innovative activities as measured by patent data has been strong, especially in recent years, but China may well have some lessons to learn from India and in particular from the Republic of Korea. We will argue that rising wages and expanding markets are among the important drivers behind China's patent explosion.

Third, because the Chinese economy continues to have a nontrivial share of state-owned enterprises, we investigate possible resource misallocation in the innovation space. Although state-owned enterprises have received more subsidies from the government, their performance in innovation is lackluster compared to private enterprises. Furthermore, the elasticity of patent filing or patents granted with respect to expenditures on research and development is significantly higher for private sector firms than for state-owned enterprises. We interpret these data patterns as existence of misallocations in public fiscal resources. Interestingly, we find that China's state-owned enterprises often face higher realized tax burdens (the sum of corporate income tax and value-added tax as a share of sales or value added). To improve the efficiency of resource allocation, the direction of policy reforms should perhaps put weight on leveling the playing fields for firms across all ownership types with simultaneous reductions in discretionary subsidies and taxes.

Sources of China's Growth since 1980 and the Moderation of Growth since 2012

China's rapid growth in the past several decades has been driven by a combination of two sets of factors: a) market-oriented policy reforms to let market-determined output prices and factor prices replace administrative prices, to introduce and strengthen property rights, and to reduce barriers to international trade and investment; and b) economic fundamentals, including in particular a favorable demographic structure and a low initial level of labor cost. Here, we offer an overview of these factors and how they have evolved in the last 36 years.

The Chinese growth miracle started with the rural sector reform known as the "rural household responsibility system" in the early 1980s. Instead of collective farming and selling all output to a national procurement plan at a price set by the plan (usually substantially below the would-be market price), farmers were granted land user rights and allowed to sell what they produced in excess of the official quota at a market price. Agricultural production and rural incomes witnessed a dramatic increase in the ensuing years (Lin 1992). In a few years, hundreds of millions of farmers were released from their land and many started to work in factories, providing the nonfarm sector with a seemingly unlimited labor supply. In the 1980s, China's labor cost was among the lowest in developing countries, lower than in India and the Philippines and indeed lower than 114 out of 138 non-OECD

Table 1
Number of Registered Firms in China (1995–2014)

	Firm count at year end	Private (%)	State-owned firms (%)	Foreign (%)
1995	4,598,604	71	24	5
2000	5,875,706	76	19	5
2005	7,980,991	85	10	5
2010	11,150,201	90	5	5
2014	18,178,921	94	3	3
Annual growth rate (%)				
1995–2005	6	8	−3	5
2005–2014	10	11	−5	3
1995–2014	8	9	−4	4

Source: Tabulated by authors based on China Firm Registry Database.
Note: Firms ownership classification is based on the ownership information on firm registration. Foreign-invested firms include both fully foreign-owned and sino-foreign joint ventures. All firm ownership types other than "state" or "foreign" are grouped in "private."

economies. The vast majority of these workers were restricted to living in rural areas by the *hukou* system, with many working during the 1980s for township and village-owned enterprises, which were manufacturing firms located in rural areas. These enterprises provided a way for a reallocation of labor from low-productivity farm activities to higher-productivity manufacturing activities, at a time before restrictions on internal migrations were relaxed.

During the 1990s, the government launched reforms of the township and village enterprises and of the state-owned enterprise sector. Most township and village enterprises were privatized, de jure or de facto. By 2011, the township and village enterprise sector had almost disappeared, with employment plummeting from 129 million in 1995 to merely 6 million in 2011 (Xu and Zhang 2009). Among state-owned enterprises, which were overwhelmingly in urban areas, employment fell by about half from 113 million in 1995 to 67 million in 2011. The number of state-owned firms declined from 1,084,433 (or 24 percent of the total number of firms) in 1995 to 521,503 (or 3 percent of the total) in 2014 (according to our tabulations based on the China Firm Registry database in Table 1). The much larger drop in the number of state-owned enterprises than in their employment was part of a deliberate policy of "grasping the large and letting go of the small"—that is, privatizing small state-owned enterprises and consolidating bigger ones (Hsieh and Song 2015).

The reform was painful in the short run, in that tens of millions of urban workers had to leave their former state-owned employers. Remarkably, the country avoided a big spike in the unemployment rate. The key is that the de facto privatization was accompanied by aggressive reforms to lower entry barriers faced by private sector entrepreneurs. The inefficiency of the previously centrally planned, state-dominated economic system, together with very high barriers to entry, meant huge unexplored or underexplored profitable opportunities. As a result, almost all

of the lost jobs in township and village enterprises and state-owned enterprises were offset by new job opportunities in the dynamic private sector. As shown in Table 1, the number of private enterprises increased by nearly five-fold to about 17 million (18,178,921 × .94) in the period 1995–2014. By 2011, 193 million people worked in private enterprises (including self-employed) (CNBS 2012). This represents the largest de facto privatization program in world history in terms of the number of workers who move from state-sector to private sector employment, and one that was accomplished without massive unemployment.

Through this period, the growth in the Chinese economy has become driven overwhelmingly by the growth in the private sector aided by an expansion in the number of entrepreneurs. This pattern is especially true for the manufacturing sector, which has been growing faster than either the agricultural or service sectors. Indeed, Wei and Zhang (2011b) have documented two "70 percent rules" using manufacturing firm census data in 1994 and 2005: First, approximately 70 percent of the growth in industrial value added came from private sector firms between these two census years. Second, approximately 70 percent of private sector growth in value added came from growth in the count of new private sector firms (the extensive margin), while the remaining 30 percent came from growth of existing firms (the intensive margin).

China also carried out a number of other reforms intended to incentivize local governments to pursue growth-friendly policies. For example, under the fiscal arrangement introduced in the early 1980s, local governments and the central government follow a pre-determined revenue formula (though varying across regions as a function of local bargaining power), which stimulates the incentives of local officials to create a more business-friendly environment. More generally, in spite of the political centralization by the Communist Party, the country has implemented a system of fiscal and economic decentralization that grants local governments sufficient decision-making power—and more importantly incentives—to compete with each other. The local economic growth rate is used as a key performance indicator for the career advancement of officials. The delegation of economic policy authority to local governments, which have better knowledge of local information, and competition for investment and tax base among local governments in the Chinese style of federalism have provided a useful check on the temptation of local government officials to expropriate local firms. As a result, firms acquire some de facto security of property rights, even if the formal property rights institutions are problematic (Qian and Weingast 1997; Xu 2011).

China's government also set up numerous special economic zones and special development zones in the coastal provinces to attract foreign direct investment in the 1980s and 1990s. These zones help the government to meet two challenges. First, public funding for infrastructure was limited, especially in the early days of the reform era. The government was able to concentrate limited public funding to provide adequate roads, power supply, waste treatment and other infrastructure to the firms within the zones, even when it was not able to improve the infrastructure nationally at the same speed. Second, policy reforms within these zones were politically easier than doing the same things on a national scale. The success in these

zones in terms of economic growth, employment, and tax revenues in turn facilitated similar market-oriented reforms outside the zones. Foreign direct investment rose rapidly in China, especially since 1992, and these zones played an important role in attracting international firms. Foreign-invested firms were and continue to be an important channel for transfer of technology and management ideas from advanced economies to China.

China's integration with the global economy was accelerated after the country joined the World Trade Organization in December 2001. Foreign-invested firms have often accounted for half of the country's total exports. China's trade expanded fast: While China's GDP approximately doubled once every seven years, its export value in US dollar terms doubled once every four years. By 2004, China had come to be known as the World's Factory, a label describing not only the sheer volume of its cross-border trade, but also the breadth of its sector coverage (as discussed in Feenstra and Wei 2010). China's growth in both imports and exports, along with foreign investment coming to the country, is also an important channel for domestic firms to acquire technological knowhow.

While the deep cause of growth and development is institutional changes engendered by policy reforms and embrace of globalization, the proximate drivers of economic growth include improvement in productivity as a crucial component. The increase in productivity stems from innovations within sectors and the reallocation of resources (mainly workers) from lower-productivity to higher-productivity sectors, such as from the state sector to the private sector and from the agricultural sector to nonagricultural sectors (Zhu 2012). Sectoral productivity and structural change accounted for 42 and 17 percent of economic growth during 1978–1995 (Fan et al. 2003).

For three decades following the start of market-oriented reforms, China appeared to have an inexhaustible amount of "surplus labor" (which can be thought of as conceptually the same as low-productivity labor in rural areas). But signs of labor shortage started to emerge in the first decade of the 2000s. According to Cai and Du (2011) and Zhang, Yang, and Wang (2011), wages for unskilled workers showed double-digit growth starting in 2003–2004. The exact timing of a sharp increase in the wage rate of unskilled workers is subject to debate. Wang, Huang, Zhang, and Rozelle (2011) report a turning point as early as 2000. On the other hand, Knight, Deng, and Li (2011) and Golley and Meng (2011), for example, point out that barriers to internal migration, especially a rigid household registration system that prevents rural households from moving freely to urban areas, imply additional scope for rural-to-urban migration if and when the remaining barriers can be dismantled. In any case, China is a low-wage country no more.

Two features of demographic transition have also been a powerful driver of China's growth in the past three and a half decades. The first feature is a favorable dependency ratio. China's sharp decline in fertility rate has meant fewer young dependents to support for a given size of the working cohort. The fraction of prime-age people in total population rose steadily for three decades, creating an unusually large demographic dividend, which in turn contributed to economic growth (Cai and Wang 2008; Wei 2015).

The second feature of demography that affects growth is the gender ratio imbalance of the premarital cohort. This less-studied factor may have a quantitatively significant effect as well. The one-child policy has yielded an unintended consequence in distorting the sex ratio in favor of boys. As the one-child generation enters the marriageable age, young men face a very competitive marriage market. In order to attract potential brides, families with sons choose to work harder, save more, and take on more risks, including exhibiting a higher propensity to be entrepreneurs (Wei and Zhang 2011a, b; Chang and Zhang 2015; Wei, Zhang, and Liu forthcoming). It is estimated that increasing marriage market competition due to sex ratio imbalances has contributed to about two percentage points of economic growth per year (Wei and Zhang 2011b).

It is important to point out that the additional growth due to an unbalanced sex ratio is of an immiserizing type: social welfare is likely to have become lower even though the GDP growth accelerated. The logic is explained in Wei and Zhang (2011b): The extra work effort and extra risk-taking that produce a higher GDP growth rate are motivated by a desire to improve one's chance (or one's children's chance) of success in the marriage market. Yet the fraction of young men who will not get married in the aggregate is determined by the sex ratio, and not by the economy-wide work effort, risk-taking, or GDP growth rate. In this sense, the extra work effort and risk-taking are futile; households collectively would have been willing to give up this part of income growth in exchange for no sex ratio imbalance.

Thus, from 1980 to 2011, China was experiencing a relatively low wage, a large workforce with a favorable dependency ratio, and an increasingly unbalanced sex ratio in the premarital cohort. But starting in 2012, China's age cohort of 15–60 started to shrink in absolute size. Policy changes to postpone the official retirement age or to encourage more female labor force participation will at best moderate the resulting decline in the workforce. Because the female labor force participation was very high under the central planning regime before the 1980s, higher than most non-Communist countries in the world, such as the United States, Japan, Germany, India, and Indonesia, the participation rate of women in the labor force has in fact come down over time. The recent relaxation of the family planning policy in November 2015 from the limit of one child per couple to two children per couple, while motivated by a desire to improve the demographic pattern for the economy, will make the dependency ratio worse for the next decade-and-a-half rather than better by adding to the number of children without altering the size of the workforce. After all, no couple can produce a 16-year old right away (Wei 2015). The sex ratio at birth started to become less unbalanced in 2009, and the contribution to growth from an unbalanced sex ratio will become weaker over time.[1]

---

[1] Beside a moderation of growth since 2012, China has to deal with challenges associated with income inequality, regional disparity, environmental degradation, and corruption. For perspectives on these challenges, see Fan, Kanbur, Wei, and Zhang (2014).

Figure 1
Contributions to GDP Growth of Physical Capital, Human Capital, and Total Factor Productivity, 1979–2015
(share)
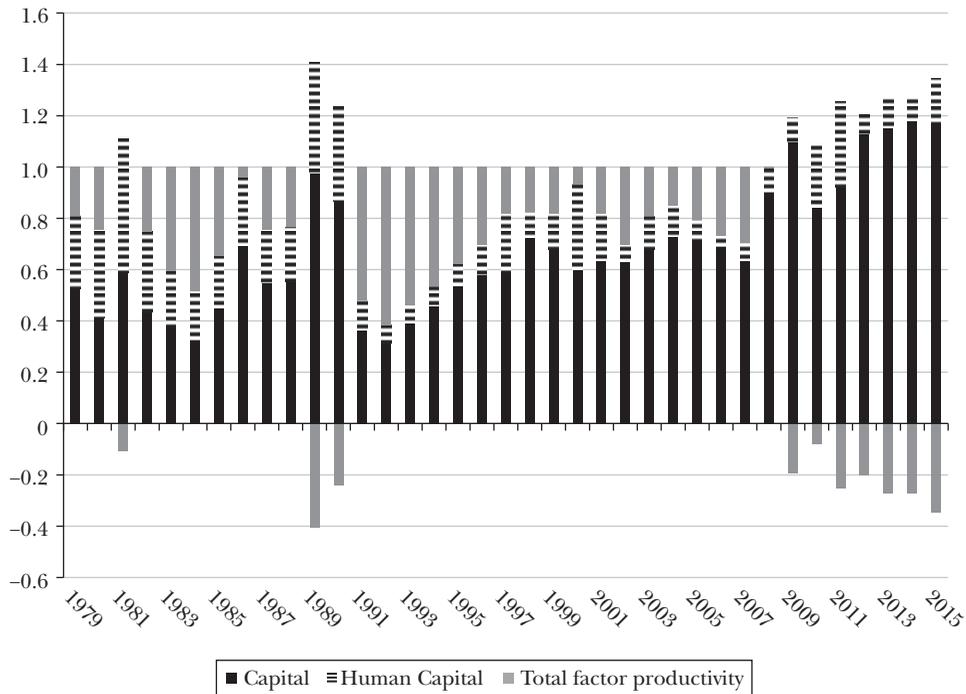
Source: Authors' calculations based on the methodology and data sources detailed in Online Appendix A.
Note: See Appendix for details of the estimation.

Evolution of Aggregate Productivity

To see how the growth of physical capital, human capital (workforce adjusted for average years of schooling), and total factor productivity each contributes to China's GDP growth, we perform a simple decomposition based on an aggregate production function approach.[2] Figure 1 summarizes the result. A few features are worth noting. First, investment in physical capital has always been important for China's growth, accounting for 67.9 percent on average throughout this period. The relative share of contribution from physical investment increased to 107 percent after 2009, which resulted from the government stimulus package in response to the global financial crisis. Second, the contribution from the growth of human capital has been positive, at 12.5 percent during 1999–2008 and 16 percent during 2009–2015.

---

[2] The computation method and data sources are explained in online Appendix Part A, available with this paper at http://e-jep.org.

Because of the outsized role of physical investment in the Chinese economy, the contribution of human capital is smaller than what one typically finds from growth decomposition for an OECD economy. Third, the growth of total factor productivity was a major contributor to GDP growth before 2008, often accounting for 20 percent or more to the total growth. (An exception was the period of 1989–1991, a time of domestic political turbulence and international sanctions.)

Strikingly, the contributions from the growth of total factor productivity have turned persistently negative since 2009.[3] Upon reflection, this is perhaps not overly surprising. The Chinese government's response to the global financial crisis that started in 2008 was to encourage physical investment through an aggressive fiscal (and bank lending) program, but there were no ambitious structural reforms pursued during this period that could have raised aggregate efficiency, and yet GDP growth started to moderate after 2012—and this combination is a recipe for negative growth in total factor productivity.

The Chinese economy is at a crossroads. Structural factors in the form of less-favorable demographics and a higher cost of labor imply a lower potential growth rate. To achieve robust future growth, raising the growth of total factor productivity is a must.

One way to raise future productivity growth is to pursue more structural reforms. These include removing barriers to labor mobility from rural to urban areas (the *hukou* system) and leveling uneven access to bank loans by firms of different ownership. Another way to raise productivity growth is via innovation. Innovation can take the form of creating new products, new ways of using existing products, new designs, new processes for producing existing products that are more efficient and cost-effective, new ways of organizing business, and new ways of branding and marketing the products or services.

Can China transition from a world assembly line to an innovation powerhouse? It's easy to list reasons to be skeptical. There is no shortage of news stories of intellectual property rights violations by Chinese companies.[4] There is criticism that the Chinese school system puts too much weight on rote learning and not enough on creative and critical thinking. On the other hand, more optimistic examples are available, too. Tencent, the company that provides the popular communication tool, WeChat, which combines group chat, voice calls, video sharing, and financial exchanges, is generally regarded as among the most innovative internet companies in the world. Huawei, the telecom equipment producer, is said to take out more

---

[3] The baseline calculation assumes a capital income share of 0.50. We vary the share from 0.4 and 0.55 and find that the broad pattern of the evolution of total factor productivity stays the same. Our finding is broadly consistent with Wu and Conference Board China Center (2014). For example, they reported negative total factor productivity from 2007 to 2012, while our estimate indicates such a decline from 2009 to 2014. One difference between their study and our growth decomposition is that they obtain a larger contribution from human capital, which may be related to the way the schooling adjustment is made.

[4] Fang, Lerner, and Wu (2016) provide evidence that regional variations in the strength of intellectual property rights protection in China are correlated with propensity to innovate for privatized formally state-owned firms.

patents a year than either Apple or Cisco. The world's first quantum satellite was launched by China in August 2016. To address whether such examples of innovation are exceptions or the norm, we offer a systematic look at the data in the next section.

It is hard to quantify with precision the relative contributions to total factor productivity growth from different sources. From the *China Statistical Yearbook on Science and Technology*, we compute and compare investment made by firms in the survey in (a) importing and digesting foreign technologies, (b) buying and digesting technologies from other domestic firms, and (c) developing their own in-house technological improvement. In 2000, the survey firms collectively spent nearly 20 percent of their technology improvement budget on importing and digesting foreign technology, about 2 percent on buying technologies from other domestic sources, and 78 percent on developing their own in-house technological improvement. Over time, the share of the first item declines, whereas the last two items expand. By 2014, the survey firms collectively spent 11 percent of their technological improvement budget on importing and digesting international technologies, about 5 percent on buying technologies from domestic sources, and the remaining 84 percent on developing their own in-house technological advancement, with the last two categories showing a significant increase over the shares in 2000 (see online Appendix Figure A1). These numbers indicate in an indirect way the improvement in the domestic innovation capacity in China's manufacturing sector.

Research and Development: Investment and People

Innovative leaders at both the corporate and national levels tend to invest heavily in research and development. The United States, Japan, and Germany, the largest three rich economies, invested more than 2.7 percent of their GDP in research and development in 2014, which is almost 50 percent more than an average OECD country (about 1.9 percent in 2014), and about three times as much as most developing countries. If China makes the transition to a more innovative economy, it needs to make a commitment to research and investment spending as well.

In 1991, when systematic data on this subject started to be collected, China invested 0.7 percent of GDP in research and development. This was much lower than technological leaders like the United States, Japan, and Germany, but not out of line with big developing economies such as India, Brazil, or South Africa. Indeed, because China's competitiveness at this time was based on exploiting its vast cheap labor and making use of technologies developed elsewhere, domestic research and development and innovation were not an imperative at this time.

A comparison of research and development spending between China and other economies is provided in Figure 2. For all countries in the world other than China, we plot research and development expenditure as a share of GDP in the latest possible year (which is 2014 for most countries). Clearly, higher-income countries tend to have a higher ratio of research and development spending to GDP. For China, we plot the same ratio using corresponding data for China from 1995 to

Figure 2
Research and Development Intensity Comparison
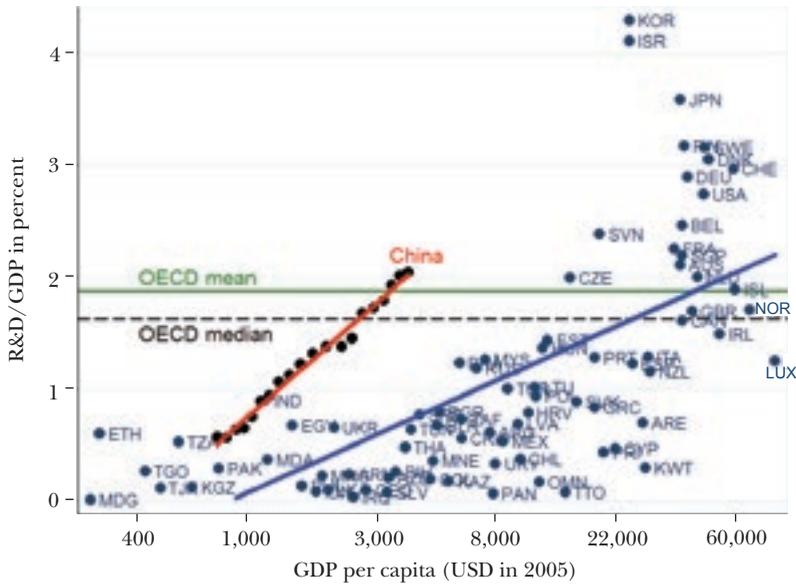
Source: World Bank, OECD database (https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm).
Note: Data for China are from 1995 to 2014, and data for all other countries are for 2014 or the latest year available.

2014. By 2010, China had reached the median value of research and development as a share of GDP. By 2012, its spending had caught up with the OECD average (at 1.88 percent of GDP in that year) even though China's income level was still less than one-fifth of the OECD average. By 2014, China's research and development spending had reached 2.05 percent of GDP. From an aggregate R&D spending viewpoint, China is an overachiever.

Another indicator of innovation effort is the share of researchers in the population. In 1996, China had 443 researchers per million people. In comparison, the shares for the United States, Japan, and Korea were 3,122, 4,947, and 2,211 per million, respectively. The Chinese ratio in 1996 was comparable to Brazil (420 per million in 2000) and better than India (153 per million in 1996), though much lower than Russia (3,796 per million in 1996). By 2014, the share in China had grown to 1,113 researchers per million population.[5] Because China's research and development expenditure has grown faster than the number of researchers, research and development expenditure *per researcher* has grown over time as well.

---

[5] For more cross-country comparisons, see online Appendix Figure A2, available with this paper at http://e-jep.org.

Table 2
Patent Applications and Patents Granted (1995–2014)

Year	Number of patent applications at China's State Intellectual Property Office (SIPO) (1)	Number of patents granted by China's SIPO (2)	Distribution of patents granted by type of patents (3)			Share of patents granted to applicants from outside China (%) (4)	Number of patents granted by foreign patent offices to China-based applicants (5)
			Invention (%)	Utility model (%)	Design (%)		
1995	83,045	45,064	8	68	25	8	75
2000	170,682	105,345	12	52	36	10	88
2005	476,264	214,003	25	37	38	20	398
2010	1,222,286	814,825	17	42	41	9	2,587
2014	2,361,243	1,302,687	18	54	28	7	10,603
Annual growth rate in different periods (%)							
1995–2005	19	17	31	10	22	28	18
2005–2014	19	22	18	27	18	9	44
1995–2014	19	19	25	18	20	18	30

Source: Authors' tabulation based on data from China's State Intellectual Property Office's webpage (http://www.sipo.g ov.cn/tjxx/) for columns 1–4 and the World Intellectual Property Office for Column 5.

The Pace of Innovation as Measured by Growth in Patents

Not all dimensions of innovation are equally well measured. The output of innovation can take the form of patents, commercial secrets, improvement in business processes or business models, and others. Innovation can also take place in areas outside the commercial space, such as culture. Since innovation in the form of patents is relatively well measured, we will pay special attention to patents by firms. Our conjecture is that innovation across all dimensions is positively correlated.[6]

The number of Chinese patents has exploded. Table 2 presents some summary statistics. The number of patent applications filed in China's State Intellectual Property Office (SIPO) rocketed from 83,045 in 1995 to more than 2.3 million in 2014, at an annual growth rate of 19 percent (column 1). In 2011, China overtook the United States as the country with the most patent filings in the world that year (based on data from WIPO).

What explains the explosion of Chinese patents? Could it be easy approval or low-quality patents? Some straightforward comparisons across countries suggest not.

One simple metric for judging ease of patent approval is the ratio of the number of patents granted in year t to the number of patent applications in year $t − 1$, which we will call the patent approval rate. Based on data from the World Intellectual Property Organization, the patent approval rate in China in recent years is 30–40

---

[6] A simple regression of firm-level total factor productivity (estimated using the Olley–Pakes method based on data from the Annual Survey of Manufacturing Firms) on the cumulative number of patents yields a positive slope coefficient. In other words, firm-level total factor productivity and the stock of patents are positively correlated. Fang, He, and Li (2016) also show a positive association between firm-level total factor productivity and patent count. They interpret this as patent innovation raising productivity; such an interpretation would need an instrumental variable approach to back it up.

Table 3

Number of Patents Granted by the US Patent and Trademark Office (USPTO) to International Corporate Applicants

Year	China	Brazil	India	Russia	South Africa	Germany	Japan	Republic of Korea
1995	62	63	37	98	123	6,600	21,764	1,161
2000	119	98	131	183	111	10,234	31,296	3,314
2005	402	77	384	148	87	9,011	30,341	4,352
2010	2,657	175	1,098	272	116	12,363	44,814	11,671
2014	7,236	334	2,987	445	152	16,550	53,849	16,469
Annual growth rate in different periods (%)								
1995–2005	21	2	26	4	−3	3	3	14
2005–2014	38	18	26	13	6	7	7	16
1995–2014	28	9	26	8	1	5	5	15

Note: Computed by authors based on data from the United States Patent and Trademark Office (USPTO).

percent, which is essentially in the middle of the approval rates across countries. For example, the Chinese approval rate is higher than those in India and Brazil, which are close to 20 percent, but lower than those in the United States and Korea, which are in the range of 50–60 percent. Therefore, the Chinese patent approval ratio does not seem to be unusually high.

Among the three types of patents (invention, utility model, and design), the fraction of approved invention patents, arguably the most technically intensive category, rose from 8 percent in 1995 to 18 percent in 2014 (column 2 of Table 2). In 2005, patents granted to foreign applicants accounted for more than 20 percent of China's total approved patents, but dropped to 7 percent in 2014, suggesting an increasing role of indigenous innovations in the Chinese economy since 2005. As Table 2 shows, both total Chinese patents filing and approvals show a rapid growth.[7]

One way to consider the quality of Chinese patents is to examine patents applied by and granted to Chinese firms in other countries. As noted earlier, the rate of patents approved by China's patent office grew at an annual rate of 19 percent from 1995 to 2014. During that period, the number of patents granted to Chinese applicants by patent offices in developed countries was rising even faster at 30 percent per year (see last column in Table 2).

Of particular interest is a comparison of the number of patents granted by the US Patent and Trademark Office (USPTO) to Chinese firms with those to firms from other countries. As shown in Table 3, the number of patents granted by the USPTO to Chinese corporate applicants rose from 62 in 1995 to 7,236 in 2014. The annual growth rate was 21 percent in the first half of the period (1995–2005) but accelerated

---

[7] The online Appendix available with this paper at http://e-jep.org includes more detail on patent data. For example, Appendix Tables A2 and A3 provide more detail on Chinese patent filings and approvals, while Appendix Figure 3 provides more details on cross-country comparisons of patent approval rates.

Figure 3
Patents Granted in US Patent and Trademark Office (USPTO) to Different Countries
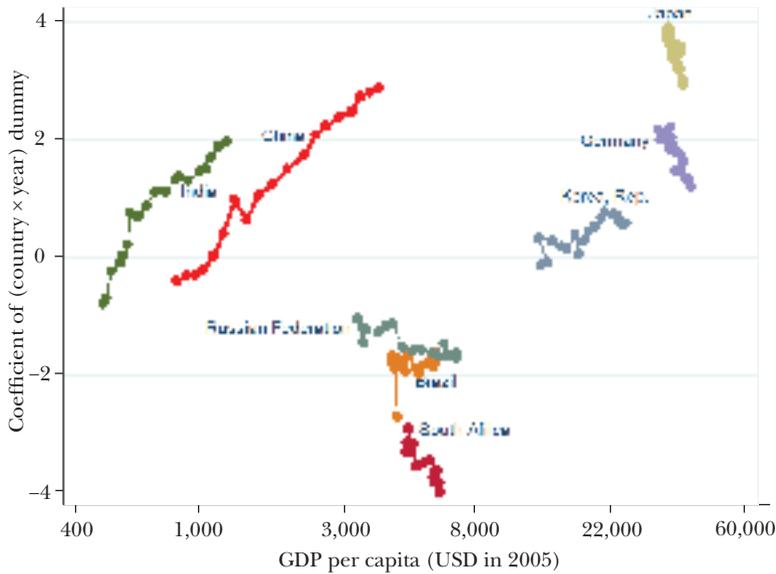
Source: Authors using data from World Intellectual Property Office (WIPO).
Note: Conditional plot by controlling for population, population squared, and country and year fixed effects. One can interpret the coefficients indicated on the y-axis as showing how a given country does relative to the average international experience based on its population size.

to 38 percent a year during the latter half of the period (2005–2014). Of the comparison countries—Brazil, Russia, India, South Africa, German, Japan, and Korea—only India had a similar rate of growth in corporate patents in the United States.

Two natural adjustments are to consider a country's population size and income level. To this end, we run cross-country regressions with log number of patents granted to applicants from various comparison countries by the US Patent and Trademark Office as the dependent variable. As explanatory variables, we use the log of population, squared log of population, and country times year (country × year) fixed effects. Figure 3 plots the estimated coefficients for the interaction term between county and year fixed effects for selected counties. These coefficients can be interpreted as how a given country does relative to the average international experience based on its population size. China shows steady gains in patents even with these adjustments. Of the comparison countries, India also shows gains over time after these adjustments, but Japan, Germany, the Republic of Korea, the Russian Federation, Brazil, and South Africa do not. Overall, Chinese firms collectively do better in their patent count than what the country's population size and income level would have suggested.[8]

---

[8] Details of the regressions are available in an online Appendix available with this paper at http://e-jep.org. See Appendix Table A5.

One can also look at foreign citations of Chinese patents (granted by China's State Intellectual Property Office). The count of foreign citations of Chinese invention patents grew at the rate of 33 percent a year during 1995–2005, but accelerated to 51 percent a year from 2005 to 2014. The growth of citations of Chinese utility model patents is similar, at 36 percent per year during 1995–2014. After adjusting for population size and income, Chinese firms perform well.[9] This pattern is consistent with international recognition of rising scientific and innovative ideas out of China.

Overall, not only has the number of Chinese patents exploded, but a variety of comparisons suggest that Chinese patent quality also exhibits a real and robust improvement over time that is quite favorable relative to international experience. There is no reason to be pessimistic about the intrinsic ability for Chinese firms to innovate.

Patterns of Innovation Growth

By looking at patterns of patents across different categories of industries, we can gain insight into some of the factors as potential drivers of innovation, including the rise in relevant market size, industrial competition, market size, and change in relative prices (such as rising wages). We merge the Chinese patent database with the Annual Survey of Industrial Enterprises in China (ASIEC). The ASIEC database covers all the state-owned enterprises and private firms with sales exceeding 5 million yuan from 1998 to 2009, including ownership information.[10] The patent database contains all patents granted by China's State Intellectual Property Office between 1985 and 2012. One pattern that emerges is that state-owned enterprises in general perform worse than private firms in generating patents. During the period 1998–2009, the number of patents granted to private firms in China grew by 35 percent per year, overtaking the number of patents given to state-owned and foreign firms by a comfortable margin. The drop in the share of patents by state-owned enterprises is partly due to the shrinkage of that sector, as described earlier. In 1998, state-owned enterprises accounted for 30 percent of total firms in the ASIEC database, while they dropped to 2 percent by 2009. Clearly, private firms have become the engine of innovation in China.

Market size has been regarded as a key driver of innovation in the literature (Acemoglu and Linn 2004). In other words, firms aiming at larger global markets should be more innovative. In past decades, the Chinese economy has become

---

[9] We perform cross-country regressions similar to those described in Figure 3 with the forward citation of Chinese firms' patents by all patent applicants in the United States as the dependent variable. The online Appendix Table A5 provides more detail on the extent of forward citation across countries regression analysis, and Appendix Figure A4 shows the coefficients on the interaction term between country and year fixed effects against log income. Overall, relative to a country's population size and income level, the Chinese firms do well in terms of forward citations of their patents. See also Xie and Zhang (2015) for an analysis of the growth of patents in China.

[10] While ASIEC data for 2010–2014 seem to be available on the gray market, the quality appears suspect. To be conservative, we do not use these data in this paper.

increasingly integrated with the world economy, in particular since China joined the World Trade Organization in 2001. In this data, exporting firms in China are indeed more innovative than nonexporting firms.

Since 2003, real wages in China have grown by more than 10 percent a year. Some reckon that China has passed the so-called "Lewis turning point," which means that an era of ultra-low-wage production is over (for example, Zhang, Yang, and Wang 2011). While patents are rising for both capital- and labor-intensive firms, the fraction of patents granted to labor-intensive firms increased from 55 percent in 1998 to 66 percent in 2009. Rising labor costs may have induced labor-intensive sectors to come up with more innovations to substitute for labor.[11]

We can connect the discussions on total factor productivity and on innovation. We separate all firms in the ASIEC sample into those with no patents during 1998–2007, those with a cumulative patent count of 1–4 patents during the same period, and those with a cumulative count of more than 4 patents. We compute the growth of total factor productivity for each individual firm. We find that firm-level productivity tends to grow faster in the group that engages in more innovation. This suggests that to reverse China's negative levels of total factor productivity, it would be helpful for China to facilitate conditions that expand both the number of firms that engage in innovative activities and the intensity of innovation per innovating firm.

Misallocation of Innovation Resources

The innovation gap between China and leading advanced economies such as the United States, Japan, and even Korea is still wide. On the list of 2015 Thomson Reuters' Top 100 Global Innovators, Japanese and US firms lead the way, while no single Chinese firm makes the list. More systematic data confirms the continued gap in innovation (Shen, Wang, and Whalley 2015). The numbers of US patents received by either Japanese, German, or Korean firms are still more than twice as many as those obtained by Chinese firms in spite of their smaller population size (as shown earlier in Table 3). Part of the gap reflects different stages of development: as we have shown, both investment in R&D and innovation measured by patent count are strongly positively related to GDP per capita. However, another contributor to the gap could be resource misallocation in the innovation space. We turn to this topic next.

Following China's reforms in the late 1990s, the share of state-owned enterprises in total firms dropped significantly from 24 percent in 1995 to 3 percent in 2014, as discussed earlier. However, most of the surviving state-owned enterprises are relatively big, and are in upstream industries or strategically important sectors

---

[11] The descriptions in these paragraphs are based on bivariate correlations, and as such are, of course, only suggestive. In order to evaluate the relative importance of these factors' contributions to firm innovations in a more rigorous manner, we run multivariate regressions using a hybrid binomial estimation method proposed by Allison (2005). The details are available in the online Appendix, Part B, available with this paper at http://e-jep.org (see Appendix Tables A7–A11). Overall, the findings confirm the importance of rising labor cost.

(Hsieh and Song 2015). They are typically subject to less competition than private enterprises. Thus, China's state-owned firms both absorb nontrivial resources, including government subsidies, and still command nontrivial political weights. Part of China's move to becoming an innovative economy must be to improve the efficiency of resource allocation between state-owned and private firms. China's state-owned firms continue to receive considerable financial support from the government, including access to low-cost bank loans and research and development subsidies. In the aftermath of both the 1997 Asian financial crisis and the 2008 global financial crisis, the Chinese government launched stimulus packages which often involved credit expansion and which disproportionately went to state-owned enterprises. The more favorable policies and injection of massive stimulus funds have reduced the returns to capital of state-owned enterprises since 2008 (Bai and Zhang 2014), caused a decline in their total factor productivity (Wu 2013), and provided a lifeline for inefficient zombie firms (Tan, Huang, and Woo 2016). The returns to capital of state-owned enterprises are much lower than their private counterparts (Hsieh and Song 2015). Moreover, state-owned enterprises lagged behind private firms in total factor productivity (Brandt 2015). These patterns suggest a misallocation of government support between state-owned and private enterprises. Government subsidies for research and development can promote firm innovations in China (as reviewed in Boeing 2016 and confirmed by our own firm-level regressions). Government subsides can be defended on the ground that research and development by firms generate positive externalities. Indeed, most advanced countries subsidize research and development as well. The question is not whether subsidies can be justified at all, but rather whether China's allocation of such subsidies is consistent with economic efficiency.

Based on simple averages, it would appear that a greater fraction of state-controlled firms are innovative (that is, they have patents) than domestic private sector firms. Indeed, some state-controlled firms receive many patents in a year. But the simple averages are misleading both because state-controlled firms are much larger on average (and larger firms tend to invest more in research and development), and because they tend to receive more subsidies from various levels of the government. Subsidies from local governments to local government-controlled firms are especially noteworthy.

We examine firm-level data for evidence of effectiveness of research and development spending in generating innovations. Based on firms in the ASIEC sample during 2005–2007, for every 10 million yuan of firm-level investment in research and development, domestic private- sector firms and foreign-invested firms generate 6.5 and 7.6 patents, respectively. In comparison, the same investment by state-owned firms yields a more meager 2.2 patents. We may obtain a more informative picture by sorting firms by size and ownership. In Figure 4, on the horizontal axis, all Chinese firms are sorted into ten size deciles based on the sum of the sales during the period, with the first decile being the smallest and the 10th being the largest. Within each size decile, firms are then sorted by ownership. "State" refers to all firms in which the state (either the central or the local governments) have controlling shares (50 percent or

Figure 4

Patent Count per Million Yuan of R&D Investment as a Function of Ownership and Size
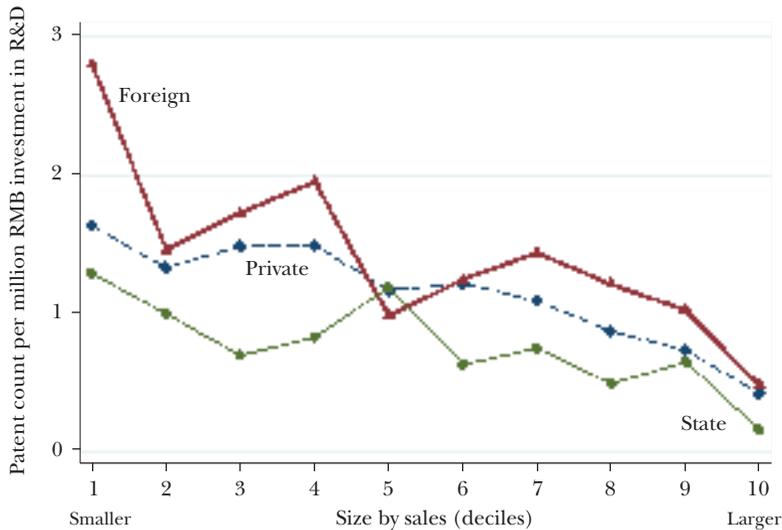
Source: Data from Annual Surveys of Above-Scale Manufacturing Firms, 2005–2007.
Notes: We used a 2005–2007 sample. Firms are first sorted by size (sales) deciles, and, within each size decile, are subdivided into three ownership groups: "State" refers to firms for which the state (central or local governments) has 50 percent share or more; "Foreign" refers to firms for which foreign entities have at least 10 percent of the share and the state has less than 50 percent share; "Private" includes all other firms. Firms that report positive patents but no R&D spending (or spending of less than 100 yuan) are excluded.

more); "foreign" refers to all firms in which foreign entities, including investors from Taiwan, Hong Kong, or Macao, have a 10 percent share or more but the state has no more than 50 percent of the shares. All other firms are in the "private" category.

Table 4 presents statistics on domestically granted patents by firm ownership and size during 2005–2007 when all relevant data are available. In most of the size categories, domestic private sector firms and foreign-invested firms invest more in research and development and generate more patents than their state-owned counterparts.

Inspecting Figure 4 and Table 4, several patterns are especially noteworthy. First, the returns to research and development spending—as measured by the number of patents per million yuan of research and development spending on the vertical axis—tend to decline with firm size. Because large firms tend to spend more on research and development, this pattern is consistent with the idea that diminishing returns apply to investment in innovation. Second, across most size deciles, we see that foreign-invested firms and domestic private sector firms tend to have higher returns to investment in research and development. Third, we do not observe a connection between firm subsidies (relative to sales) and effectiveness at converting research and spending into innovative outcome as measured by patents. Instead, we

Table 4
Patents, Research and Development Expenditure, and Subsidies by Firm Type and Size

		Size by sales quantile				
	Firm type	*0–20%*	*20–40%*	*40–60%*	*60–80%*	*80–100%*
Number of patents	Private	1,107	2,630	4,003	7,585	64,586
	Foreign	226	579	876	3,031	44,178
	State	46	87	177	351	9,116
R&D expenditure (million RMBs)	Private	769	1,763	3,335	7,933	143,848
	Foreign	122	312	760	2,333	86,946
	State	41	112	210	595	51,172
(Subsidies/sales) × 100	Private	0.22	0.25	0.25	0.24	0.31
	Foreign	0.13	0.11	0.11	0.11	0.11
	State	0.84	0.86	0.71	0.74	0.27

Source: Authors' calculation based on Annual Surveys of Above-Scale Manufacturing Firms, 2005–2007.
Note: We used a 2005–2007 sample; divided into 5 groups by sales; and drop observations that invest less than 100 Yuan but have positive patents. "Foreign" refers to firms for which foreign entities have at least 10 percent of the share and the state has less than 50 percent share. "Private" includes all other firms.

see that state-controlled firms tend to have much higher subsidies (relative to sales) than either domestic private firms or foreign-invested firms. Interestingly, because small and medium state-owned firms are mostly owned by local governments, they receive more subsidies from the local governments than large state-owned firms.

König, Song, Storesletten, and Zilibotti (paper in progress) argue that, in theory, the most productive firms should pursue innovation and less-productive firms should just imitate. Against this theoretical benchmark and also compared to the data patterns in Taiwan, they find that less-productive firms in China engage in too much research and development spending—and the more-productive firms may not do enough. Based on their calibrations, if the R&D misallocation can be reduced (so that the association between productivity and R&D spending resembles that in Taiwan), the aggregate productivity growth in Chinese manufacturing during 2001–2007 could have grown by about one-third to one-half.

In sum, there is prima facie evidence that the pattern of subsidies across China's firms represents resource misallocation. China's economy-wide innovative outcomes would have been higher if the subsidies were more evenly spread across firm ownership.[12] The sensible policy reforms would be to provide subsidies only

---

[12] The Appendix available online with this paper at http://e-jep.org offers some exploratory regressions that tend to confirm the intuition in the text. In particular, we regress patent count on firm R&D expenditure by controlling for firm sales, firm fixed effects, and year fixed effects. In order to evaluate whether private firms and state-owned enterprises have different elasticity regarding R&D expenditure, we interact firm ownership with R&D expenditure in the regressions. The interaction term between the state-owned enterprise dummy and R&D variable is statistically negative, indicating that the elasticity of

in cases where the social returns exceed private returns (such as certain innovative activities) without regard to firm ownership.

Conclusions

China's past success in economic growth means that its real manufacturing wage has increased by about 14-fold from 1980 to 2015. In addition, China's shrinking workforce since 2012 has added to the wage pressure. By necessity, China has to move to a growth model that is based more on innovation and productivity increase than in the past.

Can China rise to the challenge? One sometimes hears an argument for the "middle-income trap hypothesis," which claims that only in exceptional cases can a middle-income country ever manage to become an innovative high-income economy. Indeed, the challenges facing China have often been expressed in the context of a possible middle-income trap by both the government of the country and some scholars (for example, OECD 2013; Ma 2016). Han and Wei (2015) do not find support for an unconditional notion of the middle-income trap hypothesis, using both a transition matrix analysis and a nonparametric analysis (by regression trees). Nonetheless, they identify certain conditions under which growth in a middle-income country could stagnate or even regress.

We have argued that Chinese firms have demonstrated a capacity to become more innovative in response to wage pressure and global opportunities. The data on Chinese patents, both from a quantity and a quality perspective, appear encouraging enough that we should not be that pessimistic about China's prospects for a successful transition to a more innovation-based growth model.

If China finds effective ways to embrace a shift to a more innovative economy, it can realize faster its dream of moving into the high-income club. The government subsidies tend to favor state-owned firms, and yet both domestic private sector firms and foreign-invested firms are more effective in converting investment in R&D to innovation outcomes as measured by patents. One meaningful step along those lines would be for China to reduce its apparent misallocation of innovation resources by leveling the playing field for firms of all ownership types, limiting the government's discretion in subsidies for research and development, and assuring that private sector firms have a fair chance at receiving those subsidies. This will be a helpful structural reform that will complement the reforms in stronger protection of intellectual property rights and in the education system.

---

patents granted with respect to R&D expenditures is significantly higher for private firms than for SOEs. This finding is consistent with the view that state-owned enterprises have not spent R&D resources as efficiently as private firms.

References

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119(3): 1049–90.

Allison, Paul D. 2005. *Fixed Effects Regression Methods for Longitudinal Data Using SAS.* Cary, NC: SAS Institute.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333.

Bai, Chong-en, and Qiong Zhang. 2014. "Zhongguo de Ziben Huibaolv Jiqi Yingxiang Yinsu Fenxi (Returns to Capital and Their Determinants in China)." *World Economy* [in Chinese] 10: 2–30.

Boeing, Philipp. 2016. "The Allocation and Effectiveness of China's R&D Subsidies: Evidence from Listed Firms." *Research Policy* 45(9): 1774–89.

Brandt, Loren. 2015. "Policy Perspectives from the Bottom Up: What Do Firm-Level Data Tell Us China Needs to Do?" Paper presented at the Asia Economic Policy Conference on "Policy Challenges in a Diverging Global Economy," Federal Reserve Bank of San Francisco, November, 19–20. In *Policy Challenges in a Diverging Global Economy* edited by Reuven Glick and Mark M. Spiegel, 282–302. Available at http://www.frbsf.org/economic-research/events/2015/november/2015-asia-economic-policy-conference/.

Cai, Fang, and Yang Du. 2011. "Wage Increases, Wage Convergence, and the Lewis Turning Point in China." *China Economic Review* 22(4): 601–610.

Cai, Fang, and Meiyan Wang. 2008. "A Counterfactual Analysis on Unlimited Surplus Labor in Rural China." *China & World Economy* 16(1): 51–65.

Chang, Simon, and Xiaobo Zhang. 2015. "Mating Competition and Entrepreneurship." *Journal of Economic Behavior & Organization* 116(C): 292–309.

China National Bureau of Statistics (CNBS). 2012. *China Statistical Yearbook.* Beijing: China Statistical Press.

Feng, Shiliang. 2014. "Factor Allocation and Economic Growth in China: Analysis across Regions and Sectors from 1978 to 2012." Dissertation, Peking University. [In Chinese.]

Fan, Shenggen, Ravi Kanbur, Shang-Jin Wei, and Xiaobo Zhang. 2014. *The Oxford Companion to the Economics of China.* Oxford University Press.

Fan, Shenggen, Xiaobo Zhang, and Sherman Robinson. 2003. "Structural Change and Economic Growth in China." *Review of Development Economics* 7(3): 360–77.

Fang, Jing, Hui He, and Nan Li. 2016. "China's Rising IQ (Innovation Quotient) and Growth: Firm-level Evidence." Unpublished paper, IMF.

Fang, Lily, Josh Lerner, and Chaopeng Wu. 2016. "Intellectual Property Rights Protection, Ownership, and Innovation: Evidence from China." Unpublished paper, INSEAD, Harvard University, and Xiamen University.

Feenstra, Robert C., and Shang-Jin Wei. 2010. *China's Growing Role in World Trade.* University of Chicago Press.

Golley, Jane, and Xin Meng. 2011. "Has China Run Out of Surplus Labour?" *China Economic Review* 22(4): 555–72.

Han, Xuehui, and Shang-Jin Wei. 2015. "Re-examining the Middle Income Trap Hypothesis (MITH): What to Reject and What to Revive?" CEPR Discussion Paper 10989.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403–48.

Hsieh, Chang-Tai, and Zheng (Michael) Song. 2015. "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China." *Brookings Papers on Economic Activity* 50(1): 295–366.

Knight, John, Quheng Deng, and Shi Li. 2011. "The Puzzle of Migrant Labour Shortage and Rural Labour Surplus in China." *China Economic Review* 22(4): 585–600.

König, Michael, Zheng M. Song, Kjetil Storesletten, and Fabrizio Zilibotti. In progress. *"From Imitation to Innovation: Where is All that Chinese R&D Going?"*

Li, Bin. 2011. "Comparative Analysis of Estimates on Capital Stock of China" [in Chinese]. *Journal of Quantitative & Technical Economics* [in Chinese] (12): 21–36.

Lin, Justin Yifu. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82(1): 34–51.

Ma, Damien. 2016. "Can China Avoid the Middle Income Trap?" *Foreign Policy*, March 12.

OECD. 2013. *The People's Republic of China— Avoiding the Middle-Income Trap: Policies for Sustained and Inclusive Growth.* September. Paris: OECD.

Qian, Yingyi, and Barry R. Weingast. 1997. "Federalism as a Commitment to Preserving Market Incentives." *Journal of Economic Perspectives* 11(4): 83–92.

Qu, Yue, Fang Cai, and Xiaobo Zhang. 2013. "Has the Flying Geese Paradigm Occurred in China?" *China Economist* 8(6): 18–31.

Shen, Keting, Jing Wang, and John Whalley. 2015. "Measuring Changes in the Bilateral Technology Gaps between China, India and the U.S. 1979–2008." NBER Working Paper 21657.

Tan, Yuyan, Yiping Huang, and Wing Thye Woo. 2016. "Zombie Firms and the Crowding-Out of Private Investment in China." *Asian Economic Papers* 15(3).

Wang, Xiaobing, Jikun Huang, Linxiu Zhang, and Scott Rozelle. 2011. "The Rise of Migration and the Fall of Self-employment in Rural China's Labor Market." *China Economic Review* 22(4): 573–84.

Wei, Shang-Jin. 2015. "Freer Choice, Lower Growth, and Higher Welfare: Recent Changes in China's Population Control Policy and the Impact on Its Economy." Keynote (Dinner) Address at the Asia Economic Policy Conference "Policy Challenges in a Diverging Global Economy," Federal Reserve Bank of San Francisco, November 19–20. In *Policy Challenges in a Diverging Global Economy,* edited by Reuven Glick and Mark M. Spiegel, 129–46. Available at http://www.frbsf.org/economic-research/events/2015/november/2015-asia-economic-policy-conference/.

Wei, Shang-Jin, and Xiaobo Zhang. 2011a. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings in China." *Journal of Political Economy* 119(3): 511–64.

Wei, Shang-Jin, and Xiaobo Zhang. 2011b. "Sex Ratios, Entrepreneurship and Economic Growth in the People's Republic of China." NBER Working Paper 16800.

Wei, Shang-Jin, Xiaobo Zhang, and Yin Liu. Forthcoming. "Home Ownership as Status Competition: Some Theory and Evidence." *Journal of Development Economics.*

World Intellectual Property Organization (WIPO). 2012. *World Intellectual Property Indicators.* Geneva: WIPO.

Wu, Harry X. 2013. "Measuring and Interpreting Total Factor Productivity in Chinese Industry" [in Chinese]. *Comparative Studies* [in Chinese] 69(6). http://magazine.caixin.com/2013/cs_69/.

Wu, Harry X., and Conference Board China Center. 2014. "China's Growth and Productivity Performance Debate Revisited: Accounting for China's Sources of Growth with a New Data Set." Economics Program Working Paper no. 04-01, Conference Board. https://www.conference-board.org/pdf_free/workingpapers/EPWP1401.pdf.

Xie, Zhuan, and Xiaobo Zhang. 2015. "The Patterns of Patents in China." *China Economic Journal* 8(2): 122–42.

Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature* 49(4): 1076–1151.

Xu, Chenggang, and Xiaobo Zhang. 2009. "The Evolution of Chinese Entrepreneurial Firms: Township-Village Enterprises Revisited." IFPRI Discussion Paper 854, International Food Policy Research Institute.

Zhang, Xiaobo, Jin Yang, and Shenglin Wang. 2011. "China Has Reached the Lewis Turning Point." *China Economic Review* 22(4): 542–54.

Zhu, Xiaodong. 2012. "Understanding China's Growth: Past, Present, and Future." *Journal of Economic Perspectives* 26(4): 103–124.

A New Era of Pollution Progress in Urban China?

Siqi Zheng and Matthew E. Kahn

C hina's rapid and sustained growth since the start of its economic reforms in the late 1970s is an economic miracle, but it has also brought environmental costs in the form of air and water pollution. For example, air pollution from particulates is conventionally measured as PM_{10} or $PM_{2.5}$, which refers to particles with an aerodynamic diameter less than either 10 or 2.5 microns (commonly abbreviated as μm). In 2013, 99.6 percent of China's population was exposed to $PM_{2.5}$ air pollution levels exceeding the guidelines of the World Health Organization (Brauer et al. 2015). The Asian Development Bank reports that fewer than 1 percent of the 500 cities in China meet the air quality standards recommended by the World Health Organization, while seven of China's cities are ranked among the top ten most polluted cities in the world (Zhang and Crooks 2012).

Even at the beginning of China's economic reform in the 1980s, Chinese cities already suffered from black smoke from heavy industry and high levels of coal-burning by power plants and winter heating units. This activity created extremely high levels of acid rain pollution in southern cities (He, Huo, and Zhang 2002). But the pollution challenge grew worse after the 1980s as Premier Deng Xiaoping launched a new economic development strategy. China's air pollution problems

■ *Siqi Zheng is Samuel Tak Lee Associate Professor at the Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts; and Visiting Professor at Hang Lung Center for Real Estate, Tsinghua University, Beijing, China. Matthew E. Kahn is Professor of Economics, University of Southern California, Los Angeles, California, and a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are zhengsiqi@gmail.com and kahnme@usc.edu.*

Figure 1
Coal Consumption in China and the Rest of the World
(million short tons)
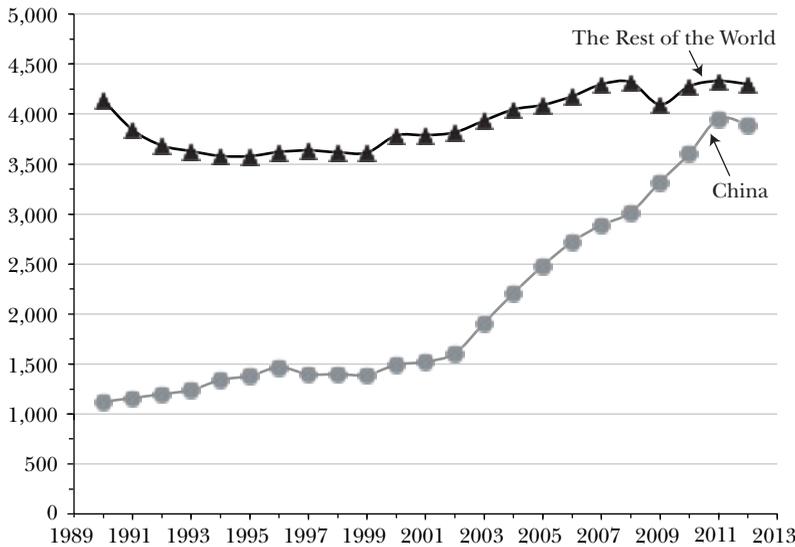
had become acute by the 21st century, when consumption of coal took off. As one vivid example of the linkage from economic growth to air pollution, China's annual consumption of coal increased from less than 700 million to almost 4 billion tons between 1980 and 2012. Figure 1 shows that in the late 1990s, China consumed about one-third as much coal as the rest of the world combined. By 2012, China was consuming almost as much coal as the rest of the world combined. Such coal burning has caused China's aggregate carbon dioxide emissions to soar. China's share of global carbon dioxide emissions has increased from 12.8 percent in 2000 to 23.6 percent in 2012.

China's combination of coal-burning and rapid industrialization has increased a number of pollutants, including sulfur dioxide and nitrogen oxides, but for simplicity we will focus much of our discussion on particulate matter (Guan et. al. 2014; Chen, Ebenstein, Greenstone, and Li 2013). In 2016, 56 percent of China's population lives in cities, and this percentage will grow over time. China's hundreds of cities differ with respect to their environmental quality. For example, Xiamen is a coastal city with relatively clean air, while Beijing has severe fog and haze: the average $PM_{2.5}$ concentrations in these two cities in the 2013–2014 winter were 45 and 108 $\mu g/m^3$, respectively. For an international historical perspective, Beijing's ambient total suspended particulates was 118 $\mu g/m^3$ in 2011, while Pittsburgh's 1960 concentration for total suspended particulates was 160 $\mu g/m^3$ (Davidson 1979).

In this essay, we begin with an overview of the broad institutional background and drivers that led to China's pattern of urban development. We look at the rise in air pollution over recent decades, and the perhaps surprising finding that in many of China's urban areas, PM_{10} levels have been decreasing during the last 10–15 years. We then turn to the costs and tradeoffs of air pollution, including costs to human health, reductions in worker productivity, and how people are seeking to reduce their exposure to pollution as shown by compensating differentials in real estate prices and purchases of masks and air filters. We discuss how rising incomes tend to raise the demand for environmental amenities, and thus increase political pressure for environmental protection, and then we turn to the policy tools that China has used to reduce pollution. We conclude by arguing that as China's government is preparing for an additional 300 million people to move to urban areas over the next 30 years, it will have a number of opportunities for China to reduce pollution through a shift from manufacturing to services, along with various steps to improve energy efficiency and resource conservation. Overall, we will argue that China is on track to improve its environmental performance in the years ahead.

The Rise of China's Highly Polluted Industrial Cities

In the early years of the Chinese Communist Party (CCP), its leaders sought to minimize transportation costs by locating factories for heavy manufacturing industries close to the Soviet Union, China's main trading partner at that time. Environmental protection goals were not prioritized. During the First Five-Year Plan, between 1953 and 1957, 156 national industrial projects were built with the help of the Soviet Union. Those projects were mainly located in the cities in the Northeast Region and the Central Region, shown by black dots in Figure 2A. The Central Region has a large endowment of coal. The Northeast Region had a better pre-existing industrial base, along with infrastructure left over from the Japanese occupation during World War II.

As foreign relations soured between China and the Soviet Union during the late 1950s and into the 1960s, China became concerned about bolstering national defense. Mao initiated the "Third Front" construction movement which included investing in 13 provinces within its core area in the Northwest and Southwest, shown as the shaded area in Figure 2A. More than 1,100 large and medium-sized projects, such as Panzhihua Iron and Steel, and Second Auto Works, were established during the Third Front period (1964–1978). Dozens of cities, such as Mianyang, Deyang, and Panzhihua in Sichuan, Guiyang in Guizhou, and Shiyan in Hubei, emerged as major industrial cities. By deciding to specialize in heavy industry while using coal as the major energy source, China's central planners chose a pollution-intensive growth policy. These initial conditions persist and influence the economic geography of China's pollution today.
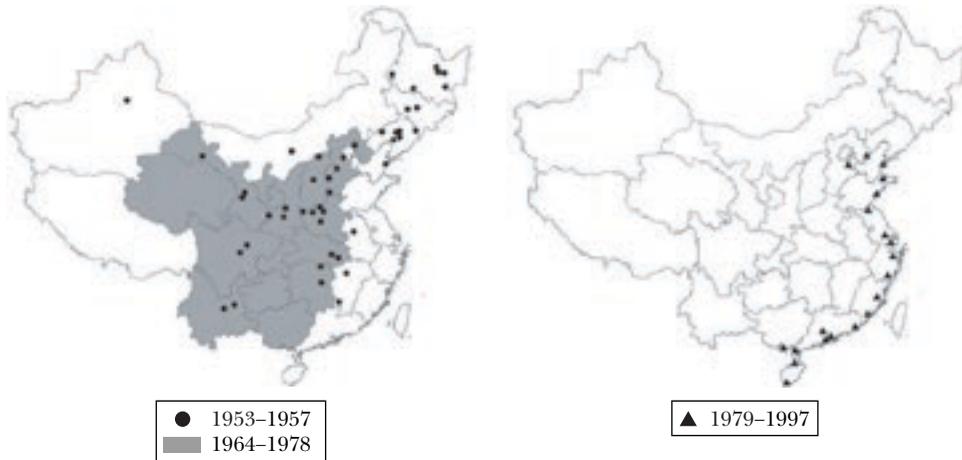
Starting in the early 1980s, China transitioned from a centrally planned economy to a market economy. Transportation costs relative to global markets came

Figure 2
The Geography of Industrial Production in China after 1949

A: Industrial Cities/Regions Developed 1953–1978   B: Industrial Cities Developed 1979–1997

● 1953–1957
▧ 1964–1978

▲ 1979–1997

Source: Authors using information from Wikipedia.

to play an increasingly important role in determining the locational choice for a profit-maximizing firm. China's coastal cities in the Eastern Region feature access to global markets and thus became the most attractive locations for private manufacturing firms (Zheng, Sun, Qi, and Kahn 2014c). The black triangles in Figure 2B represent the "coastal open cities" and "special economic zones" specified by Deng Xiaoping. In the mid-2000s, more than 65 percent of manufacturing employment in above-scale industrial firms (with annual sales above 5 million RMB) was located in the coastal region, as compared to 42 percent in 1980. The geographic concentration of this production has meant that certain industrial cities are the epicenters of pollution hot spots.

Figure 3 presents a time series of coal and steel production since the establishment of the People's Republic of China (1949–2015). Both production indicators climb steadily after the early 1980s when the economic reform was initiated and then exhibit very rapid growth after China joined the World Trade Organization in 2001. In recent years, such production has declined.

Air Pollution

During the last decade or so, China's major cities—such as Beijing and Shanghai—have experienced improvements in urban air pollution. Figure 4A shows the average particulate matter level as measured by PM_{10} in these cities between 2001 and 2013. During these years—when both cities experienced population growth and significant per-capita income growth—Beijing's PM_{10} levels have declined by 39 percent and Shanghai's by 20 percent. On the other hand, some

Figure 3
Coal and Steel Production in China (1949–2015)
(100 million tons)
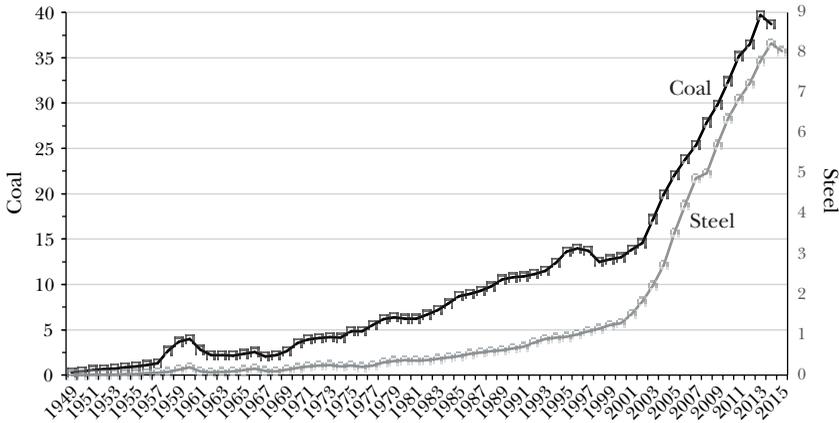
Source: China *Automotive Industry Yearbook,* China *Energy Statistical Yearbook,* and China's *Economic Statistics Bulletin.*

Figure 4
Air Pollution in Selected Chinese Cities, 2001–2013

A: PM_{10} in Beijing and Shanghai
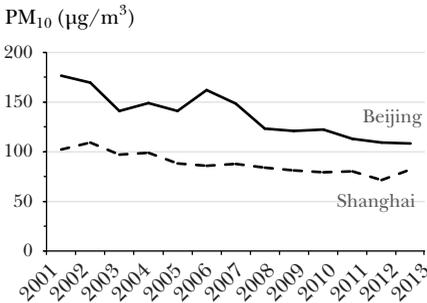
B: PM_{10} in Shijiazhuang (Hebei) and Zaozhuang (Shandong)
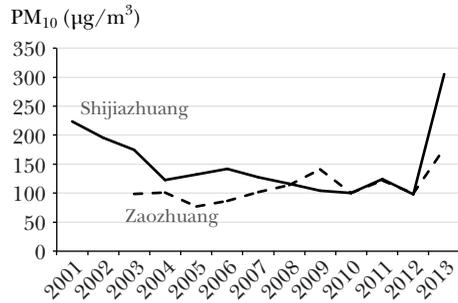
Source: The PM_{10} concentration data are calculated from the official Air Pollution Index (API) use the API calculation formula. Both API and its formula are obtained from the Data Center of the China's Ministry of Environmental Protection (http://datacenter.mep.gov.cn/).

heavy industrial cities such as Shijiazhuang in Hebei Province and Zaozhuang in Shandong Province, shown in Figure 4B, saw a rise in PM_{10}. But overall, Figure 5 presents the population-weighted average PM_{10} concentration for 85 Chinese cities, and also this measure for the two cities ranked at the 25th and 75th percentiles of the PM_{10} distribution across those cities. These three numbers have all declined during this period.

Figure 5
Chinese Urban Population PM$_{10}$ Exposure for 85 Cities
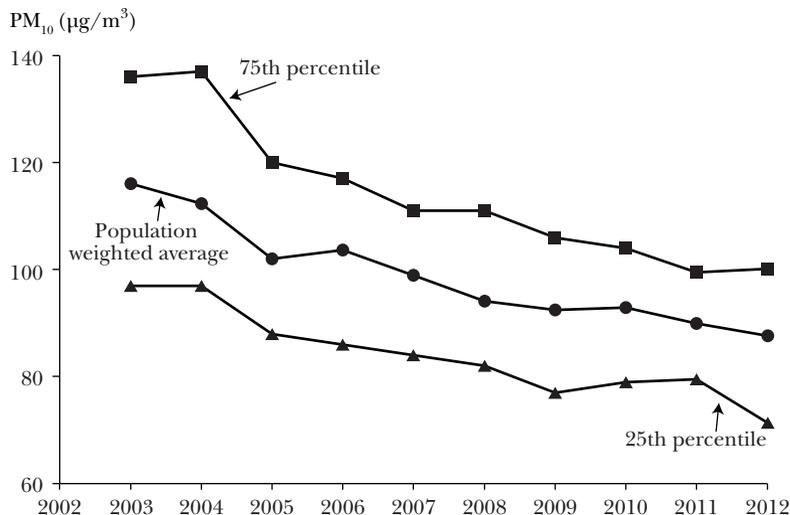
Source: Same as Figure 4.
Note: Figure 5 presents the population-weighted average PM$_{10}$ concentration for 85 Chinese cities, and also this measure for the two cities ranked at the 25th and 75th percentile of the PM$_{10}$ distribution across those cities.

In recent years, PM$_{2.5}$ has become the primary air pollutant of concern in many large Chinese cities. China's Ministry of Environmental Protection (MEP) started to report PM$_{2.5}$ readings in major Chinese cities in 2013. The US Embassy in Beijing has been releasing PM$_{2.5}$ data around its location (in downtown Beijing) since 2008. Figure 6 presents three time series—the long time series from the US Embassy in downtown Beijing, and two short time series for the average PM$_{2.5}$ readings in Beijing and Shanghai from MEP. The two time series for Beijing have similar trends in the short time period (the downtown US Embassy's reading is slightly higher than the Beijing average). Shanghai's pollution level is much lower than Beijing's. Beijing's PM$_{2.5}$ level reached its lowest level in 2008, when the Olympic Games were held in Beijing and the government implemented many short-run regulations to control the pollution. After that, the PM$_{2.5}$ concentration has averaged around 100μg/m^3, and it is worse in the winter due to coal-based heating.

Air pollution can drift across regions, just as industrial water pollution can flow downstream. In major urban cities such as Beijing and Hong Kong, much of the air pollution is caused by emissions from nearby regions. This cross-boundary spillover problem is most severe when a city is adjacent to a very dirty neighbor, such as the Hebei Province near Beijing and the Guangdong Province near Hong Kong. According to the latest report on PM$_{2.5}$ sources conducted by a research group within the Ministry of Environmental Protection, Hebei Province, which produces a significant chunk of the entire world's steel, contributed to 18 percent of the PM$_{2.5}$

Figure 6

PM$_{2.5}$ in Beijing and Shanghai

PM$_{2.5}$ (μg/m^3)
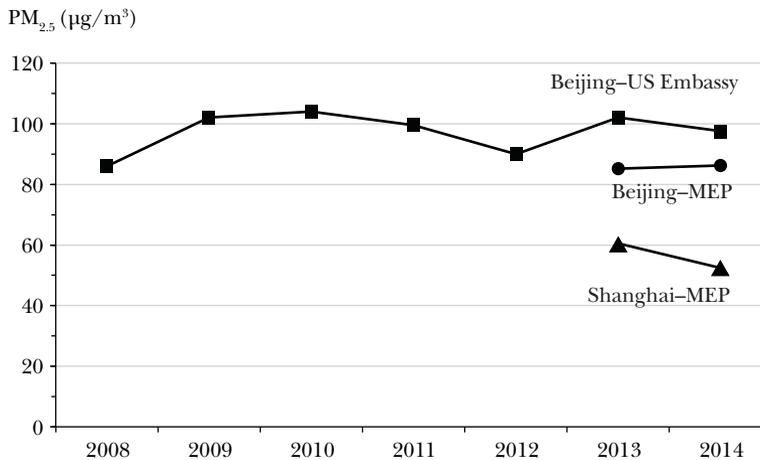
Source: The PM$_{2.5}$ concentration data are obtained from the Data Center of the China's Ministry of Environmental Protection (MEP) (http://datacenter.mep.gov.cn/) and US Embassy.

in Beijing and 20 percent in Tianjin (Ecns.cn 2016). Hong Kong has also suffered from nearby industrial emissions drifting into its airspace. In the Pearl River Delta, about two-thirds of air pollutant emissions are from industrial cities in Guangdong Province such as Dongguan and Foshan. Indeed, since 2008, Hong Kong has paid manufacturers in Guangdong Province about $150 million every year to install pollution-reducing equipment (Hong Kong Environment Bureau 2013).

Soil and Water Pollution

Although our discussion will focus mostly on air pollution, as has much of the research in this area, soil and water pollution deserve some mention. During China's rapid urbanization process, many old, dirty firms moved away from the central areas of large cities; as a result, land contamination has emerged as a new and serious issue in recent years. A World Bank study estimated that there are over 300,000 total brownfield parcels in China (Xie and Li 2010). At the same time, high levels of heavy metal pollution can be found throughout China's farmlands. It is estimated that the country loses US$3 billion per year to soil pollution, and that between 40 to 70 percent of China's soil is already contaminated with heavy metals and toxic fertilizers (as discussed in Guilford 2013). These facts highlight that a "web of nature" connects the urban economy to its natural surroundings. If polluting factories leave the city and relocate in the countryside, urbanites enjoy a reduced direct exposure to industrial pollution, but they still face the challenge of brownfield remediation and increased risk of food contamination as dirty industry moves near farmland.

Some recent research has studied water pollution along China's major rivers. Kahn, Li, and Zhao (2015) document evidence of higher pollution levels at provincial

Figure 7
Life Expectancy Trends in Three Asian Nations
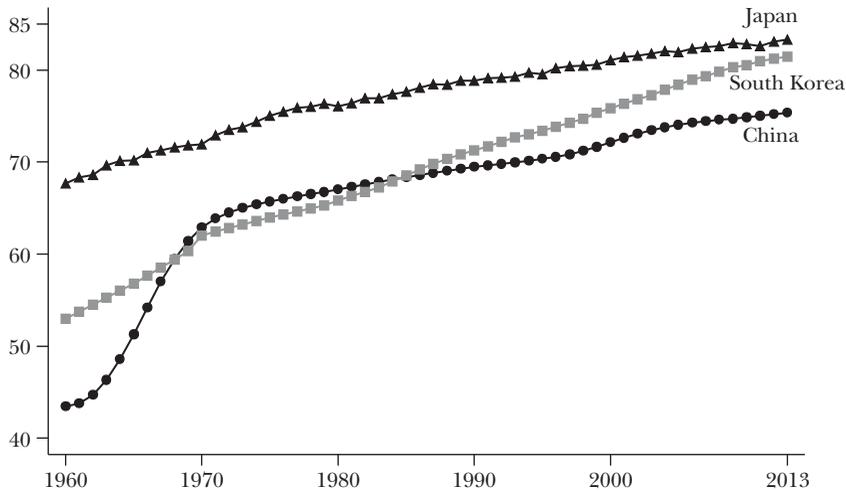
borders. Polluting industries such as pulp and paper mills cluster in these areas such that the social costs of this type of activity flow downstream. The central government recognizes the challenge of free riding at political borders (Sigman 2002) and has introduced regulations to reduce transboundary water pollution levels. Such regulations have helped to reduce pollution for targeted criteria but did not reduce such spillovers for nontargeted pollutants, like heavy metals (Kahn, Li, and Zhao 2015).

Costs and Consequences of High Pollution Levels

Elevated pollution levels have a variety of costs. Perhaps most obvious and severe are the immediate costs to human health. But in addition, pollution reduces worker productivity, alters real estate values, and affects the local quality of life, and it has contributed to global environmental concerns like climate change.

The Health Costs of China's Pollution

Life expectancy trends for China, Japan, and South Korea from 1960 to 2013 are reported in Figure 7. During the 1960s, China's life expectancy converged with South Korea's. At that time, many "barefoot doctors," trained with basic medical knowledge, were sent to work in rural areas. That program effectively reduced the high mortality of newborn infants and the morbidity of contagious diseases. But then, starting in the late 1980s, the two nation's life expectancies diverged, and in recent years South Korea's life expectancy has converged with Japan's.

The relatively flat profile for China's life expectancy during a time of sharp growth after the 1980s hints that some factor, like rising pollution levels, may be hindering life expectancy. Ebenstein et al. (2015) use several datasets to document that China's growth in life expectancy is less than what would be predicted given the nation's per capita income growth. They study the relationship between income, pollution, and mortality in China from 1991–2012. They document a positive association between city-level GDP and life expectancy. They also find a negative association between city-level particulate air pollution exposure and life expectancy that is driven by elevated cardiorespiratory mortality rates: for example, a 100 $\mu g/m^3$ increase in PM_{10} exposure is associated with a decline in life expectancy of 1.5 years at birth and 2.3 years at age five. Hanlon and Tian (2015) use cross-city data in modern China and in late 19th-century England to study the relationship between polluting industry agglomeration and mortality. In both settings, there is a positive correlation between heavy industry and death risk, although modern China's gradient is less steep than the relationship found in historical England.

There are other studies on the relationship between air pollution and life expectancy in China. Chen et al. (2013) find that China's Huai River policy, which provides free winter heating via the provision of coal for boilers in cities north of the Huai River but denies heat to the south, results in about 184 $\mu g/m^3$ higher ambient concentrations of total suspended particulates in the north. This pollution exposure is associated with a reduction in average life expectancy of 5.5 years in the North, due to an increased incidence of cardiorespiratory mortality.

The health costs of air pollution is also directly reflected in mortality and morbidity. Aunan and Pan (2004) find that an increase of one $\mu g/m^3$ in PM_{10} and sulfur dioxide will result in a 0.03 and 0.04 percent increase in all-cause mortality, respectively, along with a 0.07 and 0.19 percent increase in hospital admissions due to cardiovascular diseases, respectively, while the coefficients for hospital admissions due to respiratory diseases are 0.12 and 0.15 percent, respectively. The impact of long-term PM_{10} levels on the prevalence of chronic respiratory symptoms and diseases per $\mu g/m^3$ are a 0.31 percent increase in adults and 0.44 percent in children.

Such health damages caused by pollution in China have created a substantial burden for its economy. According to Matus et.al. (2012), the ozone and particulate matter concentrations beyond background levels have led to a loss of US$16 billion to US$69 billion in consumption (or 7 to 23 percent) in 2005. In another study, Kan and Chen (2004) find that the total economic cost of health impacts due to particulate air pollution in Shanghai in 2001 was approximately $625 million in US dollars, accounting for 1 percent of the gross domestic product of the city. Wang and Mauzerall (2006) estimate that the health costs due to year 2000 anthropogenic emissions in Zaozhuang, a city in eastern China heavily dependent on coal, are 10 percent of that city's GDP.

The costs imposed by ambient air pollution can be reduced by investing in self-protection. Chinese urbanites can choose where to live in a city and what private

self-precautions to take, such as purchasing masks or air filters, and how much time to spend outside.

These adaptation strategies have increased in importance over time in China. In the past, China's domestic passport or *hukou* system limited the ability of individuals to vote with their feet and move across cities or from rural to urban areas. Within a city, firms provided cheap housing close to the workplace to minimize commutes. Moreover, there was not a strong understanding of how pollution affects health and productivity, and there were few self-protection products that a concerned susceptible person could purchase. But in modern urban China, urbanites have at least some control over their air pollution exposure through location decisions and self-protection measures, and the urban wealthy in particular may be able to reduce their exposure substantially (Kahn and Zheng 2016).

Effects of Pollution on Worker Productivity and Local Growth

In an open system of cities that compete for workers and firms, geographic pollution hot spots can influence the spatial distribution of firms, workers, real estate prices, and wages. Hanlon (2016) develops a spatial equilibrium theory incorporating local pollution effects that can explain how the evolution of city pollution levels, driven by local industrial growth, can affect overall local economic development. He applies the theory to study the cities of 19th century Britain, where some cities specialized in heavy coal-using industries such as iron and steel production, resulting in very high levels of local pollution. As these industries grew (because of world demand), the cities that specialized in such dirty industries experienced a rise in local coal use and hence in local air pollution. Hanlon shows that the increase in local pollution acted as a substantial drag on local employment growth. Industries in more polluted cities had to pay higher "combat pay" to attract workers at the same time that the elevated pollution reduced worker and firm productivity. A study in China also shows that higher levels of air pollution decrease worker productivity by reducing the number of calls that workers complete each day at call centers (Chang, Zivin, Gross, and Neidell 2016).

The spatial equilibrium dynamics described in Hanlon (2016) did not play out in Mao's era because of the binding *hukou* system in China, in which citizens were designated as living in a certain area and not allowed to move without official permission. With the de facto relaxation of the *hukou* system and the liberalization of the labor and land markets starting in the 1980s, Chinese urbanites have more choice over where to work and live, both within and across cities. Given that Beijing is the nation's capital city, it continues to attract highly skilled workers and floating workers because its overall bundle of opportunities and benefits exceeds the disamenity costs of its relatively high local pollution. Now that China's cities form an open system, Hanlon's (2016) model offers relevant predictions. The major heavy industrial cities Mao established in the Northeast Region and the "Third Front Region" are now lagging behind because they feature dirty and declining industries and cannot attract high-skilled clean industries.

How Households Seek to Reduce their Pollution Exposure

With the relaxation of the *hukou* system and the liberalization of the labor and land markets, people are also able to "vote with their feet." The theory of compensating differentials yields insights about the pricing of real estate across cities and the resulting spatial allocation of different households and firms across China's cities (Rosen 1979; Roback 1982). Using standard revealed preference methods, several studies have documented the rising demand for environmental quality in China's cities.

In a compensating differentials study (Zheng and Kahn 2008), we find that all else equal, a 10 microgram per cubic meter increase in PM_{10} particulate pollution reduces home prices by 4.1 percent in Beijing. In an intercity study of 35 Chinese large cities (Zheng, Kahn, and Liu 2010), we find that home prices are lower in cities with higher ambient pollution levels, and the marginal valuation for green amenities is rising over time. In Zheng, Cao, Kahn, and Sun (2014a), we exploit the fact that the particulate matters that are imported into a city depend on the dominant wind direction and emissions from nearby cities, and the sandstorms from Inner Mongolia. Using wind and sandstorms as instrumental variables, we find that on average, a 10 percent decrease in imported neighbor pollution is associated with a 0.76 percent increase in local home prices. For such cross-city variation in disamenities to be capitalized into real estate prices requires that migrants both are aware of the differences in local public good quality and that such differentials influence their locational decisions.

Another method that households can use to reduce air pollution exposure is to purchase masks and air pollution filters. Using a dataset of Internet purchases in 35 Chinese cities, in Sun, Kahn, and Zheng (forthcoming), we document that Chinese households invest more in masks and air filter products when ambient pollution levels exceed key alert thresholds. Those with higher incomes are more likely to invest in air filters, which are much more expensive than masks but also more effective. This finding suggests that richer people are exposed to less air pollution than poorer people. When the outdoor air is polluted, people will prefer to drive private cars (rather than walk on the street), and they will also decrease their time spent outdoors (Neidell 2009). Based on the estimate in Chen and Zhao (2011), the indoor concentration of particulate matter is on average about 80 percent of the outdoor concentration in Chinese cities, so people can breathe less-polluted air when they are indoors on polluted days. More-educated workers are more likely to work in such indoor jobs.

Global Environmental Costs

Although our primary focus here is on costs experienced by Chinese from China's pollution, we should mention that China's contribution to greenhouse gas emissions is now the largest of any country. We are not aware of a study that puts the China-specific costs in context. China is now the world's leading greenhouse gas producer. However, Muller, Mendelsohn, and Nordhaus (2011) estimate that the global external damage for US petroleum and coal product manufacturing is 35 percent of the industry's value added. Also, China's pollution has imposed costs on nearby nations such as South Korea (Baek, Altindag, and Mocan 2015; Jia and Ku 2016).

Higher Incomes, Greater Awareness, and Pressures to Reduce Pollution

A number of economic arguments suggest that as incomes rise, demand for environmental protection also rises. As an example of US-based evidence on this subject, Costa and Kahn (2004) use data on wages and risks of fatalities on a range of jobs, using Census data from 1940 to 1980. They estimate the "value of a statistical life"—in this case, the amount that workers needed to be compensated for taking on riskier jobs. They estimate value of life with respect to per capita GNP in the range of 1.5 to 1.7; that is, the value of a statistical life rises faster than national per-capita income.

An array of evidence suggests that this pattern is occurring in China, as well. As one example, the evidence in the previous section about household spending to mitigate exposure to pollution, as shown by compensating differentials in real estate prices and by spending on masks and filters, clearly shows that as incomes rise, people are seeking a lower level of pollution.

Perhaps the best-known relationship between income and pollution is the "environmental Kuznets curve," which represents a reduced form relationship between a geographic location's pollution and per-capita income (Grossman and Krueger 1995; Harbaugh, Levinson, and Wilson 2002). In its simplest form, this curve posits that a location's pollution level is an increasing but concave function of local per-capita income: that is, as income rises, pollution first rises, but then falls. The earlier evidence that particulate matter concentrations have been falling in China's urban areas, even as economic growth has continued and coal consumption has risen dramatically, also offers a prima facie case that higher incomes are accompanied by pressure for lower pollution levels.

We investigated the environmental Kuznets curve in China more systematically using data for 83 cities for the years 2003 to 2012. For our dependent variable, we used the city's air pollution level as measured by the PM_{10} concentration in that year. For explanatory variables we used the log of per capita income, as well as the log of per capita income squared and cubed to allow curvature in the result. As other control variables, we also used population for the urban area, share of manufacturing in total output, average years of schooling, rainfall and a temperature index, longitude and latitude, and then in different specifications either a time trend or dummy variables for years.[1]

Table 1 reports our regression results. The first and second columns are the same except that one uses time trend and the other uses dummy variables for years. The results are quite similar. Chinese cities featuring a larger population and a

---

[1] City-level variables are obtained from the China *Statistic Yearbooks* and the China *City Statistical Yearbooks*. The variables include; GDP (adjusted by inflation), city population (nonagricultural population size), the employment share of manufacturing industry, average years of schooling in 2000, annual rainfall in 2007, and the temperature discomfort index in 2007. See Zheng, Kahn, and Liu (2010) for the definitions and descriptive statistics of these variables.

Table 1
Correlates of Urban Air Pollution in China

			Dependent Variable: $\log(PM_{10})$			
Sample	*All*	*All*	*Above-median-educated cities*	*Below-median-educated cities*	*2003–2007*	*2008–2012*
Variables	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
$\log(GDP\ per\ capita)$	–0.434***	–0.424***	–0.161	–0.616***	–0.524***	0.0505
	(–3.37)	(–3.32)	(–0.31)	(–3.30)	(–3.03)	(0.23)
$(\log(GDP\ per\ capita))^2$	0.300***	0.296***	0.137	0.392***	0.316***	0.0944
	(4.02)	(4.00)	(0.57)	(2.69)	(2.99)	(0.81)
$(\log(GDP\ per\ capita))^3$	–0.0596***	–0.0592***	–0.0355	–0.0641*	–0.0586***	–0.0331*
	(–4.42)	(–4.42)	(–0.98)	(–1.89)	(–2.97)	(–1.68)
$\log(population)$	0.164***	0.164***	0.248***	0.0812***	0.169***	0.161***
	(11.95)	(12.01)	(13.62)	(3.64)	(7.77)	(10.09)
$\log(manufacturing\ share)$	0.0498	0.0450	0.0387	0.0410	0.115*	–0.00172
	(1.25)	(1.14)	(0.71)	(0.65)	(1.71)	(–0.04)
$\log(average\ years\ of\ schooling)$	–0.918***	–0.926***	–1.883***	–0.606**	–0.876***	–0.937***
	(–6.43)	(–6.53)	(–7.18)	(–2.11)	(–3.86)	(–5.64)
$\log(rainfall)$	–0.0987***	–0.0977***	0.0554	–0.113*	–0.187***	–0.000664
	(–2.84)	(–2.83)	(1.20)	(–1.93)	(–3.40)	(–0.02)
$\log(temperature\ index)$	0.391***	0.394***	0.672***	0.208**	0.431***	0.330***
	(5.29)	(5.37)	(5.25)	(2.16)	(3.62)	(3.87)
T	–0.0316***					
	(–10.11)					
Year dummies	No	Yes	Yes	Yes	Yes	Yes
Constant	–1.095**	–1.252***	–0.353	–1.020	–0.929	–2.011***
	(–2.47)	(–2.86)	(–0.44)	(–1.22)	(–1.31)	(–3.85)
Turning point (thousand RMB in 2012)	100	99	53	204	100	85
Observations	846	846	428	418	421	425
F-statistic for joint significance of income polynomial	7.25***	7.35***	6.47***	9.17***	3.52**	8.72***
R^2	0.432	0.444	0.568	0.428	0.440	0.443

Note: The latitude and longitude of each city are controlled for in all columns.

higher manufacturing share of total employment are more polluted. The negative time trend during the years 2003–2012 from column 1 highlights the potential for technological change to reduce pollution. Based on our regression estimates, we find that wealthier cities are reaching the turning point at about 100,000 yuan (about US\$15,000) in terms of GDP per capita (2012 constant RMB).

We partition the observations into higher-educated cities and lower-educated ones (based on the average years of schooling) in columns 3 and 4, and then into the early period (2003–2007) and late period (2008–2012) in columns 5 and 6. The higher-educated cities have earlier turning points than the set of cities whose education is below the cross-city median, which is consistent with the hypothesis that citizens with a higher education level will tend to be more concerned about environmental costs and will tend to strategically locate, or exert political pressure, to experience lower environmental costs. Comparing the results in columns 5 and 6, we find that cities are reaching the environmental Kuznets curve turning point

at a lower level of per-capita income in the later period, which is consistent with a hypothesis that concern over environmental issues is rising over time.

This latter finding is consistent with the claim of Dasgupta, Laplante, Wang, and Wheeler (2002, in this journal) that the environmental Kuznets curve shifts down and in over time. This claim means that, over time, geographic areas suffer less environmental damage as they develop and that they reach the per-capita turning point earlier. Endogenous technological change and the diffusion of cleaner technology provides one micro-foundation for this observed pattern (Acemoglu, Aghion, Bursztyn, and Hemous 2012; Acemoglu, Akcigit, Hanley, and Kerr 2016).

An advantage of the kind of approach shown in Table 1 is that it emphasizes within-China variation in ambient air pollution production and exposure. The cross-national research on an environmental Kuznets curve implicitly averages the cities within a country. This removes the within-nation urban variation that arises due to the sorting of heterogeneous industries and workers and the policies implemented by local leaders.[2]

A Suite of Environmental Protection Policies

Starting in the early 2000s, the Chinese central government has increasingly emphasized pollution reduction, including climate change mitigation. China's five-year plans provide blueprints that foreshadow the medium-term goals of the Chinese Communist Party. We counted the words related to environment, energy, and ecology in each of the 13 five-year plans the central government made after 1949. The content of the five-year plans sends clear signals to local officials about what goals they should prioritize. The first three five-year plans had almost no mention of environmentally oriented terms; indeed, the 3rd Five-Year Plans had literally no mention of these terms. From the 5th to the 9th Five-Year Plan, from 1.8–2.5 percent of the words were related to environmental issues. In the 11th through 13th Five-Year Plans—the most recent ones—about 5 percent of the language is devoted to environmental issues. We view this simple word count as a starting point. Future work could use more machine-trained semantic analysis algorithms from computational linguistics to quantify the details and trends of those "green" words in the five-year plans and other Chinese government documents. According to the

---

[2]We recognize that our within-China evidence shown in Table 1 concerning a cross-city environmental Kuznets curve does not test the claim that China is on a more benign curve than the rest of Asia. To test this claim, one would have to use cross-national panel data for each Asian nation j at time t and run a regression of the form

$$Pollution_{jt} = B_j + D_t + f(income_{jt}) + B \times China_j \times time\ trend_t + U_{jt}$$

One test of whether China is on a different pollution/income path than the rest of Asia would be if, controlling for nation fixed effects, year fixed effects, and a per-capita income polynomial, B is less than zero where B is the coefficient on the China specific time trend. Given data limitations, we cannot estimate this regression.

statistics released by Chinese government, the 11th FYP met or exceeded all of the stated goals, including an energy intensity reduction target (Price et al. 2011; see also Casey and Koleski 2011). By 2014, the energy intensity reduction number set in the 12th FYP (2011–15) had been 82.5 percent fulfilled. At the political level, one motivation for the Chinese Communist Party to emphasize "green progress" is that the Party seeks to build legitimacy by signaling to both domestic constituents and international actors that it cares about the quality of life for everyone (Wang 2013).

China's government has a suite of policy tools that it has been using to improve environmental quality: command and control, incentives for local officials, reducing subsidies for energy and water, and direct investments in pollution mitigation.

Command and Control Regulations

The Chinese Communist Party has often relied on command-and-control regulation to achieve its urban externality mitigation goals. Examples include shutting down firms, imposing driving restrictions, and mandating car registration lotteries. Based on daily data from multiple monitoring stations, Viard and Fu (2015) find that, air pollution falls 19 percent during an every-other-day driving restriction (during the 2008 Beijing Olympics) and 8 percent during one-day-per-week restriction (October 2008 to December 2009).

In contrast, a study by Sun, Zheng, and Wang (2014) finds that Beijing's driving restriction has some effect on mitigating road congestion but an insignificant effect on reducing air pollution. Li (2015) compares the two types of car license allocation mechanisms in China's major cities. To combat worsening traffic congestion and urban pollution, Beijing requires that potential car buyers participate in a random lottery, while Shanghai auctions vehicle registration licenses. In both cities, the licenses obtained are nontransferable. Li (2015) finds that the nontransferable lottery is inefficient compared to an auction mechanism. This within-nation variation in policy raises the possibility of social learning as local experimentation offers new insights about the effectiveness of different policies. As other cities learn, they have an option to adopt policies that have achieved environmental improvements in other Chinese cities.

Incentives for Local Officials

China has a strong one-party central government, but hundreds of local governments act as competing enterprises. Upper-level governments promote or demote local leaders based on performance evaluation. Such a political tournament raises the possibility that the central government can provide strong incentives for local officials, and indeed, this has been a tool for encouraging local officials to promote economic growth in the past. However, the central government has been changing the performance evaluation and promotion criteria for local officials, moving beyond purely output-based criteria to include more environmental goals in the performance metrics (Zheng, Khan, Sun, and Luo 2014b; Kahn, Li, and Zhao 2015).

We have studied the promotion propensities for mayors in 83 Chinese cities during the years 2004 to 2010. We find that local GDP growth continues to be the

most important factor in determining promotion rates, but that declines in local air pollution and reductions in local industry energy intensity are also statistically significant correlates of promotion chances, especially in relatively richer cities on the east coast (Zheng et al. 2014b).

Chen, Li, and Lu (2016) evaluate the effectiveness of China's "Two Control Zones" (TCZ) policy. This policy intends to reduce sulfur dioxide levels in targeted cities. They employ a difference-in-differences econometric approach and find that, compared to the control group (non-TCZ cities), local bureaucrats in TCZ cities exerted more effort to reduce sulfur dioxide emissions after the emissions quota was built into their performance evaluation in 2006. Local officials appear to devote more effort to reducing pollution when their performance evaluation is partially based on environmental performance.

Local mayors, especially those in richer cities with a more educated populace, face pressure from both the central government and the public who are increasingly expressing their pollution concerns (Kahn and Zheng 2016). Mayors of poorer cities are aware of their city's industrial structure and the need for sheer economic growth. In our interviews with these mayors, some of them admit that they are implementing policies targeted to increase industrial production, and they recognize that this will increase local pollution levels.

The two cities of Dalian and Tangshan offer an example of how these dynamics can unfold. Both cities are on the coastal line of the Bohai Sea and have similar climate conditions, but they implemented quite different development strategies. Tangshan is rich in iron and close to coal mines, so it chose to develop energy-intensive industries such as steel, construction materials, and chemicals. In 1984, Dalian was selected to be one of the 14 coastal "open cities" by Deng Xiaoping and thus attracted many high-tech firms. Dalian also regards itself as a tourist city, so a clean environment is viewed as offering direct economic benefits.

As China invests in bullet trains, the Chinese people now have the opportunity to decentralize and live in nearby second-tier cities if the megacities of Beijing and Shanghai suffer from low quality of life. For example, the construction of China's high-speed rail system means that nearby second-tier cities such as Tianjin (40 miles from Beijing) offer a type of safety valve if quality of life declines in China's megacities. Transport innovation has been associated with rising real estate prices in secondary cities near first-tier cities such as Shanghai and Beijing (Zheng and Kahn 2013).

Reducing Price Distortions

China has removed the most egregious distortions and subsidies in its energy sector: for example, prices for electricity, gasoline, and natural gas generally reflect costs over the long run, and some are at or even above international market levels (World Bank and DRC 2014).

For gasoline, oil prices were under tight state control and were set very low before 1998. With rising dependence on oil imports, China had to adjust its domestic oil prices. In June 1998, China reformed the oil pricing system to set domestic oil price in accordance with the global oil price (Hang and Tu 2007). Since then,

China's National Development and Reform Commission (NDRC), which regulates petroleum prices, has used international oil prices as the benchmark for domestic prices. Price subsidies offered during the early years contributed to sharp increases in energy consumption in the 1990s before falling for the first time in 1998 when energy reforms began to take effect (Poon, Casas, and He 2006). The pricing system has adjusted several times since 1998, and the current policy was implemented since 2009. Gas prices in China are no longer low. In April 2016, the price of gasoline in China was roughly 50 percent higher than in the United States.

However, "economic and social stability" is also considered by the NDRC when it adjusts gas prices. If the international benchmark price exceeds $130 a barrel, diesel and gasoline prices are not raised or are raised only by a small margin (World Bank and DRC 2014). When the central government raises gasoline prices, it faces trade-offs between interest groups. The fossil fuel market is dominated by three major state-owned enterprises—PetroChina, SINOPEC, and CNOOC—that would benefit from gasoline price increases. But the public will be angry and may protest against such price increases, which is a potential threat to social stability. Studying the consequences of this tension between interest groups in determining the enforcement of Chinese environmental regulations remains an open research topic.

For electricity, industrial users pay around RMB 0.70 ($0.10) per kilowatt-hour on average ($0.17 per kilowatt-hour on a purchasing power parity basis), while the average rate in the OECD countries is $0.11 per kilowatt-hour in 2010 (IEA 2012). However, residential electricity prices in China are quite low compared with many developed countries. For example, in Beijing the rate was $0.08 in 2011, which was about one-third of that in New York ($0.20) and even lower in comparison to Berlin or Copenhagen ($0.40) (World Bank and DRC 2014). The cross-subsidy from industrial to residential electricity use in China has been widely criticized.

Residential water prices in Chinese cities are quite low by international standards: in Beijing the residential tariffs ($0.54 in US dollars per cubic meter) are less than one-tenth of those in Berlin or Copenhagen. However, the current industrial water tariff in Beijing (6.21 RMB) and Tianjin (7.85 RMB) are higher than those in Canada, the United States, and other developed countries (World Bank and DRC, 2014).

Direct Government Investment

In 2016, the general public budget of China's Ministry of Environmental Protection (MEP) increased by 18 percent over the previous year, from 3.2 billion RMB in 2015 to 3.8 billion RMB in 2016.[3] The final budget in 2015 indicates that 63 percent of the general public expenditure was on "energy saving and environment protection," such as purchasing equipment to construct the national environmental monitoring web, hiring workers to supervise the enforcement of environmental laws, investigating and evaluating countrywide underground water environment conditions, developing educational activities on environment

[3] The source is in the Chinese language, at the Ministry of Environmental Protection webpage http://www.mep.gov.cn/xxgk/zwgk_1/czzj.

protection, and other activities. The other 32 percent of the expenditure was spent on "science and technology, including funding public scientific research and major science and technology projects on environment protection.

The central government provided a 10.6 RMB billion special fund to support air pollution control in key areas including the Beijing–Tianjin–Hebei area, Yangtze River Delta, and Pearl River Delta. With the funding, a number of new regulatory actions have been taken. For example, 1.26 million "yellow label cars"—the name given to heavy-polluting vehicles—were phased out. The funding was used to support strict inspection on road and gas quality, and subsidies to those car owners to upgrade to lower-polluting cars. Subsidies were also given in rural areas to encourage a switch from burning crop straw to using (cleaner) coal.

## Future Environmental Progress

China's future environmental progress in reducing air pollution can be viewed as a race. On the one side, even if China's economy does not grow at the fast pace of the last couple of decades, it seems plausible that the economy could continue growing at 6 percent per year. On the other side, even with a growing economy, technological progress and industrial composition shifts could help to offset the amount of pollution emitted and its effects. Over the medium term, China should be able to reduce both its energy intensity (energy use/GDP) and its pollution intensity (emissions/GDP).

The ratio of energy use/GDP in China has declined dramatically over the last 25 years, as shown in Figure 8. During the last decade, it dropped by nearly 25 percent, reaching close to the level of the whole world and South Korea by the year of 2013.

The relative size of China's manufacturing and construction industries is shrinking—its share in GDP fell from 45.4 percent in 2000 to 40.5 percent in 2015, while the size of what is referred to as "tertiary industry" increased from 39.8 percent to 50.5 percent. China's share of power generated from coal decreased from 80.3 percent in 2000 to 73 percent as of 2015. Indeed, China's overall coal use declined 3.7 percent in 2015, which followed a decline of 2.9 percent in 2014 (as reported in Yeo 2016). This reduction took place during a time that China's economy grew by 6–7 percent annually. Green and Stern (2015) predict that China's coal use has peaked and is likely to plateau over the next five years and to continue to decline. As China's cities transition from industry to cleaner services, and as more power is produced from cleaner energy sources, ambient air and water pollution is likely to decline.

As hundreds of millions of Chinese move to new cities over the next decades, there is an opportunity to take advantage of cutting-edge technology in building energy-efficient real estate, transportation systems, and power generation infrastructure. Energy consumption in buildings accounts for about 30 percent of China's total energy use, but China has a program for green building certification since 2008, and the central government started to offer subsidies for green buildings in 2012. China became the largest producer of solar photovoltaic cells in the world in

Figure 8
Energy Use per Unit of GDP in China, South Korea, and World
(kg of oil equivalent per constant 2011 PPP $1000)
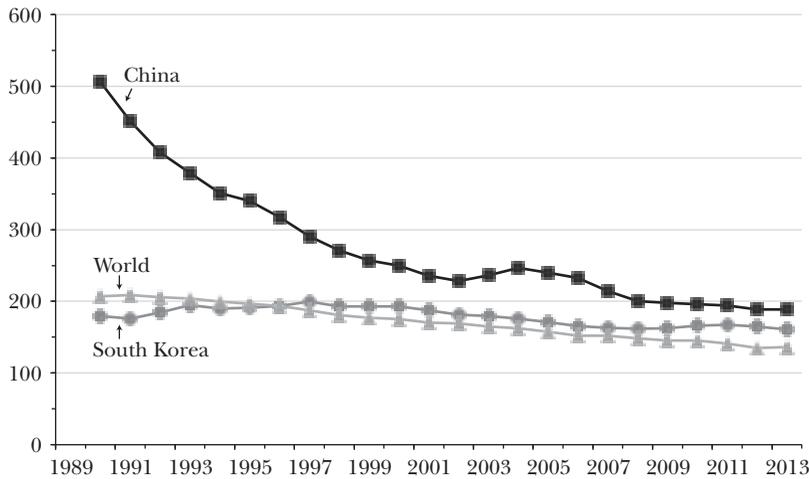
Source: The World Bank Data, http://data.worldbank.org.

2007 (as reported in PR Newswire 2015), and in 2015, China overtook Germany as the nation with the largest installed photovoltaic capacity (as reported in Hill 2016). China is aiming to increase its non-fossil-fuel energy capacity to 15 percent of total primary energy consumption.

In September 2016, China ratified the Paris Climate Change Agreement. It has promised that its greenhouse gas emissions will peak by the year 2030. It further has pledged to reduce its carbon emissions per unit of GDP by 60 percent from its 2005 level by 2030 and to increase its share of power generated by renewables to 20 percent. As the world's high-income nations invest in green technology, China will be able to import new technologies. The investments in basic research at China's universities suggest the possibility that China could become a producer and exporter of green technology innovation (Freeman and Huang 2015). Several promising trends, especially in China's rich coastal cities, are now unfolding that suggest significant environmental progress could take place. Manufacturing is land-intensive. As urban rents rise, such industrial activity is leaving the major cities. In addition, the inland cities have a cost advantage in electricity price. Given these push factors, the second-tier and third-tier inland cities have become increasingly attractive destinations for labor- and energy-intensive industries (Zheng, Sun, Qi, and Kahn 2014c). During a time of rapid urbanization, environmental progress can occur if a city's industrial composition shifts from heavy industries to cleaner services (Kahn 1999).

Taken together, these political and economic factors strongly suggest the possibility that China's pollution per unit of economic activity could sharply decline,

but this dynamic may unfold unevenly across the country. China's leaders face the challenge that macroeconomic growth is slowing. The government has responded with strategic subsidies to foster profitability of heavy industries that employ many low-skill laborers. If the Chinese Communist Party faces a strict "jobs versus the environment" tradeoff, it is likely to choose protecting jobs. Poorer cities in China's western region, such as Baotou and Panzhihua, continue to rely on heavy industry as the area's major employer. In contrast, the rich coastal cities such as Shanghai feature a cleaner set of industries and high-human-capital workers who seek out clean air, both for the amenity value but also to raise their own and their children's productivity (Zivin and Neidell 2013). This "tale of two cities" could give rise to increasing environmental inequality across China's regions.

References

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The Environment and Directed Technical Change." *American Economic Review* 102(1): 131–66.

Acemoglu, Daron Ufuk Akcigit, Douglas Hanley, and William Kerr. 2016. "Transition to Clean Technology." *Journal of Political Economy* 124(1): 52–104.

Aunan, Kristin, and Xiao-Chuan Pan. 2004. "Exposure-Response Functions for Health Effects of Ambient Air Pollution Applicable for China—A Meta-analysis." *Science of the Total Environment* 329(1–3): 3–16.

Baek, Deokrye, Duha T. Altindag, and Naci Mocan. 2015. "Chinese Yellow Dust and Korean Infant Health." NBER Working Paper 21613.

Brauer, Michael, et al. 2015 "Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013." *Environmental Science & Technology* 50(1): 79–88.

Casey, Joseph, and Katherine Koleski. 2011. *Backgrounder: China's 12th Five-year Plan.* US–China Economic and Security Review Commission. Available at: http://www.uscc.gov/Research/backgrounder-china%E2%80%99s-12th-five-year-plan.

Chang, Tom, Joshua Graff Zivin, Tal Gross, Matthew Neidell. 2016. "The Effect of Pollution on Worker Productivity: Evidence from Call-Center Workers in China." NBER Working Paper 22328.

Chen, Chun, and Bin Zhao. 2011. "Review of Relationship between Indoor and Outdoor Particles: I/O ratio, Infiltration Factor and Penetration Factor." *Atmospheric Environment* 45(2): 275–88.

Chen, Yuyu, Avraham, Ebenstein, Michael Greenstone, and Hongbin Li. 2013. "Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy." *PNAS* 110(32): 12936–41.

Chen, Yvonne Jie, Pei Li, and Yi Lu. 2015. "Accountability, Career Incentives, and Pollution: The Case of Two Control Zones in China." Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2703239.

Costa, Dora L., and Matthew E. Kahn. 2004. "Changes in the Value of Life, 1940–1980." *Journal of Risk and Uncertainty* 29(2): 159–80.

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler. 2002. "Confronting the Environmental Kuznets Curve." *Journal of Economic*

Perspectives 16(1): 147–68.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, Peng Yin, and Maigeng Zhou. 2015. "Growth, Pollution, and Life Expectancy: China from 1991–2012." *American Economic Review* 105(5): 226–31.

Ecns.cn. 2016. "A Victim Itself, Hebei Contributes to 18% of Beijing PM2.5." *ECNS Wire*, July 11. Editor: Mo Hong'e. http://www.ecns.cn/cns-wire/2016/07-11/217681.shtml.

Freeman Richard B., and Wei Huang. 2015. "China's 'Great Leap Forward' in Science and Engineering." NBER Working Paper 21081.

Global Water Intelligence. 2011. "Global Water Market 2011—Meeting the World's Water and Wastewater Needs until 2018." http://www.giiresearch.com/report/gwi108050-lobwater-wastew.html (accessed December 2013).

Green, Fergus, and Nicholas Stern. 2015. "China's 'New Normal': Structural Change, Better Growth, and Peak Emissions." *Policy Brief,* Centre for Climate Change Economics and Policy (CCCEP), University of Leeds, June.

Grossman, Gene M., and Alan B. Krueger. 1995. "Economic Growth and the Environment." *Quarterly Journal of Economics* 110(2): 353–78.

Guan, Dabo, Xin Su, Qiang Zhang, Glen P. Peters, Zhu Liu, Yu Lei, and Kebin He. 2014. "The Socioeconomic Drivers of China's Primary $PM_{2.5}$ Emissions." *Environmental Research Letters* 9(2): 024010.

Gwynn Guilford. 2013. "Half of the Rice in Guangzhou Is Polluted." *Atlantic,* May 21. http://www.theatlantic.com/china/archive/2013/05/half-of-the-rice-in-guangzhou-is-polluted/276098/.

Hang, Leiming, and Meizeng Tu. 2007. "The Impacts of Energy Prices on Energy Intensity: Evidence from China." *Energy Policy* 35(5): 2978–88.

Hanlon, W. Walker. 2016. "Coal Smoke and the Costs of the Industrial Revolution." http://www.econ.ucla.edu/whanlon/papers/hanlon_coal_and_the_costs_of_the_IR.pdf.

Hanlon, W. Walker, and Yuan Tian. 2015. "Killer Cities: Past and Present." *American Economic Review* 105(5): 570–75.

Harbaugh, William T., Arik Levinson, and David Molloy Wilson. 2002. "Reexamining the Empirical Evidence for an Environmental Kuznets Curve." *Review of Economics and Statistics* 84(3): 541–51.

He, Kebin, Hong Huo, and Qiang Zhang. 2002. "Urban Air Pollution in China: Current Status, Characteristics, and Progress." *Annual Review of Energy and the Environment* 27(1): 397–431

Hill, Joshua S. 2016. "China Overtakes Germany to Become World's Leading Solar PV Country." *Clean Technica,* January 22. http://cleantechnica.com/2016/01/22/china-overtakes-germany-become-worlds-leading-solar-pv-country/.

Hong Kong Environment Bureau. 2013. "A Clean Air Plan for Hong Kong." http://www.enb.gov.hk/en/files/New_Air_Plan_en.pdf.

IEA (International Energy Agency). 2012. *World Energy Outlook 2012.* Paris: OECD.

Jia, Ruixue, and Hyejin Ku. 2016. "Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust." http://www.sole-jole.org/16345.pdf.

Kahn, Matthew E. 1999. "The Silver Lining of Rust Belt Manufacturing Decline." *Journal of Urban Economics* 46(3): 360–76.

Kahn, Matthew E., Pei Li, and Daxuan Zhao. 2015. "Water Pollution Progress at Borders: The Role of Changes in China's Political Promotion Incentives." *American Economic Journal: Economic Policy* 7(4): 223–42.

Kahn, Matthew E., and Siqi Zheng. 2016. *Blue Skies over Beijing: Economic Growth and the Environment in China.* Princeton University Press.

Kan, Haidong, and Bingheng Chen. 2004. "Particulate Air Pollution in Urban Areas of Shanghai, China: Health-based Economic Assessment." *Science of the Total Environment* 322(1–3): 71–79.

Li, Shanjun. 2015. "Better Lucky Than Rich? Welfare Analysis of Automobile License Allocations in Beijing and Shanghai." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2349865.

Matus, Kira, Kyung-Min Nam, Noelle E. Selin, Lok N. Lamsal, John M. Reilly, and Sergey Paltsev. 2012. "Health Damages from Air Pollution in China." *Global Environmental Change* 22(1): 55–66.

Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *American Economic Review* 101(5): 1649–75.

Neidell, Matthew J. 2009. "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources* 44(2): 450–78.

Poon, Jessie P. H., Irene Casas, and Canfei He. 2006. "The Impact of Energy, Transport, and Trade on Air Pollution in China." *Eurasian Geography and Economics* 47(5): 568–84.

PR Newswire. 2015. "China Photovoltaic (PV) Research Report 2015–2019." September 3. http://www.prnewswire.com/news-releases/china-photovoltaic-pv-research-report-2015-2019-300137770.html.

Price, Lynn, Mark D. Levine, Nan Zhou, David Fridley, Nathaniel Aden, Hongyou Lu, Michael McNeil, Nina Zheng, Yining Qin, and Ping

Yowargana. 2011. "Assessment of China's Energy-Saving and Emission-Reduction Accomplishments and Opportunities during the 11th Five Year Plan." *Energy Policy* 39(4): 2165–78.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6): 1257–78.

Rosen, Sherwin. 1979. "Wages-based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter M. Mieszkowski and Mahlon Straszheim. Baltimore: John Hopkins University Press.

Sigman, Hilary. 2002. "International Spillovers and Water Quality in Rivers: Do Countries Free Ride?" *American Economic Review* 92(4): 1152–59.

Sun, Cong, Matthew E. Kahn, and Siqi Zheng. Forthcoming. "Self-Protection Investment Exacerbates Air Pollution Exposure Inequality in Urban China.**"** *Ecological Economics*.

Sun, Cong, Siqi Zheng, and Rui Wang. 2014. "Restricting Driving for Better Traffic and Clearer Skies: Did It Work in Beijing?" *Transport Policy* 32: 34-41.

Viard, V. Brian, and Shihe Fu. 2015. "The Effect of Beijing's Driving Restrictions on Pollution and Economic Activity." *Journal of Public Economics* 125: 98–115.

Wang, Alex L. 2013. "The Search for Sustainable Legitimacy: Environmental Law and Bureaucracy in China." *Harvard Environmental Law Review* 37(2): 365–440.

Wang, Xiaoping, and Denise L. Mauzerall. 2006. "Evaluating Impacts of Air Pollution in China on Public Health: Implications for Future Air Pollution and Energy Policies." *Atmospheric Environment* 40(9): 1706–21.

World Bank and DRC (State Council Development Research Center in China). 2014. "Urban China: Toward Efficient, Inclusive, and Sustainable Urbanization." Open Knowledge Repository, World Bank. https://openknowledge.worldbank.org/handle/10986/18865.

Xie, Jian, and Fasheng Li. 2010. "Overview of the Current Situation on Brownfield Remediation and Redevelopment in China." Sustainable Development—East Asia and the Pacific Region Discussion Papers, World Bank. September.

Yeo, Sophie. 2016. "Analysis: Decline in China's Coal Consumption Accelerates." *Carbon Brief,* February 29. https://www.carbonbrief.org/analysis-decline-in-chinas-coal-consumption-accelerates.

Zhang, Qingfeng, and Robert Crooks. 2012. "Toward an Environmentally Sustainable Future: Country Environmental Analysis of the People's Republic of China." Asian Development Bank.

Zheng, Siqi, Jing Cao, Matthew E. Kahn, and Cong Sun. 2014a. "Real Estate Valuation and Cross-Boundary Air Pollution Externalities: Evidence from Chinese Cities." *Journal of Real Estate Finance and Economics* 48(3): 398–414.

Zheng, Siqi, and Matthew E. Kahn. 2008. "Land and Residential Property Markets in a Booming Economy: New Evidence from Beijing." *Journal of Urban Economics* 63(2): 743–57.

Zheng, Siqi, and Matthew E. Kahn. 2013. "China's Bullet Trains Facilitate Market Integration and Mitigate the Cost of Megacity Growth." *Proceedings of the National Academy of Sciences* 110(14): E1248–53.

Zheng, Siqi, Matthew E. Kahn, and Hongyu Liu. 2010. "Towards a System of Open Cities in China: Home Prices, FDI Flows and Air Quality in 35 Major Cities." *Regional Science and Urban Economics* 40(1): 1–10.

Zheng, Siqi, Matthew E. Kahn, Weizeng Sun, and Danglun Luo. 2014b. "Incentives for China's Urban Mayors to Mitigate Pollution Externalities: The Role of the Central Government and Public Environmentalism." *Regional Science and Urban Economics* 47(C): 61–71.

Zheng, Siqi, Cong Sun, Ye Qi, and Matthew E. Kahn. 2014c. "The Evolving Geography of China's Industrial Production: Implications for Pollution Dynamics and Urban Quality of Life." *Journal of Economic Surveys* 28(4): 709–724.

Zivin, Joshua Graff, and Matthew Neidell. 2013. "Environment, Health, and Human Capital." *Journal of Economic Literature* 51(3): 689–730.

A Real Estate Boom with Chinese Characteristics

Edward Glaeser, Wei Huang, Yueran Ma, and Andrei Shleifer

Between 2000 and 2010, American housing went through a spectacular boom/bust cycle, which made and broke fortunes and left financial wreckage in its wake. Real prices of US homes grew by 5 percent per year between 1996 and 2006, and then declined by 6.4 percent per year between 2007 and 2012, according to Federal Housing Finance Agency data. Annual construction exceeded 1.9 million housing units in 2005 and 2006, and then dropped to an average of 688,000 units per year between 2009 and 2013.

Yet this US housing cycle looks stable and dull relative to the great Chinese real estate boom. In China's top cities, real prices grew by 13.1 percent annually from 2003 to 2013 (Fang, Gu, Xiong, and Zhou 2015). Real land prices in 35 large Chinese cities increased almost five-fold between 2004 and 2015 (Wu, Gyourko, and Deng 2015). As prices rose, so did construction. Between 2003 and 2014, Chinese builders added 100 billion square feet of floor space, or 74 square feet for every person in China (National Bureau of Statistics of China 2014; Chivakul et al. 2015). During this time, China built an average of 5.5 million apartments per year. In 2014, 29 million people worked in China's construction industry, or 16 percent of urban employment. By comparison, construction industry accounted for 8 percent of total

■ *Edward Glaeser is the Fred and Eleanor Glimp Professor of Economics, Wei Huang is a PhD student in Economics, Yueran Ma is a PhD student in Economics, and Andrei Shleifer is Professor of Economics, all at Harvard University, Cambridge, Massachusetts. Their email addresses are eglaeser@harvard.edu, huangw@nber.org, yueranma@g.harvard.edu, and ashleifer@harvard.edu.*

employment in the United States and 13 percent of that in Spain at the peak of their most recent housing booms.

We begin by reviewing these facts and two other striking differences between the housing booms in the United States and China. Unlike in the United States, high vacancy rates are a distinct feature of Chinese housing markets. Vacancies include both completed units unsold by developers, and purchased units that remain unoccupied. We estimate that this stock of empty housing now adds up to at least 20 billion square feet. The other profound difference is the much greater and quite distinctive public sector engagement with real estate in China compared to the United States. For example, all urban land in China is owned by the government, and private individuals only buy and sell "use rights," which were originally bought from local governments and run for a maximum of 70 years. It is unclear what the government will do when those use rights expire. While many US local governments, especially on the coasts, fight against new development, Chinese local governments encourage building to boost local output and employment, as well as to raise revenue from land sales. The US government subsidizes home purchases through tax and credit policies, but does not have nearly as aggressive a home-building promotion program as its Chinese counterpart.

The dramatic size of the Chinese housing boom has left the world wondering whether Chinese real estate is a bubble waiting to burst. We approach this question by analyzing the determinants of demand and supply of housing in China. On the demand side, we examine the economic, demographic, cultural, and speculative factors shaping demand. We also consider valuations of apartment prices in China as a gauge of frothiness. Fang et al. (2015) find that price-to-income ratios range from 6 to 10 in their sample of Chinese cities—high but not unusually so. They also show that Chinese housing price growth is not faster than income growth, but their data stop in 2012 and rely on mortgage applicants who tend to have relatively high incomes. Since then, prices continued to rise dramatically in some cities. As of 2016, a 90-square-meter apartment in Beijing or Shanghai fetches more than 25 times average household income. Even so, conclusions on the valuation of Chinese housing based on demand factors turn critically on difficult-to-forecast national growth rates. If China continues to grow very fast, the existing high prices can be justified.

The supply-side approach to housing fundamentals is to compare housing prices to the physical costs of construction and an assessment of the long-run price of land. Building costs can be estimated with reasonable precision, but land values are hard to predict in markets where supply is constrained by geography and regulation. The supply approach to Chinese housing suggests that prices are far too high, especially outside the few top cities. Construction costs are typically less than one-third of the selling price of finished space in these cities. Absent regulation, land is not scarce in the country as a whole, especially away from the coast. The rapid pace of construction and the vast number of vacant units support the view that supply is currently elastic.

To put our empirical results together, we simulate possible prices in 20 years. We assume that speculative demand for housing cannot be sustained in the long run

and home prices stabilize at about 10 times income, a high but not implausible level by global standards. If income growth is robust and housing supply growth is effectively muted, then equilibrium prices in 20 years for apartments in top tier cities could exceed 2 million RMB. But even if income growth stays high, real returns on housing will be negative if future supply is as ample as it has been in the past.

We conclude that a housing crash is not inevitable, but that the outcome depends on decisions made by the Chinese government. Only if new construction is sufficiently restricted can prices remain high. Yet the social costs of restricting new housing supply could be significant, because of both the lower employment and reduced growth of China's hyper-productive cities. This is indeed a real estate boom with Chinese characteristics: like much of modern China, its fate rests ultimately with government decisions.

Four Differences between the Housing Booms in China and the United States

In this section, we compare four aspects of the great housing booms of China and the United States: housing prices, construction, vacancies, and the roles of respective governments. While the United States may offer a less natural comparison for China than other middle-income countries (like Brazil), the US housing boom is well measured and particularly well studied. It offers a benchmark that helps to put the Chinese boom into perspective and highlight its unique features.

Housing Price Growth in China and the United States

Housing price growth and decline between 1996 and 2012 in the United States are well documented, typically using repeat sales indices prepared both by private industry, such as the Case-Shiller index, and by the Federal Housing Finance Agency. Two facts about the US housing prices are largely uncontroversial: First, housing prices grew dramatically after 1996 and then fell dramatically after 2007. Second, there was considerable heterogeneity in price growth across metropolitan areas, which is typically related to supply elasticities (Saiz 2010; Glaeser 2013).

Official statistics in China show a "mere" doubling of housing prices between 2007 and 2014 (Chivakul et al. 2015) and some degree of moderation in housing price rises more recently. These statistics are often questioned as understating actual price growth. Fang et al. (2015) present the most compelling work to date on quality-adjusted housing price growth in China. Their data for new homes uses development project fixed effects to control for changing quality. They thus estimate price changes by comparing sales prices of units of comparable size in the same development project sold in different years. While this approach is only likely to yield price estimates for the newer housing stock, it seems far more reliable than the alternatives.

China's cities are typically divided into four categories or tiers, based on the level of economic development. Tier 1 includes only the four most developed

metropolitan areas: Beijing, Shanghai, Shenzhen, and Guangzhou. In these areas, demand is most robust and supply is most likely to be restricted. Tier 2 includes most provincial capitals and some very developed prefecture cities. These are typically large, industrialized, and have relatively strong local economies. Tier 3 includes prefecture cities that have medium to high levels of income, which are smaller but still large by western standards. Tier 4 cities are further down in economic development and size, but still very populous compared to the average western city. As these categories are the norm in Chinese real estate analysis, for the comparison with the United States we create corresponding tiers of US cities. We first rank US cities according to income per capita in 1990, prior to the start of the US housing boom. We then draw tiers based on income and require that each tier has the same population share as its Chinese counterpart.

Fang et al. (2015) report real annual price growth of 13.1 percent in Tier 1 Chinese cities between 2003 and 2013, despite a slowdown during 2008 and 2009. Even the Tier 3 Chinese cities have seen an average of 7.9 percent real annual price growth between 2003 and 2013. Figure 1 shows the real price growth in the top three tiers of cities in China between 2003 and 2013. It compares this growth with real housing price growth in the top three tiers of US cities between 1996 and 2006. For each line, the start year is given a value of one. The Chinese growth in housing prices is far more dramatic. Housing price growth in US Tier 1 cities was impressive from 1996 to 2006, but it was slower than the growth of Tier 3 cities in China over their boom period and only 40 percent of the growth of top tier cities in China during their boom.

With such dramatic price growth in the past 15 years, apartments in the top Chinese cities are currently not much cheaper than in major American cities. According to Soufun data, the average price per square foot has risen to $550 in Beijing and Shanghai and $770 in Shenzhen. Zillow data currently shows San Francisco apartments costing $1,000 per square foot, and Boston units costing $600 per square foot. This similarity in prices is remarkable given that annual per capita income in Beijing, Shanghai, and Shenzhen is about $7,500, compared with more than $50,000 in San Francisco and $40,000 in Boston. The price levels in China's second and third tier cities are about one-half and one-quarter of those in the first tier cities, respectively, but average incomes are also lower by a similar magnitude.

The Size of the Building Boom in China and the United States

A construction boom of epic proportions has accompanied the price boom in China. Across China as a whole, construction value added came to 6.9 percent of GDP in 2013. Sixteen percent of China's urban employment, or 29 million people, worked in the construction sector in 2014.

We compare the construction of physical, typically residential, real estate in China and the United States. In the United States, household surveys provide estimates of the total housing stock on an annual or quarterly basis. In addition, the US Bureau of the Census provides measures of the flow of new housing construction, such as the flow of new building permits and estimates of total units completed.

Figure 1
Price Growth by Tier: China and the United States
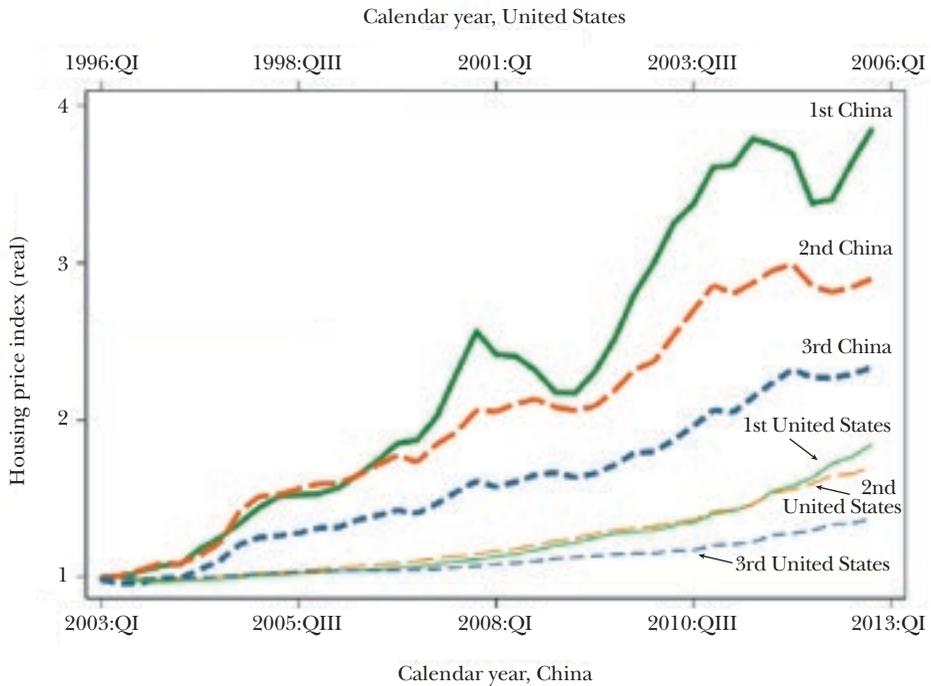
Source and Note: Housing price index data for China are from Fang et al. (2015). Housing price index data for metropolitan statistical areas (MSAs) in the US are from the Federal Housing Finance Agency. We construct tiers in the US by ranking US MSAs based on 1990 income per capita and assign tiers so that each tier has the same population share as that in China; the richest MSAs are assigned to 1st tier and so on and so forth. We calculate the real price index using the nominal price index adjusted for inflation. Price indices in each tier are weighted averages of city-level price indices.

Chinese census data provide snapshots of the housing stock at points in time but provide less-comprehensive information than their US equivalents. There are also estimates of China's new construction activity at the country and provincial levels on an annual basis.

Figure 2 shows the growth in floor-space in China and the United States. For China, we use the official statistics on housing area completed nationally. For the United States, we multiply the number of completed single-family homes by the average size of new single-family homes in that year. We then add the number of multi-family homes multiplied by the average size of new multi-family homes in that year. Figure 2 shows a large difference in total construction, which translates into similar scales on a per capita basis. For instance, between 2011 and 2014, China built 45.9 billion square feet of residential floor space, or 33.8 square feet per person. We estimate that the United States produced 16 billion square feet between 2003 and 2006, or about 55 square feet per person. In other words, although China is much

Figure 2
Construction (Square Footage) in China and the United States
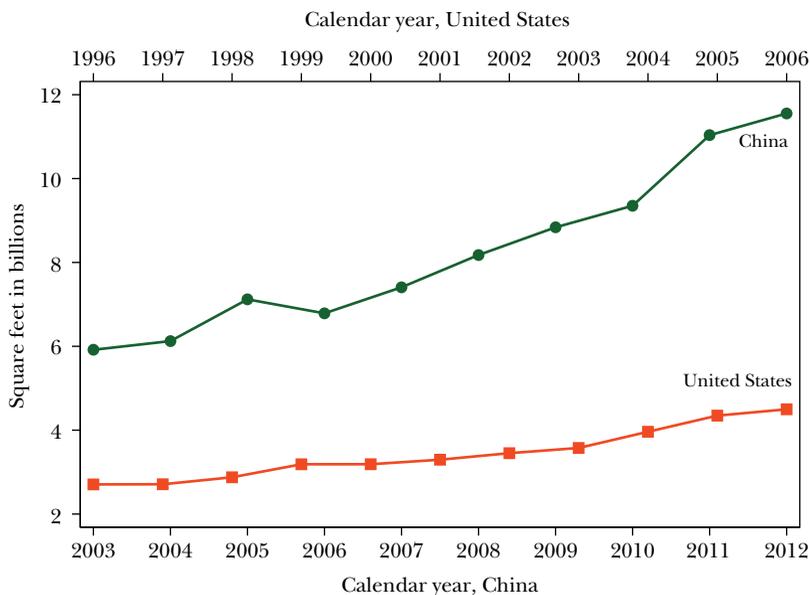

Source and Note: Annual completed residential space for China is from *China Statistical Yearbook*. Annual completed residential space for the US is calculated as units of single family homes completed multiplied by average square footage of new single family homes, plus units of multi-family homes completed multiplied by average square footage of new multi-family homes

poorer and its housing units are much smaller, it is producing about as much floor-space per person as the United States.

One striking fact about US construction growth is that more housing was built in less-productive areas, which creates a mismatch between productivity and employment discussed by Hsieh and Moretti (2015). Figure 3A shows a −0.29 correlation between per capita income in 1990 and housing supply growth between 1990 and 2010. A major reason for this misallocation of construction is supply restrictions in the most productive cities, such as zoning regulations (Glaeser, Gyourko, and Saiz 2008). Figure 3B shows the same striking negative relationship across Chinese cities between 2000 and 2010, with a slope coefficient of −0.45. The Chinese story here may be similar to the American one: local governments with the weakest economies have been the most aggressive in promoting new housing supply, perhaps as a means of generating economic growth, while local governments in the richest cities have been slowing down construction.

We have also used data provided by Matthew Khan and Siqi Zheng to examine whether air quality is correlated with urban growth and house price growth over the past decade. There is no correlation between initial air quality and subsequent price growth. There is a positive correlation between initial air quality and population growth among Tier 1 and Tier 2 cities, but not across the full sample of cities.

Figure 3
Construction and Income across US and Chinese Cities
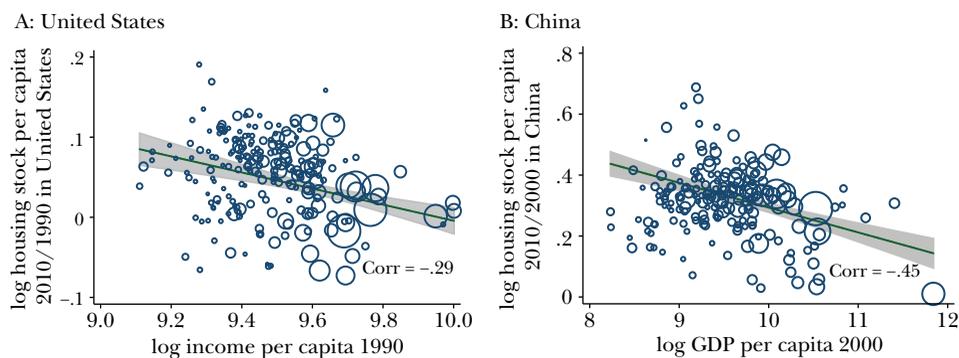
Source and Note: In panel A, housing stock per capita and income per capita for US metropolitan statistical areas (MSAs) are from the decennial census. In panel B, housing stock per capita for Chinese prefecture level cities is from 2000 and 2010 censuses and is restricted to counties that are designated as urban in the 2000 census. Log GDP per capita in 2000 is from *China City Statistical Yearbook 2000* for city proper. The correlations are weighted by initial population.

Housing Inventories and Vacancies in China and the United States

The death knell of the US housing boom was sounded by large inventories in Las Vegas that were brought swiftly to market in 2007. When developers offered the many units they owned for quick sale, prices quickly moved downwards and perhaps found a new equilibrium. Compared to the United States, inventories held by developers and unoccupied housing held by households are much larger in the Chinese boom. Unsurprisingly, this raised concerns about the viability of the boom.

At the height of the US boom, developers owned 573,000 unsold and unoccupied homes, according to the National Association of Home Builders. The vacancy rate among owner-occupied homes reached a high of 3 percent in 2008, and this statistic does not include "temporarily" vacant units. But, overall, relative to the stock of 130 million American homes as of 2010, the United States has relatively few vacant homes.

For China, our primary source for developers' inventory is Soufun, which compiles data from local housing bureaus (Fangguanju). The data covers developers' inventories for 32 major Chinese cities. We combine this data with statistics on vacant homes held by households, which is based on the Chinese Urban Household Survey (UHS). Figure 4 shows estimates of inventory by tier over time. We estimate inventories for each tier using the tier's inventory per capita in the Soufun data multiplied by the tier's urban population. The total inventory numbers show an increase from just over 4 billion square feet in 2011 to over 10 billion square feet in 2015. The bulk of this growth occurred in the lower-tier cities. For example, inventories in Tier 1 cities grew from 310 million square feet in 2011 to 390 million square feet in 2014, followed by a slight decline in 2015. Total estimated inventory in

Figure 4
Developer Inventory Estimates by Tier
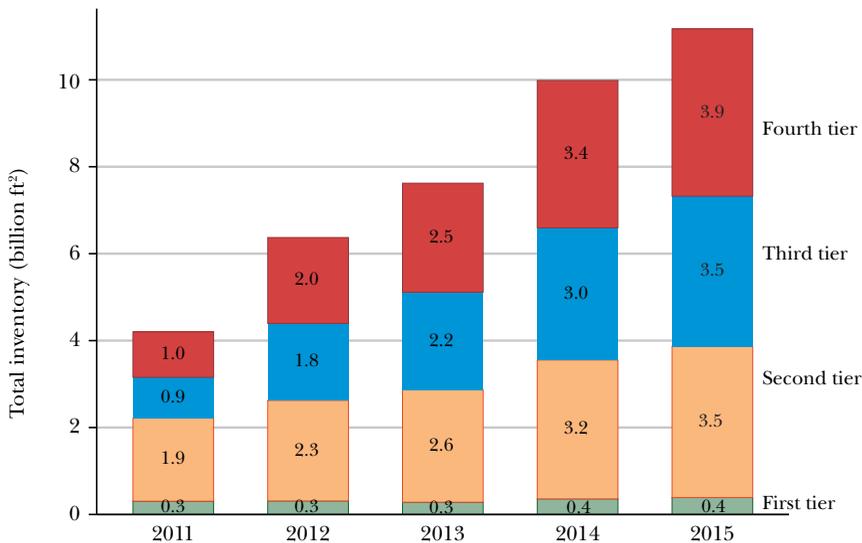

Source and Note: Developer inventory data is from local housing bureau (Fangguanju), and compiled by Soufun. Because the data does not cover all the cities, our estimates of inventory in each tier are based on inventory per capita by tier multiplied by urban population in each tier. Because the data do not have many Tier 4 cities, we also assume the Tier 4 cities have the same inventory per capita as the Tier 3 cities. In total, the estimates cover 262 prefecture level cities: 4 cities in the first tier, 34 in the second tier, 84 in the third, and 140 in the fourth.

Tier 3 cities increased from 940 million square feet to 3.9 billion square feet from 2011 to 2015.

To put these quantities into perspective, China's occupied housing space in urban areas is 111 billion square feet based on the 2010 census, with 7.7, 32.1, 32.2, and 39.5 billion square feet in the first to fourth tier cities respectively. We can also compare developer inventory in China and the United States on a per capita basis. The average new US home in 2006 was 2,500 square feet, and the US population was almost 300 million. The 527,000 new housing units in the developers' inventory in the United States in 2006 thus translate to about 4.5 square feet per capita. By 2015, developer inventory reached about 6 square feet per capita in China's Tier 1 cities, and 20 square feet per capita in Tier 2 and Tier 3 cities. The developers' stock is much higher in China than in the United States, and is concentrated not in the most successful cities, but in the less-developed ones.

Visitors to Chinese cities are sometimes struck by the number of vacant apartments. Many of these units are owned by household investors who just leave their properties empty. We can get some sense of this stock of purchased but vacant units by looking at the Urban Household Survey. We compute a vacancy rate across 36 cities by calculating the total number of vacant units owned by the residents of each city in the sample, and then dividing by the total number of housing units owned

Figure 5
Household Vacancy Rates 2001–2012
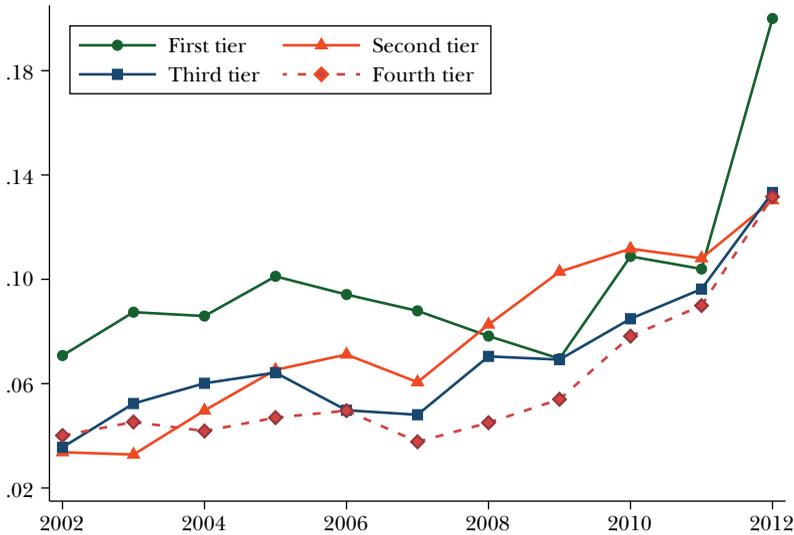
Note: We compute a vacancy rate across 36 cities by calculating the total number of vacant units owned by the residents of each city in the sample, and then dividing by the total number of housing units owned or occupied by the city's residents in the sample. We keep the 36 cities with observations throughout 2002–2012.

or occupied by the city's residents in the sample. Anecdotal evidence suggests that owning properties outside one's own metropolitan area is rare.

Figure 5 shows these sold-but-vacant figures from 2002 to 2012 for a sample of 36 cities. The vacancy rates look relatively flat from 2002 to 2008 and then rise sharply after 2009. In 2012, the vacancy rate drew close to 20 percent in Tier 1 cities. The rates for lower-tier cities were about 13 percent.

In Table 1, we combine estimates of total developer inventories with estimates of total household vacancies. To ensure representativeness, we use vacancy estimates based on the 2009 Urban Household Survey.[1] We take the average vacant space per capita in each tier and estimate total vacant space in that tier by multiplying by its urban population. This estimate is conservative, as vacancy rates have increased since then. We pair the vacancy estimates with inventory estimates for 2014 to provide an up-to-date picture, as total developer inventories almost doubled from 2011 to 2014. Data for 2014 also cover the largest set of cities in our developer inventory dataset.

The estimates imply that there are about 16 square feet per capita of vacant real estate in Tier 1 cities, of which 5.5 square feet is owned by developers, and 10 square

---

[1] Up to 2009, China's Urban Household Survey provides data on household vacancies for up to 120 cities (but data are only available for 36 cities from 2010 to 2012).

Table 1
Estimates of Inventory and Vacancies

Tier of city (1)	Developer inventory per capita (ft²) (2)	Household vacancy per capita (ft² per capita) (3)	Urban population (million) (4)	Predicted developer inventory (billion ft²) (5)	Predicted vacancy (billion ft²) (6)	Developer inventory + vacancy (billion ft²) (7)
1	5.5	10.4	57.7	0.32	0.60	0.92
2	17.1	19.9	166.9	2.85	3.33	6.17
3	14.5	16.0	185.1	2.69	2.97	5.66
4	-	12.9	206.5	3.00	2.67	5.67
Total	**13.2**	**15.4**	**616.2**	**8.86**	**9.56**	**18.42**

Source: Inventory area per capita is calculated from local housing bureau (Fangguanju) data for 2014 compiled by Soufun, and vacant area per capita is calculated based on 2009 Urban Household Survey.
Notes: Urban population is from the 2010 census. Column 5 is column 2 multiplied by column 4, and column 6 is column 3 multiplied by column 4. Column 7 is the sum of columns 5 and 6. We assume 4th tier cities have the same inventory per capita as 3rd tier cities.

feet is owned by household investors. The total would be 20 square feet per capita if we instead used the Urban Household Survey from 2012. In Tier 2 cities, there is 37 square feet per capita of vacant real estate, split more or less evenly between developers and household investors. There are 30 square feet per capita in Tier 3 cities. In the cross section of cities, both developer inventories and household vacancies on a per capita basis are highly positively correlated with past price growth.

These data provide a picture of an enormous stock of unoccupied new housing in China, which dwarfs the US equivalent in 2006 (4.5 square feet per capita). The combination of crowded cities and empty housing stock is a highly distinctive feature of Chinese housing today.

A Regulated Bubble? The Greater Role of the Public Sector in China

The American real estate sector is hardly a bastion of laissez-faire capitalism. A web of local land use controls regulate the industry and restricts supply, while national policies, including the Low Income Housing Tax Credit, the home mortgage interest deduction, and the government-sponsored enterprises like Fannie Mae and Freddie Mac subsidize home purchases. Still, the shifting dynamics of the US housing industry are only modestly shaped by the national and local governments.

In China, the public sector retains far more control over future prices and construction. Most importantly and dramatically, all land in China is owned by the state, and the central government decides which land can be auctioned to private developers. Land auctions do not confer permanent ownership, but rather the right to use the land for several decades (up to 70 years). Article 70 of the Chinese Property Rights Law of 2007 does grant the ability to own the structures on the land but not the land itself. Article 149 suggests that use rights will be automatically renewed, and so far, China's government has not extracted additional fees for the relatively unconstrained rights to use the land after the first land sale, but it is unclear what compensation local governments will require for use right renewals in the future.

Anyone buying real estate for the long haul must trust the Chinese government to roll over their land leases at a reasonable cost.

Back in 2000, 70 percent of Chinese floor space was built by state-owned enterprises and housing collectives. In 2013, such public entities only built 13 percent of China's floor space. But the decline of formal public construction activity does not mean that the private construction operates without public guidance. Both the central and local governments are deeply involved in the housing market, and government incentives play a key role in shaping market outcomes.

Since Deng Xiaoping declared that "to get rich is glorious," China's leaders have promoted economic growth with remarkable success, with construction playing a pivotal role in these efforts. Leaders of local governments are promoted or penalized depending upon their ability to attain national growth objectives (Li and Zhou 2005). With ample credit and low construction costs, developers build towers that add to the GDP. The construction activity itself is measurable economic output, and ideally the buildings will then attract commercial and residential tenants who further boost local growth.

Another important incentive that pushes Chinese governments, particularly the local governments, towards real estate development comes from the Chinese tax system. The tax system in China leaves local governments with a limited stream of ongoing fiscal revenues. A key source of fiscal income is land sales revenue, which is collected in the transfer of state-owned land to developers. Taxation of real estate development and apartment sales also contributes to local fiscal revenue. In some years, revenues from land sales accounted for as much as 40 percent of local government fiscal revenue (Wu 2015). This revenue system helps explain why Chinese local governments are very enthusiastic about new construction.

Since the late 2000s, China's central government has increasingly shifted focus to housing market stability, and has held the local governments accountable for price stability targets. Under the central government's guidance, local governments have sought to stabilize housing prices, using restrictions or subsidies on home purchases and tightening or loosening controls on new development. The goals of economic growth and price stability, however, are often in conflict: maintaining high prices by restricting supply undermines the growth coming from construction.

Government interventions impact all players in the Chinese housing market. When it comes to developers, the Chinese local governments, like their US counterparts, regularly impose land use restrictions such as maximum floor-area ratios, although developers sometimes manage to exceed the originally mandated limit (Cai, Wang, and Zhang forthcoming). Unlike in the United States, the Chinese central government also frequently intervenes to regulate local land development. Such interventions restrict the conversion of land from agricultural to urban uses, limit the use of land for golf courses and luxury housing, and mandate development of apartments with certain sizes.

When it comes to homebuyers, American and Chinese governments have pursued very different policies. Despite some local land regulations, the US promotes homeownership through tax and credit policies. China is much more

cautious. Interest rates on mortgages, as well as minimum down payments, tend to be high. To curtail speculation, Chinese governments have also tried to directly restrict second and third home purchases, and to limit the resale of homes in less than five years. Property taxation on homeowners is also strikingly different. US governments tax real estate on an annual basis, which creates a sizable cost to holding real estate. Chinese real estate taxes are imposed only on transactions, creating an incentive to buy-and-hold.

When it comes to banks, state-owned institutions have historically dominated China's banking system, and at every level, banks' policies towards the housing market follow the government's lead. The financing for developers by government-controlled banks is critical for understanding housing in China. Historically, China had very favorable credit policies toward developers, and banks are pushed to lend. Developers are highly leveraged, and to some extent enjoy the "heads I win, tails the bank loses" situation. On the mortgage side, banks are also under government direction to adjust mortgage rates to discourage (or occasionally to subsidize) home purchases, as part of housing market stability management. But banks have not been required to systematically support home ownership, as they have been required to support construction. Although China's government has become increasingly careful of risks in development loans and tightened them in recent years, the stockpile of debt is still enormous. According to official statistics, Chinese developers had $6.2 trillion of debt by 2014, whereas households had $3.8 trillion.

In sum, the main US housing market policies encourage mortgage borrowing and restrict new development with local land use regulations. The most important Chinese housing market interventions boost construction and restrict mortgage borrowing. With these policies, it is unsurprising that China builds much more than the United States. It is perhaps more surprising that Chinese housing price growth has been so much greater, given the discouragement of home borrowing and the encouragement of new supply. This strong demand for housing, as we discuss below, is mostly coming from Chinese households who flood the real estate market with their savings and investment demand, as opposed to a levered play on a "house of debt."

Does China Have a Real Estate Bubble?

Will China's real estate boom end in a bust? We discuss two approaches to assessing whether housing is in a bubble. The demand-side approach focuses on the motivations and the behavior of buyers, but also considers whether prices seem reasonable given proxies for the demand for housing services, such as income. The supply-side approach asks whether prices seem reasonable given the cost of delivering new real estate. Here we present some facts bearing on these two ways of thinking about China's housing boom, as before, frequently comparing China and the United States.

The Demand-Side Approach to Assessing Bubbles

In the United States, many homebuyers during the boom were young and highly leveraged. Mian and Sufi (2008) document the role of subprime lending fueling the US housing boom. Glaeser, Gottlieb, and Gyourko (2010) use Data-Quick loan data and find that the median US homebuyer in 2006 had a loan-to-value ratio of 90 percent. Over one-quarter of the buyers put no money down, often by using multiple mortgages.[2] The US boom featured a huge rise in the share of single homebuyers, and in 2006, almost 60 percent of loan applications came from one-person households. The National Association of Realtors data show that in 2005 and 2007 in the United States, 67 percent of new homes and 40 percent of all homes were bought by people under the age of 35.

In China, housing investment is fueled by middle-aged households saving for retirement or for their children's marriage. Home ownership has always played a prominent role in the Chinese society, and is remarkably high, standing at roughly 90 percent compared to about 65 percent in the United States.

Chinese households are well known for high saving rates, motivated by many factors from economics to culture (Chamon and Prasad 2010). Chinese savers have been pushed towards housing as a crucial form of investment, not because it is easy to borrow but because it is difficult to invest. Chinese mortgage rates have often been much higher than in the United States, and down payment requirements have rarely fallen below 20 percent. For Chinese savers buying real estate as investment assets, however, the more relevant cost of capital is not the mortgage rate but the opportunity cost of putting their funds elsewhere. The average real interest rate on Chinese bank deposits has been about zero over the past decade. The Shanghai Stock Exchange Index has had average annual real returns of zero from 2001 to 2016, which are anemic relative to more than 10 percent annual real returns to Shanghai real estate over the same period. Tight capital controls also restrict investing abroad. Moreover, traditional agrarian culture in China cultivates a strong preference for owning land and real estate. Painter, Yang, and Yu (2003) find that home ownership is 18 percentage points higher among Chinese than among white households in the United States, after controlling for household characteristics.

In addition to investing for wealth preservation, many Chinese households find it important to purchase real estate for their children, as home ownership is considered close to a necessity for marriage (and required especially for successful marriage of young men). This tradition is reflected in a high home ownership rate among young Chinese households, most of whom bought their properties with funds that to a significant extent came from their parents. The homeownership rate for Chinese households below 35 in 2013 is around 55 percent, based on data from the China Household Finance Survey, compared to 37 percent in the United States. Wei and Zhang (2011) provide detailed evidence of a competitive savings motive for

---

[2] From 1998 to 2007, the 75th percentile of the loan-to-value distribution for new buyers never fell below 95 percent in the United States. The mean loan-to-value level was 74 percent in 2006 (over the first three mortgages on the home), reflecting the significant number of US buyers who borrowed far less.

children's marriage in China, and the China Household Finance Survey also indicates that ownership of multiple residences and vacant residences is higher among families with unmarried young men.

Both the investment motive and the marriage motive contribute to Chinese households buying second homes if they can afford them, even without occupying them. The Chinese Urban Household Survey for 2009 contains data on ownership of second homes. The data show that these owners are richer and better-educated than the average survey respondent. Owners of vacant homes have 34 percent higher earnings than average, and 0.9 more years of education. Chinese investors are mature and prosperous. The average age of the investors is 49, and 95 percent of them are married. Estimates from several sources suggest that housing accounts for 70 to 85 percent of urban households' total assets in China (Huang 2013; Xie and Jin 2015). Despite the heaving investment in real estate, the ratio of total Chinese household debt to GDP was still under 40 percent by the end of 2015, according to data from the Bank of International Settlements (BIS Statistics Explorer, Table F3.1, Q1 2016, http://stats.bis.org/statx/srs/table/f3.1). In comparison, in the first quarter of 2008, US household debt reached 99 percent of GDP, according to data from the International Monetary Fund (reported at FRED Economic Data, https://research.stlouisfed.org/fred2/series/HDTGPDUSQ163N).

Are these Chinese real estate investors overly optimistic? In the US context, Case, Shiller, and Thompson (2012) use survey evidence to document the extremely strong overoptimism of home buyers. Simple extrapolative models do seem to fit the stylized facts of the US real estate market, such as the strong positive serial correlation in price changes, remarkably well (Glaeser and Nathanson forthcoming). There is also a robust correlation between past price growth and stated expectations about future price growth in the United States (Williams 2013). On the other hand, US optimism about real estate prices during the housing price boom had little justification other than past price growth. Chinese investors in their 50s have lived through a national economic growth miracle. Their optimism about future price growth is surely understandable, although not necessarily correct. Whether or not their buying decisions are based on overoptimism, the portfolios of the Chinese middle class are highly skewed towards real estate. Significant changes in housing values will have a considerable impact on their future well-being.

An alternative way to ask whether demand for housing is speculative and excessive is to assess whether the price of the asset is roughly equal to the value of the flow of services that the asset will deliver. In finance, this approach compares prices with flows of earnings or dividends (Shiller 2000). High prices compared to earnings are a possible candidate for bubbles.

In the case of housing, the approach typically compares housing prices to rents, assuming that rents capture the flow of services from owning the house. When interest rates are low, small differences in expectations about future growth rates can generate wild swings in the predicted price to rent ratio. As a result, some concluded that US prices in 2005 were compatible with current rents and rationality

(in this journal, Himmelberg, Mayer, and Sinai 2005). In the Chinese case, renting is so rare that rental data is unavailable for units comparable to new properties.

An alternative to rent is to compare prices to income, which *The Economist* (2016) has referred to as "a gauge of frothiness." Fang et al. (2015) use this approach in their balanced discussion of whether China has "a housing bubble that is about to burst." They note both that "housing prices are indeed expensive to the income of many households," and that "price appreciation has been mostly accompanied by equally impressive income growth." They report "price-to-income ratios of around eight in second- and third-tier cities and, in some years, even over ten in first-tier cities." These ratios are quite high by global standards, but the parallel trends of income and price growth remind us that optimistic Chinese buyers look far more sensible than their American counterparts who paid sky-high prices during the US housing price boom despite only moderate US economic growth during those years.

But a constant price-to-income ratio is neither necessary nor sufficient to establish the existence of a bubble. If houses, or any product, can be built for a fixed cost of $100,000, then even if house prices temporarily rise above $100,000 and keep pace with rapid income growth, they may still ultimately fall down to the cost of supply. Alternatively, prices can rise faster than income growth in places with restricted supply, perhaps because amenities are increasing, and there is no reason for these high prices to later crash.

In China, educational and medical resources are heavily concentrated in major cities, and the *hukou* system restricts access to many resources to local residents. Home purchases and permanent residence in top tier cities is one step towards becoming a local resident. Consequently, high housing prices could in part reflect a premium on urban amenities and expectations about future demand for urban amenities.

Overall, price-to-rent or price-to-income ratios offer remarkably little precision about whether housing prices are too high or too low. Slight changes in projected growth or interest rates (Chinese mortgages are typically variable rate) cause the implied price to rent ratio (or price to income ratio) to change substantially. Because we do not know the future value of urban amenities, it seems hard to have much confidence about what the demand should be for urban space.

The Supply-Side Approach to Assessing Real Estate Bubbles

The demand-side calculations, taken alone, ignore the power of supply over housing prices. Glaeser (2013) argues that in the history of US real estate bubbles, the most common mistake for investors is to ignore the role that supply plays in determining prices of housing and land. Investors buying in London's Belgravia or Paris' Ile St. Louis can probably rest secure that these areas will not be flooded with new housing supply. But we have already documented the enormous amount of new construction and massive quantities of vacant homes in Chinese cities. Supply seems still elastic in many Chinese cities, especially outside the first tier, although that could change quickly with a shift in public policy.

The physical cost of construction in China appears to be quite low, and fairly similar across different cities. We use construction cost estimates from two standard

sources. The first source is the China Engineering Cost Network, directed by the Ministry of Housing and Urban-Rural Development. The second source is the *China Real Estate Yearbook.* As of 2014, the China Engineering Cost Network provides a cost of 1,362 RMB per square meter for multistory housing and 1,817 RMB per square meter for high-rise housing. This would represent a cost of between $19 and $26 per square foot for building, which is astonishingly low by US standards. The total physical cost at completion from the *China Real Estate Yearbook* is somewhat higher, at 2,730 RMB per square meter, about $38 per square foot. Their estimates appear to include some softer cost and additional expenses. Developers typically report that 1,800 RMB per square meter is a reasonable industry standard for hard costs in recent years. Total cost has gone up substantially since the early 2000s, based on the *Yearbook* data. These rising costs may reflect increasing labor costs or improving quality of construction, but they remain low relative to prices. Moreover, technological progress in construction is likely to keep future construction costs low. Modular skyscraper design seems to have enabled one developer to build a 57-story apartment building in 19 days (as reported by Baer 2015).

Construction costs in China are typically given on a gross square footage basis (that is, including halls, stairwells, and the like), while sales prices refer to net square footage. As gross square footage can readily be 40 percent higher than net, construction costs range from 2,000 RMB per square meter to 3,800 RMB per square meter of usable space. These costs should be compared with the current prices of over 35,000 RMB per square meter in Shanghai and Beijing, 18,000 RMB per square meter in Guanzhou, and 10,000 RMB per square meter in Wuhan. Soufun's 100 city index shows an average sale price of 12,000 RMB per square meter by mid-2016. Construction costs appear to be less than 30 percent of sale prices in typical Chinese cities, and perhaps as little as 15 percent in the top tier cities.

The large gap between construction costs and prices does not necessarily imply that developers are earning huge profits. Land is expensive and taxes are high. Wu, Gyourko, and Deng (2015) have built a remarkable index of land prices in China based on repeat land sales (available at http://real.wharton.upenn.edu/~gyourko/chineselandpriceindex.html). They find that real land prices in 35 Chinese cities increased five-fold between 2005 and 2015. For 2014, Wu, Gyourko, and Deng (2015) data show an average land price of 23,385 RMB per square meter of floor area in Beijing (based on 31 parcel sales) and 20,620 RMB per square meter in Shanghai (based on 41 sales). Lower-tier cities have lower land prices. In a relatively well-off provincial capital like Wuhan, land prices average 4,839 RMB per square meter (based on 40 sales). Most Tier 2 and Tier 3 have land prices of around 1,000 RMB per square meter (Deng, Gyourko, and Wu 2012) and average sales price of around 5,000 to 10,000 RMB per square meter.

There are two ways in which the land cost contribution to building costs could decline and bring prices closer in line with construction costs. First, Chinese cities could permit higher floor-to-area ratios, which determine the ratio of built floor space to land in an entire parcel, and thus reflect both building height and vacant land left around buildings. Despite the height of many Chinese buildings, actual

floor-to-area ratios are relatively low. Cai, Wang, and Zhang (forthcoming) show average floor-to-area limits for new residential construction of two in many cities (for example, 2.12 in Beijing and 1.65 in Shanghai). Assume that in the long run China's total population rises to 1.5 billion with an urbanization rate of 70 percent. At a floor-to-area ratio of 2, assuming that urban Chinese consume 40 square meters of net living space or 56 square meters of gross living space per capita, in the long run an urban population of 1.05 billion will require 29.4 billion square meters of urban residential land. This is still less than China's 40 billion square meters of urban residential land.[3] Cities eager for more development and for less sprawl could ratchet floor-to-area ratios upward. At a floor-to-area ratio of 10 as opposed to 2, accommodating 1.05 billion urban Chinese will require only seven billion square meters of urban land.

An alternative path to lower land costs is to expand urban space in new cities. Shenzhen was a small town in 1980. Now it is a thriving megalopolis of 11 million. China's urban space could expand substantially, and the land constraint could be relaxed by freeing land up for urban purposes.

It is well within the power of the Chinese government to allow either enough new cities or higher floor-to-area ratios, either of which could create an ocean of new housing supply. In the next section, we will explore the effect that new supply could have on housing prices in 20 years.

China's Housing Prices in 20 Years

The Chinese investors who buy and hold empty properties are investing for the long run. In this section, we assess whether today's prices in Chinese cities will look reasonable in 20 years. We consider a range of possible scenarios and compare price estimates both with the price today, and the price today plus 3 percent annual growth. If investors require a return of 1 percent per year over and above maintenance and depreciation costs (which we assume are roughly 2 percent), then housing prices need to rise by 80.6 percent in total over 20 years. Our estimation of future price derives from estimates of demand for and supply of housing in the long run. We produce estimates separately for Tier 1, Tier 2, and Tier 3 cities. While Chinese cities have significant heterogeneity, each tier groups cities that are relatively similar in economic development.

Our demand calculations begin by assuming that the willingness to pay for housing will eventually settle on 10 times income, high by US standards but not unusual in Europe.[4] In other words, we assume that all individuals with local incomes

---

[3] This estimate for China's urban residential land is from Karen C. Seto and is available in slides titled "Urban Growth in China: Challenges and Prospects," https://web.stanford.edu/class/cee320/CEE320B/Seto.pdf.

[4] To justify this with a spatial equilibrium, assume that income for individual i in a city is denoted y_i, and income in the reservation locale (rural China) is $\theta y_i + (1 - \theta) \underline{y}$. If the rural price is fixed by construction

above one-tenth of the price (P) are willing to live in the location. The buyers are assumed to be the richest potential urbanites. Consequently, if the income distribution is characterized by a cumulative distribution $F(Y)$, then total demand for housing is $N(1 - F(Y^*))$ where N is the size of the "potential" urban population in this locale and $1 - F(Y^*)$ denotes the share with incomes above the cutoff value Y^*, who will be homeowners. The equilibrium price P^* will equal ten times Y^* by assumption.

Hence, we predict prices by using the equilibrium condition:

$$\text{Housing Supply} = N(1 - F(P^*/10)).$$

This equation requires us to estimate the future distribution of income, growth in the potential urban population, and growth in the housing stock. Low housing supply and high population growth will both lower the fraction of households served. High income growth means that the willingness to pay for the same marginal buyer will be higher.

For the distribution of future household income, we start with the current distribution of income in each tier, based on data from China's Urban Household Survey. We assume that the distribution is log-normal, which is consistent with the data. We then estimate the income distribution in 2030 based on assumptions of income growth rates. In the official statistics, real incomes for urban Chinese in the past ten years have grown by about 10 percent per year. Our baseline assumption is that average future growth in the next two decades will be one-half that size, but Figures 6A, B, and C will show results for a range of income projections.

We project that the growth in future potential urban population will mirror past growth in actual urban population, recognizing that past urban population reflects both changes in regional population patterns and changes in the urbanization rate. The Tier 1 cities have high and stable urbanization of 90 percent, but they have experienced much higher total population growth than the national average. The population inflow comes from these cities' advantages in creating high-paying jobs and the fact that they offer educational, medical, and many other social resources, and these advantages seem unlikely to change. For the Tier 2 cities, population growth has been at the national average, and the urbanization rate has been rising steadily. For the Tier 3 cities, total population growth has been lower than the national average, due to labor outflow, but they are experiencing very rapid urbanization of the local population.

We assume that the first tier cities will continue to be highly urbanized and that the growth in potential population will be higher than the projected national population growth rate. In second tier cities, we assume that potential population

---

costs at $\underline{P}$, then the willingness to pay is $P + \frac{(1-\theta)}{r-g_i}(y_i - \underline{y})$, where g_i is the growth rate of this income gap. As long as $\underline{P} - \frac{(1-\theta)}{r-g_i}\underline{y}$ is small, then willingness to pay is $\frac{(1-\theta)}{r-g_i}$ times income. Our approach assumes that $\frac{(1-\theta)}{r-g_i}$ approximately equals 10.

will increase at above the national rate because of increasing urbanization. The third tier cities are assumed to have regional population growth below the national average but growing urbanization rate.

For supply of housing, we start with current occupied housing stock and then add inventories and vacancies. New completion of floor space nationwide has risen to about 10 billion square feet per year in the past five years, up from around 6 billion square feet in the mid- 2000s. We consider a range of supply scenarios, from a lower bound of zero to an upper bound of about 8 billion square feet per year nationally. We assume that the existing housing stock depreciates by 2 percent per year.

Will there be enough land space to accommodate that additional housing supply? While China is a very populous country, density in Chinese cities is typically lower than major American metropolitan areas. For example, the first tier cities have population density varying from 1,000 to 2,000 people per square kilometer in 2010, whereas New York City has density of 10,000 people per square kilometer and the top 100 American metropolitan statistical areas all have density above 4,000 people per square kilometer. Our most extreme high-supply scenario will less than double housing stock in the first tier cities, and the more probable range we consider will add less than 25 percent to existing housing space. It is also unlikely that population growth in the first tier cities can significantly exceed our estimate of 40 percent in the next 20 years, and this amount of population growth rate will still put top tier Chinese cities at lower density than many American cities.

Table 2A summarizes the specific parameter values we assume. Since our data and projections of housing stock are based on the urban core in each city, which has roughly 60 percent of total urban population, we also multiply by .6 to capture the core's share of total population. This potential urban population may not all be homeowners in the city. Those who cannot afford housing may rent or live outside of the urban core. Finally, we assume that per capita housing consumption in 2030 will be 40 square meters (or 430.56 square feet), up from 30 square meters in China's 2010 census.

Table 2A discusses two specific cases for each tier of cities. Here we focus on 5 percent annual income growth. We consider a conservative case where annual new supply is roughly cut in half compared to the average from 2000 to 2010, and an aggressive case where the supply is higher. These examples confirm that very high price growth seems unlikely, but as long as income growth remains robust, the downside in the eventual long-run equilibrium also appears to be modest.

Figures 6A, 6B, and 6C show long-run price estimates for each tier as a function of income growth and new supply. They present a range of possible future scenarios. Here we translate our price estimate into an average annualized house price growth from 2010 to 2030. The smallest increases in housing prices will be in the upper left corner of these figures—that is, large supply and slow income growth—while the largest increases will conversely be in the bottom right-hand corner. The horizontal line shows the average annual supply from 2000 to 2010 in each tier. The contour lines on the figures show what combination of income and supply will produce a specific rate of return: for example, what combination will produce the average

Table 2
Estimates of Housing Prices in 20 Years, by City Tier

A: Key Inputs and Assumptions

City tier:	1st	2nd	3rd
Total housing in urban area 2010 (billion ft^2)	8.6	37.3	36.1
Used	7.7	32.1	32.2
Developer inventory	0.3	1.9	0.9
Household vacancy	0.6	3.3	3.0
Total population growth 2000–2010	50%	26%	7%
Urbanization rate 2010	90%	70%	50%
Assumptions			
Depreciation rate of housing	2%	2%	2%
Housing per capita	403.56 square feet (or 40 square meters)		
Long-run price to income ratio	10	10	10
Total population growth 2010–2030	40%	20%	10%
Urbanization rate 2030	90%	80%	70%

B: Selected Scenarios

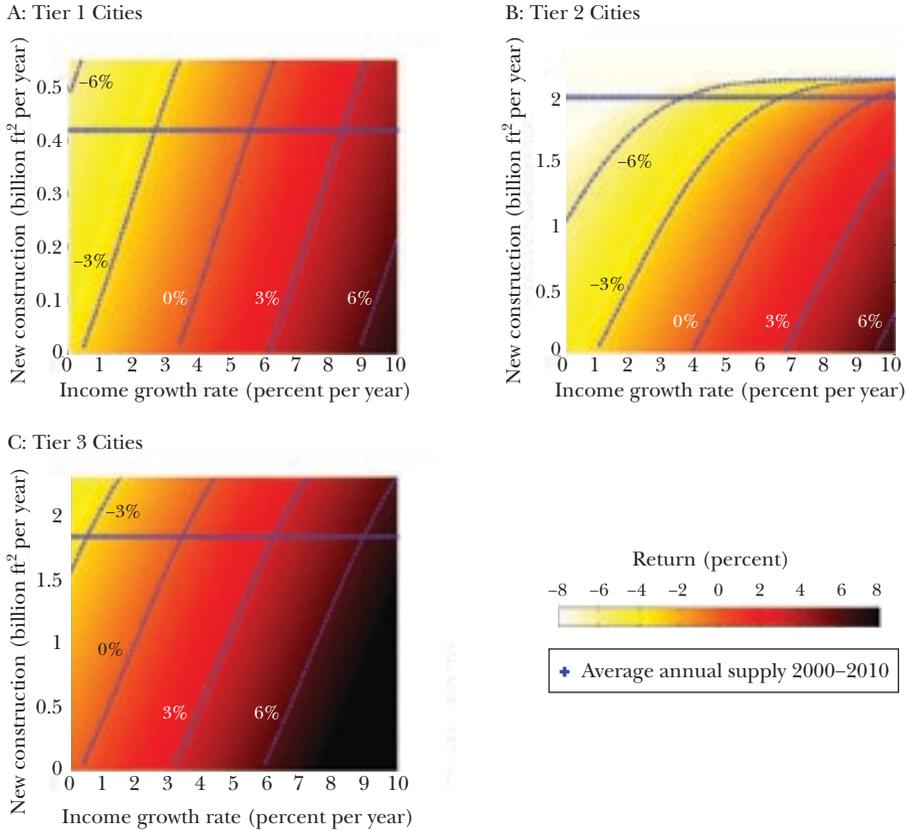| City tier	1st		2nd		3rd	
Scenarios	Conservative	Aggressive	Conservative	Aggressive	Conservative	Aggressive
Additional assumptions						
Annual housing supply 2010–2030 (billion ft^2)	0.17	0.48	0.72	2.00	0.90	2.83
Annual income growth	5%	5%	5%	5%	5%	5%
Price in 2030 (2014 RMB per ft^2)	2,165	1,531	1,039	410	1,017	403
Annual price growth	0.82%	−0.91%	−0.29%	−4.82%	3.37%	−1.30%
Annual housing supply 2000–2010	0.42		1.98		1.84	
Price in 2010 (2014 RMB per ft^2)	1,840		1,102		524	

Source and Notes: Housing in urban area in 2010 is from the 2010 census, covering the city proper. Inventory is estimated using Soufun data, and vacancy is estimated using Urban Household Survey data. Household income data are also drawn from Urban Household Survey. Annual housing supply in 2000–2010 is estimated as housing in city proper in the 2010 census minus that in the 2000 census times 0.82 (assuming housing stock in 2000 depreciates by 2 percent per year), then divided by 10. Based on current data, 60 percent of urban population live in the city proper area. Price estimates are in 2014 RMB. We consider a conservative case where annual new supply is roughly cut in half compared to the average from 2000 to 2010, and an aggressive case where the supply is higher.

3 percent annual real return. Different tiers can end up with different housing market outcomes, depending on economic growth and housing supply.

For both Tier 1 and Tier 2 cities, 3 percent annual real return in housing prices over the next 20 years is quite difficult to achieve. It requires annual real income of at least 6–7 percent, which is very optimistic for the average pace in the next 20 years, together with new housing supply close to zero. If the growth of housing supply persists at the average level of 2000 to 2010, returns will be minimal or negative unless income growth turns out to be spectacular. For the Tier 3 cities, our simulation provides some room for future price growth as long as supply does not overwhelm the market. The results are driven by the fact that these cities have the largest scope for increased urbanization, and their current prices are generally low.

Figure 6
Estimates of Average Annual Housing Price Growth, 2010—2030

A: Tier 1 Cities

B: Tier 2 Cities

C: Tier 3 Cities

Return (percent)

−8 −6 −4 −2 0 2 4 6 8

+ Average annual supply 2000–2010

Source and Notes: Plots of estimated average annual real housing price growth from 2010 to 2030, as a function of annual real income growth (x-axis) and annual new construction (y-axis). Darker color denotes higher returns. The solid curved line is the frontier where annual real returns are 3 percent. The dashed curve lines indicate (from left to right) returns of −6, −3, 0, and 6 percent annual returns, respectively. The horizontal line shows the level of average annual supply from 2000 to 2010. Annual housing supply in 2000–2010 is estimated as housing in urban area in the 2010 census minus that in the 2000 census times 0.82 (assuming housing stock in 2000 depreciates by 2 percent per year), then divided by 10.

Our assumptions of urban growth could have been generous for Tier 3 cities. These cities are also most likely to land in the high-supply scenario, so the outlook may not be as optimistic as Figure 6C suggests.

Overall, these simulations show a possible range of prices, depending on income growth, potential urban population growth, and housing supply growth. Of these three determinants of future prices, the Chinese government has the most capacity to shape housing supply growth. With sufficiently controlled housing supply, current prices can be maintained, but if housing supply continues to aggressively deliver new space, prices will fall.

Conclusion

In many respects, China looks like a classic housing bubble. Housing prices have soared. New construction is enormous. Vacancies are large and pervasive. It is tempting to view these events from afar and conclude that a price drop is imminent.

As we have tried to demonstrate, this scenario is far from certain. Chinese homebuyers appear to be investing for the long run and are unlikely to sell voluntarily even if home prices decline. Nor are they heavily leveraged, so repossessions and liquidations of homes are unlikely. Chinese developers are more leveraged, but are cozy with state banks, so their loans are likely to be restructured if necessary. Even if banks repossess properties from developers, they are unlikely to dump them on the market. Compared to Chinese stocks, more inertia is built into China's housing market.

In addition, there is the critical role of the Chinese government in housing markets. The demand for urbanization in China is large, so if the government acts to sharply restrict new supply, it can probably maintain prices at close to current market levels. If the government buys up excess housing inventory—perhaps to convert it to social housing as it started to do in 2015—it can further bolster softening prices. There does appear to be a feasible public path towards housing price stability.

Yet that path may create significant social costs. Construction employment would plummet. Millions of Chinese may lose the apparent productivity advantages associated with living in Chinese cities (Chauvin, Glaeser, Ma, and Tobio forthcoming). Local governments would lose the financial autonomy from land sales and taxes that has been their institutional basis.

The alternative for the Chinese government is to accommodate high levels of construction and housing supply. As we have showed, this will lead to very low or negative expected returns to investment in housing. The welfare of potential new buyers will rise, but current owners will suffer losses.

Bursting real estate bubbles have traditionally done great harm when they are associated with financial crises. Bubbles that burst without banking meltdowns, as in 1980s Los Angeles, are temporary events that seem to cause little long-run damage. Going forward, an important step is to secure China's financial system, rather than focus solely on maintaining high housing costs in Chinese cities.

References

Baer, Drake. 2015. "A Simple Design Innovation Let a Chinese Entrepreneur Build a 57-Story Skyscraper in 19 days." *Business Insider*, June 19. http://www.businessinsider.com/chinese-skyscraper-built-in-19-days-with-modular-design-2015-6.

Bank of International Settlements. No date. BIS Statistics Explorer, Table F3.1 titled "Total credit to households (core debt)," Q1 2016. http://stats.bis.org/statx/srs/table/f3.1.

Cai, Hongbin, J. Vernon Henderson, and Qinghua Zhang. 2013. "China's Land Market Auctions: Evidence of Corruption?" *Rand Journal of Economics* 44(3): 488–521.

Cai, Hongbin, Zhi Wang, and Qinghua Zhang. Forthcoming. "To Build Above the Limit? Implementation of Land Use Regulations in Urban China." *Journal of Urban Economics.*

Case, Karl E., Robert J. Shiller, and Anne Thompson. 2012. "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets." NBER Working Paper 18400.

Chamon, Marcos D., and Eswar S. Prasad. 2010. "Why Are Saving Rates of Urban Households in China Rising?" *American Economic Journal: Macroeconomics* 2(1): 93–130.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. Forthcoming. "What Is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics.*

Chivakul, Mali, W. Raphael Lam, Xiaoguang Liu, Wojciech Maliszewski, and Alfred Schipke. 2015. "Understanding Residential Real Estate in China." IMF Working Paper no. 15/84, International Monetary Fund.

Deng, Yongheng, Joseph Gyourko, and Jing Wu. 2012. "Land and House Price Measurement in China." NBER Working Paper 18403.

Economist, The, **The Data Team.** 2016. "American House Prices: Reality Check." Daily Chart. August 24. http://www.economist.com/blogs/graphicdetail/2016/08/daily-chart-20.

Fang, Hanming, Quanlin Gu, Wei Xiong, and Li-An Zhou. 2015. "Demystifying the Chinese Housing Boom." Chap. 2 in *NBER Macroeconomics Annual 2015*, vol. 30, edited by Martin Eichenbaum and Jonathan Parker. University of Chicago Press.

Glaeser, Edward L. 2013. "A Nation of Gamblers: Real Estate Speculation and American History." *American Economic Review* 103(3): 1–42.

Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko. 2010. "Did Credit Market Policies Cause the Housing Bubble?" Policy Brief, Harvard Kennedy School, May.

Glaeser, Edward L., Joseph Gyourko, and Albert Saiz. 2008. "Housing Supply and Housing Bubbles." *Journal of Urban Economics* 64(2): 198–217.

Glaeser, Edward L., and Charles G. Nathanson. Forthcoming. "An Extrapolative Model of House Price Dynamics." *Journal of Financial Economics.*

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions." *Journal of Economic Perspectives* 19(4): 67–92.

Hsieh, Chang-Tai, and Enrico Moretti. 2015. "Why Do Cities Matter? Local Growth and Aggregate Growth." NBER Working Paper 21154.

Huang, Zhen. 2013. "A Study of Household Finance in China." Doctoral dissertation, Lingnan University, Hong Kong. Retrieved from http://dx.doi.org/10.14793/econ_etd.25.

Li, Hongbin, and Li-An Zhou. 2005. "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China." *Journal of Public Economics* 89(9–10): 1743–62.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124(4): 1449–96.

National Bureau of Statistics of China. 2014. *China Statistical Yearbook.* http://www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm.

Painter, Gary, Lihong Yang, and Zhou Yu. 2003. "Homeownership Determinants of Chinese Americans: Assimilation, Ethnic Concentration, and Nativity." Unpublished paper.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics* 125(3): 1253–96.

Shiller, Robert J. 2000. "Irrational Exuberance." Princeton University Press.

US Bureau of the Census. 2007. *Census Atlas of the United States.* https://www.census.gov/population/www/cen2000/censusatlas.

Wei, Shang-Jin, and Xiaobo Zhang. 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119(3): 511–64.

Williams, John C. 2013. "Bubbles Tomorrow and Bubbles Yesterday, But Never Bubbles Today?" Speech for presentation to the National Association for Business Economics, September 9. http://www.frbsf.org/our-district/press/

presidents-speeches/williams-speeches/2013/
september/asset-price-bubbles-tomorrow-
yesterday-never-today/.

**Wu, Jing, Joseph Gyourko, and Yongheng
Deng.** 2015. "Evaluating the Risk of Chinese
Housing Markets: What We Know and What We

Need to Know." NBER Working Paper 21346.

Wu, Xun. 2015. "An Introduction to Chinese
Local Government Debt." Unpublished paper.

Xie, Yu, and Yongai Jin. 2015. "House-
hold Wealth in China." *Chinese Sociological
Review* 47(3): 203–229.

Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda

Bei Qin, David Strömberg, and Yanhui Wu

Around midnight on March 29, 2014, some Chinese internet night owls noticed that the hazard factor of P-Xylene (PX) had been changed from "low" to "high" on Baidu Encyclopedia—the Chinese equivalent to Wikipedia. The next morning, hundreds of protestors assembled in Maoming—a city in southern China's industrial heartland—where a large-scale PX plant was planned. At 8:38 am, a message with pictures of the protest was posted on Sina Weibo—the Chinese equivalent to Twitter. Tens of thousands of people joined the protests, demanding responses from local officials, burning a car, and throwing bottles until police dispersed the protesters with tear gas and batons. The next day, pictures of bloodied protestors circulated online. Thousands of posts debating the PX project and condemning the government's action appeared on various social media platforms.

In the era of advanced information technology, social media can in some cases provide a huge information shock to a country like China, in which information and public communication has been limited by government control. How does such an information shock generated by social media affect the participation of Chinese citizens in political events? And how does the Chinese government respond

■ *Bei Qin is Assistant Professor at the School of Economics and Finance, Faculty of Business and Economics, University of Hong Kong, Hong Kong. David Strömberg is Professor at the Institute for International Economic Studies, Stockholm University, Stockholm, Sweden. Yanhui Wu is Assistant Professor of Finance and Business Economics, Marshall School of Business, University of Southern California, Los Angeles, California. Their email addresses are beiqin@hku.hk, david.stromberg@iies.su.se, and yanhuiwu@marshall.usc.edu.*

to this information shock? A first step in addressing these questions is to document the information published and communicated on China's social media.

In this paper, we document basic facts regarding public debates about controversial political issues on Chinese social media. Our documentation is based on a dataset of 13.2 billion blog posts published on Sina Weibo—the most prominent Chinese microblogging platform—during the 2009–2013 period. We also discuss the implications of our findings.

Our primary finding is that a shockingly large number of posts on highly sensitive topics were published and circulated on social media. For instance, we find millions of posts discussing protests such as the anti-PX event in 2014, and these posts are informative in predicting the occurrence of specific events. We find an even larger number of posts with explicit corruption allegations, and that these posts predict future corruption charges of specific individuals.

This type of social media content may increase the access of citizens to information and constrain the ability of authoritarian governments to act without oversight. In China, social media clearly have the potential to play such a role because of their immense popularity. Nearly one-half of the population has access to the internet, and two of every ten Chinese actively use Weibo. Every day, millions of blog posts are produced, exchanged, and commented upon. Many of these posts reach thousands or even millions of readers. It would be very costly for China's government to monitor millions of users, especially when they are sometimes posted in large surges in the middle of the night.

However, social media also provide authoritarian governments with new opportunities for political control as noted by Egorov, Guriev, and Sonin (2009), Shirky (2011), Morozov (2012), and Lorentzen (2014). Social media messages are transmitted in electronic form through an infrastructure that is typically controlled by the government. Recent advances in automated text analysis, machine learning techniques, and high-powered computing have substantially reduced the costs of identifying critical users and censoring messages (Edmond 2013). Governments can use these methods to track and analyze online activities, to gauge public opinion, and to contain threats before they spread.

In an autocracy, how effective can a surveillance system based on social media be? We evaluate this possibility by exploring a very simple surveillance tool, and we find that social media can be very effective for protest surveillance. Most of the real-world protests and strikes that we study can be predicted one day in advance based on social media content. The method of detection is simple, and the cost is low.

Indeed, Chinese government agencies across the country have invested heavily in surveillance systems that exploit information on social media. The result is illustrated by another anti-PX event, which played out differently, a year earlier. Some citizens in Chengdu posted messages on Sina Weibo calling for demonstrations against the launch of a local PX project on Saturday, May 4, 2013. This call was picked up by the Chengdu government, which immediately took measures including making Saturday and Sunday of that week into working days and requiring students to be in school on those days. Consequently, the demonstration in Chengdu was

limited to a few hundred participants, whereas a similar anti-PX protest in Kunming on the same day drew more than 10,000 participants. In this case, surveillance of social media content helped the government contain a protest.

Another important surveillance function of social media is to monitor local governments and officials. In China, many political and economic decisions are delegated to local governments. These decisions need to be monitored, but local news and internal reports are likely to be distorted because local politicians control the local press and administration. In contrast, national politicians regulate social media. In social media, relentless complaints about local officials are abundant. Posts exposing officials who wore Rolex watches, lived in mansions, or had inappropriate girlfriends have resulted in investigations and dismissals. Not surprisingly, we observe millions of posts with explicit corruption allegations in our data.

We find that social media posts related to corruption topics are effective for corruption surveillance. These posts help identify when and where corruption is more prevalent. Furthermore, we can predict which specific politicians will later be charged with corruption, up to one year before the first legal action. This result indicates that social media information can be a useful instrument to identify corruption, which probably holds in other countries as well.

It is not without risk to post information about protests and negative information about local leaders, and some users have been punished for doing so (Freedom House 2012; Reporters Without Borders 2013). However, as far as we know, there is no previous systematic research on the extent and effectiveness of this form of policing. We investigate the scale of these practices by tracking regular users who post on sensitive issues and seeing whether their accounts are subsequently closed.

Government can also use social media as a propaganda channel. In the above anti-PX examples, governments conducted vigorous propaganda campaigns via their accounts on social media. In one case, government accounts continuously blogged that "PX is no more carcinogenic than coffee." The extent of this type of government postings is difficult to know because they emanate from a plethora of accounts across regions and government levels. In 2012, Sina Weibo reported that approximately 50,000 accounts were operated by government offices or individual officials, but there are no external estimates of this kind of government presence on social media.

We seek to identify government accounts from user names and text analysis of the posts in our data. Based on this approach, we estimate that there are 600,000 government-affiliated accounts, which contribute 4 percent of all posts about political and economic issues on Sina Weibo, using a measure that includes government organization, mass-organization, and media users. Even when limited to the most restrictive definition of government user, Sina Weibo's reported number substantially underestimates the government presence on Sina Weibo.

Our findings challenge a popular view that an authoritarian regime would relentlessly censor or even ban social media. Instead, the interaction of an authoritarian government with social media seems more complex. From the government point of view, social media is not only (1) unattractive as a potential outlet for organized social protest but is also (2) useful as a method of monitoring local

officials and (3) gauging public sentiments, as well as (4) a method for disseminating propaganda. From the point of view of citizens, any perceived benefits of social media need to be evaluated in a context of (5) possible pervasive policing, punishment, and (6) censorship of such media. As mentioned above, this complex interaction has been discussed by researchers. However, rigorous empirical study on this subject is scant. An exception is Enikolopov, Makarin, and Petrova (2016), who find that social media affected protests in Russia in 2011. The other exception is censoring, which has been studied extensively (for example, Bamman, O'Connor, and Smith 2012; Fu, Chan, and Chau 2013; King, Pan, and Roberts 2013, 2014; Zhu, Phipps, Pridgen, Crandall, and Wallach 2013). In contrast, we will not discuss censoring, but present evidence key to all the other five points listed above.

We begin with an overview of the development of social media in China, followed by a description of the data. We proceed to analyze protests, corruption, and the government presence on Sina Weibo. Finally, we discuss the implications of social media, given our results.

Background on the Development and Regulation of Social Media in China

By 2013, there were 618 million Chinese internet users, accounting for approximately 46 percent of the Chinese population. This rate is slightly higher than the global average of 39 percent (China Internet Network Information Center 2014; International Telecommunication Union 2013). Of China's internet users, 281 million (45 percent) actively participated in microblogging, which refers to social media outlets that focus on short messages, individual images, or perhaps video links (as opposed to social media like Facebook that have the potential for longer-form or more detailed communication).

The popularity of microblogs is a recent phenomenon. In 2006, Chinese people became aware of Twitter; the next year, major Chinese counterparts—Fanfou, Digu, and Jiwai—were launched. The number of microbloggers grew slowly at first. After the Urumqi riots in July 2009, the Chinese government not only blocked Twitter and Facebook but also shut down most domestic microblogging services. The microblog market in China was then essentially vacant until Sina Weibo appeared in August 2009, and NetEase, Sohu, and Tencent followed in 2010. The number of microblog users surged from 63 million at the end of 2010 to 195 million by mid-2011 (China Internet Network Information Center 2011).

Sina Weibo is a hybrid of Twitter and Facebook: up to 140 Chinese characters per tweet, embedded pictures or videos are allowed, and users can send private messages, comment, and repost. With its easy access and use, Sina Weibo soon became the most popular microblogging platform in China. By 2010, it had 50 million registered users, and this number doubled in 2011, reaching a peak of over 500 million at the

end of 2012. Since 2013, Sina Weibo has lost some ground to WeChat, a cellphone-based social networking service, but has remained an influential platform.[1]

In recent years, microblogs have provided some of the most extensive and vivid discussions and debates in China. According to the Reports on Public Opinion 2010–2013, for example, in 2012, the two most popular Facebook-type social media platforms in China—Renren and Kaixin—covered the top 20 public events listed by the Public Opinion Monitoring Agency (which is run by the government newspaper *People's Daily*) in 20 million posts. However, Sina Weibo—the leading microblog site at the time—covered the same events in more than 230 million posts.

The Chinese central government has the power and tools to limit sensitive content on social media, if it wishes. The sensitive social media content that we find is unlikely to be the result of inability to clean up social media. Instead, the government must perceive benefits from leaving this content visible. Two primary tools are used to limit content.

The first tool is policing—to punish users who post sensitive content to induce self-censorship and to avoid content being posted. Policing is carried out by tens of thousands of information officers and internet monitors who are active at all levels of government (Chen and Ang 2011). Local politicians may use their own internet police to suppress negative information about the regions under their administration, even if blogging about this information is tolerated or encouraged by the central government. Users who post undesired content may receive warnings, have their accounts shut down, and even be imprisoned. Reporters Without Borders (2013) documented a total of 69 netizens in Chinese jails as of February 2013, although the number of unreported cases may be much bigger. There is no previous systematic research on the extent of policing, as far as we know. We will investigate the scale of these practices by tracking regular users who post on sensitive issues to see whether their accounts are subsequently closed.

Of course, personal punishments can occur only if a user is identified. The Chinese government initially allowed users on Sina Weibo to post anonymously. In March 2012, the media control authority required users to reveal their identities to social media providers. However, three years later, service providers had yet to implement this regulation in its entirety.

The other control tool is censoring. Censorship is regulated by the national Propaganda Department of the Chinese Communist Party, as well as by a number of national media control offices. However, in practice, censorship is implemented largely by private service providers who are registered in Beijing. The estimated extent of censorship of Sina Weibo ranges from 0.01 percent of posts by a sample of prioritized users, including dissidents, writers, scholars, journalists, and VIP users (Fu, Chan, and Chau 2013) to 13 percent of posts on selected sensitive topics (King, Pan, and Roberts 2013). King, Pan, and Roberts find that the Chinese government

---

[1] Our data for this study ends in 2013. In that year, the number of Weibo users dropped by almost 28 million and the utilization ratio dropped by 9.2 percentage points, according to the China Internet Network Information Center (2014).

allows criticism of officials and bureaucrats but censors information about collective action more strictly (27 percent of posts censored). More generally, Bamman (2012) and Fu, Chan, and Chau (2013) find that internet censorship in China focuses on political and minority group issues. Zhu et al. (2013) find that the implementation of censorship is speedy: 30 percent of deletions occur within the first half hour and 90 percent within 24 hours. Unlike these studies, our paper examines the content that is available on microblogs rather than what is removed.

The government can affect debates and sentiments on social media by actively posting their own content. Chinese governments at all levels have opened microblog accounts in an effort to steer public opinion. In 2012, Sina Weibo reported that approximately 50,000 accounts were operated by government offices or individual officials. Governments at different levels also hire internet trolls, nicknamed "the 50-cent party" because some are paid at a piece-rate of 50 cents per post. Some commentators, at the behest of local politicians, may post fake positive reviews about the politicians or the regions under their administration.

Data on Sina Weibo Posts, 2009–2013

Our primary data, Sina Weibo posts, were collected by Weibook Corp. During the 2009–2013 period, this company executed a massive data collection strategy to download the posts of active users. First, the firm identified 200–300 million authentic active persons using Sina Weibo based on the individual's information and interaction with other users. Second, they categorized users into six tiers based on the number of followers. They downloaded the microblogs of the top-tier users at least daily, the second and third tiers every 2–3 days, and the lowest tier downloaded on a weekly basis. Thus, the data include at least some posts that are later censored. For each post, they provided the content, posting time, and user information (including self-reported location).

In total, the dataset that we study contains 13.2 billion posts published from 2009 to 2013. According to our estimation, the Weibook data contains approximately 95 percent of the total posts published on Sina Weibo.[2]

From this Weibook database, we extract microblogs mentioning any of approximately 5,000 keywords that are related to social and political topics. The keywords fall into two groups. The first group refers to categories of issues, including political positions from the central to the village levels, names of top political leaders, social and economic issues (such as corruption, pollution, food and drug problems, disasters and accidents, and crimes), and collective action events (such as strikes, protests, petitions, and mass conflicts). Some words occur at a very high frequency.

---

[2]Using the Sina Weibo public API, we downloaded all posts containing the neutral words "ya" or "hei" during four five-minute intervals each day and then divided by the average share of posts that contained these words and the average share of posts contained in these five-minute intervals in a day. We were not able to do this for later years because the public timeline API denied access.

We collect a random sample that consists of only 10 percent of the posts mentioning these words. The second group of keywords refers to specific collective action events that we have recorded, including those events noted in censorship directives issued by the Chinese media control authorities and a large number of massive collective action events from 2009 to 2013. In total, our extracted data contain 202 million posts from 30.6 million different users.[3]

Conflicts, Protests, and Strikes

We analyze 545 large collective action events that took place in mainland China between 2009 and 2012. The list of events was derived from coverage by Radio Free Asia, a nonprofit radio station based in Washington, DC. We classify these collective action events into four categories, ranked by sensitivity. The first category contains the most sensitive events, which involve direct confrontations between government and the public, including those involving riots and violence. The second category contains protests, including street demonstrations and mass protests, which are typically more expected and organized, less violent, and even often approved by the government. In several cases, protests evolved into riots, as in the Wansheng Event in Chongqing in 2012; we code such events as "conflicts." The third category contains strikes, including strikes in factories and schools and among taxi drivers. The last category includes anti-Japan demonstrations.

We select keywords that identify posts about each event type and extract all posts that mention these keywords from the entire Weibook dataset. The method for extracting keywords is described in the online Appendix.

Content and Users

We initially thought that coverage of these events on social media would be very limited. As just noted, it is well-documented that Chinese internet users have been punished after posting about protests and other collective action events (for example, Freedom House 2012) and that these types of posts are censored (King, Pan, and Roberts 2013). But to our surprise, we found a large number of posts covering even the most sensitive collective action events based on our classification. In our data, we identify 382,000 posts in the "conflict" category and over 2.5 million posts in the "protest" category. As this finding attracted some doubt when we presented the paper, we use various approaches to examine it further.

As a starting point, we characterize the "hot topics" in posts about collective action. These topics are identified by words that are used more often in collective

---

[3] To analyze word frequencies in the Chinese text, we use the Stanford Word Segmenter to segment the words in each microblog post. We remove stopwords, punctuations, URLs, usernames, and non-Chinese characters (except meaningful English abbreviations) from the text. We exclude words with more than 30 characters and words occurring fewer than five times. We obtain 3.2 million distinct words and 6.0 billion "tokens" (or word occurrences).

action posts than in the entire sample of posts. More precisely, we compare the frequency of each word in a given category with the overall frequency of this word in our dataset, as in Kleinberg (2006). Table 1 presents the hot-topic words in order of statistical significance. For example, in the conflict category, "suppression" has the most abnormally high use. Note that the topic ranking is not based on the absolute frequency of the words, but on the use of the word relative to its general usage. For example, "tear-gas bomb" is ranked above "government" because the latter word is more commonly used in general. Other topic words in this category include "police" "violence," "revolt," and "opening fire."

To characterize these data further, we investigate a random sample of 1,000 posts for each of the first three collective action categories in Table 1 and for anti-Japan demonstrations. We manually code whether and how the posts cover a particular type of event, with the results shown in Table 2. Out of our 1,000 post samples, the share of posts that actually cover the events ranges from 50.4 percent for the anti-Japan category to 31.2 percent for the strike category. The more sensitive events like conflict and protests receive more coverage in the form of general and retrospective comments. Here are a few examples to convey a sense of our coding.

"I saw hundreds of policemen armed with weapons. Fire was everywhere after some gas containers were bombed." [Conflict, ongoing]

"A big crowd is gathering in front of the government building, holding 'No Forced Demolition of House' signs." [Protest, ongoing]

"The money from selling lands all went into the pockets of officials. They are nothing but gangsters. We have no choice but to rebel." [Protest, general]

"Seriously? Taxi-drivers strike again!" [Strike, ongoing]

"Low wages, cheap labor. We make tons of Made-in-China, but receive little in return. Migrant workers, strike!" [Strike, general]

"We will march towards the Japanese Embassy today. Gathering at the People's Square at 10 am. Anyone wanna join?" [Anti-Japan, forthcoming]

This Weibo content predicts real-world collective action events. Table 3 reports the average number of posts for each event type published by users in the prefecture where an event took place on the day of the event and on the day before. Suppose, for example, that a strike took place in a given prefecture and day. We then count the number of posts that contain any of our keywords related to strikes by users from this prefecture on the same day as the strike and on the day before the strike. We do this for all strikes and report the averages in the table. The average number of posts is much higher on the day of and the day before a collective action event than on other days. To make sure that the posts the day before really do predict these events,

Table 1
Hot Topics by Category

A. Conflict (sensitivity: very high) # of posts: 382,232			B. Protest (sensitivity: high) # of posts: 2,526,325		
Frequency	*Word*	*Translation*	*Frequency*	*Word*	*Translation*
322,797	镇压	suppression	647,711	示威	demonstration
32,117	冲突	conflict	534,784	静坐	sit-in
19,124	警民	police and people	430,112	自焚	self-immolation
17,460	催泪弹	tear-gas bomb	260,574	讨薪	ask for compensation
31,161	矛盾	contradictory	346,836	游行	parade
40,286	警察	police	164,367	请愿	petition
14,271	官民	officials and people	113,936	示威者	demonstrators
31,935	暴力	violence	109,339	堵路	stops up the road
130,036	被	by	166,600	抗议	protest
74,391	政府	government	101,845	集会	assembly
12,002	宽恕	forgiveness	118,262	农民工	migrant workers
12,764	武力	military force	103,975	思	thinking
18,951	军队	army	80,481	静静	static
29,566	民众	populace	60,237	闲谈	chat
14,701	叙利亚	Syria	58,318	人非	shortcomings of people
20,170	抗议	protest	72,753	民工	laborers
60,068	人民	people	63,719	白宫	White House
21,521	村民	villagers	130,198	坐	sitting
10,264	起义	revolt	60,957	己	oneself
10,150	开枪	gunfire	37,904	玩火自焚	being made to pay for one's evil doings

C. Strike (sensitivity: medium) # of posts: 1,348,964			D. Corruption # of posts: 5,326,897		
Frequency	*Word*	*Translation*	*Frequency*	*Word*	*Translation*
1,361,854	罢工	strike	1,455,878	贪污	embezzlement
69,068	罢课	student strike	1,658,687	腐败	corrupt
101,887	工人	workers	681,055	公款	government money
98,822	电脑	computer	674,503	受贿	take bribe
65,557	出租车	taxi	556,609	贿赂	give bribe
164,549	泪	tears	975,187	官员	officials
46,219	工会	trade union	393,125	廉政	honest government
91,051	抓狂	driven nuts	639,293	利益	benefit
55,687	司机	drivers	1,002,491	政府	government
48,845	集体	collective	245,606	挪用	diverting
52,066	员工	staff	512,006	集团	group
157,937	今天	today	201,891	吃喝	food and drink
24,477	的士	taxi	153,731	职权	authority
22,559	法国人	French	572,569	钱	money
51,479	上班	going to work	247,942	贪官	corrupt officials
16,290	罢市	merchant strike	156,363	滥用	abusiveness
40,827	抗议	protest	291,309	原	former
86,612	手机	cell phone	288,287	干部	cadres
17,679	罢	strike	123,827	行贿	bribery
41,586	工资	wages	126,820	情妇	mistress

Note: This table presents the hot-topic words, ordered by how abnormally high the frequency of the word is in the posts that cover a particular category (for example, "conflict") compared to that in the entire dataset.

Table 2
Collective Action Posts

	Total posts containing keywords	Out of a random 1,000 post sample:				
		Posts actually about defined topic	Forthcoming event	On-going event	Past event	General comments
Conflict	382,232	398	1	11	156	230
Protest	2,526,325	317	2	19	172	124
Strike	1,348,964	312	5	178	39	90
Anti-Japan	2,506,944	504	9	188	42	265

Note: For each category of hot topic, out of the total instances of the words used, we investigate a random sample of 1,000 posts. We manually coded whether and how the posts cover a particular type of event.

Table 3
Event Prediction and Detection
(# per prefecture and day)

Variables	Conflict	Protest	Strike	Anti-Japan	Coal mine Accident
# Weibo posts day of event	6.1	62.6	167.3	2,036.6	3.0
# Weibo posts day before event	3.4	54.3	48.1	924.6	0.7
# Weibo posts on day with no event	0.7	4.4	2.5	4.5	1.2

Note: For each event type, this table reports the average number of posts published by users in the prefecture where an event took place on the day of the event (first row) and on the day before it (second row). The third row shows the analogous average number of posts on days when no such event took place.

and do not arise for spurious reasons such as miscoding of the posting dates, the final column of Table 3 examines coal mine accidents. These should clearly not be predicted by microblogging posts. We obtain data on the locations and days of 253 coal mine accidents during the 2010–2012 period from the State Administration of Coal Mine Safety. We search for word strings related to coalmine accidents in our dataset. While coal mine accidents are covered much more on the day of the accident, they are not discussed more frequently on the day before the accident than on other days. This finding is related to Acemoglu, Hassan, and Tahoun (2014) who find that the number of protesters in Tahrir Square on any given day was predicted by the number of tweets with Tahrir hashtags. Of course, media control in Egypt in 2011 was much less strict than in China.

　　We also carry out a more detailed analysis that examines how well microblog discussion, compared to newspaper coverage, predicts when and where these collective action events would occur. We use news reports from 62 general-interest newspapers that covered at least one of these events during the 2010–2012 period.

While microblogs are highly significant in predicting where and when collective action events took place, newspaper coverage of this event type is uninformative.[4]

In sum, we find literally millions of posts on Sina Weibo that discuss sensitive collective action events. Many of these posts were posted before or concurrent with the events and predict the events. The fact that people began discussing events before they happen indicates that Sina Weibo may be used to organize or at least to coordinate collective action events.

To investigate whether the users who post this type of sensitive content are identified and perhaps punished, we examine the subsequent posts of the users who blogged about collective action events. Sixteen percent of these posts are the last post published by a user in the data that contain any of the 5,000 keywords. In the "conflict" and "protest" categories, the corresponding rates are 17 and 23 percent, respectively. The share of users who exit from our data within five or ten more posts is slightly higher in the full data (38 and 49 percent) than in the conflict and protest categories (33–34 and 41–42 percent). In short, we find that users who posted on these topics continued to post to a similar extent as other users, indicating that their accounts were not more likely to be closed nor were their posts muted.

Another way to investigate whether users are concerned about censorship is to see whether posts on sensitive topics tend to come from user accounts with relatively few posts—which could be a sign that users create separate Sina Weibo accounts for controversial messages, perhaps even with hidden IP addresses. However, the average number of posts from users who blog on sensitive topics is not significantly lower than that of a randomly drawn comparison sample of users (drawn using the number of posts by each user as sampling weights). The bottom line is that, although there are documented cases of people being punished after posting sensitive content, this does not seem to happen on a large scale. Broadly speaking, people do not seem afraid to post on these topics.

Surveillance

We examine how effective social media information is for government surveillance of collective action events. As noted earlier, government agencies across China have invested heavily in software to track and analyze online activities, to gauge public opinion, and to contain threats before they spread (as reported in Epstein 2013). Presumably, these government agencies desire an early warning system for collective action events. We imagine that such a system might work in two steps. The first step is automated: just have software be alert for days when mentions of certain events spike on social media. Second, hire actual human beings to read the posts published on those days.

---

[4]Details of this regression analysis are reported in Table 3 in the working paper version of this paper (Qin, Strömberg, and Wu 2017). We regress an indicator for an event occurring on the number of Weibo posts from users in a prefecture that mention the event keywords on the event day, in one set of regressions, or on the previous day, in another set of regressions.

We explore this automation-plus-manpower method. We study 316 prefectures for all days from January 1, 2010, to December 31, 2012. We study anti-Japan events and strikes because these events are less likely to be censored and hence we have similar Weibo post information as the government. For the first step, we have the software alert us all days when a user from a city writes a post mentioning any of our event keywords. One day in advance, this alarm would identify 42 of the 43 anti-Japan events and 115 of the 130 strikes. This would require investigating 109,726 and 104,241 prefecture-days, respectively (out of 346,020). It is very easy to improve this crude method, both to find more events and at the same time reduce the number of observations searched.

For the second step, we manually read the strike-related social media posts in the 100 prefecture-days with the most strike-related posts, which took us about two hours in total. (Remember, these are very short messages!) Thus, our estimated time-cost of analyzing the 104,000 prefecture-days necessary to discern 115 strikes one day in advance is 2,080 person-hours. This would be the aggregate time cost for all prefectures spread over three years. From that perspective, the cost is very small. The bottom line is that collective action events that are large enough to pose potential threats to the regime are likely to be easily detected using social media data, and they can be detected one day in advance.

Several other interesting points emerge from the above analysis. One is that, when manually reading all of the strike-related social media posts for the top 100 days, we detected 23 strikes during these prefecture-days that were also in our original dataset. However we also identified 14 additional strikes. Our procedure thus shows how social media can be used as a data-collection device in countries where data on relevant social outcomes are scarce but data from social media are abundant.

Monitoring Local Politicians

Can social media provide information relevant to holding local politicians accountable to higher-level politicians? We will first describe the content on Sina Weibo related to corruption. We then analyze 200 corruption cases involving high-ranking Chinese government or Communist Party of China leaders, drawing on examples of corruption from the Central Disciplinary Committee of the Communist Party and the Ministry of Supervision, as well as news reports published by the Chinese government's mouthpiece news agency, Xinhua News. We find that Weibo posts predict a number of corruption cases one year in advance.

To examine coverage of corruption on social media, we combine two types of microblog posts: those mentioning politicians or political positions and those mentioning corrupt behavior. For the first category, we retrieve posts that mention any major political position at the central, provincial, prefectural, county, or village levels. We obtain over 11 million total posts in this category. Column 1 of Table 4 shows the number of posts covering each position or top leader. The table is sorted by number of posts per position, shown in Column 2—(for example, there are 31

Table 4
Coverage of Politicians

Position	# Posts (1)	# Posts per position (2)	% Posts discussing specific corruption cases (3)	Measure of sentiment (4)
Xi Jinping	1,374,780	1,374,780	0.23	0.88
Wen Jiabao	1,318,345	1,318,345	0.15	0.51
Li Keqiang	401,451	401,451	0.14	0.81
Hu Jintao	326,621	326,621	0.10	1.16
Provincial governor	728,386	23,496	1.88	-0.19
Provincial Party secretary	403,074	13,002	1.91	0.52
City mayor	3,431,471	10,305	1.39	0.17
City party secretary	718,856	2,159	2.81	0.28
County governor	719,634	251	1.21	-0.70
County Party secretary	324,522	113	4.40	-0.88
Village chief	1,053,346	25	0.65	-0.51
Village Party secretary	144,742	3	4.26	-1.40

Note: Column 1 shows the number of posts covering each position or top leader. The table is sorted by column 2—the number of posts per office. Column 3 shows the estimated percentage of posts mentioning a leader's position that discuss specific corruption cases. Column 4 presents a broader measure of people's sentiments towards the leader or type of leader. See text for details.

offices for provincial-level positions). Xi Jinping, the current president of China and the general secretary of the Communist Party of China, is the most discussed leader, with over 1.3 million posts mentioning his name, followed by Wen Jiabao, the former prime minister of China. In general, officials at higher levels are more extensively discussed, and executive positions are covered more than are party secretaries.

Regarding the second category of posts mentioning corrupt behavior, we search for words that are widely used to describe corrupt behavior, wrongdoing, and punishment of officials. The hot topic words in this category are "embezzlement," "corrupt," "government money," and "give bribes/take bribes" (as shown earlier in Table 1, Column 4). We identify over 5.3 million posts in this category.

To characterize posts about corruption, we manually inspect 1,000 randomly selected posts that use these terms. Most of these posts make general comments on corruption. Of the 419 posts that discuss specific corruption cases, 293 were written after the government had taken action. However, 126 posts discuss instances of corruption before government action. These 126 posts can be divided into two main types. One type targets specific government officials, illustrated by the following two examples.

"XXX, the Party secretary of XXX village, misused the money transferred from the central government for low-income villagers to pay his family members and relatives."

"XXX, the chief officer of XXX county, embezzled public money by awarding all major government project contracts to his brother's company. Even worse, he hired gangsters to stab people who reported his corruption to the upper-level government."

The other type of post conveys resentment of and anger toward certain corrupt officials. In most cases, these posts talk about positions and government divisions without specifying the names of the officials. Several examples are documented as follows.

"The black market for government positions in XXX prefecture is rampant. The price is getting higher and higher, the top officials in this prefecture are becoming richer and richer, and corruption will be more and more severe because the buyers need to make sufficient money to cover their costs."

"Without support from the prefecture party secretary and the vice governor, how dare these prefecture officials sell government positions? Crack down on the tigers!"

"Billions of money went into the pockets of local officials and their business partners! President Xi, Premier Li, and Secretary Wang in the Central Discipline Inspection Department, do you read our microblogs? Can you hear our voice? Please eradicate these corrupt officials! Right now!"

Column 3 of Table 4 shows the estimated percentage of posts mentioning a leader's position that discuss specific corruption cases. Specifically, we predict the probability that a post about a position discusses specific cases of corruption based on the frequencies of words used. For example, the last row shows that over 4 percent of all posts that mention village or county party secretaries also mention specific corruption cases. To obtain a broader measure of people's sentiments towards their leaders, we subtract the number of negative words from the number of positive words in all posts mentioning these leaders (using the National Taiwan University Sentiment Dictionary).

Column 4 of Table 4 shows that county and village party secretaries are the subject of the most negative sentiment and are associated with the largest share of corruption posts. One interpretation of this finding is that these two types of officials are usually viewed as the most powerful low-ranked politicians who have opportunities to be corrupt. An alternative view is that they are the most vulnerable officials in anti-corruption campaigns because they are at the bottom of the Chinese government hierarchy.

Estimating from our random sample of 1,000 posts, our dataset contains approximately 668,000 posts that discuss specific instances of corruption before government action. This provides a wealth of information for upper-level governments seeking to hold lower-level politicians accountable. Clearly, posts of this type are not censored by the central government.

We find some posts that do explicitly criticize top national leaders, although these posts do not contain explicit corruption allegations. Such posts claim, for example, that democracy and social stability decreased under Hu Jintao's reign, that the campaign against Bo Xilai was initiated by Xi Jinping as part of a political fight, and that Wen Jiabao shifted capital to Wenzhou to help the children of some top leaders.

Table 5
Mean Number of Posts, by Corruption Charge

| | 2–7 month lag | | 12–23 month lag | |
	Name	*Corruption*	*Name*	*Corruption*
Corrupt official	49.0	3.9	148.3	4.7
Noncorrupt official	44.4	0.4	121.1	1.8

Note: To investigate whether social media posts predict future corruption charges, we study a sample of 200 corruption charges. For comparison, we construct a matched control sample of 480 politicians who were not charged with corruption. We count the number of posts mentioning the name of each of these 680 politicians and the number of posts that mention both the politician and any word in our corruption category. We calculate the number of posts 2–7 months (as well as 12–23 months) before a corruption charge.

We find no evidence that users who post this type of sensitive content are systematically identified and punished. Users who post about corruption continue to post to the same extent as other users. It also seems that people are not afraid of posting concrete corruption allegations implicating powerful local politicians, as these posts are not generated from special accounts with few posts. One possible reason is that, even if local governments can identify bold users, it is risky for them to take action against these users because this may trigger an even bigger response in social media, which may further jeopardize the implicated politicians' career.

To investigate whether social media posts predict future corruption charges, we study a sample of 200 corruption charges: 15 at the national level, 39 at the provincial level, 114 at the prefecture level, and 32 at the county level. For comparison, we construct a matched control sample of 480 politicians who were not charged with corruption. The control politicians hold similar political positions to and are located in areas geographically near to the charged politicians.

We count the number of posts mentioning the name of each of these 680 politicians and the number of posts that mention both the politician and any word in our corruption category. We calculate the number of posts 2–7 months (as well as 12–23 months) before a corruption charge. Table 5 shows that corrupt and noncorrupt officials are mentioned in roughly the same number of posts 2–7 months before a corruption charge: 49 and 44.4 posts, respectively. However, corrupt officials appear much more frequently in posts that mention our corruption words (3.9 compared to .4). A similar pattern is found in posts published 12–23 months before a charge. Given the substantial difference in the number of corruption posts, it is not surprising that in more formal regression analysis, these posts are highly predictive of corruption charges.[5]

[5] In the working paper version of this paper (Qin, Strömberg, and Wu 2017), Table 5b presents the results of a regression of the corruption-charge indicator on the number of posts mentioning an official's name and corruption. The regression also includes the number of posts mentioning only the official's name, but this variable is never significant. Columns 2, 4, and 5 include dummy variables for case indicators, which are assigned the same value for an official charged with corruption and his or her matched

However, a significant number of corrupt officials fly under the social media radar. In particular, 133 officials later found to be corrupt were never mentioned in a corruption post two months or more before the first government action against them. From the perspective of the Chinese central government, which aims to crack down on corruption, a simple rule would be to investigate all officials with at least one corruption post. In our sample, this rule would lead to the investigation of 192 officials, 67 of whom were later charged with corruption.

Social media posts predict which top politicians will be charged with corruption one year before the first legal action. The reason may be that these individuals are indeed more corrupt. Another reason may be that the central government lifts censorship of corruption posts, or even plants its own corruption posts, about leaders who have lost political support and who will later be charged with corruption. To investigate the planting of stories, we examined a well-reported scandal involving Bo Xilai, a high-ranking official. We find that there was blanket censoring of posts mentioning Bo Xilai between the start of investigation on March 15, 2012, and the ultimate action undertaken by the Communist Party on September 28, 2012. We find no evidence that censorship focused on posts that were supportive of Bo Xilai or that there was a trend in corruption stories prior to his downfall.

In summary, a massive volume of Sina Weibo posts discuss corruption. These posts help identify the political positions, regions, time, and individuals involved in instances of corruption. The lack of censorship shows that for the central Chinese government, improved monitoring of lower-level officials outweighs the negative publicity of corruption coverage. The results also suggest that local politicians are at least not fully effective in imposing self-censorship on users or otherwise distorting the information.

Propaganda

Propaganda posted on social media is largely generated by government-affiliated users: government departments; mass organizations, such as schools and hospitals and industrial associations that are part of the public sector; and state-owned media (note that, per regulation, all general-interest media that are allowed to publish political content are owned and supervised by the government). We will study these types of high-powered users. We do not study internet trolls hired by the government to distract the public debate (King, Pan, and Roberts 2016). We use two approaches to identify government posts. On a small scale, we manually code posts published by randomly selected users; on a large scale, we use machine learning techniques to discern the language patterns used by well-known government users and thus predict which accounts are affiliated with the Chinese government. We then investigate the goals of these government-affiliated users.

officials. Corruption charges are strongly predicted by the number of posts mentioning corruption 2–7 and 12–23 months before the first legal action.

Table 6
Government Presence on Sina Weibo

	Users			Posts	
Type	*Percent*	*Estimated #*	*Standard deviation*	*Percent*	*Standard deviation*
Government	.5	149,746	66,801	.2	.1
Media	.5	149,746	66,801	2.3	1.6
Mass organization	1.0	299,491	94,233	1.1	0.5
Government-affiliated	2.0	598,982	132,590	3.6	1.6
Others	98.0	29,350,118	132,590		

Note: Based on 1,000 Sina Weibo users randomly selected from our entire database of 30 million users. A user is classified as a government user if the posts explicitly reveal the user's identity or are mostly related to the activities of a government function; mass-organization users are analogously coded. An account is classified as a media account if the posts reveal that the user is a media outlet or a division. "Government-affiliated" is the sum of "government," "media," and "mass organization."

Volume

In 2012, Sina Weibo reported that approximately 50,000 accounts were operated by government offices or individual officials. Our estimation shows that even when limited to the most restrictive definition of government user (excluding mass-organization and media accounts), this reported number substantially underestimates the government presence on Sina Weibo.

We manually code a sample of 1,000 Sina Weibo users randomly selected from our entire database of 30 million users. A user is classified as a government user if the posts explicitly reveal the user's identity or are mostly related to the activities of a government function; mass-organization users are analogously coded. An account is classified as a media account if the posts reveal that the user is a media outlet or a division. Table 6 shows the result. In our random sample of 1,000 users, 0.5 percent are government users, implying that there are approximately 150,000 (with a standard deviation of 67,000) government users in the entire dataset. State-owned media and mass-organization users contribute an even larger share. In total, these three types of government-affiliated users comprise 2 percent—or 600,000—users.

Thus, we estimate that the government-affiliated accounts contribute 3.6 percent of all posts in our database (with bootstrapped standard errors of 1.6 percent); see the right panel in Table 6. This percentage is greater than the 2 percent of government-affiliated users because these users publish more posts than others do. Note that these estimates are restricted to the sample of posts that mention words related to political and economic issues. Because we do not include users who write on other topics, the total number of government-affiliated accounts on Sina Weibo is likely to be higher than our estimates. However, the share of government posts may be substantially lower on topics outside politics and economics.

Identifying Government Affiliation by Language

We also use a linguistics-based approach to predict the probability that a user is affiliated with the government. We restrict our attention to the 5.6 million users

who publish more than five posts in our dataset. These users contribute more than three-quarters of all posts. First we identify 1,042 official, government-affiliated, and 538 newspaper accounts by manually inspecting the blogs of thousands of users with user names typically associated with these functions. These accounts comprise only a small subset of all government accounts, but we can use them to determine what words are characteristic of government accounts and then use this information to estimate the total number of government accounts. Second, we adopt a widely used classification algorithm (Support Vector Machine (SVM)) to identify this type of user from a one-percent sample (28,440) of randomly drawn users based on the frequencies of certain words in their posts.[6] Based on performance in other classification tasks, SVMs have been identified as one of the most efficient classification methods (Dumais, Platt, Heckerman, and Sahami 1998; Joachims 1998; Sebastiani 2002). In the SVM classification, a large number of words are important. However, just to give a sense of the classification, the words with the highest weight are "Communist Youth League," "Municipal Party Committee," and "Convention." Third, we use the estimated result from this test sample to predict the probability that each of the 5.6 million users is government-affiliated. We compute the average of this probability in total, by province, and by prefecture. This average probability provides a measure of the share of government-affiliated users across geographic regions.

At the national level, we estimate that 3.1 percent of the 5.6 million users are government affiliated (with a standard error of .8 percent). This is higher than the 2 percent in the overall sample, but the two estimates are nonetheless roughly consistent, because government users contribute more posts and are thus more strongly represented in the sample of users with more than five posts. The estimated share of posts published by government-affiliated users in this sample is 3.9 percent (with a standard deviation of 1.0 percent).
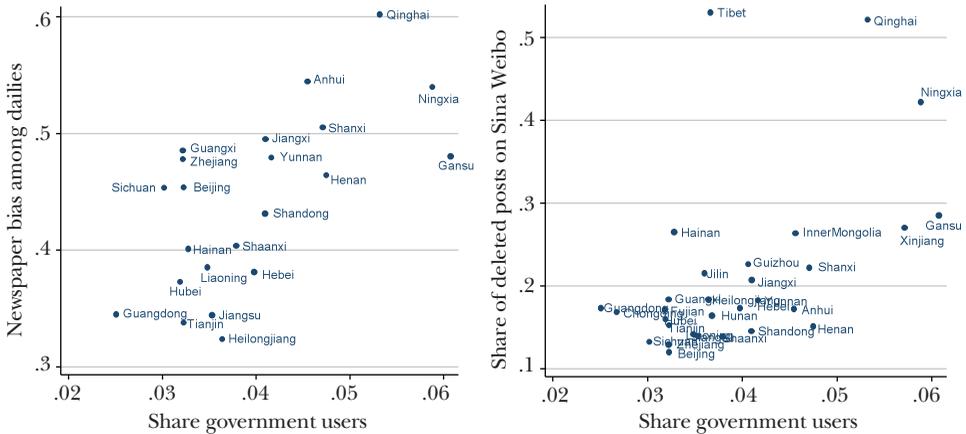
Goals of Government Users

Government users of social media may provide neutral information or propaganda. Several patterns can help us to distinguish between the two possibilities. For example, in areas where the government perceives that the need for influence is high, we should observe more of both censorship and propaganda and a strong positive correlation between them. We should also observe a positive correlation between posts from government users and pro-government bias in traditional media, which are subject to greater government control than social media. Conversely, these correlations should be absent if government users mainly provide neutral information.

Earlier research suggests some other theses about propaganda, as well. Propaganda may be more effective on audiences that share the message sender's view,

---

[6]The word frequencies in each post are computed after the preprocessing described at the earlier section in Footnote 4. As inputs to the Support Vector Machine, we use term-frequency inverse document frequencies. We use the software SVM-light (Joachims 1999). Using a new random sample of 500 users, we estimate a probit model of the probability of being a government account conditional on the SVM parameter. See the online Appendix for details.

Figure 1

The Share of Government Users on Sina Weibo across Provinces versus Newspaper Bias and Censorship

Note: Each dot represents one province in China. The left panel plots estimated share of government users against the measure of media bias in the daily newspapers strictly controlled by the Communist Party (from Qin et al. 2016). The right panel plots estimated share of government users against the measure of censorship developed by Bamman et al. (2012), the share of deleted posts.

while the effect of propaganda may be negative when the audience holds opposing views. For example, Adena, Enikolopov, Santarosa, and Zhuravskaya (2015) find that Nazi radio in the 1930s was most effective in places where anti-Semitism was historically high and had a negative effect on support for Nazi policies in places with historically low levels of anti-Semitism. Similarly, in a laboratory experiment, DellaVigna, Enikolopov, Mironova, Petrova, and Zuravskaya (2014) find that Serbian radio exposure caused anti-Serbian sentiment among Croats. If the Chinese regime believes in this argument, then we would expect to find more government-affiliated accounts in Communist Party strongholds.

Finally, propaganda is likely to reduce consumers' valuation of social media. To the extent that service providers can affect the amount of propaganda, we should see fewer official accounts in areas where the advertisement market is valuable and where competition for consumers is high. Although we lack direct measures of these factors, they are likely to be positively related to local incomes or GDP per capita.

We test these hypotheses using our own measure of government users on Sina Weibo across provinces of China against a measure of censorship and against a measure of bias in Chinese newspapers. The left panel of Figure 1 plots the estimated share of government users against the measure of media bias in the daily newspapers strictly controlled by the Communist Party (from Qin, Strömberg, and Wu 2016). The latter measure is based on nine content categories, including leader mentions, citations of the official Communist Party news agency, and coverage of stories that criticize the regime. The right panel of Figure 1 plots the estimated share of government users against a measure of censorship developed by Bamman,

Table 7
Patterns in the Share of Government Users Across Prefectures
(Dependent Variable: Share of Government Users)

	I
GDP	-0.849***
	(0.103)
CPC stronghold	0.533**
	(0.236)
Treaty port	-0.079
	(0.166)
Distance to Beijing	-0.464***
	(0.165)
Population	0.366***
	(0.129)
Latitude	0.052***
	(0.016)
Longitude	-0.037***
	(0.014)
Observations	259
R^2	0.358

Note: The unit of observation is the prefecture. The result is obtained by cross-sectional ordinary least-squares regression. GDP and Population values are from 2010, which is the first year Sina Weibo was in use. Robust standard errors are in parentheses. "CPC" is "Communist Party of China." "Treaty port" is a proxy for Western influence.
  *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

O'Connor, and Smith (2012): the share of deleted posts. Guangdong has the lowest share of government users (2.5 percent), whereas Ningxia and Gansu have the highest share (6 percent). The graph looks virtually the same if we use the share of posts published by government users instead of the share of government users.

The two panels show that the estimated share of government users is strongly correlated with both the share of deleted posts and newspaper bias (the correlation coefficient is 0.7 in both cases). This positive correlation is consistent with the hypothesis that censorship, newspaper bias, and official accounts on Sina Weibo are used for the same propaganda purpose. Note that in Figure 1, Tibet has fewer government users than expected. Perhaps this is an indication that propaganda is not viewed as particularly effective in Tibet because of weaker underlying support for the Chinese central government.

Other correlations are also consistent with the belief that government users of social media are engaging in propaganda. Table 7 offers an illustrative cross-section regression. The dependent variable is the share of government users of Sina Weibo across prefectures of China. We use GDP as a measure of economic development. We include a variable, "CPC stronghold," indicating areas where the Communist Party is comparatively more appreciated due to historical reasons (Qin, Strömberg, and Wu 2016). Conversely, some areas have a history of Western influence, notably, the areas that were part of a treaty port controlled by Western powers during the

1840–1910 period (Jia 2014). Other variables in the regression include the distance to Beijing, latitude, longitude, and population.

Our estimated share of government users across prefectures is significantly lower in areas with high levels of GDP and is higher in Communist Party of China strongholds. The latter result is consistent with the view that propaganda is more effective in areas where the audience shares the ideology of the sender. The estimated share of government users also appears higher in areas that are closer to Beijing and in areas that are more populous. To sum up, these patterns are consistent with propaganda being the main goal of this government social media content.

Conclusions and Discussion

We use a large dataset of blog posts from the most prominent Chinese microblogging platform—Sina Weibo—over the 2009–2013 period to document a number of basic facts. We now discuss how to make sense of these facts and what they may imply for outcomes that we are ultimately interested in, such as corruption, regime stability, local and central accountability, and the central–local balance of power.

Given the immense efforts to police and censor social media, it may seem puzzling that we find so much sensitive material available on social media. What are the private returns to posting this material? Why does the regime not punish users on a large scale, and why does the regime not censor all the sensitive material? We suggest the following explanation.

The central government limits its censoring because only a small fraction of potentially sensitive material is likely to pose a meaningful threat to the regime. Although diverse and even dissenting public opinion may displease the regime, a complete cleanup of sensitive content can impair the regime's ability to learn from bottom-up information and to address social problems before they become threatening. Therefore, there exists a subtle trade-off in information control faced by an authoritarian regime.

Social media users have an incentive to speak out about local problems as they expect the central government to address these problems. In the corruption posts, we find frequent explicit appeals for central government action. The posts about protests and strikes may help organize the events, but, importantly, they also make these events visible to the central government. This may force local leaders to deal with the problems that cause the protests and strikes. This line of reasoning may also explain why protests and strikes are rising in China[7] even though we find it is simple and cheap to use social media to identify these events one day in advance. Local officials may not want to suppress them through police and violence because this will trigger an even larger response on social media.

---

[7] See, for example, Mark Magnier, "China's Workers Are Fighting Back as Economic Dream Fades," *Wall Street Journal*, December 14, 2015 (http://www.wsj.com/articles/chinas-workers-are-fighting-back-as-economic-dream-fades-1450145329).

The visibility-enhancement effect of social media may also explain why we find no evidence of large-scale punishment of dissent users. One might think that this finding just reflects the consistency between censorship and policing—the posts that we observe were not sensitive enough to be censored, and people are thus not punished for posting them. However, the censors and the internet police are not the same people. The central government implements censorship while local governments can rely only on policing. So the question is why local politicians do not punish users whose posts may jeopardize their career, for instance, those who accuse them of corruption. One reason is that local politicians are not able to identify these users. Another reason is that social media can make local politicians' punishment of dissenting users visible to the national leaders, which exposes local politicians to great political risks.

Given that the central government uses information from social media to monitor local officials, it is not surprising that local officials actively run their microblogs on Sina Weibo. Local politicians can signal their loyalty to the central government by engineering propaganda that promotes the party line while also deflecting corruption charges and demonstrating their ability to identify and solve local problems. However, it is costly to produce high-quality propaganda to domi-nate a discussion among millions of users. We estimate that government-affiliated accounts contribute around 4 percent of all posts about politics and economics on Sina Weibo. While this is substantially larger than Sina's official numbers, it is not overwhelming. In this light, it is not surprising that we find that the local leaders are unable to effectively dilute negative information about, for example, corruption and collective action.

What are the implications for the outcomes that we ultimately care about? Our findings suggest that social media in China primarily affects those outcomes in which the central regime and general users share a common interest. For example, the regime and social media users both benefit from fighting local corruption and other abuse of power by local leaders. In this aspect, Chinese social media seem to play a positive role in public affairs at the local level, improving the public's access to information, engagement in public debate, and their ability to coordinate mass actions and respond to local problems.

Conversely, outcomes in which the central regime and users have opposing interests are less likely to be affected. For instance, a very limited number of posts discuss top Chinese national leaders in a negative manner. Similarly, social media coverage of large-scale conflicts is muted, either by censorship or by self-censorship. Consequently, one might expect social media to increase the incidence of small and medium-sized protests, while decreasing the probability of large regime-threatening protests. Thus, it is far from clear that social media is capable of restraining the Chinese central government.

Our findings finally suggest that social media in China is likely to increase the power of the central government at the expense of local governments, which will be more closely monitored and have less discretion. Eventually, it may also lead to increased centralization of power as the central government's incentive

to decentralize decisions is diminished by its improved ability to acquire local information.

Given the advancing information technology and the changing Chinese political landscape, it is natural to ask the question of how relevant the results of this paper are outside our sample period that ends in 2013. Since 2012, Sina Weibo has lost ground to other services, most notably WeChat and Qzone. Overall, user behavior and government censoring on these services are similar to that on Sina Weibo, but some differences should be noted. According to Ng (2015), a smaller share of posts are censored on WeChat than on Weibo. WeChat provides a somewhat less-effective tool to organize large-scale collective action or to expose corrupt officials, since only subscribers to an account can see a post. Thus, it is not surprising that, while WeChat is the most popular platform for instant messaging, Sina Weibo and Qzone are the most popular platforms for public information sharing. A survey-based report issued by CNNIC shows that in 2015, the number of social media users who listed Sina Weibo as their first choice to follow current news events and social issues was twice as many as those who listed Qzone.[8]

Since 2012, Chinese governments have escalated their efforts to control social media. As part of this endeavor, since March 2012, users have been required to reveal their true identities to social media providers. Because our sample period spans until the end of 2013, we can provide a glimpse of the effect of tightening control and increased competition from WeChat. Our data show a fall in the number of total Weibo posts from 2012 to 2013 by around 30 percent. However, the number of posts about sensitive topics (conflict, protests, strikes, and corruption) changed only marginally. This result indicates that the traffic that moved to WeChat and other services was concentrated on nonpublic topics such as personal life and private chatting. It also suggests that the stricter regime has still found it not in its interest to fully censor posts about the sensitive topics that we study.

References

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun. 2014. "The Power of the Street: Evidence from Egypt's Arab Spring." NBER Working Paper 20665.

Adena, Maja, Ruben Enikolopov, Veronica Santarosa, and Katia Zhuravskaya. 2015. "Radio and the Rise of Nazis in Prewar Germany." *Quarterly Journal of Economics* 130(4): 1885–1939.

Bamman, David, Brendan O'Connor, and Noah A. Smith. 2012. "Censorship and Deletion Practices in Chinese Social Media." *First Monday* 17(3).

China Internet Network Information Center. 2011. "The 28nd Statistical Report on Internet Development in China." July 2011, Beijing.

China Internet Network Information Center. 2013. "The 32nd Statistical Report on Internet Development in China." January 2013, Beijing.

China Internet Network Information Center. 2014. "The 34th Statistical Report on Internet Development in China." January 2014, Beijing.

Chen, Xiaoyan and Peng Hwa Ang. 2011. "The Internet Police in China: Regulation, Scope and Myths." In *Online Society in China: Creating, Celebrating, and Instrumentalising the Online Carnival*, edited by David Kurt Herold and Peter Marolt, 40–52. New York: Routledge.

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya. 2014. "Cross-border Media and Nationalism: Evidence from Serbian Radio in

[8]The Chinese version of the report can be downloaded from http://www.cnnic.cn/.

Croatia." *American Economic Journal: Applied Economics* 6(3): 103–32.

Dumais, Susan, John Platt, David Heckerman, and Mehran Sahami. 1998. "Inductive Learning Algorithms and Representations for Text Categorization." Proceedings of the 7th International Conference on Information and Knowledge Management, 48–155. ACM Digital Library.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies* 80(4): 1422–1458.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review* 103(4): 645–68.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2016. "Social Media and Protest Participation: Evidence from Russia." Available at SSRN: https://ssrn.com/abstract=2696236.

Epstein, Gady. 2013. "China's Internet: A Giant Cage." *The Economist,* April 6. http://www.economist.com/news/special-report/21574628-internet-was-expected-help-democratise-china-instead-it-has-enabled.

Freedom House. 2012. "Freedom on the Net: China." https://freedomhouse.org/report/freedom-net/2012/china.

Freedom House. 2015. "2015 Freedom of the Press Data." https://freedomhouse.org/report/freedom-world/freedom-world-2015#.WFxACX3CB1A.

Fu, King-wa, Chung-hong Chan, and Marie Chau. 2013. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy." *IEEE Internet Computing* 17(3): 42–50.

International Telecommunication Union. 2013. "The World in 2013: ICT Facts and Figures." Geneva. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf.

Jia, Ruixue. 2014. "The Legacies of Forced Freedom: China's Treaty Ports." *Review of Economics and Statistics* 96(4): 596–608

Joachims, Thorsten. 1998. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." 10th European Conference on Machine Learning, Vol. 1398 of Lecture Notes in Computer Science, 137–142. Berlin: Springer Verlag.

Joachims, Thorsten. 1999. "Making Large-Scale Support Vector Machine Learning Practical." Chap. 11 in *Advances in Kernel Methods: Support Vector Learning,* edited by Bernard Schölkopf, Christopher J. C. Burges and Alexander J. Smola. MIT Press.

King, Gary, Jennifer Pan, and Margaret E.

Roberts. 2013. "How Censorship in China Allows Government Criticism But Silences Collective Expression." *American Political Science Review* 107(2): 1–18

King, Gary, Jennifer Pan, and Margaret E Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science* 345(6199): 1–10.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2016. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." Unpublished paper, Harvard University.

Kleinberg, Jon. 2006. "Complex Networks and Decentralized Search Algorithms." Proceedings of the International Congress of Mathematicians (ICM).

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2): 402–414.

Morozov, Evgeny. 2012. The Net Delusion: The Dark Side of Internet Freedom. Reprint edition. Public Affairs.

Ng, Jason Q. 2015. "Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform." Unpublished paper, University of Toronto.

Platt, John C. 1999. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." *Advances in Large Margin Classifiers* 10(3): 61–74.

Public Opinion Monitoring Agency. Various years. Reports on the Online Public Opinion (2010– 2013). Published by *People's Daily.*

Qin, Bei, David Strömberg, and Yanhui Wu. 2016. "Media Bias in China." Unpublished paper.

Qin, Bei, David Strömberg and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda." CEPR Discussion Paper DP11778.

Reporters Without Borders. 2013. "2013 World Press Freedom Index: Dashed Hopes after Spring." https://rsf.org/en/news/2013-world-press-freedom-index-dashed-hopes-after-spring.

Sebastiani, Fabrizio. 2002. "Machine Learning in Automated Text Categorization." *ACM Computing Surveys* 34(1): 1–47.

Shirky, Clay. 2011. "The Political Power of Social Media: Technology, the Public Sphere, and Political Change." *Foreign Affairs,* January/February.

Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall, and Dan S. Wallach. 2013. "The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions." arXiv preprint, arXiv:1303.0597.

The Evolution of China's One-Child Policy and Its Effects on Family Outcomes

Junsen Zhang

I n 1979, China introduced its unprecedented one-child policy, under which households exceeding the birth quota were penalized. However, estimating the effect of this policy on family outcomes turns out to be complicated. China had already enacted an aggressive family planning policy in the early 1970s, and its fertility rates had already dropped sharply before the enactment of the one-child policy. The one-child policy was also enacted at almost the same time as China's market-oriented economic reforms that triggered several decades of rapid growth, which would also tend to reduce fertility rates. During the same period, a number of other developing countries in East Asia and around the world have also experienced sharp declines in fertility. Overall, finding defensible ways to identify the effect of China's one-child policy on family outcomes is a tremendous challenge, one that the literature has tended to underemphasize. Along with the issues just mentioned, most of the studies have not taken sufficient care to address the endogeneity of public policy (as emphasized by Besley and Case 2000). For example, regional variations in the enforcement of the one-child policy may also reflect prevailing family fertility preferences and regional socioeconomic conditions. However, compared with empirical studies on other developing countries that were mainly descriptive (for example, Alba and Potter 1986; Cho 1996), those on China appear to be much more rigorous.

This essay begins with a discussion of the evolution of the one-child policy in China, covering the lead-in to its inception, major modifications over time, and

■ *Junsen Zhang is the Wei Lun Professor of Economics, Chinese University of Hong Kong, Shatin, Hong Kong. His email address is jszhang@cuhk.edu.hk.*

heterogeneity in enforcement between urban and rural areas and across provinces. I then turn to the general trends of fertility and development in China and other selected developing countries. Subsequently, I expound the main empirical approaches to the identification of the effects of the one-child policy, with an emphasis on their underlying assumptions and limitations. I then turn to empirical results in the literature. I discuss the evidence concerning the effects of the one-child policy on fertility and how it might affect human capital investment in children. Numerous studies have considered these two effects, but research on other potentially important effects of the one-child policy in China, including its potential effects on divorce, labor supply, and rural-to-urban migration, is lacking. For these issues, I will offer some exploratory and preliminary estimates.

The paper concludes with some remarks about China's future demographic challenges. Although the Chinese government has recently replaced the one-child policy with a universal two-child policy, the legacy of the one-child policy, which has affected millions of people for over 30 years, continues to be of great interest.

Evolution of Family Planning Policy in China

Mao Zedong, the supreme Chinese leader between 1949 and 1976, believed in the principle: "More people, more power." Such thinking may be due to Mao's military career and rise to power. For example, Mao wrote in 1949, "Any miracle can be created as long as there are people under the leadership of the Communist Party" (as quoted in Mao 1966, p. 1516). However, Mao acknowledged the hardships of overpopulation on many occasions during the 1950s and 1960s, and there were several times during these decades when it seemed as if China might adopt a national family planning policy.

For example, Deng Xiaoping, who was Vice Premier in 1952, called for increased contraceptive use. At that time, China's population size was 600 million, according to the first census conducted in 1953. When the famous Chinese economist Ma Yinchu (President of Peking University), first recommended family planning in China during 1955–1957, Mao was initially supportive and even considered setting up a national family planning commission. However, during the Great Leap Forward movement in 1958, Mao reverted to a belief that a larger population was better. Ma's population control theory was heavily criticized nationwide and he was labeled as a "rightist" (Tian 2009).

After the famine during the Great Leap Forward of 1958–59, China's fertility rebounded significantly, and China's total fertility rate reached more than six births per woman in the early 1960s (Banister 1987). In 1962, the central government first announced a position advocating "birth planning in urban areas and densely populated rural areas." Shortly afterward, family planning commissions were established at national and provincial levels (Tian 2009; Ebenstein 2010).

However, the Cultural Revolution in 1966 halted these early family planning operations. Deng was purged for a time, and would not rise again to full power

until after Mao's death in 1976. By the end of 1969, China's population exceeded 800 million. Economic growth stagnated. The deterioration in China's standard of living was often discussed in terms of overpopulation, and rather than reconsidering China's economic institutions, leaders in China again began to focus instead on the urgency of population control. In the early 1970s, Mao instructed: "Population growth must be controlled" (as quoted in Peng 1996).

A serious family planning campaign began in China in 1971, commencing with the propaganda theme: "One child isn't too few, two are just fine, and three are too many." In July 1973, the State Council established the Leading Group for Family Planning, which was responsible for calling a national birth planning conference in December that year. That conference endorsed the slogan: "Later, Longer, and Fewer." The term "Later" meant late marriage requirements of 23 years for women and 25 years for men. "Longer" signified a birth planning rule of more than three years between the first and second child. "Fewer" implied that a couple could have two children at most. This family planning campaign was successful; and China's overall fertility rate declined by half between 1971 and 1978.

China's family planning campaign in the 1970s was technically voluntary, but it had a number of coercive elements. As Whyte, Wang, and Cai (2015) pointed out: "[T]he state bureaucratic hierarchy in charge of enforcing birth control then oversaw grass-roots birth planning workers in each village, urban work unit and neighborhood. These birth planning enforcers kept detailed records on each woman of child-bearing age under their responsibility, including past births, contraceptive usage and even menstrual cycles, in many reported instances becoming 'menstrual monitors' who tried to detect out-of-quota pregnancies at an early stage." Whyte et al. also provided evidence that IUD insertions, sterilizations, and abortion increased sharply in the 1970s. However, although China's family planning campaign in the 1970s did have a number of coercive elements, it was significantly less coercive than the one-child policy that followed.

It's useful to place China's 1970s family planning policy in the context of its time, when global concerns about population growth were especially high. At that time, countries like India, Bangladesh, and Indonesia also had family planning programs with significant elements of coercion.[1] On the one side, China was a largely isolated or closed state at that time, and it was not clear how much China's policy was influenced by international concerns over population (Tian 2009). On the other side, China was at least aware of the global concerns and initiatives. Wang, Cai, and Gu (2013) noted that during the first UN-organized World Population Conference in

---

[1] Wang, Cai, and Gu (2013) observed (citations omitted from quotation for brevity): "India's notorious forced sterilization campaign was carried out between 1975 and 1976, during which more than 8 million sterilizations were performed. The campaign led to the collapse of Indira Gandhi's government in January 1977 … In a number of other countries with large populations and rapid growth rates, such as Bangladesh and Indonesia, coercive methods of a lesser degree were also used. … Between 1976 and 1996, the number of governments viewing their population growth rate as too high increased from 55 to 87. By launching a nationwide policy of limiting each couple to only one child, China, of course, established an unprecedented level of government control of births."

1974 in Bucharest, as well as other international forums, China denounced calls for family planning as part of an imperialist agenda. However, back at home, China was simultaneously implementing its population control policies.

Mao died in 1976. By 1978, Deng Xiaoping rose to leadership. Deng and other senior leaders stressed the importance of birth control policies for China (Peng 1996). Hesketh, Lu, and Zhu (2005) explained the reason for introducing the one-child policy: "At the time, China was home to a quarter of the world's people, who were occupying just 7 percent of world's arable land. Two-thirds of the population were under the age of 30 years, and the baby boomers of the 1950s and 1960s were entering their reproductive years. The government saw strict population containment as essential to economic reform and to an improvement in living standards."

Deng's position in support of population control had been consistent over many decades. Thus, when he assumed power in 1978, it was not surprising that a strong population control policy was enacted. Deng may have viewed population control as a method of raising the GDP per capita of China. For example, when Deng met with the Japanese Prime Minister in 1979 in Beijing, he referred to a development goal of quadrupling GDP per capita by the end of the century. He intended to raise the GDP through economic reforms and the "Open Door Policy" of reorienting China's foreign trade away from the heavy dependence on the USSR. It was in the context of both a neo-Malthusian concern over population and ambitious economic development goals that the Chinese government launched the coercive one-child policy in 1979 (Peng 1991).

Scharping (2003) has carefully documented China's birth control campaign from 1979 to 1983. In a circular addressed to all members of the Communist Party in January 1980, the Secretariat of the Central Committee called for legal, economic, and administrative measures in favor of one-child families. The Party Central Committee addressed an open letter to all members of the party and the Communist Youth League and reaffirmed the one-child policy on September 25, 1980. During this period, when the central government evaluated the performance of local officials, economic growth, social stability, and enforcement of the one-child policy were consistently among the highest priorities (Birney 2014). Provincial and prefectural officials had strong incentives to require or even force local residents to comply with the one-child policy (Hardee-Cleaveland and Banister 1988).[2]

Enforcement of the one-child policy did waver in the mid-1980s. Many rural families, particularly those with only one female child, strongly resisted the policy. Given the practical difficulties, the central government relaxed the policy to make it more feasible in rural areas (with the philosophy of "open a small hole to block

---

[2] Hardee-Cleaveland and Banister (1988) pointed out that, during 1979 and 1982, required abortions in the second and third trimester of pregnancies were mandated in some provinces. In 1983, the implementation of such compulsory measures reached a peak, and "China carried out a nationwide campaign of mandatory sterilization, abortion, and IUD insertion characterized by a degree of heavy handedness in family planning unprecedented in the world." They also pointed out that, for those who escaped the forced abortion and continued an unapproved pregnancy, heavy fines were imposed, and those households often had to pay 10–20 percent of their family income for 7 to 14 years.

a big gap") during 1984–1985. A list of 14 types of cases eligible for second-child permits was drawn up, with the most important rule being that rural couples with only one daughter could have a second child. Many local governments regarded this change as a signal that the central government attached less importance to the policy, and China's fertility rebounded considerably from 1984 to 1986. In 1986, the State Council reiterated the one-child policy and retightened the policy across the country from 1986 to 1990. After 1990, the one-child policy became relatively stable.

Although the one-child policy was applicable throughout China, local implementation of the policy, such as the actual penalties for above-quota births, often varied across rural and urban areas, regions, provinces, and even ethnicities. In particular, the government controlled the behavior of urban residents more easily than that of rural residents. Chinese urban residents can be more closely and directly affected by government policies and measures. For example, many urban residents worked in state-owned enterprises or institutions, and by having more than one child, they could lose their jobs and access to social welfare payments. In addition, sex-selection technology was somewhat available in urban areas for those with an especially high demand for sons. By contrast, rural residents received fewer benefits from the government than urban residents did. Furthermore, those carrying out agricultural work had a distinct need for a larger family, and sex selection technologies were mostly unavailable. A common punishment for rural residents who exceeded the birth quota was a one-time fine, and many rural families were so poor that they could not pay the fine, anyway. These complications resulted in a de facto two-tier policy: that is, urban couples were only allowed to have one child, whereas rural couples were often allowed to have a second child, particularly if the first child was female.

The policy variation was also considerable across regions. For example, in the less-developed western regions of China, the fertility level was significantly higher than that in the more-developed eastern regions before the one-child policy was introduced. Therefore, strictly implementing the one-child policy in the western regions meant enforcing a larger social change and was thus more difficult. Even within eastern regions, the enforcement of the one-child policy varied significantly. For example, Zhejiang and Jiangsu Provinces were close to each other and shared many similar characteristics, but the one-child policy was more strictly implemented in Jiangsu than in Zhejiang. Another variation was that the practical application of the one-child policy differed for the ethnic majority Han population and ethnic minorities. Specifically, Han women in urban areas were generally allowed to have only one child, whereas minority women were normally allowed to have two or more children.

Trends in Total Fertility Rates in China and Selected Countries: 1950–2010

A number of countries experienced sharp declines in fertility in recent decades. Does the timing or speed of China's fertility decline—whether in urban or rural

Figure 1

Total Fertility Rate for Rural and Urban China and Four Selected Countries (1960–2010)
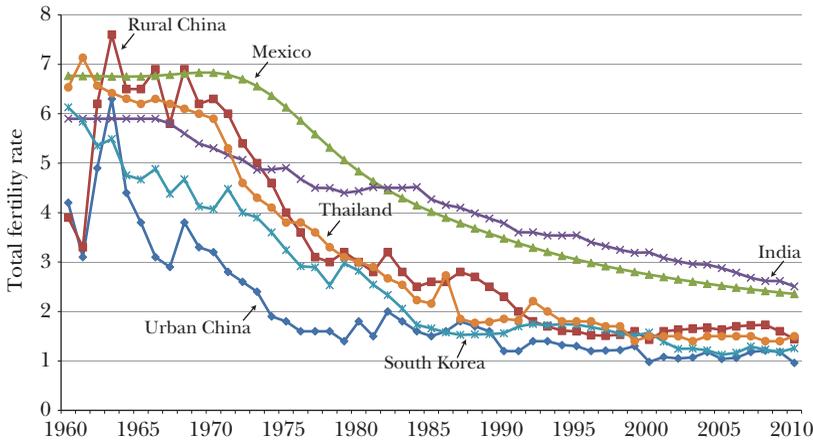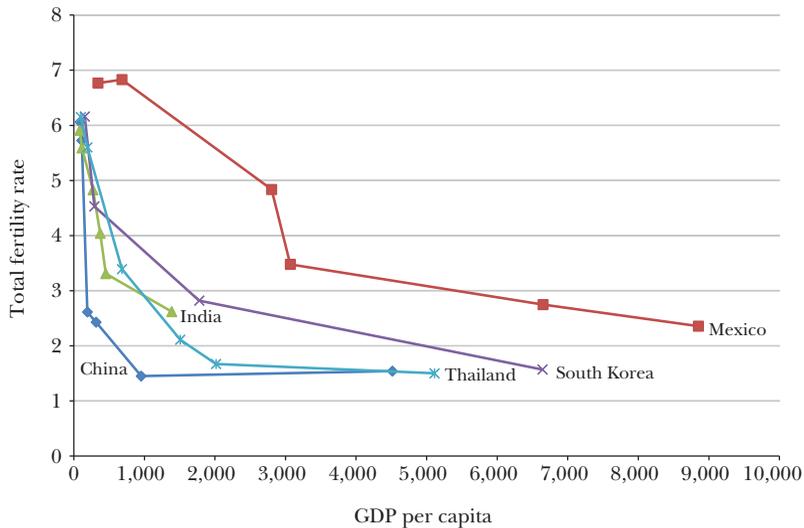

Source: The total fertility rate data for each country are from the World Bank (2016). China's urban and rural total fertility rate data are obtained from several sources: the 1960–1992 data are adopted from Peng and Guo (2000); the 1993–1999 data are from the 2001 National Fertility and Reproductive Health Survey; the 2001–2009 data are adopted from Hao and Qiu (2011); and the 2000 and 2010 data are obtained from the national population censuses in the respective years.

areas—look dramatically different from that in other developing countries like South Korea, Thailand, Mexico, or India?

Figure 1 illustrates total fertility rates between 1960 and 2010 for rural China, urban China, and a comparison group of four other developing countries that, like China, also had high fertility levels in the 1960s. The figure suggests that all these developing countries, even without a compulsory population control policy such as the one-child policy in China, exhibited rapid declines in fertility after 1970. South Korea and Thailand are Asian countries that had high economic growth rates after 1970 and implemented voluntary family planning programs. Their fertility rates fell more slowly than those for urban China in the 1960s and 1970s, but by 2010 their fertility was as low as that of China. Thailand is especially worth noting because it had fertility and GDP per capita similar to China's by 2010. Mexico had a weaker family planning campaign, but its fertility rate still reached a low level by 2010. Economic development seems to be a fundamental cause of such fertility decline. India's development level was similar to that of China in 1978. However, India's family planning initiative after 1978 was not as strong as China's, and its economic development was significantly less advanced than that of South Korea and Thailand. Nonetheless, India also saw a marked downward trend in fertility—although the decline was not as fast as other Asian countries in the figure. Given that the economic growth rate of China was higher than those of the four other countries from 1979, I believe that even without the one-child policy, China's fertility rate would have declined substantially after 1978.

Figure 2

Total Fertility Rate and GDP per capita for China and Four Selected Countries (1960–2010)

Sources: The total fertility rate and GDP data for each country are from the World Bank (2016), and the unit of GDP per capita is US dollars (2014 constant prices).

Notes: The figure plots each country's total fertility rate against its real GDP per capita, showing decadal points from 1960 to 2010. Each point represents, from left to right, the total fertility rate and GDP per capita of each country in 1960 (for China 1962 was chosen instead of 1960 to avoid the unusual famine effect in China), 1970, 1980, 1990, 2000, and 2010. When GDP per capita of a country exceeded $10,000, that country actually stepped up to become a developed economy. Thus, I dropped two points with GDP per capita levels above $10,000 for South Korea to make the figure compact.

Focusing on the trend of China's fertility decline, we can get some hints of the effect of the one-child policy. Given the vast socioeconomic differences between urban and rural areas in China, Figure 1 illustrates these areas separately. The relatively low rates of fertility in China in 1960 were due to the Great Famine. Fertility rebounded after the famine and maintained a relatively high level in the subsequent ten years, but it declined significantly and persistently from 1971 to the late 1970s. By 1978, the total fertility rates of the urban and rural areas dropped to 1.55 and 2.97 per woman, respectively. After the one-child policy was implemented, in 1979, fertility rates continued to decline. However, since the early 1990s, both rural and urban fertility rates have roughly stabilized over time. This pattern suggests that the one-child policy may have had a small short-term effect on fertility around 1979, but little or no additional long-term effect.

As this discussion has suggested, economic growth is also linked to declines in fertility. Figure 2 illustrates this relationship by plotting each country's total fertility rate against its real GDP per capita, showing decadal points from 1960 to 2010. Figure 2 indicates a negative correlation between total fertility rates and GDP per capita; indeed, for China, India, South Korea, and Thailand, their fertility rates had declined

sharply before their GDP per capita levels reached $3,000. Mauldin (1982) argued that family planning programs played a significant role in the sharp decline of fertility rates in the four countries during 1960–1980, before they reached high levels of economic development. The fertility decline in these countries preceded economic development. By contrast, the pace of fertility decline was slower in Mexico, where the family planning campaign was much weaker than that in other four countries.

Focusing on China, it is important to note that the magnitude of the fertility decline in the post-1979 era was significantly smaller than that during the 1970s. Specifically, China's total fertility rate for all regions declined from 5.8 in 1970 to 2.7 per woman in 1978, whereas from 1979 to 1995 under the one-child policy, China's fertility rate decreased from 2.8 to 1.8 per woman. China's economic development was almost stagnant during the 1970s (as shown in Figure 2), so economic development was not the main cause of the fertility decline during that time. Instead, the birth planning campaign during that period was widely believed to be the main contributor to reducing China's fertility rate by half, which was a success that some hailed as representing an "induced fertility transition" (Mauldin 1982).

Empirical Approaches to Identifying the Effect of the One-Child Policy

Identifying the causal effect of the one-child policy requires care and caution. After all, the adoption of the policy was to some extent prompted by the prevailing socioeconomic considerations. It coincided with a change in leadership philosophy, including the beginning of economic reforms and opening to foreign trade. In this section, I discuss four main methodological approaches that have been used in the literature and consider some concerns about them. Some studies use several of these approaches. The next section then discusses the empirical findings on the effects of the one-child policy.

A first approach sought to compare the birth behavior of China's women before and after the one-child policy implementation (Ahn 1994; Ding and Hesketh 2006; Whyte et al. 2015). For example, using data from the 1985 China's In-depth Fertility Survey, Ahn (1994) studied the probability of the second and third births before and after 1979. He entered a policy dummy for before and after 1979 to capture the policy effect. Ahn (1994) also interacted the policy dummy with a set of family characteristics (which is equivalent to allowing the effect of the family characteristics on fertility to differ between before and after 1979). He also controlled a linear time trend (based on the calendar year at the previous birth) to account for the effect of economic development on fertility decline. Essentially, this approach makes a rather restrictive assumption that the fertility patterns for women giving first birth before and after 1979 had a common linear trend. Thus, any deviation from the linear time trend after 1979 was attributed to the one-child policy.

A second approach used cross-sectional and temporal variations on fines and/or the eligibility for a second birth (McElroy and Yang 2000; Liu 2014; Huang, Lei, and Sun 2016). For example, McElroy and Yang (2000) studied regional

variations in fines and rural fertility. However, using this variation in fines as an exogenous measure of the one-child policy is clearly problematic, because local governments may set the amount of fines according to local financial situations and local fertility demand. Huang et al. (2016) also used a similar approach, but argued that fines may be exogenously determined by the characteristics of local officials and used the fines as a measure for the one-child policy implementation to estimate its reduced-form effect on the educational attainment of girls. To support their argument, Huang et al. (2016) observed that the largest changes in the fertility penalties were associated with the arrival of new provincial officials immediately after the central government announced in the end of 1989 that the performance of the one-child policy implementation would be used as an additional factor to evaluate local officials for promotion. (As noted earlier, the implementation of the one-child policy was part of the evaluation of local officials before 1989, but the implementation of the policy fluctuated several times during the 1980s.) Although it is plausible that the arrival of new local officials may serve as an exogenous shock, a part of the fines remains unaccounted for by local officials' characteristics, which suggests that fines may remain endogenous in the schooling equation. An alternative approach that can avoid the endogeneity problem is to regress girls' education on the fines, with the turnover and characteristics of local officials as instrumental variables for the fines. This approach requires the exclusion restriction that the set of instrumental variables—in this case, the characteristics of local officials, such as how recently they assumed their position—does not directly affect child education other than through the level of fertility fines.

A third approach compared the fertility rates of Han and minority women in a difference-in-differences framework. In Li, Zhang, and Zhu (2005), we argued that Han women were required to follow the one child policy strictly, whereas ethnic minority women were often allowed to have two children. Thus, comparing the change in Han women's fertility or family behavior before and after 1979 with that of the minorities would identify the effect of a two child versus the one child policy. The underlying common trend assumption on fertility or family behavior would be easier to hold if women from minority groups had similar family behavioral patterns to those of the Han before the one-child policy. However, economic reforms may affect the Han and the minorities differently. Thus, the common trend assumption may be difficult to maintain. Along the same lines, in Li and Zhang (2007), we used the share of the minorities as an instrumental variable for fertility to estimate the fertility effect on economic growth in China.

A fourth approach is based on a recent work by Bingjing Li and Hongliang Zhang (forthcoming), which builds on a similar idea in Poston and Gu (1987). In this case, the empirical identification relies on exploring the heterogeneity of the intensity of the one-child policy implementation across prefectures and the differential length of exposure to the one-child policy of distinct birth cohorts. Specifically, they construct a measure based on excess births to Han women over and above the one-child rule in each prefecture. By looking only at Han women, they avoid the issue of differential enforcement of the one-child policy across ethnic groups.

They then control for differences in pre-existing fertility and community socioeconomic characteristics, seeking to take into account the local fertility demand and local government's response to the local fertility demand in setting this policy. The remaining residual after these controls is a proxy measuring largely exogenous regional policy stringency in the one-child policy implementation, which is taken as the pure policy variable. In the next section, I use this approach to estimate the effect of the one-child policy on various family outcomes. This approach is based on a common trend assumption that the strict and less-strict prefectures would exhibit the same family pattern in the absence of the one-child policy. The assumption implies more specifically that any unobserved prefecture-specific shocks to fertility or other family outcomes over time are uncorrelated with the cross-sectional measure of the one-child policy enforcement intensity.[3] However, this strong assumption is more likely to hold when looking only at Han women, under the excess fertility residual approach, as opposed to doing a Han-minority comparison.

An ideal scenario for estimating the aggregate effects of the one-child policy would have some extremely strong conditions; for example, it would have been useful for research purposes if China had randomly selected a cross-section of prefectures to implement the one-child policy and then forbidden interprefecture migration. However, such conditions are unrealistic. Thus, the practical research approaches are based on key underlying assumptions and on answering specific questions only.

The first approach assumes that the introduction of the one-child policy was exogenous—that is, its introduction was not correlated with the change of socioeconomic factors that may directly affect fertility and other family outcomes. The third approach attempts to purge the confounding change of socioeconomic factors by using the minorities as a comparison group for the treated Han women, which requires a common trend assumption in the specific outcome for post-1979 Han women and minorities. The second and fourth approaches do not estimate the aggregate effects of a situation with the one-child policy versus a situation without such a policy; instead, what they estimate is the differential effect of the varying one-child policy implementation across regions or over time.

Micro Evidence

In this section, I discuss the empirical results in the literature on the effects of the one-child policy, largely drawing on the range of methodologies in the previous section.

---

[3] As noted in Bingjing Li and Hongliang Zhang (forthcoming), this specification is analogous to Duflo's (2001) difference-in-differences specification, which measured the regional intensity of the school construction in Indonesia by the number of schools constructed, controlling for the number of children in primary school age. In this study, we treat 1982 as a base year because of data availability. We can then look at the change in enforcement over time and the change in an outcome variable.

Effect of the One-Child Policy on Fertility

In general, very different views exist on how the one-child policy affected fertility: one group of studies argued that the one-child policy had a significant or decisive effect on fertility in China, while another group argued that socioeconomic development played a key role in China's fertility decline. A plausible reconciliation of these views is that the one-child policy accelerated the already-occurring drop in fertility for a few years, but in the longer term, economic development played a more fundamental role in leading to and maintaining China's low fertility level. To put it more bluntly, China's fertility might well have dropped to the current low level with rapid economic development, even without the one-child policy, although the timeline of the decline would not appear quite the same.

In an early study of the one-child policy, focusing on the short-term, Poston and Gu (1987) examined the relationships among socioeconomic development, family planning, and fertility across 28 subregions of China. They found that although the implementation of family planning programs had a large effect on fertility (a one-standard-deviation increase in the contraceptive prevalence/effectiveness rate would lead to a roughly 0.5-standard-deviation decrease in the total fertility rate), the effect of socioeconomic development on fertility decline was substantial even in the mid-1980s.

Other studies of the one-child policy carried out in the late 1980s and early 1990s, within a relatively few years after the launch of the policy, had similar findings. For example, Ahn (1994) estimated the effects of the one-child policy on second and third births in the provinces of Hebei and Shaanxi and the municipality of Shanghai. He found that urban families were less likely to have a second or third child than rural families, and this urban–rural gap was widened after the one-child policy: the birth likelihood of a second child in urban areas fell by 50 percent in Hebei and Shanxi and by 86 percent in Shanghai (almost no family had a second child in urban Shanghai after the one-child policy). McElroy and Yang (2000) found that China's population policies played a considerable role in fertility. Specifically, the family size is smaller in counties with higher penalties on above-quota births: an additional penalty of approximately $177 would reduce the family size by 0.33 to 0.47 with different estimation methods. Ebenstein (2010) documented high ratios of males to females in China. By exploiting regional and temporal variation in fines levied on above-quota births, he found that higher-fine regimes (stricter fertility control) discouraged fertility and were associated with higher sex ratios (of males to females). In Li, Zhang, and Zhu (2005), we found a large effect of the one-child policy on the probability of not having a second child.

Although the one-child policy reduced China's fertility, particularly in the early period of its introduction, some researchers have argued that China's below-replacement fertility in the long run was mainly driven by socioeconomic development rather than the government policy. A study by Yong Cai (2010) is an example of those that emphasized economic development and compared China's fertility experience to that of other countries. This study points out that given a negative correlation between the total fertility rate and GDP per capita in the global

context, as well as China's rapidly growing economy over the past several decades, the country's fertility would have also reached a relatively low level even without the one-child policy. He exploited the differences in the policy intensity between urban and rural residents in two provinces and found that the pronounced difference in the one-child policy implementation intensity between Jiangsu and Zhejiang did not translate into substantially different fertility levels and that socioeconomic factors had significantly stronger effects on fertility. He concluded that the one-child policy may have accelerated China's drive to reach below-replacement fertility, but it was not the key driving force for the low fertility that emerged. Similar views have been shared by Wang, Zhao, and Zhao (2013) and Whyte et al. (2015). Whyte et al. (2015) argued that the total fertility rate in China had dropped from 6 per woman in 1970 to approximately 2.7 per woman at the end of the 1970s, which meant that roughly three-quarters of the decline in fertility since 1970 up to the present occurred before the introduction of the one-child policy. They attributed the substantial fertility decline to the "Later, Longer, and Fewer" campaign in the 1970s, and believed that China's rapid economic development since 1980 contributed significantly to China's additional fertility reduction.

Effect of the One-Child Policy on Child Education

The effects of the one-child policy may be expected to affect other family outcomes as well. If families expect to have only one child, they may invest more in the human capital of the child (Becker 1991).

Families that have twins can be considered as a natural experiment in the context of the one-child policy. In Rosenzweig and Zhang (2009), we exploited the incidence of twinning to estimate the effect of the fertility change induced by the one-child policy on other outcomes. Using the Chinese Child Twins Survey conducted in 2002 in Kunming City, we observed that the one-child policy had been strictly enforced in the urban area of Kunming: approximately 95 percent of households without twins had only one child, and over 99 percent of households with twins stopped at two children.[4] In comparing one-child households to households with twins, we showed that the fertility decline induced by the one-child policy had a moderately positive effect on children's human capital at most. Even using the upper-bound estimates to compute the impact of the one-child policy on human capital development in China, the results suggested that the policy increased children's schooling attainment by at most 4 percent, the probability of attending college by less than 9 percent, school grades by 1 percent, and the incidence of good or excellent health by less than 4 percent.

Using twins as a natural experiment is a neat and worthwhile approach, but it has some limitations. For example, the health status of twins may not be comparable to singletons: for example, twins are more likely to have low birthweight, which is

[4]In the rural area of Kunming, a two-child policy was strictly enforced. Hence twinning at the second-parity was an exogenous break of the two-child policy. For brevity, I discuss here only the results for the urban sample.

often used as a simple measure of health status at birth. The Rosenzweig and Zhang (2009) dataset has birthweight information, which is often unavailable in other datasets. While the sample in the study is locally representative, a nationally representative twins survey with valid birthweight information is not yet available in China.

In a different approach, Liu (2014) made use of the variation in fertility induced by cross-community variations in the one-child policy, which as noted earlier was rigorously enforced in the early 1980s and then relaxed after 1984, to identify the effect of child quantity on child quality. He used the eligibility for having two children, fines for unsanctioned births, and their interactions as instruments for fertility. He found a quantity–quality trade-off in children's height: children born into settings with stricter enforcement of the one-child policy were taller. However, using three educational outcomes as quality measures—school enrollment, middle school graduation status, and years of schooling (normalized for age and gender)—he found the effect on educational attainment to be statistically insignificant in most cases, particularly for boys.

In a study mentioned earlier, Bingjing Li and Hongliang Zhang (forthcoming) provided evidence on the quantity–quality trade-off by exploiting regional differences in the one-child policy implementation intensity as an exogenous source of variation in family size. They constructed a quantitative indicator of the extent of local violation of the one-child policy and then used this measure as an instrumental variable for fertility. Using the 1 percent sample of the 1982 and 1990 Chinese Population Censuses, they found that prefectures that enforced the one-child policy more strictly experienced larger declines in family size and also greater improvements in child education. However, the magnitude of the effect was not large.

Education levels in China have been rising. However, the existing studies indicate either a modest or minimal effect of the fertility change induced by the one-child policy on children education. This is consistent with evidence from other developing countries. For example, Fitzsimons and Malde (2014) did not find evidence that family size had a detrimental effect on different measures of girls' education attainment in rural Mexico. They argued that households with more children might adjust on other margins such as increasing maternal labor supply. In Guo, Yi, and Zhang (2016), we presented a rationed fertility theory and some evidence showing that a forced reduction below an optimal fertility generated a utility loss and thus involved a negative income effect, which attenuated the schooling adjustment on the margin.

Effect of the One-Child Policy on Other Family Outcomes

Although the discussion here has focused on fertility and human capital investment in children, studies have also looked at a range of other family outcomes. Tuljapurkar, Li, and Feldman (1995) pointed out that the preference for sons was leading to a gender imbalance: for example, they found an "imbalance in the first-marriage market" with around one million more males than females by 2010. Zhang and Spencer (1992) investigated who complied with the one-child policy in China and why; for example, wife's education and husband's occupation were associated with compliance, but husband's education and woman's occupation were not.

Goodkind (2011) estimated that 19 percent of all children born in China at ages 0–4 were unreported to authorities in 2000, more than double the level in the 1990 census, following the wave of increased enforcement and penalization under the one-child policy that occurred in 1991.

However, a number of other outcomes might plausibly be influenced by the one-child policy, including marital status, labor supply, and migration. Here, I offer some exploratory results along these lines using the excess fertility residual approach discussed earlier, which is to construct a proxy for differences in enforcement of the one-child policy across prefectures in China.

Specifically, the approach proceeded in several steps. First, I constructed a measure of excess fertility for Han women above the one-child rule in each prefecture of China in 1982 and regressed it on a set of average household-level variables for that prefecture, including the mother's age at first birth, the first child's age, the mother/father's education level, and the mother/father's employment sector. By netting out the influences of these control variables, I obtained the excess fertility residual. Second, for the outcome variables, I conducted the same procedure on each outcome variable (in its first difference over 1982 and 1990 or 2000) and obtained the residual for the difference in the outcome variable as for the excess fertility residual. Third, I ran a reduced-form regression using the excess fertility residual—that is, the proxy variable showing the level of enforcement of the one-child policy—as the explanatory right-hand-side variable, and the residual for the differenced outcome variable as the dependent left-hand-side variable. For marital status and labor supply, I looked at the change in each of them from 1982 to 1990. For migration, because the migration information in the 1990 census was not comparable to that in the 1982 census, I compared 1982 and 2000 census data for the change in migration.

Table 1 shows the results.[5] The first row shows the estimated result for the likelihood of divorce for females older than 19 years and males older than 21 years. The dependent variable is a dummy indicating whether an individual is divorced. The coefficient is negatively significant, indicating that a more relaxed one-child policy is associated with a lower likelihood of divorce. A possible explanation for this pattern is that a family with more children is less likely to divorce. But although the estimate is statistically significant, the magnitude is small. The standard deviation of the excess fertility rate residual is 2.55 percent, and its interquartile range equals 2.85 percent. A tightening of the one-child policy in terms of one interquartile range decrease of the excess fertility rate residual is associated with an increase in the probability of divorce by 0.0153 percentage points (that is, 0.0285×0.00538).

The second row reports the results with female labor supply as the dependent variable. The dependent variable is a dummy indicating whether a female household head (or the spouse of the household head) participated in the labor force.

---

[5] The online appendix available with this paper at http://e-jep.org offers more detail on these results, including a series of graphs that provide a visual representation of a difference-in-differences estimate of the effect of the variations in enforcement of the one-child policy.

Table 1

Regression Results for Changes in Enforcement of the One-Child Policy on Several Family Outcomes

(independent variable is the excess fertility residual; a higher residual indicates weaker enforcement)

Dependent variable:	Coefficient	Standard error	Observations
Divorce	−0.00538***	0.00139	586
Female Labor Supply	0.037	0.0731	586
Male Labor Supply	−0.0387***	0.00942	584
Rural Migration	−0.279***	0.052	530

Notes: I ran a reduced-form regression using the excess fertility residual—that is, the proxy variable showing the level of enforcement of the one-child policy—as the explanatory right-hand-side variable, and the residual for the differenced outcome variable as the dependent left-hand-side variable. For all of these regressions, the excess fertility rate residual was calculated using 1982 census data, and all the regressions used prefectural-level data weighted by the number of observations in each prefecture in 1982. The regression on divorce and labor supply used 1982 and 1990 census data, while the regression on migration used 1982 and 2000 census data.

The coefficient (0.037) is statistically insignificant. The third row of Table 1 shows the result for the labor force participation indicator of male household heads. The coefficient is negatively significant, indicating that a more relaxed one-child policy would reduce labor force participation of males. Again, the size of the effect remains rather small. A tightening of the one-child policy in terms of one interquartile range decrease of the excess fertility rate residual is associated with an increase in the probability of labor force participation of a male household head by 0.117 percentage points (that is, 0.0302×0.0387).

The fourth row shows the results with rural migration as the dependent variable: specifically, the dependent variable is the ratio of the number of individuals not living in the household for more than one year to the number of the labor force in a household in the rural areas. No migration information was available in the 1990 census in the same way as in the 1982 census. Thus, I looked at the 2000 census for the variable with the same form as the 1982 census question on migration. The coefficient has the expected negative sign—that is, having fewer children tends to be associated with a higher level of migration—and is statistically significant with a moderate-sized effect. A tightening of the one-child policy in terms of one interquartile range decrease of the excess fertility rate residual can increase the rural migration rate by 0.823 percentage points in 2000 (that is, 0.0295×0.279).

There has been little study of the effect of the one-child policy on these outcomes in China. As other countries consider the effects of declining birth rates or in some cases the effects of broad-based planning policies, further research on these topics seems especially useful. Overall, these findings suggest that when thinking about reductions in birth rates, effects of increasing male labor force supply and increasing migration may be especially worth taking into account.

Concluding Remarks

Although the literature on China's one-child policy has generally found a statistically significant effect on reducing fertility, most of the studies found the effect to be rather small, especially in the longer-term (Wang et al. 2016). However, this small effect must be understood in the context of the aggressive family planning program in the early 1970s. Given China's extremely high economic growth after 1979 and the fertility transition experienced in other East Asian countries, China's further decline of fertility after 1979—which was significantly smaller than what had already occurred during the 1970s—cannot be fully or even mainly attributed to the effect of the one-child policy. Using data from 16 countries with similar birth rates to that of China in 1970, Wang et al. (2013) found that the birth rate of these countries declined significantly and much more sharply than what the Chinese government predicted would have happened in China without the one-child policy, and concluded that China's government exaggerated the effect of the one-child policy. To further analyze the counterfactual scenario without the one-child policy, Wang et al. (2013) applied a Bayesian model that United Nations (2011) used for its population projections and established that "fertility in China would have continued to decline if the country's rapid fertility decline in the 1970s offers any hint about the country's future fertility trajectory. According to the Bayesian model, the decline would have continued after 1980. By 2010, fertility would have fallen to its currently observed level of around 1.5 children per woman."

Although the enforcement of the one-child policy may have mildly accelerated the fertility transition in China, it also brought substantial costs, including political costs, human rights concerns, a more rapidly aging population, and an imbalanced sex ratio resulting from a preference for sons. In retrospect, one may question the need for introducing the one-child policy in China.

What lesson may other developing countries draw from China's one-child policy? If the country is already experiencing a decline in fertility rates, then a one-child policy doesn't seem useful. However, a country with a total fertility rate in the window of 5 or 6 children per women might be tempted. Implementing a coercive family planning program such as the one-child policy in a developing country that is not governed by an exclusive power such as the Chinese government would be difficult. Rather than focusing on a one-child policy, a developing country may want to consider an aggressive but less-coercive family planning program such as the policy implemented in China in the 1970s.

China has now experienced moderately low fertility for more than 30 years and will face the challenge of a declining and aging population in the near future. Numerous studies documented the falling youth dependency ratio in the short term and the rising elderly dependency ratio in the long term (Poston 2000; Hesketh et al. 2005; Zhang and Goza 2006; Hu and Yang 2012; Liu 2013). However, because fertility was in sharp decline before the one-child policy, economic growth was significant after the one-child policy, and longevity has been rising in China as

elsewhere, it seems plausible that China was going to experience a demographic transition at some point even without the one-child policy.

China now appears heading for a future of negative population growth. The *World Population Prospects* published by United Nations offers authoritative global demographic estimates and projections that are often used by researchers in this area. According to its medium fertility variant projection, China's total population will peak around 2030 with 1,416 million people, and then it will decrease to 1,348 million in 2050 (United Nations 2015). Using the data from the 2008 revision of *World Population Prospects* (United Nations 2009), Fang Cai (2010) found that China's working-age population would peak significantly earlier in 2015 with 998 million people and then begin to decrease. Based on the shortage of migrant workers and their increasing wages in China, the author concluded that China would have to face the challenge of getting old before getting rich. Yong Cai (2012) found that the proportion of the aging population in China would quickly increase to 30 percent or higher in the near future, and it would remain around 35 percent until the end of this century. Hu and Yang (2012) even argued that China's dependency ratio—that is, the number of children and the elderly divided by the number of working-age adults—would increase to more than 50 percent.

In 2016, China's government officially replaced the one-child policy with the two-child policy. However, this change seems unlikely to alter China's population decline. Wang et al. (2016) found that the increase in fertility in response to the two-child policy would be mild, within the range of 0.3 to 0.7 more births per woman. Even under the high-fertility variant in which fertility increases by 0.7, their simulation results showed that China's total population would still peak around 2030 at approximately 1,460 million. Wang et al. (2016) also simulated the proportion of the aging population under a one-child and a two-child policy for the next 30 years and found only a small difference under the two policies.

Given that the brand-new two-child policy cannot be expected to reverse the trend of a declining and aging population, the Chinese government will need to adapt to the predicted demographic changes in other ways. For example, China should speed up its discussion of postponing the usual retirement age. It should also move forward with the removal of the *hukou* registration system, in which people are designated as "rural" or "urban" and then hindered from migrating. The one-child policy has often been blamed for a labor shortage in China's urban areas. However, if the Chinese government would relax the *hukou* restrictions and phase in later retirement ages, the labor shortage in urban areas will be substantially improved. Finally, as China's population ages, the government needs to pay more attention to the reform of the pension systems in the urban and rural areas.

I have focused on the effects of the one-child policy on selected family outcomes. It would be valuable to analyze some important underexplored issues in future research. First, for a given supply of land, how did the one-child policy affect housing prices, the cost of living, and real incomes? Notably, the cost of living may affect family formation and ultimately family outcomes. Second, how did the one-child policy influence labor productivity, given labor supply? Higher schooling

(though a modest increase) should increase labor productivity. Third, how did the policy affect government budgets, given labor supply and productivity? An older population implies higher outflows, particularly for pensions and health.

References

Ahn, Namkee. 1994. "Effects of the One-Child Policy on Second and Third Births in Hebei, Shaanxi and Shanghai." *Journal of Population Economics* 7(1): 63–78.

Alba, Francisco, and Joseph E. Potter. 1986. "Population and Development in Mexico since 1940: An Interpretation." *Population and Development Review* 12(1): 47–75.

Banister, Judith. 1987. *China's Changing Population.* Stanford University Press.

Becker, Gary S. 1991. *A Treatise on the Family.* Harvard University Press.

Besley, Timothy, and Anne Case. 2000. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110(467): F672–94.

Birney, Mayling. 2014. "Decentralization and Veiled Corruption under China's Rule of Mandates." *World Development* 53: 55–67.

Cai, Fang. 2010. "Demographic Transition, Demographic Dividend, and Lewis Turning Point in China." *China Economic Journal* 3(1): 107–119.

Cai, Yong. 2010. "China's Below-Replacement Fertility: Government Policy or Socioeconomic Development?" *Population and Development Review* 36(3): 419–40.

Cai, Yong. 2012. "United Nation's Prediction: China Goes Rapidly to Aging." *Global Economic Review* [in Chinese] 2012(01): 73–95.

Cho, Nam-Hoon. 1996. *Achievements and Challenges of the Population Policy Development in Korea.* Korea Institute for Health and Social Affairs.

Ding, Qu Jian, and Therese Hesketh. 2006. "Family Size, Fertility Preferences and Sex Ratio in China in the Era of the One Child Family Policy: Results from National Family Planning and Reproductive Health Survey." *BMJ* 333(7564): 371–73.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91(4): 795–813.

Ebenstein, Avraham. 2010. "The 'Missing Girls' of China and the Unintended Consequences of the One Child Policy." *Journal of Human Resources* 45(1): 87–115.

Fitzsimons, Emla, and Bansi Malde. 2014. "Empirically Probing the Quantity–Quality Model." *Journal of Population Economics* 27(1): 33–68.

Goodkind, Daniel. 2011. "Child Underreporting, Fertility, and Sex Ratio Imbalance in China." *Demography* 48(1): 291–316.

Guo, Rufei, Junjian Yi, and Junsen Zhang. 2016. "Rationed Fertility: Theory and Evidence." Unpublished paper, Chinese University of Hong Kong.

Hao, Juan, and Changrong Qiu. 2011. "A Comparative Analysis of the Fertility of the Urban and Rural in China since 2000." *South China Population* [in Chinese] 26(5): 27–33.

Hardee-Cleaveland, Karen, and Judith Banister. 1988. "Fertility Policy and Implementation in China, 1986–88." *Population and Development Review* 14(2): 245–86.

Hesketh, Therese, Li Lu, and Wei Xing Zhu. 2005. "The Effect of China's One-Child Family

Policy after 25 Years." *New England Journal of Medicine* 353(11): 1171–76.

Hu, Naijun, and Yansui Yang. 2012. "The Real Old-Age Dependency Ratio and the Inadequacy of Public Pension Finance in China." *Journal of Population Ageing* 5(3): 193–209.

Huang, Wei, Xiaoyan Lei, and Ang Sun. 2016. "When Fewer Means More: Impact of One-Child Policy on Education of Girls." Working paper, Harvard University.

Li, Bingjing, and Hongliang Zhang. Forthcoming. "Does Population Control Lead to Better Child Quality? Evidence from Regional Variation in the Enforcement Intensity of China's One-Child Policy." *Journal of Comparative Economics*.

Li, Hongbin, and Junsen Zhang. 2007. "Do High Birth Rates Hamper Economic Growth?" *Review of Economics and Statistics* 89(1): 110–17.

Li, Hongbin, Junsen Zhang, and Yi Zhu. 2005. "The Effect of the One-Child Policy on Fertility in China: Identification Based on the Differences-in-Differences." Working Paper.

Liu, Haoming. 2014. "The Quality–Quantity Trade-off: Evidence from the Relaxation of China's One-Child Policy." *Journal of Population Economics* 27(2): 565–602.

Liu, Huangsong. 2013. "China's Potential Economic Growth Rate in the Next 20 Years Based on the Analysis of Demographic Factors." *Shanghai Journal of Economics* [in Chinese] 2013(1): 13–21.

Mao, Zedong. 1966. *The Selected Works of Mao Zedong*. People's Publishing House.

Mauldin, W. Parker. 1982. "The Determinants of Fertility Decline in Developing Countries: An Overview of the Available Empirical Evidence." *International Family Planning Perspectives* 8(3): 116–21.

McElroy, Marjorie, and Dennis Tao Yang. 2000. "Carrots and Sticks: Fertility Effects of China's Population Policies." *American Economic Review* 90(2): 389–92.

Peng, Peiyun. 1996. *A Complete Book of China's Family Planning* [in Chinese]. Beijing: China's Population Publisher.

Peng, Xizhe. 1991. *Demographic Transition in China: Fertility Trends since the 1950s*. Oxford: Clarendon Press.

Peng, Xiezhe, and Zhigang Guo. 2000. *The Changing Population of China*. Blackwell Publishers.

Poston, Dudley L., Jr., and Chengrong Charles Duan. 2000. "The Current and Projected Distribution of the Elderly and Eldercare in the People's Republic of China." *Journal of Family Issues* 21(6):

714–32.

Poston, Dudley L., and Baochang Gu. 1987. "Socioeconomic Development, Family Planning, and Fertility in China." *Demography* 24(4): 531–51.

Rosenzweig, Mark R., and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's 'One-Child' Policy." *Review of Economic Studies* 76(3): 1149–74.

Scharping, Thomas. 2003. "Birth Control in China 1949–2000: Population Policy and Demographic Development." *Routledge*.

Tian, Xueyuan. 2009. *Sixty Years of China's Population Policies* [in Chinese]. Social Sciences Academic Press.

Tuljapurkar, Shripad, Nan Li, and Marcus W. Feldman. 1995. "High Sex Ratios in China's Future." *Science* 267(5199): 874–76.

United Nations, Department of Economic and Social Affairs, Population Division. 2009. *World Population Prospects: The 2008 Revision*. Available at: https://esa.un.org/unpd/wpp/.

United Nations, Department of Economic and Social Affairs, Population Division. 2011. *World Population Prospects: The 2010 Revision*. Available at: https://esa.un.org/unpd/wpp/.

United Nations, Department of Economic and Social Affairs, Population Division. 2015. *World Population Prospects: The 2015 Revision*. Available at: https://esa.un.org/unpd/wpp/.

Wang, Feng, Yong Cai, and Baochang Gu. 2013. "Population, Policy, and Politics: How Will History Judge China's One-Child Policy?" *Population and Development Review* 38(Special Issue: Population and Public Policy: Essays in Honor of Paul Demeny): 115–29.

Wang, Fei, Liqiu Zhao, and Zhong Zhao. 2016. "China's Family Planning Policies and Their Labor Market Consequences." IZA Discussion Paper 9746.

Whyte, Martin King, Wang Feng, and Yong Cai. 2015. "Challenging Myths about China's One-Child Policy." *China Journal* 74: 144–159.

World Bank. 2016. *World Development Indicators*.

Zhang, Junsen, and Byron G. Spencer. 1992. "Who Signs China's One-Child Certificate, and Why?" *Journal of Population Economics* 5(3): 203–215.

Zhang, Yuanting, and Franklin W. Goza. 2006. "Who Will Care for the Elderly in China? A Review of the Problems Caused by China's One-Child Policy and Their Potential Solutions." *Journal of Aging Studies* 20(2): 151–64.

The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops

Claudia Goldin and Joshua Mitchell

For US women born before the 1950s, labor force participation over the life cycle followed a distinct inverted-U-shape—or what could be termed a "hump." Specifically, labor force involvement for birth cohorts of women before the 1950s rose steeply from the time the women were in their 20s to when they were in their late 40s. Participation rates then decreased after women reached their early 50s. However, the hump has now disappeared.

A new life cycle of women's labor force participation has emerged. For cohorts born since the mid-1950s, the female labor force participation rate is high during the decade or so after schooling ends. Labor force participation rates then decrease somewhat when women are in their 30s and early 40s, a feature we term the "sagging middle." Participation then increases a bit, before phasing out as cohorts move into their 60s and beyond. We cannot yet observe more recent cohorts in their older years, but for earlier cohorts, labor force participation has greatly expanded for women in their 60s and 70s relative to previous cohorts. Thus, it seems plausible that later stages of the life cycle will involve a more prolonged phasing out of work— an expanding top relative to the past. The full new life cycle of women's labor force participation now looks relatively high and fairly flat, vaguely like that of men's but with a somewhat lower level and squishier middle.

■ *Claudia Goldin is Henry Lee Professor of Economics, Harvard University, and Research Associate, National Bureau of Economic Research, both in Cambridge, Massachusetts. Joshua Mitchell is a Research Economist, US Bureau of the Census, Washington, DC. Their email addresses are cgoldin@harvard.edu and joshua.w.mitchell@census.gov.*

The combination of the sagging middle and expanding top has produced a "twist" in the labor force rates of cohorts at the start, middle, and completion of their working lives. What we mean by a twist is that historically, more recent cohorts have had the highest participation rates at each age and earlier cohorts the lowest. It remains true that for women in their 20s and 30s, the most recent cohorts have the highest labor force participation levels, followed by the earlier cohorts, and so on in almost perfect chronological order. A similar pattern is found for women older than around 55 years: that is, the most recent cohorts that can be followed to those ages also have the highest levels of labor force participation and the earlier cohorts have lower levels in strict year-of-birth order. But in the middle years, the most recent cohorts of women have somewhat lower labor force participation than some of the earlier ones. Thus, there has been a twist in the ordering.

Our description of changes in labor force participation relies on three customary effects: period (year), cohort (year of birth), and life-cycle (age). These three effects are linearly related (for example, the current year = year of birth + age). Yet despite the inherent difficulty for researchers in identifying their separate influences, there are reasons to believe that there are different forces are at work. Period effects influence all individuals in a year, independent of their age. Wars or recessions, for example, could lead individuals at all ages to increase or decrease their desired labor supply during a given period. Cohort effects determine the intercept of a life-cycle path, in effect shifting life-cycle labor force participation of a cohort up or down. Each cohort can have a similarly shaped labor force path but be above (or below) the others. The life-cycle effect determines the shape of the function by age and can be altered by changes in the age at marriage and at first birth, among other factors. We will assume in this discussion that period or year effects are negligible and that cohort and life-cycle effects dominate.

Both the sagging middle and the expanding top have attracted attention. The observations of a sagging middle led, around a decade ago, to a disparagement that young women were "opting out." The expanding top has recently led some to comment optimistically that older women are working in greater numbers than before because they are healthier and find greater enjoyment in their jobs. Others, expressing some pessimism, have noted that many women have insufficient financial resources to enjoy their older years and end their employment (for discussion, see the papers in Goldin and Katz forthcoming B).

The changes in employment in the middle and the top of the age range may appear to be opposing trends. We will argue that they are not. The sagging middle emerged because the increases in the cohort effect have decelerated and have been trivial recently. Instead, we are now seeing the life-cycle or aging effect almost entirely. The life-cycle effect, moreover, changed with the 1950s and 1960s birth cohorts. Those cohorts of women began to marry later and have their children at older ages than did previous cohorts (on the role of the contraceptive pill, see Bailey 2010; Goldin and Katz 2002). Most of their members participated in the labor force early on and delayed childbearing. Some withdrew for a while in their middle years and later returned. A large fraction will (most probably) have a less steep decline

in employment in their later years than did previous cohorts. But since none of the more recent cohorts has yet to reach their older years, the later chapters of their life-cycle story have yet to be written.

The female population has been distinctly heterogeneous in its labor supply for some time (Goldin 1989; Heckman and Willis 1977). Labor supply heterogeneity means that women who are in the labor force remain in for a long time, while those who are out of the labor force enter as the cohort rate increases. As they enter, they, too, remain in. That is, heterogeneity in this case is based on the observation that there is considerable persistence among those currently in the labor force. In contrast, a homogeneous labor force would mean that all women work an equal fraction of the year, sometimes low and sometimes high. Persistence has implications for the role of employment early in life for that later, and it also implies consequences for lengthy spells out of the labor force to care for children and others. We first explore the general labor force trends and then examine the heterogeneity of the population as the earlier life cycle of women's employment has morphed into the new.

We first map out the general trends using synthetic, rather than actual longitudinal, cohorts that we have created based on data from the Current Population Survey (CPS). We then move to using true longitudinal data from the Survey of Income and Program Participation (SIPP) and the Health and Retirement Study (HRS), both linked to the Social Security Administration (SSA) earnings data (from 1957 for the SIPP and from 1951 for the HRS) and income tax (W-2) records (from 1978 for the SIPP and 1980 for the HRS). We estimate the distribution of years for women in the labor force and examine the heterogeneity of female labor force participants. We then turn to the changing impact of births on employment using an event study analysis and also consider the role of leave policy. We end with a discussion of the reasons why these changes have occurred and the future of US female employment.

Our bottom line is that the US female labor force has greatly expanded and evolved, but that birth events that had always produced a temporary retreat from employment are now occurring later with the delay in marriage and childbirth. They are, moreover, more apparent because of the increase in employment at younger ages. The increased employment of women in their older years appears to be a continuing trend, but only time will tell.

The Evolving Life Cycle of Women's Employment

A "synthetic" birth cohort links age groups over time for a given cohort: for example, those who were born from 1935 to 1940 will all be between ages 25 and 30 in 1965, and between ages 30 and 35 in 1970, and so on. In this way, one can track the experiences of the group over time without having data on specific individuals. One can condition on time-invariant variables such as education level (for those beyond school age) and birthplace. Synthetic labor force participation rates for

different cohorts can be created by linking data by birth cohort using the annual figures from the CPS March surveys (also known as the CPS Annual Social and Economic Supplement (ASEC)). We do this for all women and also by education level for college graduates and all others. We use only native-born women when those data begin in 1994.[1]

Our focus is on nine cohorts born during five-year intervals from 1930 to 1974, which are chosen for consistency with our later discussion of longitudinal administrative and survey data from the Survey of Income and Program Participation and the Health and Retirement Study. We begin the analysis with age 25 to avoid confusing increases in higher education with decreases in labor force participation. Because the Current Population Survey microdata starts with 1962, we cannot include information for some of the early cohorts in their younger years.

Figure 1A shows the results of this synthetic cohort labor force data for women born from 1930 to 1974. Labor force participation rates for women have generally increased with each cohort, as shown by increases in the intercept. Within cohorts, participation has often increased for a time, before declining. However, labor force participation has not uniformly increased at each point along the life cycle for each subsequent cohort. Participation has clearly not increased at each point along the life cycle across the most recent cohorts. The most recent cohorts have the highest participation rates relative to other cohorts at ages from the mid to late 20s (their lines are the highest at the upper left of the figure); they no longer do in their middle years. For the cohorts we can observe in their older years, the ordering of cohorts by participation rate returns to one that is more strictly chronological. These features have produced the sagging middle of lower labor force participation rates among women in their 30 and 40s, along with a twist in terms of the ordering of the cohorts.

Figure 1B shows labor force participation rates by cohort for college graduate women, and Figure 1C gives detail on the five most recent college graduate cohorts born in five-year intervals from 1950 to 1974. The fraction of women born in the 1980s who will be college graduates by the time they are 35 years old is today almost 45 percent.[2] Therefore, the new life-cycle labor force participation of women is tending to look more like the five cohorts in Figure 1C—beginning high, dipping down a bit in the mid-30s, and then increasing again. Because the earlier cohorts among these five did not start out as high and did not dip as much as the most recent ones, the cohort lines are no longer one on top of each other in strict chronological procession. Rather, the arrangement of the lines distinctly twists. For early

---

[1] The reason to use native-born only is because the foreign-born could enter at any age. In addition, our later use of the Health and Retirement Study and Survey of Income and Program Participation requires using the native-born since we use longitudinal information from Social Security records. See the online Appendix available with this paper at http://e-jep.org for figures that include all women regardless of birthplace.

[2] The fraction who are college graduates is about 40 percent for native-born women from the 1980 cohort who were 35 years old, and extrapolations suggest that for the 1987 cohort, 44 percent will graduate by 35 years old. Calculations use the March CPS-ASEC.

ages, the order begins chronological with the latest cohort on top with the highest participation. But the order reverses by the early 40s, with earlier cohorts having the highest participation and the more recent ones the lowest. It is still too early to know whether the ordering at later ages in the life cycle will once again return to the strict chronology, but from the slopes of the lines, it looks like it will.

In this new life cycle, for all education groups, the hump-shape of labor force participation apparent for earlier cohorts of women largely disappears. Instead, participation rates for the average woman do not change much until older ages with the phasing out of employment. One way to think about these changes is that the cohort effect has become swamped by the life-cycle effect. For earlier cohorts, each line is above its predecessor, but with much the same shape. However, participation rates for recent cohorts (1950s onwards) in their 20s are high and do not vary much over the life cycle. With a diminished cohort effect, the life-cycle effect of decreased participation in child-rearing years has become more apparent. Because the child rearing years are now later, a sagging middle has resulted. We will show the effects of child-rearing and persistence in the labor force in the next section, using true longitudinal information by mothers' cohort.

Longitudinal Data

Aggregate synthetic cohort data can demonstrate the evolution of a new life cycle of women's employment. But these data cannot reveal the degree to which specific women persisted in the labor force and whether those in the labor force earlier in their lives remained in with a greater likelihood. The synthetic cohort data cannot show how women's employment has changed by cohort over time in response to important life cycle events, such as births. Moreover, synthetic cohort data do not allow us to distinguish among women within a cohort to see the fractions of their post-schooling lives that are spent in the labor force and how that distribution changed within and across cohorts.

To make better sense of the evolution of the new life cycle of work we turn to longitudinal data from the Survey of Income and Program Participation and the Health and Retirement Study, both linked to the Social Security Administration earnings records and W-2 forms. The linkage to the SSA records provides extensive longitudinal information on the earnings of large numbers of individuals across cohorts born from the early 1930s to the mid-1970s. This section describes our longitudinal data; the next section describes the labor force patterns based on the data.

These two longitudinal datasets offer rich and complementary information. Because they are each complicated in their construction, we will summarize only those aspects pertinent to this article.[3] The Health and Retirement Study began

---

[3]See the online Appendix available with this paper at http://e-jep.org for more information on the Health and Retirement Study and the Survey of Income and Program Participation.
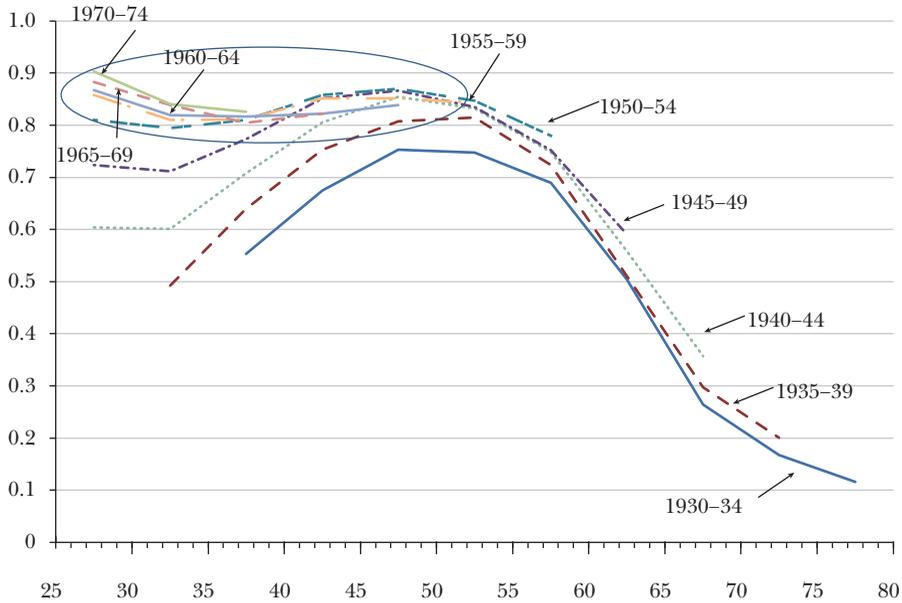
Figure 1

Female Labor Force Participation Rates by Cohorts Born from 1930 to 1974 by Five-Year Age Groups and Five-Year Birth Cohorts
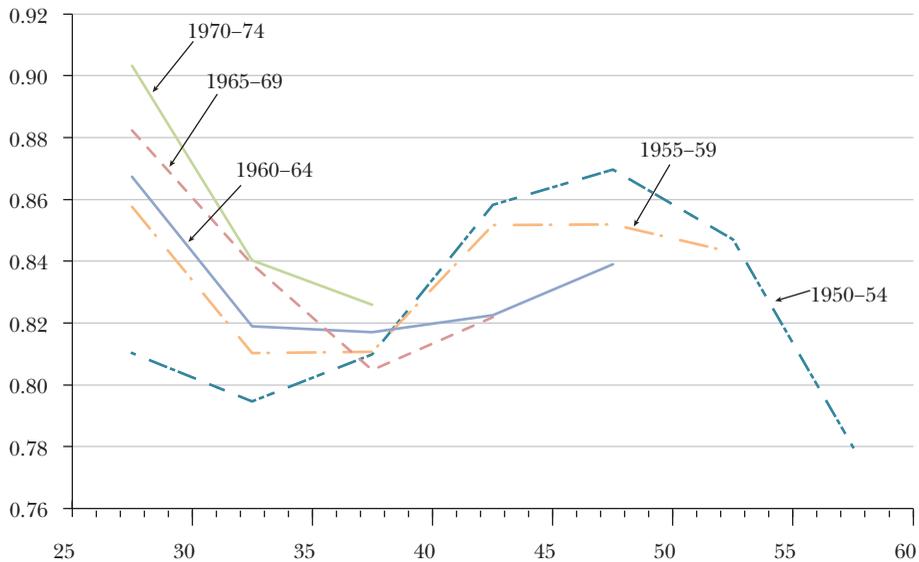
A: All Education Groups
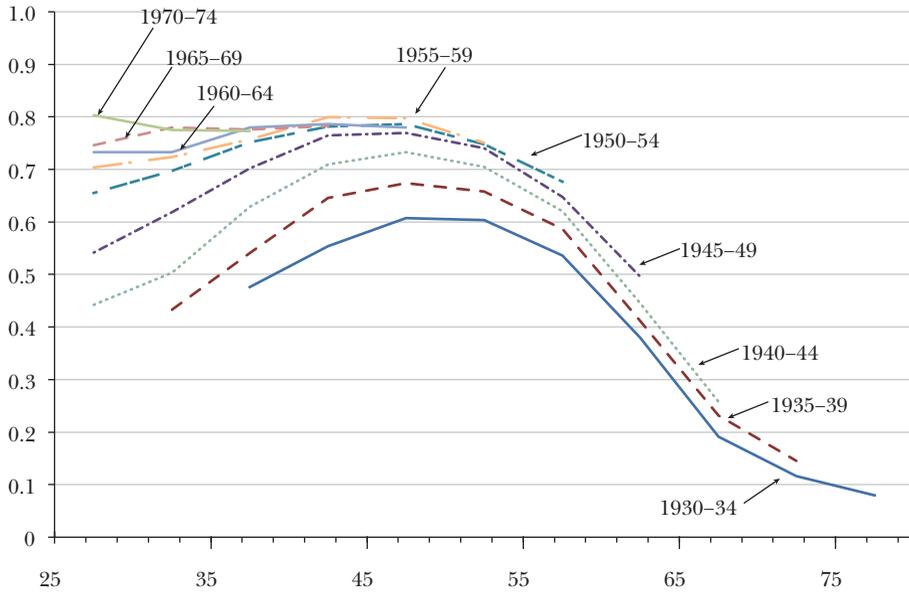
B: College Graduates

C. College Graduates, Four Recent Cohorts (Area Circled In Figure 1B)

Source: CPS-ASEC microdata, March, 1963 to 2014.
Notes: Every point on the graphs is the average of 25 cells (5 single years of age groups and 5 single years of birth cohorts). Native-born only women can be identified in the CPS-ASEC in 1994 and after. Data for native-born only are shown for the 1955–59 birth cohorts 40–44 to 50–54 years; 1960–64, 35–39 to 45–49 years; 1965–69, 30–34 to 50–54 years; and 1970–74, 25–29 to 45–49 years. The difference between the total and that for native-born only is small and is largest (0.03) for the 1970–74 birth cohort. The graphs for all women are in the online Appendix available with this paper at http://e-jep.org.

in 1992 with 51 to 61 year-olds who were then interviewed biennially. Additional cohorts were added in 1998, 2004, and 2010 for respondents who were then 51 to 56 years old. Together with the spouses of the respondents who became age-eligible at some later date, these are the main birth cohorts we use from the HRS. They span birth years from 1931 to 1959. Respondents were given the option of having their Social Security earnings records linked to their HRS surveys. Because this was done during each interview, the earlier cohorts have a higher fraction linked. Linkage rates are 80 percent on average and about 88 percent for those born before 1943.

The Survey of Income and Program Participation was begun in 1984 with new panels added in 1996, 2001, 2004, and 2008. Each panel begins with individuals who are between 30 and 60 years old and are interviewed for four consecutive years. We use the Gold Standard File, which is a harmonized set of SIPP panels linked to longitudinal earnings records.[4] Our analysis uses SIPP panels 1996, 2001, 2004, and 2008 and integrates information from the fertility history topical modules. Our overall

---

[4] Data from the SIPP Gold Standard File are confidential. All results have been formally reviewed to ensure that no confidential Census Bureau data have been disclosed.

sample begins with women who range from around 30 to 60 years old at the time of their fertility history interview.

Our primary interest is in the work history information. Each of the two datasets has a survey component and an administrative portion from the Social Security Administration earnings records. Although the administrative component is identical, the datasets differ in their coverage of retrospective information that bears on the work history. For example, the Health and Retirement Study provides retrospective information on the respondent's longest occupation and also the years when the individual worked for a government agency for upwards of two periods. The Survey of Income and Program Participation contains information that bears on whether the woman took job-protected or paid leave after having a birth, whether she returned to the same employer after that leave, and whether she quit her current job around the time of the birth event. Both the HRS and the SIPP contain variables that are (reasonably) time invariant (for example, the level of education for those beyond age 35, or children ever born by age 40) and both have time-variant longitudinal information for the duration of the surveys.

The work history information from the Social Security Administration earnings records, W-2 forms, and the survey data provide annual labor earnings, but not labor force participation, or hours and weeks of work, except for the survey years. We generate an estimate of labor force "participation" for the years we have the SSA and W-2 records by assuming that individuals are labor force participants if they earned more than some minimum amount—equivalent to 10 hours a week for 52 weeks at the federal minimum wage—in that year. Our estimated participation rates are nearly identical to those from the Current Population Survey–Annual Social and Economic Supplement (CPS-ASEC) for the overlapping years.[5]

We cobble together our data on cohort labor force participation by using each of our longitudinal datasets when it seems the most complete. For example, some occupations such as most teachers and other government employees were exempt from Social Security tax and would not have earnings reported in the Social Security Administration data even when they were employed. But they can be included after 1977 when W-2 data are available for the Survey of Income and Program Participation (SIPP) and after 1979 for the Health and Retirement Study. These workers can also be folded in for the HRS in the years the respondent listed retrospective information on government employment. Our choice of whether to use the SIPP-SSA or the HRS-SSA data is a function of the birth cohort and the age of the individual. Because we employ ten-year age intervals, our decision depends on the youngest age in the interval. The exempt worker issue is far less of a problem

---

[5] For details of the comparison between our estimate and the Current Population Survey estimate, see the online Appendix available with this paper at http://e-jep.org. The CPS labor force estimate comes from a question about whether the individual was working for pay or profit during at least one hour in the survey week or was actively searching for work. The Social Security administration data are annual and there is no obvious amount of annual income that would be equivalent to the CPS labor force question. Because most labor force respondents are working a reasonable number of hours during a survey week, we chose our definition of ten hours at the minimum wage.

for the non-college-graduate group since they would not have been teachers and are less likely to have been government employees in general. The SIPP and HRS longitudinal labor force data that we generate from the SSA records closely match each other for overlapping birth cohorts, so any measurement error introduced by comparisons between the two longitudinal datasets should be modest.

Evidence on New and Old Life Cycles from Administrative Data

Labor Force Experience

Our longitudinal data from a combination of these datasets allow us to estimate labor market experience for women born from 1935 to 1974 by age and by education. The aggregate data are given in Figure 2 for three groups—all women, college graduates, and non-college-graduates. The data are shown for the full 25 to 54 year-old group, then for the youngest group 25 to 34 years old, and finally for all women in three ten-year age groups using more high-frequency birth cohorts. Only longitudinal data can be used to construct work experience; for example, the Current Population Survey did not ask respondents how long they had been employed.[6]

The entire 25–54 year-old group, shown in Figure 2A, can be observed for (native-born) birth cohorts up to 1959. For those cohorts, mean years of work experience in that 30-year interval increased from 16.4 to 22.2. For the most recent of the cohorts in our data (1955–59), the average woman was employed for 74 percent of the 30-year period. For college graduates in the most recent cohorts, the figure is 82 percent. Much of the total increase in mean years of work experience across successive cohorts occurred in the youngest of the age groups (25–34). Figure 2B shows cumulative experience within the 25–34 year group for birth cohorts from 1935 to 1974.

The increase in mean years of work experience for the youngest age group was large: slightly more than half of the total increase from 25 to 54 years old among women born from 1935–39 to 1955–59 (3.05 of the total of 5.9 years) occurred in the 25–34 year group. But most of the increase in mean years of work experience across successive cohorts from 1935–39 to 1970–74 occurred for cohorts born before 1959. The increase for the 25–34 year-old group from the 1955–59 to the 1970–74 birth cohorts was less than an additional year.

For the 25–34 year-old group, average work experience for the 1970–74 birth cohorts reached around 8.7 years for college graduates and around 7.3 years for the non-college group. The change from the 1935–39 to the 1970–74 birth cohorts for all women, it should be noted, was almost four years, a bit greater than for the two separate educational components—the college and non-college groups—because

---

[6] In another study using a different source of longitudinal data, Attanasio, Low, and Sánchez-Marcos (2008) use the Panel Study of Income Dynamics (PSID) and analyze the different life-cycle employment among three cohorts of women, those born at the end of the 1930s, 1940s, and 1950s.
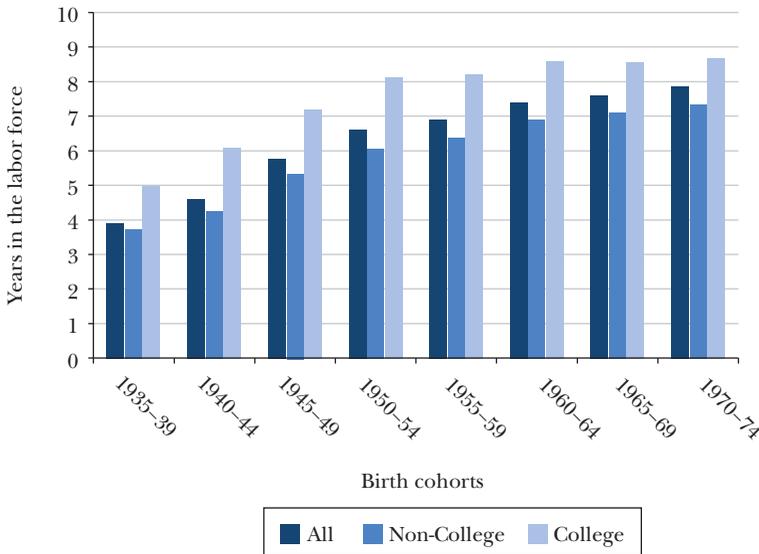
Figure 2

Labor Force Experience for Women Born 1935 to 1974 and by Education Level

A: Cumulative Experience from 25 to 54 years
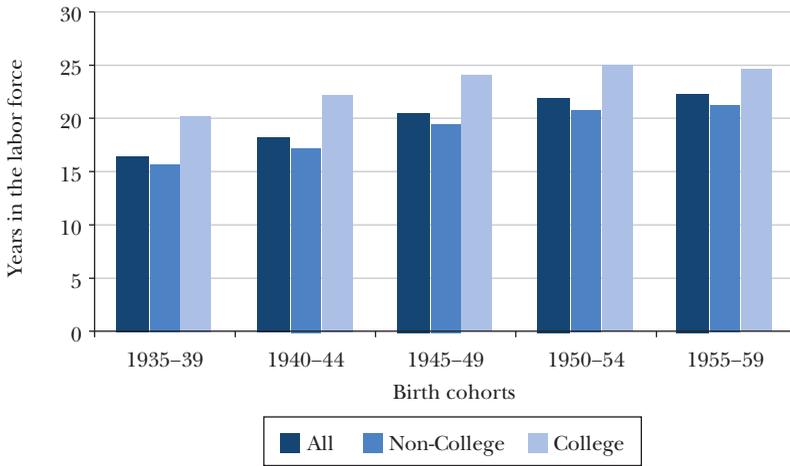
B: Cumulative Experience from 25 to 34 Years

of a relative increase in the college graduate group. But the main findings are not much different for each group separately.

Figure 2C puts the three ten-year age groups together in one graph. Total cumulative experience for the 25–34 year-old group doubled from around 3.9 years for the late 1930s cohorts to 7.85 years for the early 1970s cohorts. The other two ten-year age groups show sturdy increases until around the early 1950s cohorts. Looking

C: Labor Force Experience of All Women by Ten-Year Age Groups
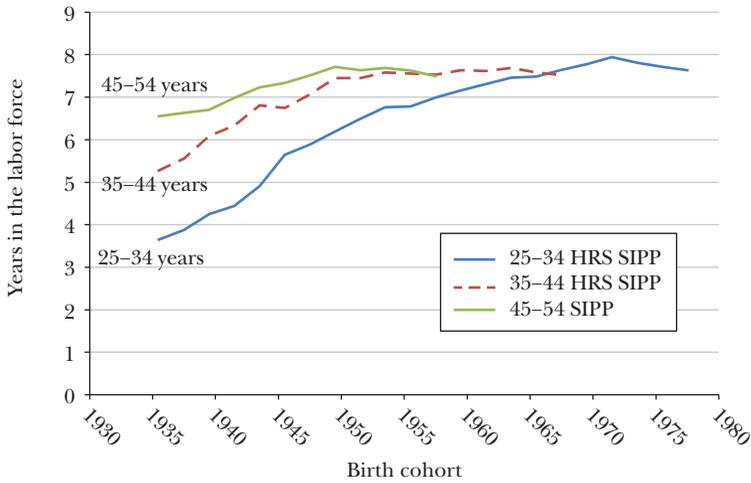
Sources and Notes: Labor force experience for an age group and birth cohort (only native-born) is computed from our estimates of participation rates (see online Appendix) from the Health and Retirement Study–Social Security Administration (HRS-SSA) and Survey of Income and Program Participation–Social Security Administration (SIPP-SSA). We use HRS-SSA data for 25–34 year-olds for birth cohorts from 1935 to 1954 and for 35–44 year-olds for the 1935 to 1944 birth cohorts. For all other cohorts, we use the data from the SIPP-SSA. To compute the totals for the 25–54 year span, we sum the separate ten-year age groups independent of the source. In all cases, we use native-born women only. The data for overlapping cohorts between the HRS-SSA and SIPP-SSA are generally very close. Our reason for using the HRS-SSA rather than the SIPP-SSA concerns the issue of exempt workers. See text.

at the cohorts born around 1965 reveals that the fraction of time that 25–34 year-old women were in the labor force began to exceed that for the 35–44 year-old group (both being around 7.5 years out of the 10). Previously the older group's fraction had exceeded the younger for all previous cohorts. Delay of childbirth, we will soon suggest, led to increased participation for the youngest group but also caused slower increases for the middle group—yet another way of understanding the appearance of a sagging middle.

Distribution of Work Years

Work experience clearly increased for women across birth cohorts from the 1930s to the 1970s. But the aggregate numbers do not reveal the distribution of work years: for example, a 60 percent labor force participation rate could mean that all women work 60 percent of the time or that 40 percent are never at work and 60 percent work full-time. The former scenario is termed "homogeneous" (because all women are the same) and the latter "heterogeneous" (because women in that scenario greatly differ). Most estimates have found considerably more heterogeneity than homogeneity, and our data will reveal the same. But as labor force participation rates rise, there is less room for heterogeneity.

To explore heterogeneity among labor force participants we compute the distribution of years in the labor force for each birth cohort by age interval. As summary statistics, we provide the fractions at the two tails of the distribution: the fraction working more than 80 percent of the period and the fraction working less than 20 percent. For the most recent cohorts that can be observed across the full 30-year period from age 25–54—that is, those born in 1957–58 in our data—53.3 percent of women were employed for more than 80 percent of the 30-year period and just 9 percent were employed for less than 20 percent, as shown in the right-hand tails of the lines in Figure 3A. For the earlier cohorts, born in the 1930s, the distribution of employment across the life cycle is far different. Employment for the 1930s cohort is almost uniformly distributed across the quintiles, with around 20 percent employed less than 20 percent of the entire period and 20 percent employed for more than 80 percent (as shown on Figure 3)—and about 20 percent employed in each of the other quintiles (not shown).

Because labor force participation for these early cohorts was only around 30 percent for the 25–29 year group and then rose to around 60 percent as the cohorts aged to 50–54 years old, the findings on the distribution of life-cycle employment for these cohorts are more consistent with a heterogeneous model of participation rather than a homogeneous one.

Figure 3B presents the same evidence, but for the 25–34 year-old group, thus also covering more recent cohorts. For the most recent cohorts shown, greater than 60 percent are employed more than 80 percent of the ten years, whereas only about 16 percent had been for the earliest cohorts shown, those born in the 1930s.

There has been considerable persistence in participation among women: that is, those who work more when young continue in the labor force when older to a greater degree than those who worked less when young. We have examined the labor force participation of two groups in the 1950–54 cohort: those who worked more than six of the years from 25 to 34 years old and those who worked five or fewer years. Among women who worked the longer period, 77 percent later worked more than 80 percent of the period from 35 to 44 years old and about the same (76 percent) did so from 45 to 54 years old. Conversely, of those who worked the shorter period when young, just 32 percent were employed more than 80 percent of the 35–44 year period and 50 percent were employed more than 80 percent from 45 to 54 years old.

There are important implications of increased life-cycle employment for continued work later in life. In related work, Goldin and Katz (forthcoming A) find that greater employment early in one's life is strongly related to employment at the later ages (they examine participation at ages 59 to 63) given education and birth cohort. Thus, the increase in life-cycle employment for those 25–54 implies delayed retirement. But even though the college educated would appear to have hit a plateau in their life-cycle employment around the 1950 cohort, Goldin and Katz (forthcoming A) caution that other factors have led to the increased employment of college graduate women at older years and will probably do the same for the more recent cohorts.
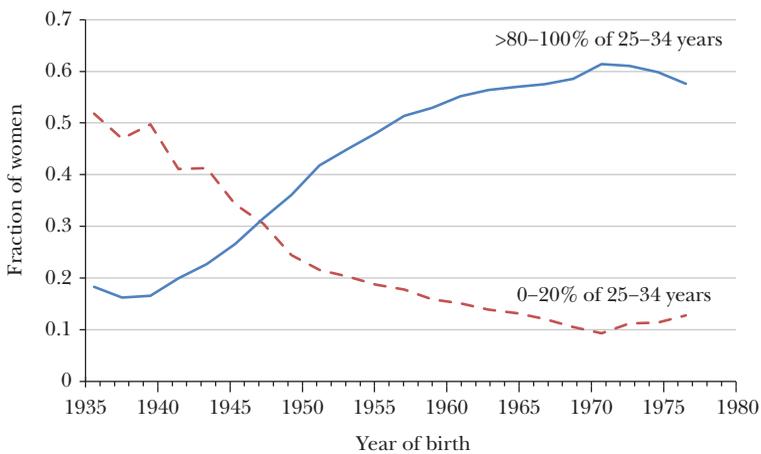
Figure 3

Heterogeneity in Life-Cycle Labor Force Participation by Age Group and Birth Year

A: All Women, 25 to 54 years
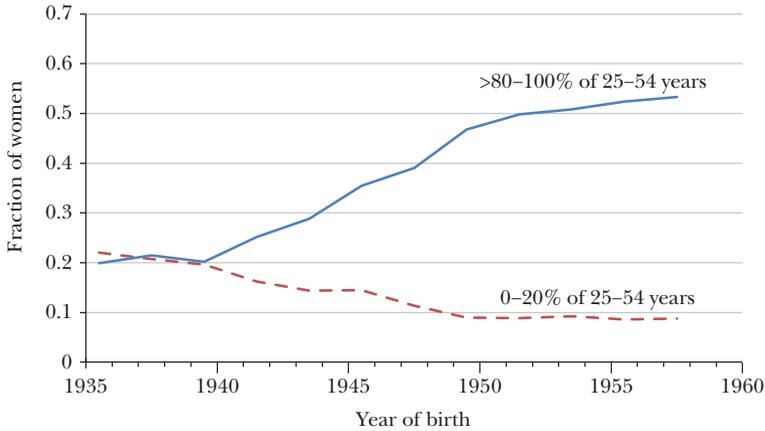
B: All Women, 25 to 34 years

Sources and Notes: Survey of Income and Program Participation–Social Security Administration (SIPP-SSA) data are for native-born women only. Health and Retirement Study–Social Security Administration (HRS-SSA) gives somewhat higher fractions in the lowest and highest groups in the earlier period. Because the fraction of college graduate women in these cohorts is not very large, we chose to use a consistent series for the distribution even though some workers in exempt occupations will be excluded prior to 1978 using the SIPP.

What about the role of delayed childbirth for the most recent cohorts? We find, using the Health and Retirement Study, that an increase in the age at which the first child is born, say from 25 years to 30 years, is correlated with increased participation in the 25–34 year interval but decreased participation in the 35–44 year interval,

even holding the number of children born constant.[7] This finding implies that a later age at first birth is an important factor in the twist in life-cycle labor force participation. It is still the case, however, that later births mean greater participation for the entire 25–44 year period probably because substantial human capital investments are made early on. We turn now to an analysis of the role of childbirth and labor force participation across cohorts.

Childbirth and Life-Cycle Labor Force Participation

The changed timing and number of children are important parts of the transition to the new life cycle of women's employment. Not only are children in more recent cohorts being born to older mothers, but also there are fewer children in these families than in the earlier cohorts. The previous norm was one in which women had their children when young, then left the labor force and re-entered employment somewhat later. In the current era, women have their children when older, have greater attachment to the labor force, take less time off, and later re-establish their employment and careers faster. The one possible exception, we will see, is college graduate women in the most recent cohorts we can track. Their participation rates at first birth are very high but do not return to those levels a decade after the first birth.

Both the Survey of Income and Program Participation and the Health and Retirement Study contain information on the year of birth for the first child and the number of subsequent births. We use the data to create event studies for all women who had a first birth. The event study evidence is given in Figure 4A for all women and separately in Figure 4B for those with a four-year college degree. We use the HRS for the 1935 to 1949 birth cohorts and the SIPP for the 1950 to 1969 birth cohorts for reasons mentioned earlier (specifically, the HRS is better at identifying workers exempt from Social Security in the pre-1978 period).

Looking first at all women in Figure 4A, the cohorts born from 1935 to 1944 had initial participation rates a bit higher than 0.5 before the first birth. These rates plummet to around 0.24 to 0.28 just after the birth and never recover to pre-birth levels in the ten subsequent years. These are "baby boom" mothers, for whom the number of children (conditional on having one) is 3.26 for the 1935–1939 cohort and 2.85 for the 1940–1944 cohort (using the CPS June Fertility Supplements microdata for women 40 years and older).[8] A more detailed look at the data shows

---

[7]The regressions are estimated for the fraction of the 25–34 year and 35–44 year intervals in the labor force across birth cohorts from 1931 to 1954 for women with at least one birth. Cohort, number of children, education, and race dummies are included. Age of the mother at the first birth is entered as a quadratic.

[8]In the online Appendix available with this paper at http://e-jep.org, Appendix Table 1 contains the mean age at first birth to ever-moms, the numbers of children eventually born to ever-moms, and the fraction of the group with zero births in the Health and Retirement Study and Survey of Income and Program Participation. Appendix Table 1 also has the numbers of children for ever-moms and fraction of the group with zero births from the much larger CPS June Fertility Supplements microdata. Data are provided for all and college graduate women (native-born in the more recent cohorts due to availability in the CPS).

A: All Women

B: College Graduate Women
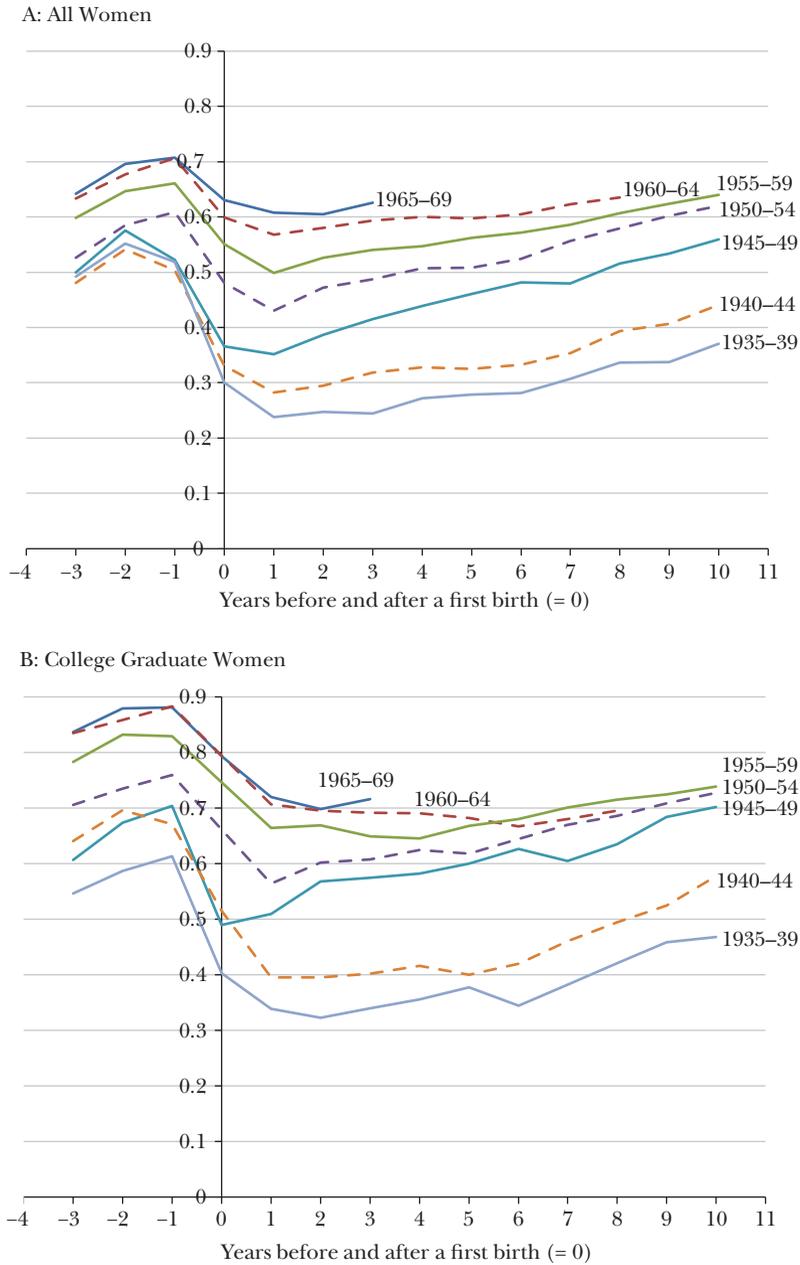
Sources and Notes: Includes only native-born women with a first birth. Health and Retirement Study–Social Security Administration (HRS-SSA) is used for the 1935 to 1949 birth cohorts; Survey of Income and Program Participation–Social Security Administration (SIPP-SSA) is used for 1950 to 1969 birth cohorts. See text.

that even for women with just one or two births until they are in their 40s, labor force participation rates never reach pre-birth levels in the next ten years. Mothers in the 1935–44 birth cohorts retreated from the labor force for some time. But since births are staggered and pre-birth participation rates for these cohorts were low, the increase in participation across the life cycle as women aged produced the hump shape seen in Figure 1.

Labor force recovery for the 1945–49 cohorts, unlike their predecessors, is complete by ten years after the first birth even though the initial participation rates are much higher. Participation rates for the cohorts born from 1950–54 to the early 1960s begin around 0.6 to 0.7, and for those that can be observed ten years out, rates equal or exceed those before the birth. The number of eventual births was 2.48 in the 1945–49 cohort and 2.29 for the 1955–59 cohort, again conditional on having at least one child.[9]

The data for the college group shown in Figure 4B are similar but noisier, due to the smaller number of observations. The levels are considerably higher than for the total group that include all education levels. As with the total group, there is a sharp break with the 1945–49 cohort. Whereas participation rates of previous cohorts did not fully recover, the 1945–49 and 1950–54 cohorts did so after ten years. Subsequent cohorts, however, have pre-birth labor force rates around 0.83 to 0.88, and those that we can observe ten years after the birth do not fully recover. In fact, the rates of labor force participation six or more years after the first birth are lower for the 1960–64 cohorts than for the 1955–59 cohorts. These facts are consistent with the crossing of the synthetic cohort participation lines for college graduates born from 1955–59 and 1960–64, shown earlier in Figures 1B and 1C.

One may have wondered why there is an increase in participation just before the birth. Because some births occur just a few years after school completion, labor force participation often increases in the three years preceding the first birth in both Figure 4 graphs. When we focus only on first births occurring after age 24, labor force participation no longer increases in the pre-birth years.

The effect of children on the employment of women in their 20s and early 30s has consequences for later employment. Thus, fewer years out of the labor force by mothers in more recent cohorts, as seen in Figure 4A, is predictive of greater employment later in their lives.

The modules in the Survey of Income and Program Participation contain information on the use of paid and unpaid leave during the year a woman's first child is born, as well as whether the woman quit her current job in that interval. Rather

---

[9] For details, see online Appendix 1. These means are conditional on having at least one birth and use the CPS June Fertility Supplement microdata for women 40 years and older. Similar findings are given in Attanasio, Low, and Sánchez-Marcos (2008, figure 7) although they present data on employment only 12 months before and 12 months after the birth. See also Angelov, Johansson, and Lindahl (2016) on Sweden and Kleven, Landais, and Søgaard (2016) on Denmark for birth event analyses that track labor force participation, hours, and earnings for at least ten years after the birth. Kleven et al. find substantial and persistent decreases in labor force participation that are greater than those we estimate for all US women, showing that the motherhood penalty is substantial even in nations with generous leave policy.

than presenting the information by mother's birth cohort, as we did for the labor force data, a more meaningful arrangement is by the birth year of the first child. All the women in our analysis sample had a birth in one of three periods—1980–1989, 1990–1999, and 2000–2007—and all reported employment in the SIPP at some point in their pregnancy.[10] Because we begin with 1980s births, the mothers are part of cohorts born since 1955 and are, therefore, part of the group of women exhibiting the new life cycle of work.

The fraction of women who reported that they quit their jobs around the time of the birth decreased from 28 percent for those having their first child in the 1980s to 19 percent in the early 2000s.[11] The fraction taking paid leave increased from 34 percent in the 1980s to 42 percent in the 2000s, and the fraction on unpaid leave stayed fairly constant at around 30 percent (29 percent in the 1980s and 31 percent in the 2000s), where paid and unpaid leave include sick, vacation, disability, maternity, and other. About 8 percent in each of the three periods made no declaration of any type of leave or quit. Leaves increased and quits decreased by 10 percentage points for the entire group.

We estimate the labor force participation of first-time mothers depending upon their leave or quit status during the pregnancy and at birth, and track their post-birth participation for ten years and their pre-birth participation for three years. Recall that the sample is defined in terms of mothers who reported in the Survey of Income and Program Participation that they worked at some point during pregnancy, but that we estimate labor force participation in each calendar year as having administrative earnings above a minimum threshold. Because the participation rates are derived from annual earnings data, we cannot precisely identify the moment of the pregnancy and only know the year of the birth. If the birth was early in the year, the pregnancy would have mainly been in the previous year and the woman could be considered out of the labor force the year of the birth. But if the birth was late in the year, the woman could be deemed in the labor force that year. Since many mothers have low annual incomes due to unpaid leaves or quits and because the leaves can be staggered for the group, the aggregate rates shown in Figure 5 for the year before the birth are less than 100 percent. For the group that eventually quit, they are considerably lower.

Interestingly, during the 33-year period observed (1980 to 2012) and conditional on leave type, new mothers did not change their behavior much in terms of employment after their first birth. Because the results for all years given leave type are similar we show the data in Figure 5 only for women who were new mothers in the 1990s.

---

[10] We begin the analysis with the 1980s to get around the problem that exempt workers pose for the SIPP data, mentioned earlier in the text.

[11] About one-quarter of the respondents gave multiple leave types (for example, paid and unpaid leave), and we proportionately allocate leaves for each birth. Changes regarding leave type in the SIPP are greater if we extend the analysis to births in the 1960s and 1970s. But the earlier data are not fully consistent with later data and the exemption issue, discussed before, would affect many government workers in the earlier cohorts. HRS data cannot be used instead.

Figure 5

Labor Force Participation Pre- and Post-Birth Years of First Child by SIPP Leave Status, 1990s
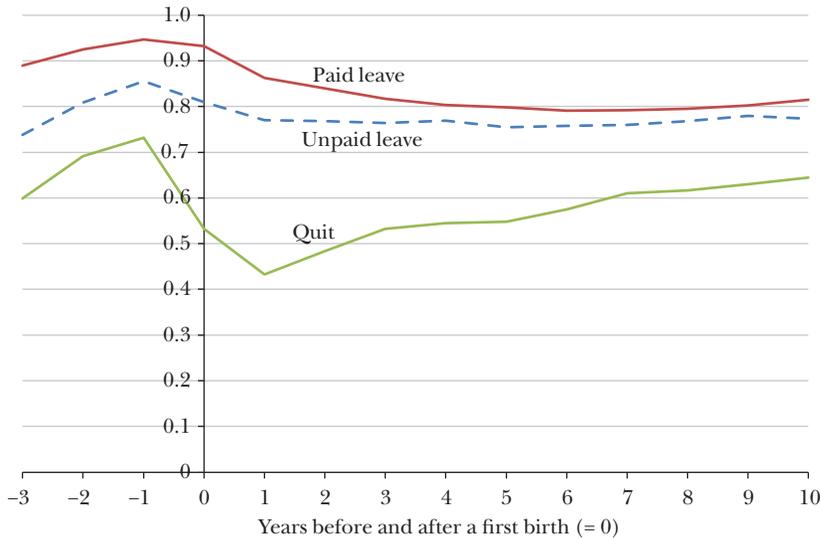
Sources: Survey of Income and Program Participation–Social Security Administration (SIPP-SSA) and SIPP Fertility History Topical Modules.
Notes: Includes native-born only women who had a first child from 1990 to 1999. Because the earnings data, from which the labor force data are derived, are for the calendar year, the precise moment of the birth cannot be linked to the earnings information. The SIPP Fertility History Topical Module lists whether the woman said she had unpaid leave, whether she had paid leave, and if she quit just before or after the birth. Some women listed more than one type and are included in more than one group.

Those on paid leave have the highest employment rates before, during, and after pregnancy, followed by those on unpaid leave. The lowest rates are for those who quit during the pregnancy, although ten years after the first birth their participation rate is 64 percent, approximately the level just before the first birth. Those on paid leave have a participation rate of 82 percent after ten years, considerably higher. Without an analysis of what determines who falls in each of these different categories, it is impossible to infer the impact that paid-leave, or longer protected-leave, policies would have on women's employment. But taking leave and staving off quits would appear to increase participation after a birth.[12]

---

[12] Olivetti and Petrongolo, in their paper in this issue, have an extensive discussion of the family leave literature. Their own empirical work shows that guaranteed and paid leave will increase women's employment to a point, but can reduce it for extensive leave policies. Only a few US states have paid leave, and protected leave is generally limited to that covered by the Family and Medical Leave Act of 1993. Rossin-Slater, Ruhm, and Waldfogel (2013) find that California's paid leave policy (which took effect in 2004) expanded leave use and had no negative employment effects and possibly positive ones.

Some International Comparisons

Current labor market participation rates of US women are low compared with those of other OECD nations. Moreover, the US rank in terms of these rates has deteriorated in the last 25 years. For women in the 25–54 year-old group among 21 OECD countries, the United States ranked sixth highest in 1990, ninth in 2000, and seventeenth in 2014. The US does much better when full-time rates are considered, because part-time work is more common in other OECD countries, particularly for women. Using the OECD common definition of full-time employment (30 hours per week in the usual job), the US was fourth (out of 18) in 1990, fifth (out of 20) in 2000, and eighth in 2014.[13] The topic of hours of work—the intensive margin—is a large and separate issue that we do not address here.

Paid and protected leaves—by definition—imply higher labor force participation rates because, in most data, individuals on leave are counted in the labor force. That factor can account for 4 to 4.5 percentage points (for the 25–54 year-old group) of the higher labor force participation among women 25 to 54 years old in Austria, Denmark, Norway, and Sweden, which are nations with very generous leave policies relative to the United States, the least generous. The actual differences in labor force participation between women 25 to 54 years old in the United States and in these countries are 10 to 14 percentage points in 2014. That is, measurement can explain around 30 to 40 percent of the difference. Among women in their 30s, the entire difference in participation rates between the United States and Denmark, Norway, and Sweden can be attributed to how women on leave are counted. But leave policy could have more than a definitional impact.

The roles played by public and private leave policies in accounting for these cross-national differences are complicated. Blau and Kahn (2013) find a positive relationship between leave policies and labor force participation, but also emphasize that the low-hours jobs women often have in these nations generally preclude careers. The fundamental question is whether new mothers who want to spend more weeks at home than allowed end up quitting their job and then have difficulty finding another position. When mandated parental leaves are greatly extended another issue is whether firms reduce demand for women in the age bracket who might be more likely to use long protected leaves (as discussed in the paper by Olivetti and Petrongolo in this issue).

Comparisons between the life-cycle employment patterns presented in this article with those for other OECD nations reveal that extensive protected and paid leave is probably not the main reason for the differences and thus not the primary reason for the slow recovery time in the birth event analysis. Life-cycle participation graphs for the United Kingdom look like those for the United States, despite the former country's longer protected and paid leave. But life-cycle employment

---

[13] The 21 countries are: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Japan, Luxemburg, the Netherlands, Norway, New Zealand, Portugal, Spain, Sweden, and the United States. The full-time measure is not always available.

patterns for France and Denmark, for example, have continued to increase with age. The lower cost and higher quality of childcare is probably the main factor (a conclusion arrived at by Olivetti and Petrongolo in their paper in this issue).

Conclusions

A new life cycle of women's labor force participation emerged with cohorts born in the 1950s. It is flatter and higher with no hump but with a dip in the middle and a phasing out that is later than for previous cohorts. High levels of female employment early in life are predictive of working longer at older ages, although no cohorts with the new life-cycle characteristics are yet old enough to observe in their 60s and 70s.

What brought about the new life cycle of women's work? The most important part of the story is the increase in participation by a succession of cohorts. That tale begins in the early twentieth century and is mainly about the impact of the growth in real wages combined with an increased importance of the substitution effect and a declining importance of the income effect (Blau and Kahn 2007). Later, young women in the late 1960s and 1970s began to have more realistic expectations of their future employment and started to make educational investments that could lead to longer and fulfilling careers (Goldin 2006). The "quiet revolution" that resulted further expanded women's employment. Together with the improved ability of young women to control the timing of childbirth (with the contraceptive "pill"), the marriage age rose and births were delayed. Motherhood came later in life and its impact on employment and careers was lessened.

More recent trends do not seem to have led to a great backpedaling in female employment. The scare about women's "opting out" of the workforce in the early 2000s was a misinterpretation of the changes in the life cycle of work. The sagging middle has been the result of a greater employment of women in their 20s together with a delay of childbirth and a negative (but smaller) employment impact of motherhood. The negative impact of motherhood is often attributed to the short-term nature of (federally) protected parental leave in the United States. Yet this explanation can only be a partial one. Many college graduates who have paid leave or more weeks of protected leave than the 12 weeks guaranteed to many workers by the 1993 Family and Medical Leave Act (FMLA) have employment that also discloses a sagging middle.

Yet relative to several other rich nations, women in the United States have been working a lower fraction of their lives when 25–54 years old. These other nations have had continued cohort effects (meaning that each cohort has higher participation rates at each age than that of the preceding cohort). It is unclear whether the difference between the United States and these other national is due to leave policy, childcare provision, or hours differences. On the plus side, however, is that US women remain in the labor force longer, what we term the expanding top.

An implication of the new life cycle of labor force participation for women is that cohorts entering their older years have more work experience, often have satisfying careers rather than just jobs, have invested more in their vocations, have more of their identity bound up in their work, and have more steeply sloped earnings trajectories. It is no wonder that employment has greatly increased at older ages, and these underlying dynamics give reason to believe that it will continue to do so.

References

Angelov, Nikolay, Per Johansson, and Erica Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics* 34(3): 545–79.

Attanasio, Orazio, Hamish Low, and Virginia Sánchez-Marcos. 2008. "Explaining Changes in Female Labor Supply in a Life-Cycle Model." *American Economic Review* 98(4): 1517–52.

Bailey, Martha J. 2010. "'Momma's Got the Pill': How Anthony Comstock and *Griswold v. Connecticut* Shaped US Childbearing." *American Economic Review* 100(1): 98–129.

Blau, Francine D., and Lawrence M. Kahn. 2007. "Changes in the Labor Supply Behavior of Married Women: 1980–2000." *Journal of Labor Economics* 25(3): 393–438.

Blau, Francine D., and Lawrence M. Kahn. 2013. "Female Labor Supply: Why Is the United States Falling Behind?" *American Economic Review* 103(3): 251–56.

Goldin, Claudia. 1989. "Life-Cycle Labor-Force Participation of Married Women: Historical Evidence and Implications." *Journal of Labor Economics* 7(1): 20–47.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." Richard T. Ely Lecture. *American Economic Review* 96(2): 1–21.

Goldin, Claudia. 2016. "Career and Family: Collision or Confluence?" Paper originally presented to the Eighth Annual Kenneth Arrow Lecture, Columbia University, December 10, 2015. (NBER Paper, forthcoming.)

Goldin, Claudia, and Lawrence F. Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110(4): 730–70.

Goldin, Claudia, and Lawrence F. Katz. Forthcoming A. *"Women Working Longer: Facts and Some Explanations."* In *Women Working Longer,* edited by C. Goldin and L .F. Katz. Chicago: University of

Chicago Press.

Goldin, Claudia, and Lawrence F. Katz, eds. Forthcoming B. *Women Working Longer.* Chicago: University of Chicago Press.

Heckman, James J., and Robert J. Willis. 1977. "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women." *Journal of Political Economy* 85(1): 27–58.

Kleven, Henrik, Camille Landais, and Jakob Søgaard. 2016. "Children and Gender Inequality: Evidence from Denmark." Unpublished paper, October.

Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel. 2013. "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes." *Journal of Policy Analysis and Management* 32(2): 224–45.

Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts

Chinhui Juhn and Kristin McCue

The ratio of female-to-male median annual earnings for full-time/full-year workers stood at 0.79 in 2014 (DeNavas-Walt and Proctor 2015). While this ratio is up from the 1960 level of 0.61, a substantial gap remains. Moreover, convergence has slowed since 2000. A recent survey article by Blau and Kahn (2016) finds that while women's gains in market skills measured by education and work experience were important in explaining convergence over the period 1980–2000, these human capital variables account for only a negligible portion of the gap that remains. In fact, women now surpass men on most measures of educational attainment, but this advantage has not filtered through to parity in wages (Goldin, Katz, and Kuziemko 2006; Murphy and Topel 2014).

The career dynamics of the gender gap for graduates of the Chicago Business School, as studied by Bertrand, Goldin, and Katz (2010), illustrate a common pattern. While women and men start their careers with similar earnings, a substantial gap arises over time, and the arrival of children is a major concurrent factor in the rising earnings gap. At least in this highly (and homogeneously) educated population, only a small share of the gender gap is due to premarket factors such as training and coursework; instead, family formation sets the gap in motion. Wood, Corcoran, and Courant (1993) and Goldin (2014) document similar patterns among graduates of the University of Michigan and Harvard law schools. Wilde, Batchelder, and

■ *Chinhui Juhn is Henry Graham Professor of Economics, University of Houston, Houston, Texas, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Kristin McCue is Principal Economist, U.S. Census Bureau, Washington, DC. Their email addresses are cjuhn@uh.edu and Kristin.McCue@census.gov.*

Ellwood (2010) offer evidence that the wage gap associated with children is larger for skilled women. An interesting and important question is whether this pattern holds true for the broader population of women as well.

In this paper, we examine the evolution of the gender gap associated with marriage and parental status, comparing cohorts born between 1936 and 1985. Women's early-in-life decisions about schooling, marriage, fertility, and work have repercussions throughout their lifetimes. This means we expect gains across cohorts to be an important part of the convergence. Those born early in this period typically began their careers with more-limited opportunities and different expectations than those born towards the end. These decisions are also clearly the outcomes of a complex and dynamic decision-making process. We do not attempt to build a structural model here, but rather document how changing associations between marriage and earnings, and between children and earnings, have contributed to the gender gap in an "accounting" sense.[1]

The model of household specialization and division of labor introduced by Becker (1981) has been the workhorse in this literature. Drawing insight from models of comparative advantage and trade, the model posits that when forming households, couples will exploit the gains from trade by having one spouse specialize in market work while the other specializes in household work. Given the historical advantage of men in the labor market, the model predicts specialization by gender and therefore an earnings advantage for married men and an earnings disadvantage for married women.[2]

Is this model of specialization useful for understanding the evolution of the gender gap across generations of women? There have been dramatic changes in both market and household production over the decades. Women's labor market opportunities have increased for a variety of reasons, making it more costly for them to stay home. Industrial change marked by a shift away from manufacturing jobs and technological change driven by advances in computing have favored cognitive over physical skills (Welch 2000; Beaudry and Lewis 2014; Weinberg 2000). The introduction of labor-saving devices such as washing machines, microwave ovens, and vacuum cleaners have freed up time that had previously been devoted to household tasks (Greenwood and Guner 2008). Improved birth control made it easier for women to control and time their fertility, making career-oriented investments less risky (Goldin and Katz 2002; Bailey, Hershbein, and Miller 2012). As a result of these and other changes, women and men have become more similar in terms of both market and household productivity. There can be gains to specialization within

---

[1] Adda, Dustmann, and Stevens (2015) build a structural model, which they estimate using German data. They find a "career cost" of children—the net present value of lifetime earnings of mothers relative to a no fertility scenario—of 35 percent, three-quarters of which result from labor supply reductions.

[2] Marriage and children can also lead to wage changes reinforced by employer discrimination, an explanation we do not address here. Correll, Bernard, and Paik (2007) present convincing evidence that this is an important contributor to the gender gap related to marriage and parenthood. In laboratory studies, they find that subjects penalize mothers and favor fathers in terms of perceived competence and make salary recommendations accordingly. A follow up audit-study involving resumes led to similar results.

the household even when women and men have identical capabilities, but there is less reason to expect specialization to remain starkly gendered—men at work, women at home. In this journal, Stevenson and Wolfers (2007) hypothesize that the returns to marriage based on production complementarities have diminished over time. In addition, greater longevity and leisure may have increased the importance of spousal consumption and leisure complementarities (Aguiar and Hurst 2007). These ongoing changes suggest that couples may increasingly match on similarities, rather than potential gains from trade. Couples are now increasingly likely to marry within education group (Schwartz and Mare 2005), and the incidence of dual-earner couples who are both highly educated, the so-called "power couples," has increased (Costa and Kahn 2000). These trends would lead to narrowing of the difference in wives' and husbands' earnings trajectories after marriage within couples, although the same trends may contribute to rising inequality across households overall.

But what about children? Women's gains in the labor market likewise predict declining gender gaps associated with children as well, but recent evidence seems to suggest otherwise. The profile of Chicago MBAs by Bertrand, Goldin, and Katz (2010) mentioned earlier suggests that even for women who have made costly investments in skills that are highly valued in market work, the extent to which women are willing or able to substitute for their time is limited when it comes to care of children. These issues have received much recent attention in popular media such as Lisa Belkin's (2003) article "The Opt-Out Revolution" and Anne-Marie Slaughter's (2012) "Why Women Still Can't Have it All."

Academic papers focusing on the "motherhood penalty" or "family gap" have shown that wages of mothers are significantly lower than those of non-mothers with similar human capital characteristics. For example, Waldfogel (1997) finds a wage gap of 6 log points for mothers with one child and 13 log points for mothers with two children. Budig and England (2001) also report a wage gap of approximately 5 log points per child for mothers versus non-mothers, even taking into account mother fixed effects.[3] Wilde, Batchelder, and Ellwood (2010) find larger wage gaps of 17 log points at ten or more years after birth, suggesting that the wage gap accumulates, particularly among more-skilled women. Each of these studies focus on hourly wages rather than annual earnings. The gaps in annual earnings are even larger as mothers are significantly more likely to work part-time and part-year.

Is there a corresponding fatherhood premium when comparing men who have children with men who do not? Researchers generally find a small positive fatherhood premium, although the results seem sensitive to specification. Lundberg and Rose (2000) find a premium of 9 percent on average, although with no effect among couples in which both parents continue working. Budig (2014) finds a fatherhood premium of about 6 percent, while Killewald (2013) finds a premium of about 4 percent—but only for married, co-resident, and biological fathers. In

---

[3] These estimates are somewhat larger than those found in earlier papers like Korenman and Neumark (1992) and Neumark and Korenman (1994), who use first difference and sibling fixed-effects models, respectively.

contrast, Wilde, Batchelder, and Ellwood (2010), find little evidence of a positive fatherhood wage premium.

In contrast, there is a well-established marriage premium for men. Cross-sectional estimates suggest that married men earn somewhere between 10 to 40 percent higher wages compared to observationally equivalent single men (for example, Hill 1979; Korenman and Neumark 1991; Antonovics and Town 2004). Korenman and Neumark (1991) find evidence of positive selection of men into marriage based on earnings—that is, men with higher earnings are more likely to marry. But they also find that much of the male marriage premium accrues from faster wage growth after marriage, which is consistent with marriage allowing men to shift towards more market work and less home production. The effect of marriage on women's wages is less well established. Waldfogel (1998), Budig and England (2001), and Killewald and Gough (2013) find that for recent cohorts, married women also command a wage premium, results consistent with what we find here.[4]

Previous studies have not, for the most part, focused on how these gaps have evolved over time.[5] We bring this longer-run perspective here. We examine whether the gaps associated with marriage and children have narrowed as the specialization model predicts. We separate the patterns associated with marriage and children because the two have become less intertwined over time with increases in single parenthood and prolonged periods of childlessness among couples due to delayed childbearing.

The Evolution of the Gender Gap

Figure 1 shows several measures of the growth of the female–male earnings ratio since the early 1960s. The figure illustrates median full-time, full-year relative earnings published in the Current Population Reports (DeNavas-Walt and Proctor 2015, Table A-4) along with two other series we constructed from 1962–2015 March Current Population Surveys (CPS) public-use files.[6] Our sample consists of prime-aged women and men, defined as those with 1 to 35 years of potential experience. Prime age differs for those with different levels of education to reflect the fact that their prime years start after their education is finished. Our earnings measure

---

[4] In Juhn and McCue (2016), we show that the positive marriage premium is entirely due to the most recent cohorts of college-educated women without children. This may indeed reflect the fact that career-minded men and women are increasingly likely to marry one another and that specialization plays a minor role when children are not involved.

[5] Exceptions include Blackburn and Korenman (1994) and Gray (1997), who find that the male marriage premium narrowed in the 1970s and 1980s. Pal and Waldfogel (2016) find a decline in the motherhood wage gap in the more recent period, 1993–2013. On the other hand, using data from the National Longitudinal Survey of Young Women and the National Longitudinal Survey of Youth, Avellar and Smock (2003) find no changes in the motherhood penalty.

[6] We use the 1962–2015 March CPS surveys downloaded from the IPUMS-CPS. The annual earnings information refers to the previous year, so the data series cover earnings for 1961–2014. Later, for our analysis using children, we use 1968–2015 surveys due to the fact that consistent measures of number and ages of children are not available in the earlier years.

Figure 1
Growth of the Female/Male Earnings Ratio
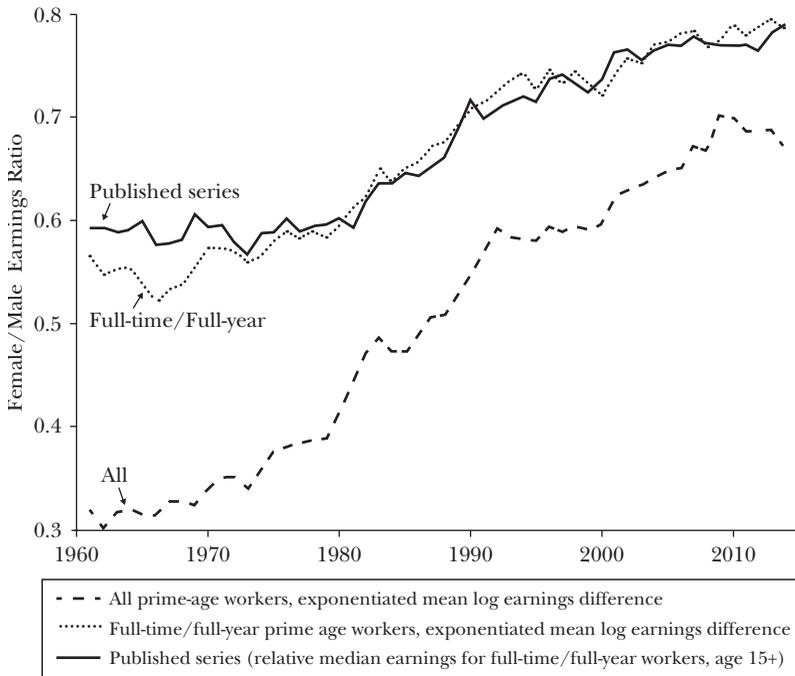
Notes and Source: The figure illustrates median full-time, full-year relative earnings published in the Current Population Reports (DeNavas-Walt and Proctor 2015, Table A-4) along with two other series we constructed from 1962–2015 March Current Population Surveys (CPS) public-use files. The published series use a direct ratio, while ours uses an approximated ratio: we use the exponent of (mean log female annual earnings – mean log male annual earnings) to approximate female/male earnings ratios. In our analysis, "Prime age" is defined as those with 1 to 35 years of potential experience. More specifically, for those who finished high school or less, the age range is 18–53 for high school dropouts and high school graduates. For those with some college, the age range is 21–56. For college graduates, the age range is 23–58. For those with post-college education, the age range is 25–60. In our analysis we define full-time/ full-year as working 35 hours or more and working 40 weeks or more. We adjust top-coded income categories by multiplying the top-code value by 1.5.

includes wage and salary, self-employment, and farm income. We use the exponent of (mean log female annual earnings – mean log male annual earnings) to approximate female/male earnings ratios. We examine these approximated female/male earnings ratios for all workers with nonzero earnings, and for a sample of full-time/ full-year workers. In our analysis we define full-time/full-year as working 35 hours or more and working 40 weeks or more.[7]

---

[7] The Current Population Report defines full-time, year-round workers in a slightly different way: "A full-time, year-round worker is a person who worked 35 or more hours per week (full time) and 50 or more weeks during the previous calendar year (year round). For school personnel, summer vacation is counted as weeks worked if they are scheduled to return to their job in the fall.

Our ratio based on average earnings of full-time/full-year workers tracks the published series fairly well.[8] The female/male earnings ratio hovers around 0.6 until 1980 and increases rapidly during the 1980s. The pattern of male/female earnings convergence is even more dramatic when we expand our sample to include part-time and part-year workers, reflecting women's gains in annual hours and weeks worked. The rapid convergence during the 1980s reflects women's gains in education and experience along with shifts in the labor market that worked to women's relative advantage (Blau and Kahn 1997). For college-educated women, occupational upgrading certainly played a role as women moved into professional occupations, especially high-wage occupations traditionally dominated by men (Black and Juhn 2000). Among less-educated workers, industrial shifts such as the decline in manufacturing jobs had a larger negative impact on earnings of men than those of women. Welch (2000) notes the striking similarity of movements in the education premium and female–male relative earnings over this period, arguing that the two phenomena both reflect rising returns to cognitive skills over physical skills.

Since the 1980s, convergence in earnings between males and females has been slower and uneven. The slowdown has been particularly pronounced at the top. Blau and Kahn (2016) report that at the 90th percentile, women essentially made no gains relative to men from 1998–2010. Similarly, among full-time/full-year workers holding a bachelor's degree and higher, women's weekly earnings as a percentage of men's earnings remained at 74.1 from 2000 to 2010 (US Bureau of Labor Statistics 2015). The notion of "swimming upstream" is likely relevant for college-educated women as overall wage inequality continued to increase at the very top. To the extent that college educated women are still positioned below college-educated men in the overall wage distribution, growth in inequality at the top puts them at a disadvantage. While women have made much progress, they are still underrepresented in the highest earnings occupations such as finance, and the science, technology, engineering, and mathematics fields. Even when women work in highly paid occupations, their earnings lag behind men (as previous studies of women with MBAs and law degrees have shown). One possible explanation is that as women enter more-demanding occupations, they have hit a "motherhood wall," and they struggle to balance work and family.[9]

Figure 2 illustrates the varying experiences of five-year birth cohorts spanning the birth years from 1936–1985. The left axis reports the log earnings difference: that is, mean log female annual earnings minus mean log male annual earnings. The right axis shows the exponent of log earnings difference, which approximates female/male earnings ratios. The top panel covers all workers, while the bottom panel covers only full-time/full-year workers. The figures highlight the large gains

---

[8] Our measure tracks the published series even better if we delete low values that fall below one-half of the 2012 minimum wage on a full-time, full-year basis. Since measurement error in earnings is less of a concern than in constructed hourly wages, however, we forego making this adjustment in our analysis.
[9] The phrase "motherhood wall" is borrowed from a 2010 book by Joan C. Williams, *Reshaping the Work-Family Debate: Why Men and Class Matter.*

Figure 2

Age Profiles of Female/Male Earnings Ratios across Five-Year Birth Cohorts, Born 1936–1985

A: All Workers

B: Full-Time/Full-Year Workers
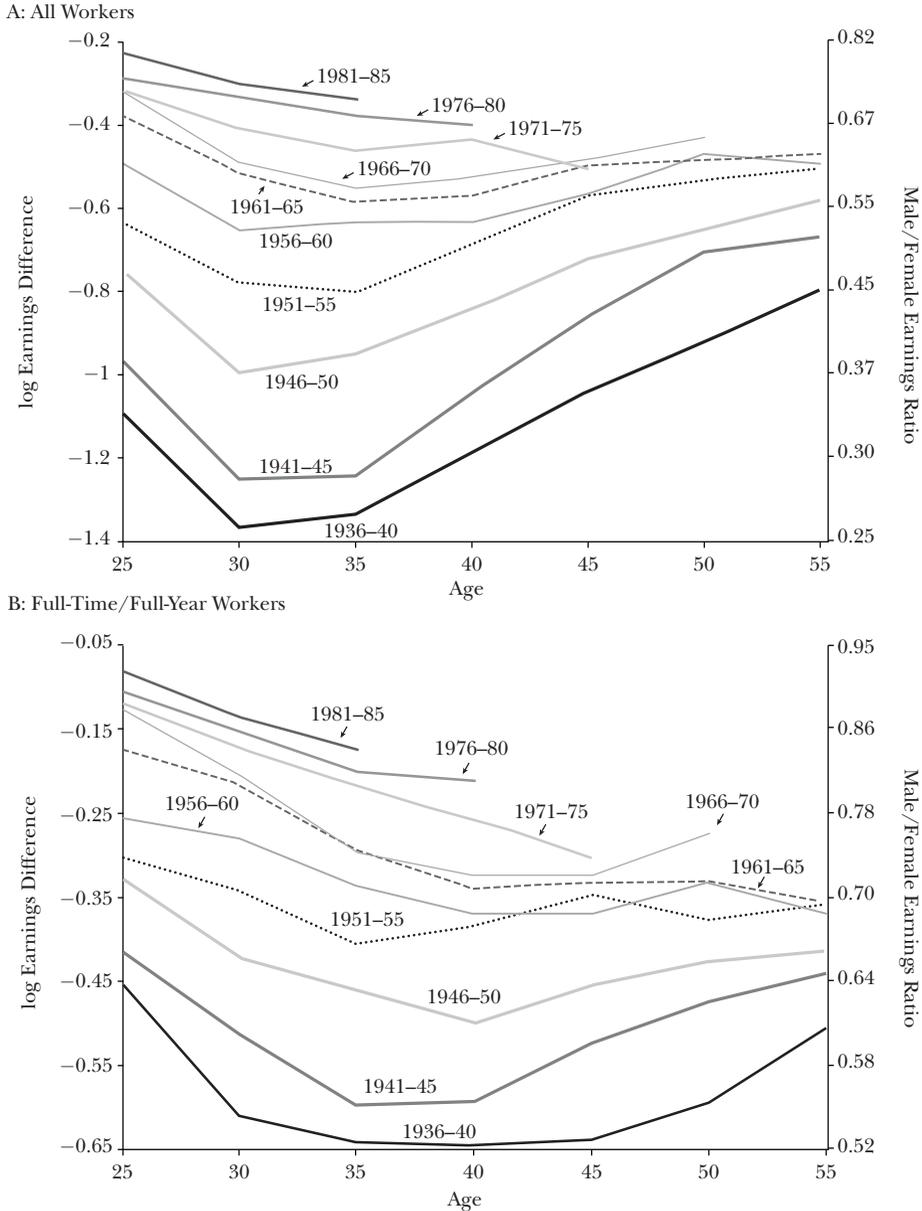
Source: 1962–2015 March Current Population Surveys.

Notes: The left axis reports the log earning difference, that is, mean log female annual earnings minus mean log male annual earnings. The right axis shows the exponent of log earnings difference, which approximates the female/male earnings ratio. The sample includes prime age workers (1–35 years of potential experience; see notes to Figure 1 for details). The sample used in the top panel includes all workers with nonzero earnings while the bottom panel sample includes only full-time/full-year workers. The figure reports averages over five-year age groups.

each new birth cohort experienced relative to the previous cohort up to the last baby-boomer cohort (those born in 1961–65). Goldin (2006) describes the "Quiet Revolution," which began with baby-boomer cohorts who entered the labor market in the early 1970s. For more-educated women in particular, the notion of a career took shape as barriers to entry fell and women entered elite schools and professional occupations. These cohorts of women were less likely to drop out of the labor force upon marriage and have children than previous generations, and when they did drop out, were more likely to return (as Goldin and Mitchell discuss in this issue). Women's increased labor force attachment is also apparent when we compare gains in annual earnings for all workers, which reflect increases in weeks and hours worked (Figure 2A) to the gains among full-time/full-year workers (Figure 2B). Note that the range of the y-axis scale is much wider for Figure 2A, covering all workers, reflecting the fact that the pace of convergence is particularly strong for this broader sample.

Why has progress slowed across more recent cohorts of women? Among full-time/full-year workers born in 1971 and later, women started at about 0.1 log points behind or at about 90 percent of male earnings at age 25. Remaining occupational differences may explain some of this gap. While women have increasingly entered traditionally male-dominated occupations, they are still underrepresented in some highly paid fields including science, technology, engineering, and mathematics (Ceci, Ginther, Kahn, and Williams 2014).

The relatively small gap upon labor market entry, however, paints an overly optimistic picture of what is to follow. Earnings of women drop relative to those of men as they progress through their career, reaching a low point somewhere around age 30 to 40. The trough occurs at older ages with each successive cohort, but it does not disappear. Among full-time/full-year workers, it is not entirely clear that earnings will recover at later ages among the most recent cohorts. In the following section, we investigate to what extent this drop in relative earnings correlates with marriage and children.

Marriage, Children, and Earnings

To investigate the association between earnings and both marital status and presence of children, we start with the following statistical model:

$$\ln Y_{it} = \beta X_{it} + \gamma^c M_{it} + \pi^c K_{it} + \varepsilon_{it}^c$$

where i indexes an individual, C indexes birth cohort, and t indexes time. The M variable is an indicator for "married." We distinguish between currently married and nonmarried individuals, implicitly assuming "divorced/widowed" and "never married" are equivalent states in terms of their effects on earnings, but our investigation of more-elaborate specifications of the relationship between marriage and earnings yielded similar qualitative results. K actually refers to two parenting-related variables: the number of children less than 18 years old and the presence of a child

younger than six. We top code the count of children at three, because in specifications in which we use a series of dummy variables to control for children, additions beyond age 3 had little additional effect on our estimates. Given this specification, the effect of having a single child aged less than six is given by the sum of the two child coefficients, while the effect of a second child under six or of an older child is given by the coefficient on the count.

In this equation, *X* refers to other controls, which here include a quartic function in age and indicator variables for education, year, race, and ethnicity. As we will see, adding interaction terms between marital status and ten-year birth cohort dummies allows us to examine how the marriage gap varies across cohorts. We include analogous interactions between children variables and cohort dummies. Prior studies on the family gap have typically included measures of actual labor market experience, thereby parceling out the effects of children that occur through reduced experience. Unfortunately, the Current Population Survey data that we are using does not collect measures of actual experience, so we are unable to do this additional decomposition exercise. The marital status and children coefficients we estimate here should thus be thought of as the gross impact, including the effect of children on women's labor market experience.

We estimate separate regressions for women and men, and for two samples—all workers with nonzero earnings and full-time/full-year workers. The detailed estimates for all coefficients are reported in the online appendix available with this paper at http://e-jep.org. We focus in Table 1 on presenting the cohort-specific associations between earnings, marital status, and the two children variables.

Panel A of Table 1 presents the earnings gap associated with marriage for women and men in our samples. The two left-hand columns of Table 1 are based on all workers with nonzero earnings while the two right-hand columns are based on full-time/full-year workers. Married male workers earn a premium of about 30 log points relative to nonmarried male workers, with only modest change in this cross-sectional relationship across our cohorts. In contrast, married women born in 1936–1945 earned 32 log points less than single women, but that marriage gap declines dramatically and then turns positive across cohorts. Among women born in years 1965–1975, married women earn a premium of 7 log points. The age range for the 1976–1985 cohort in 2014 is 29–39, so results for this group put less weight on earnings differences accompanying marriages and births at older ages than do results for other cohorts, but the available data indicates that the trend continues with that cohort. Among full-time/full-year workers, the size of the initial marriage gender gap as well as the changes across cohorts are more muted, but the basic patterns of a small decline among men and a reversal in sign among women are preserved.

Panel B of Table 1 shows the results from the same regressions but now reporting the effect of having younger children at home, while regressions in panel C report the effects of having school-aged children. Among all women workers, the earnings gap associated with young children declined dramatically from as much as 50 log points for the 1936–45 birth cohort to 23 log points for the 1966–75 birth

Table 1

Earnings Differences Associated with Marriage and Children

Birth cohort	All workers		Full-Time/Full-Year workers	
	Women	*Men*	*Women*	*Men*
A: Marriage				
1936–45	−0.316	0.331	−0.094	0.201
	(0.006)	(0.006)	(0.005)	(0.005)
1946–55	−0.130	0.368	−0.033	0.198
	(0.004)	(0.004)	(0.003)	(0.003)
1956–65	−0.023	0.340	0.008	0.179
	(0.004)	(0.004)	(0.003)	(0.003)
1966–75	0.069	0.292	0.055	0.150
	(0.005)	(0.005)	(0.004)	(0.004)
1976+	0.099	0.277	0.061	0.139
	(0.006)	(0.006)	(0.005)	(0.005)
B: Child <6				
1936–45	−0.495	0.022	−0.062	0.016
	(0.009)	(0.005)	(0.009)	(0.004)
1946–55	−0.419	0.071	−0.024	0.028
	(0.006)	(0.004)	(0.005)	(0.003)
1956–65	−0.297	0.080	−0.003	0.055
	(0.005)	(0.004)	(0.004)	(0.003)
1966–75	−0.230	0.081	−0.006	0.061
	(0.005)	(0.005)	(0.004)	(0.004)
1976+	−0.161	0.103	−0.013	0.062
	(0.006)	(0.007)	(0.005)	(0.005)
C: Child 6–17				
1936–45	−0.195	0.049	−0.052	0.032
	(0.003)	(0.002)	(0.003)	(0.002)
1946–55	−0.193	0.046	−0.043	0.037
	(0.002)	(0.002)	(0.002)	(0.002)
1956–65	−0.174	0.044	−0.043	0.045
	(0.002)	(0.002)	(0.002)	(0.002)
1966–75	−0.140	0.048	−0.043	0.043
	(0.003)	(0.002)	(0.002)	(0.002)
1976+	−0.132	0.002	−0.041	0.009
	(0.004)	(0.004)	(0.003)	(0.003)

Source: Authors' estimates based on CPS-ASEC 1968–2014 IPUMS-CPS files.

Notes: The estimates are log earnings differences. Estimates are derived from coefficients from specifications (3) and (7) in tables A1 and A2 available in the online appendix for this paper at http://e-jep.org. The samples used include working men and women with 1–35 years of potential experience. Regressions include controls for age, year, birth cohort, and education, as well as marital status and presence of children. Interactions between cohort dummies and marital/child variables are used to estimate changes across cohorts. Standard errors are reported in parentheses. See text for more detail.

cohort, and to 16 log points for the 1976–85 cohort. There have been real reductions in the motherhood gap associated with young children, particularly as women are less likely to withdraw from the labor force and cut back in terms of hours and weeks worked. The effect of having school-aged children also declines, but by a smaller amount, from about 20 log points to about 13–14 log points. Again, patterns among full-time/full-year workers are qualitatively similar although much muted in size. As we will show later, these cross-sectional patterns are likely to overstate the reductions in motherhood gaps as mothers became more positively selected in terms of earnings potential than previously.

How important is the family gap—by which we mean the gender differences in coefficients associated with marital and parental status—in explaining the gender gap? Table 2 shows an Oaxaca-style decomposition of the gender gap. The first column of the table shows the raw gender gap by cohort. The following columns show the cohort-specific predicted gap based on the differences between men and women in the associations illustrated in Table 1. Because mean age in our sample varies across cohorts, we make the comparisons using means at specific ages 30 and 40.

Thus, the first row of the table shows that at age 30 for the full sample, the gender gap in earnings is 1.34 for the 1936–45 birth cohort, (which roughly corresponds to a female/male earnings ratio of 0.26). The gender gap predicted by the difference in marriage coefficients is 0.49 while the gap predicted by differences in children coefficients is 0.52. Thus they account for about 36 and 39 percent of the overall gap, respectively. With each successive cohort, the contribution of marriage to the gender earnings gap has declined. The contributions of children to wage inequality between males and females have also declined, but not as rapidly as the overall gender gap. As a result, the contribution of the child-related variables as a share of the overall gender gap has increased from 39 percent for the oldest cohort to 56 percent for the women born in 1966–75.

The contribution of children to the gender gap is falling somewhat more rapidly at age 30 than at age 40, as shown by the second set of rows in Table 2. This pattern seems due to two effects: first, the gap associated with the presence of a preschool-aged child has fallen dramatically, and women in all cohorts are more likely to have a young child at age 30 than at age 40; second, delayed childbearing has reduced the share of women with a young child at age 30 while having less consistent effects on the child variables at age 40. The overall gender gap is falling more rapidly at age 30 than at age 40, however, so the contribution of children as a share of the total rises similarly at both ages across cohorts.

Panel B of Table 2 shows the parallel results for full-time/full-year workers. The contribution of children to the overall gender gap is much smaller in this sample, reflecting the fact that a large portion of the gender gap related to children stems from reduced hours and weeks of work. However, the stability of children's contributions across cohorts to the gender gap is notable. For women at age 40, the contribution of children is 0.10 for the 1936–45 cohort as well as for the 1965–75 cohort. Again, the predicted gap associated with children has been

Table 2
Contribution of Family Variables in Accounting for the Gender Earnings Gap

Birth cohort	Gender earnings gap	Marriage		Number of Children		Child < 6	
		Mean	Contribution to earnings gap	Mean	Contribution to earnings gap	Mean	Contribution to earnings gap
A. All Workers							
At age 30							
1936–45	1.340	0.752	0.486	1.613	0.394	0.446	0.122
1946–55	0.840	0.670	0.334	1.201	0.287	0.395	0.099
1956–65	0.575	0.631	0.229	1.100	0.240	0.407	0.064
1966–75	0.442	0.583	0.130	1.060	0.199	0.396	0.049
1976–85	0.300	0.526	0.094	0.990	0.133	0.387	0.050
At age 40							
1936–45	1.114	0.721	0.466	1.381	0.337	0.073	0.020
1946–55	0.742	0.683	0.341	1.167	0.279	0.117	0.029
1956–65	0.609	0.662	0.240	1.245	0.272	0.149	0.024
1966–75	0.484	0.643	0.144	1.299	0.243	0.199	0.024
B. Full-Time/Full-Year Workers							
At age 30							
1936–45	0.542	0.649	0.192	1.227	0.103	0.311	−0.002
1946–55	0.340	0.585	0.135	0.909	0.073	0.282	−0.006
1956–65	0.235	0.580	0.099	0.872	0.077	0.326	−0.008
1966–75	0.172	0.549	0.052	0.888	0.076	0.333	−0.004
1976–85	0.135	0.519	0.041	0.850	0.042	0.338	0.010
At age 40							
1936–45	0.632	0.654	0.193	1.187	0.100	0.059	−0.000
1946–55	0.445	0.626	0.144	1.025	0.082	0.097	−0.002
1956–65	0.358	0.618	0.106	1.116	0.099	0.128	−0.003
1966–75	0.307	0.612	0.058	1.189	0.102	0.181	−0.002

Source: Data is from the CPS-ASEC 1968–2014 IPUMS-CPS files.
Notes: Numbers result from a Oaxaca decomposition of the difference in mean log earnings between men and women at the specified age. Coefficients estimates used in the decomposition are based on samples of working men and women with 1–35 years of potential experience. Regressions include controls for age, year, birth cohort, and education, as well as marital status and presence of children. Interactions between cohort dummies and marital/child variables are used to estimate changes across cohorts. See text for more detail.

relatively stable, even as the overall gender gap has shrunk. Children, therefore, account for a growing share of the remaining gender gap.[10]

---

[10] Our findings echo the conclusions in Waldfogel (1998). She compares cohorts born 1944–1954 in data from the National Longitudinal Survey of Young Women and Young Men and cohorts born 1957–65 from the National Longitudinal Survey of Youth 1979 (NLSY79). At age 30, she finds that family variables—that is, marriage and children together—account for 35 percent of the gender gap in wages for the earlier cohort while accounting for 56 percent of the gender gap in wages for later cohort.

One way to summarize roughly how these patterns have changed across the careers of these cohorts is to take simple averages of our estimates, giving equal weight to estimates for each year of age. Using ages 25–49 for the first four cohorts (for which we have estimates to at least age 49), we find that the career ratio of women's earnings to men's earnings for our full sample rises from 33 percent to 60 percent across those cohorts. The share of the earnings gap accounted for by marriage falls from 41 percent to 28 percent, while the share associated with children rises from 26 percent to 43 percent.

What About Selection?

Up to this point, we have not addressed the possibility that the evolution of earnings gaps might stem from self-selection into marriage and parenting. For example, suppose men with the highest earnings potential were more likely to marry and have children, while women with the lowest earnings potential in the labor market were more likely to marry and have children. This combination of positive selection on earnings among fathers and negative selection on earnings among mothers would cause our estimates to overstate the impact of the family gap. Moreover, changing patterns of selection into marriage among women might plausibly account for at least some of the declining marriage gaps among women. The declining gender gap associated with marriage, for example, may reflect the fact that women with higher earnings are increasingly likely to marry. Isen and Stevenson (2010), Goldstein and Kenny (2001), and Juhn and McCue (2016) find that among recent birth cohorts, the most educated women are the ones most likely to marry—a reversal from earlier cohorts where the most educated women were least likely to marry.

Similar—and likely related—changes in selection have been found with respect to fertility. Baily, Guildi, and Hershbein (2013) find that number of children ever born has increased for the top quartile of the education distribution for women, while continuing to fall among women in the bottom quartile of the education distribution since at least the 1950 birth cohort. A related phenomenon is that childlessness—defined as not having had a live birth by age 41—has been falling among the most educated women since about the 1957 cohort. These patterns suggest that selection into motherhood may have also become more positively related to education and earnings over time.

In Juhn and McCue (2016), we investigate the possible size of such selection effects by using fixed-effects models that are closely related to the cross-sectional results presented earlier. With fixed-effects models, our estimates reflect differences in earnings *for the same individual* between periods when she is married (or has children) and periods when she is single (or has no children at home). This approach purges the estimates of persistent differences in mean earnings between married women and single women, or between mothers and non-mothers. Our specifications include controls for age, so where our fixed-effects estimates show

negative marriage gaps or motherhood gaps for women, this reflects slower wage growth for women relative to their peers during periods when they are married or have children. The difference between the cross-sectional estimates and these fixed effects estimates then provides us with the net effect of selection into marriage and motherhood.

It should be noted that fixed-effects models cannot entirely take account of selection if selection is based on differences in expected earnings trajectories rather than difference in levels of earnings. There is also a possibility of reverse causality here: that is, unanticipated shocks in earnings may sometimes cause marriage or parenthood, rather than having marriage or parenthood affect earnings. For example, men may marry after receiving a promotion or women may decide to have children if they are not promoted. To address this concern, another approach has been to conduct event-study analysis examining the trajectories of earnings before and after marriage or before and after the arrival of the first child.

Our evidence is based on data from Survey of Income and Program Participation (SIPP) panels matched to Social Security Administration earnings records from 1954–2011, and we use essentially the same selection rules that we apply here for the Current Population Survey. These data provide detailed earnings histories that allow us to estimate both cross-sectional effects and fixed-effect models, but we do not have information on hours or weeks worked, so we cannot separately examine full-time/full-year workers. Our last year of earnings data is 2011, so we did not include the 1976–85 cohort in this study.

We modify the earlier model by adding a fixed effect, as in:

$$\ln Y_{it} = \beta^c X_{it} + \gamma^c M_{it} + \pi^c K_{it} + \alpha_i^C + \nu_{it}^C$$

where α_i^C refers to a permanent (unobserved) skill component of earnings. We also incorporate interactions between our marriage variable M and the children variable K, to measure separately the association between children and the earnings of married and nonmarried women across cohorts. Again, detailed results for these regressions appear in the appendix with this paper available on-line at http://e-jep.org.

Table 3 presents the cohort-specific estimates of the marriage earnings gap that condition on parental status. A comparison of the cross-sectional and fixed-effects versions of the estimates provides evidence on the importance of selection for the observed patterns. The cross-sectional estimates, shown in the first three columns, are closest to the estimates in Table 1 though some of the details differ. We present estimates of the gap for three scenarios: women without minor children, women with one child who is less than six, and women with one child aged 6–17. For each group, the marriage earnings gap narrows over time: indeed, the youngest cohort, women born in 1966–75, has a positive gap. The fixed effect estimates, shown in last three columns, illustrate that marriage is still for the most part associated with a negative earnings gap for women. The cross-sectional

Table 3
Marriage Earnings Gap by Parental Status

Birth cohort	Cross-section			Fixed effects		
	No kids	*Child <6*	*Child 6–17*	*No kids*	*Child <6*	*Child 6–17*
1936–45	–0.237	–0.302	–0.261	–0.277	–0.390	–0.302
	(0.012)	(0.026)	(0.010)	(0.006)	(0.014)	(0.005)
1946–55	–0.122	–0.163	–0.149	–0.160	–0.296	–0.199
	(0.008)	(0.016)	(0.007)	(0.004)	(0.009)	(0.003)
1956–65	–0.034	0.023	–0.029	–0.098	–0.177	–0.136
	(0.010)	(0.017)	(0.008)	(0.005)	(0.009)	(0.004)
1966–75	0.077	0.115	0.060	0.053	–0.073	–0.011
	(0.016)	(0.025)	(0.013)	(0.010)	(0.015)	(0.008)

Source: Authors' estimates based on SIPP data linked to SSA earnings records. Estimates are derived from coefficients from specifications (4) and (9) in table A3 in the online Appendix for this paper at http://e-jep.org. The samples used include working women with 1–35 years of potential experience. Regressions include controls for age, year, birth cohort, and education, as well as marital status and presence of children. Interactions between cohort dummies and marital/child variables are used to estimate changes across cohorts. Standard errors are reported in parentheses. See text for more detail.

estimates are smaller than the fixed effect estimates, suggesting that selection into marriage typically reduces the earnings gap for women. We see a significant decline in the earnings gap across cohorts however, even in fixed effect estimates. For the youngest cohort, women without children now earn a premium upon marriage. Further investigation (described in Juhn and McCue 2016) shows that this result is driven entirely by college-educated women. There is clear evidence that among couples without minor children, marriage has become less associated with specialization.

The earnings gaps associated with motherhood, conditional on whether the mother is single or married, are presented in Table 4. The left-hand columns show the cross-sectional estimates, while the right-hand columns show the fixed effect estimates. A comparison of the cross-sectional estimates in the left-hand columns and the fixed estimates in the right-hand columns shows that the mother-hood gap is generally even larger when we take into account selection, particularly when it comes to the effect of having young children among married mothers. Among married women, mothers appear to be somewhat positively selected, while among single women, mothers appear to be somewhat negatively selected. The trends we identified earlier in Table 1 are largely preserved, however, even in the fixed effects estimates. The gap associated with young children has fallen across cohorts, but the gap associated with school-aged children has not. Here we see that it has even risen, so that the overall effect of children has remained more or less stable.

These findings suggest that the decline we observed in the motherhood effect on wages using Current Population Survey data may be somewhat overstated once selection is taken into account. With fixed effects, even among the 1966–75 birth

Table 4
Motherhood Earnings Gap by Marital Status

Birth cohort	Cross-Section		Fixed Effects	
	Single	Married	Single	Married
A: Child <6				
1936–45	−0.356	−0.421	−0.399	−0.512
	(0.023)	(0.013)	(0.013)	(0.006)
1946–55	−0.326	−0.367	−0.334	−0.470
	(0.014)	(0.008)	(0.008)	(0.004)
1956–65	−0.370	−0.313	−0.340	−0.419
	(0.015)	(0.009)	(0.008)	(0.005)
1966–75	−0.350	−0.312	−0.276	−0.402
	(0.020)	(0.014)	(0.012)	(0.008)
B: Child 6–17				
1936–45	−0.133	−0.157	−0.128	−0.153
	(0.006)	(0.004)	(0.003)	(0.002)
1946–55	−0.157	−0.184	−0.140	−0.179
	(0.005)	(0.004)	(0.002)	(0.002)
1956–65	−0.200	−0.195	−0.167	−0.205
	(0.005)	(0.005)	(0.003)	(0.002)
1966–75	−0.194	−0.211	−0.138	−0.202
	(0.008)	(0.008)	(0.005)	(0.005)

Source: Authors' estimates based on SIPP data linked to SSA earnings records. Estimates are derived from coefficients from specifications (4) and (9) in table A3 in the online Appendix for this paper at http://e-jep.org. The samples used include working men and women with 1–35 years of potential experience. Regressions include controls for age, year, birth cohort, and education, as well as marital status and presence of children. Interactions between cohort dummies and marital/child variables are used to estimate changes across cohorts. Standard errors are reported in parentheses. See text for more detail.

cohort, married women with a young child earn 40 log points less, and married women with school-aged children earn 20 log points less, compared to married women with no children.

To summarize, when no children are involved, there is little evidence that married women earn less than single women among the more recent birth cohorts. In fact, they appear to have a slight earnings advantage relative to single women, similar to what is observed for men. Married mothers have gained on single mothers. They are less likely to reduce hours and weeks as they have in the past. With regards to marriage, specialization has become less important. The presence of minor children, however, remains associated with substantially lower earnings for women, particularly those who are married. While there has been some improvement in terms of the drop associated with having very young children, the drop in earnings associated with motherhood has remained remarkably stable. In this area, it looks like specialization has remained intact.

Of course, an alternative possibility is that reduced specialization has instead taken the form of a diminished or even negative fatherhood premium so that fathers, particularly married fathers, now also experience a negative earnings gap associated with fatherhood. Unfortunately, the data from the Survey of Income and Program Participation has more-limited information on fertility of fathers, so we are not able to perform fixed effect estimates of the gap associated with children for men. But the cross-sectional estimates we present in Table 1 based on Current Population Survey data did not show much evidence of a shrinking fatherhood premium. Also, as cited earlier, recent studies, which typically control for fixed effects, find a small positive impact of fatherhood on men's earnings. Lundberg and Rose (2000) do find that among families where the mother continues working, fathers reduce labor supply. While such changes may be underway, as yet there is no consistent evidence of a changing pattern.

Previous studies have provided more evidence on the evolution of earnings gaps for highly educated women than for other groups, and we would expect career interruptions to be particularly costly for this group given large prior investments and depreciation of skills. Are motherhood gaps as large for less-educated women? In Juhn and McCue (2016), we look at marriage and motherhood gaps across three education groups: a high school degree or less, some college, and college graduate. All groups have declining marriage gaps across cohorts, but in fixed-effect estimates, marriage gaps are larger for less-educated women. With regards to the motherhood gap, the earnings gap associated with young children has shrunk most dramatically among college-educated women. The gap associated with school-aged children for this group, however, has remained surprisingly stable. The motherhood gaps are as large for women with high school or less and women with some college. While the changes have been more dramatic for college-educated women, the same general trends—declining marriage gaps and persistent motherhood gaps—also pertain to less-educated women.

Women have made remarkable progress in terms of acquiring labor market skills and in fact have surpassed men in terms of years of formal schooling. Women have also made progress in terms of labor market participation. Recent cohorts of women are much less likely to permanently leave the labor force upon having children than were previous generations of women. This means that a typical woman now has more years of labor market experience compared to previous cohorts. The life-cycle pattern of earnings is distinct from that of labor force participation however. Women may return to the labor force after having children but their earnings do not return to the promise of earlier trajectories. Since recent cohorts of women postponed childbearing and invested early in their careers, these women have in some sense more to lose in terms of labor market compensation when they are ready to have children. The experiences of the highly educated professional women catalogued in recent papers suggest that the drop in earnings reflects a combination of reduced hours, loss in human capital associated with interrupted careers, and shifts towards less-intensive career tracks.

Maternity Leave and Childcare Subsidies

In her pioneering work on the family gap, Waldfogel (1998, in this journal) pointed to family policies such as paid maternity leave and subsidized childcare as promising ways to close both the family and gender gaps. At the time, women's earnings relative to men's were much lower in the United States than in the Nordic countries with generous family leave policies and publicly provided childcare (like Sweden, Norway, and Denmark), so this argument seemed highly plausible.

But subsequent research has not always produced the expected results. Olivetti and Petrongolo (in this issue) provide a survey of this literature. The impact of family policies on reducing the gender gap appears to be stronger in cross-country studies, but these types of comparisons are likely to confound policies and prevailing social norms which affect both variables. The evidence is much weaker in microeconomic studies where policy changes within countries are investigated. Comparing across types of policies, parental leave versus subsidized child care, they find somewhat stronger case for subsidized child care. This appears to be because in many cases, particularly for longer durations of job protection and paid leave, parental leave policies may reinforce gender stereotypes and discourage hiring and promotion of women.

The experience of Scandinavian countries produces an interesting perspective. While the expansion of family policies may have increased female labor force participation, much of the increase was in part-time work, and women in these countries were less likely to be in management and professional occupations than women in the United States (Blau and Kahn 2013). Indeed, the gender gap in Sweden is larger at the upper end of the earnings distribution, consistent with the notion of underlying factors leading to a "glass ceiling" that limits women from advancing (Albrecht, Björklund, and Vroman 2003). Two recent studies using administrative data on earnings from Sweden and Denmark provide convincing evidence that mothers, but not fathers, have large reductions in relative earnings following the birth of their first child in both countries. Using Swedish data from 1986–2008, Angelov, Johansson, and Lindahl (2016) follow the within-family difference between the earnings of mothers and fathers and find that the earnings gap increases by 32 percentage points and wage gaps by 10 percentage points after birth of their first child. Using similar data for Denmark, Kleven, Landais, and Sogaard (2015) find that 10 years after childbirth, mothers' earnings have fallen 20 percent relative to non-mothers, while fathers' earnings continue to track those of non-fathers. The authors simulate the impact of these within-family gaps associated with children on the overall gender gap and find that their importance to the overall gap grew substantially between 1980 and 2011—rising from 30 percent to 80 percent of the gap. The persistence of children-related wage gaps in these countries with very generous family policies casts doubt on the notion that these policies constitute a panacea that will reduce the gender gap.

It is plausible that adopting family policies and other programs that support working families as they go about the business of bringing up children—an expensive proposition—may improve family and children's well-being. But it is not clear that such policies narrow the gender gap in earnings.

Conclusion

Given women's gains in the labor market, Becker's (1981) seminal model predicts that patterns of specialization should become less gendered. It predicts that the gender earnings gap associated with marriage should fall, and it has. However, the gender earnings gap associated with children has been more persistent, and the proportion of the remaining gender earnings gap associated with children has risen.

The persistent nature of the motherhood gap—particularly among professional women poised for high-paying careers, and among women who have access to generous leave benefits and childcare subsidies—brings home the point that women still devote much more time to child-rearing over the course of their careers than do men with similar human capital characteristics. One set of explanations put forward revolve around social norms that are slow to change and resist economic forces (Fortin 2005; Bertrand 2011; Bertrand, Kamenica, and Pan 2015). Social norms can serve as both push or pull factors. On the pull side, women may still by-and-large identify themselves as the primary caretaker of children. On the push side, work places may still be governed by norms from an earlier era of male breadwinners with stay-at-home wives. According to this set of norms, an employee is a "good" employee only if he or she is married to the job and willing to work long hours. A number of papers have shown that the gender gap is particularly large in jobs that require long hours (Goldin 2014; Gicheva 2013; Cha and Wheeden 2014; Cortes and Pan 2016).

A relevant question is whether these types of workplace practices actually lead to more-productive work places, thereby reflecting optimizing behavior by employers versus lingering practices that have outlived their usefulness. Bloom, Kretschmer, and van Reenen (2011) find that better managed firms institute family-friendly polices, but once the quality of management is controlled for, there is little correlation, either positive or negative, between family-friendly policies and productivity. More work needs to be done in this area.

The predictions of the specialization model—that specialization will be less gendered as men and women look more similar in terms of their productive capacities—point in the direction of a change that is still likely to occur over time, but it appears that closing the motherhood earnings gap may take a while longer.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2015. "The Career Costs of Children." Working paper, University College London. January.

Aguiar, Mark, and Erik Hurst. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics* 122(3): 969–1006.

Albrecht, James, Anders Björklund, and Susan Vroman. 2003. "Is There a Glass Ceiling In Sweden?" *Journal of Labor Economics* 21(1): 145–77.

Angelov, Nikolay, Per Johansson, and Erica Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics* 34(3): 545–79.

Antonovics, Kate, and Robert Town. 2004. "Are All the Good Men Married? Uncovering the Sources of Marital Wage Premium." *American Economics Review* 94(2): 317–21.

Avellar, Sarah, and Pamela J. Smock. 2003. "Has the Price of Motherhood Declined Over Time? A Cross-Cohort Comparison of the Motherhood Wage Penalty." *Journal of Marriage and Family* 65(3): 597–607.

Bailey, Martha J., Melanie E. Guldi, and Brad J. Hershbein. 2013. "Is There a Case for a 'Second Demographic Transition'? Three Distinctive Features of the Post-1960 US Fertility Decline." NBER Working Paper 19599.

Bailey, Martha J., Brad Hershbein, and Amalia R. Miller. 2012. "The Opt-In Revolution? Contraception and the Gender Gap in Wages." *American Economic Journal: Applied Economics* 4(3): 225–54.

Beaudry, Paul, and Ethan Lewis. 2014. "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980–2000." *American Economic Journal: Applied Economics* 6(2): 178–94.

Becker, Gary. 1981 [1991]. *A Treatise on the Family.* Enlarged edition. Cambridge, MA: Harvard University Press.

Belkin, Lisa. 2003. "The Opt-Out Revolution." *New York Times*, October 26. http://www.nytimes.com/2003/10/26/magazine/the-opt-out-revolution.html.

Bertrand, Marianne. 2011. "New Perspectives on Gender." Chap. 17 in *Handbook of Labor Economics.* Amsterdam: North Holland.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2(3): 228–55.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130(2): 571–614.

Blackburn, McKinley, and Sanders Korenman. 1994. "The Declining Marital-Status Earnings Differential." *Journal of Population Economics* 7(3): 247–70.

Black, Sandra E., and Chinhui Juhn. 2000. "The Rise of Female Professionals: Are Women Responding to Skill Demand?" *American Economic Review* 90(2): 450–55.

Blau, Francine D., and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics* 15(1): 1–42.

Blau, Francine D., and Lawrence M. Kahn. 2013. "Female Labor Supply: Why is the US Falling Behind?" NBER Working Paper 18702.

Blau, Francine D., and Lawrence M. Kahn. 2016. "The Gender Wage Gap: Extent, Trends, and Explanations." IZA Working Paper 9656, January.

Bloom, Nick, Tobia Kretschmer, and John Van Reenen. 2011. "Are Family-Friendly Workplace Practices a Valuable Firm Resource?" *Strategic Management Journal* 32(4): 343–67.

Budig, Michelle J. 2014. "The Fatherhood Bonus and the Motherhood Penalty: Parenthood and the Gender Gap in Pay." Report prepared for the Third Way Next Research Report series.

Budig, Michelle J., and Paula England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66(2): 204–225.

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science* 15(3): 75–141.

Cha, Youngjoo, and Kim A. Wheeden. 2014. "Overwork and the Slow Convergence in the Gender Gap in Wages." *American Sociological Review* 79(3): 457–84.

Correll, Shelley J., Stephen Bernard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5): 1297–1339.

Cortes, Patricia, and Jessica Pan. 2016. "When Time Binds: Returns to Working Long Hours and the Gender Wage Gap among the Highly Skilled." IZA Discussion Paper 9846.

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990." *Quarterly Journal of Economics* 115(4): 1287–1315.

DeNavas-Walt, Carmen, and Bernadette D. Proctor. 2015. *Income and Poverty in the United States: 2014.* U.S. Census Bureau, Current Population

Reports, P60-252. Washington, DC: US Government Printing Office.

Fortin, Nicole M. 2005. "Gender Role Attitudes and Women's Labour Market Outcomes across OECD Countries." *Oxford Review of Economic Policy* 21(3): 416–38.

Gicheva, Dora. 2013. "Working Long Hours and Early Career Outcomes in the High-End Labor Market." *Journal of Labor Economics* 31(4): 785–824.

Goldin, Claudia. 2006. "The Quiet Revolution that Transformed Women's Employment, Education, and Family." *American Economic Review* 96(2): 1–21.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4): 1091–1119.

Goldin, Claudia, and Lawrence Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110(4): 730–70.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of the American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20(4): 133–56.

Goldstein, Joshua R., and Catherine T. Kenney. 2001. "Marriage Delayed or Marriage Forgone? New Cohort Forecasts of First Marriage for U.S. Women." *American Sociological Review* 66(4): 506–519.

Gray, Jeffrey S. 1997. "The Fall in Men's Return to Marriage: Declining Productivity Effects or Changing Selection?" *Journal of Human Resources* 32(3): 481–504.

Greenwood, Jeremy, and Nizah Guner. 2008. "Marriage and Divorce since World War II: Analyzing the Role of Technological Progress on the Formation of Households." Chap. 4 in *NBER Macro Annual 2008,* Vol. 23, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford. Cambridge: University of Chicago Press.

Hill, Martha S. 1979. "The Wage Effects of Marital Status and Children." *Journal of Human Resources* 14(4): 579–94.

Isen, Adam, and Betsey Stevenson. 2010. "Women's Education and Family Behavior: Trends in Marriage, Divorce and Fertility." NBER Working Paper 15725.

Juhn, Chinhui, and Kristin McCue. 2016. "Selection and Specialization in the Evolution of Marriage Earnings Gaps." *RSF: The Russell Sage Journal of the Social Sciences* 2(4): 237–69.

Killewald, Alexandra. 2013. "A Reconsideration of the Fatherhood Premium: Marriage, Coresidence, Biology, and Father's Wages." *American Sociological Review* 78(1): 96–116.

Killewald, Alexandra, and Margaret Gough. 2013. "Does Specialization Explain Marriage Penalties and Premiums?" *American Sociological Review* 78(3): 477–502.

Kleven, Henrik, Camille Landais, Camille, and Jakob Sogaard. 2015. "Children and Gender Inequality: Evidence from Denmark." Unpublished paper, London School of Economics, March.

Korenman, Sanders, and David Neumark. 1991. "Does Marriage Really Make Men More Productive?" *Journal of Human Resources* 26(2): 282–307.

Korenman, Sanders, and David Neumark. 1992. "Marriage, Motherhood, and Wages." *Journal of Human Resources* 27(2): 233–55.

Lundberg, Shelly, and Elaina Rose. 2000. "Parenthood and the Earnings of Married Men and Women." *Labour Economics* 7(6): 689–710.

Murphy, Kevin M., and Robert H. Topel. 2014. "Human Capital Investment, Inequality and Growth." George J. Stigler Center for the Study of the Economy and State Working Paper 253, February 2014. http://www.chicagobooth.edu/assests/stigler/253.pdf.

Neumark, David, and Sanders Korenman. 1994. "Sources of Bias in Women's Wage Equations: Results using Sibling Data." *Journal of Human Resources* 29(2; Special Issue: Women's Work, Wages, and Well-being): 379–405.

Pal, Ipshita, and Jane Waldfogel. 2016. "The Family Gap in Pay: New Evidence for 1967 to 2013." *RSF: The Russell Sage Journal of the Social Sciences* 2(4): 104–27.

Schwartz, Christine R., and Robert D. Mare. 2005. "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42(4): 621–46.

Slaughter, Anne Marie. 2012. "Why Women Still Can't Have it All." *The Atlantic*, July/August. http://www.theatlantic.com/magazine/archive/2012/07/why-women-still-cant-have-it-all/309020/.

Stevenson, Betsey. 2007. "The Impact of Divorce Laws on Marriage-Specific Capital." *Journal of Labor Economics* 25(1): 75–94.

Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and their Driving Forces." *Journal of Economic Perspectives* 21(2): 27–52.

US Bureau of Labor Statistics. 2015. "Highlights of Women's Earnings in 2014." BLS Report 1058.

Waldfogel, Jane. 1997. "The Effect of Children on Women's Wages." *American Sociological Review* 62(2): 209–217.

Waldfogel, Jane. 1998. "Understanding the 'Family Gap' in Pay for Women with Children." *Journal of Economic Perspectives* 12(1): 137–156.

Weinberg, Bruce A. 2000. "Computer Use and the Demand for Female Workers." In*dustrial and Labor Relations Review* 53(2): 290–308.

Welch, Finis. 2000. "Growth in Women's

Relative Wages and in Inequality among Men: One Phenomenon or Two?" *American Economic Review* 90(2): 444–49.

Wilde, Elizabeth Ty, Lilly Batchelder, and David Ellwood. 2010. "The Mommy Track Divides: The Impact of Childbearing on Wages of Women of Differing Skill Levels." NBER Working Paper 16582.

Williams, Joan C. 2010. *Reshaping the Work-Family Debate: Why Men and Class Matter.* Harvard University Press.

Wood, Robert G., Mary E. Corcoran, and Paul N. Courant. 1993. "Pay Differences among the Highly Paid: The Male–Female Earnings Gap in Lawyers' Salaries." *Journal of Labor Economics* 11(3): 417–44.

The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries

Claudia Olivetti and Barbara Petrongolo

Among the most remarkable changes in the labor markets of high-income nations during the past century have been the rise in the female workforce and the narrowing of gender gaps in schooling and earnings. At the same time, government mandates and firm policies regarding families expanded. In some instances, legislation was preceded by great economic change, as when the spread of industrialization in the 19th century led to calls for restrictions on female work. Other legislation resulted from social and political change, as occurred during the women's movement of the 1960s and 1970s. Demographic change also played a role as nations have sought to address declining fertility or when dictatorships desired to increase population. By the early 21st century, most high-income countries have put into effect a host of generous and virtually gender-neutral parental leave policies and family benefits, with the multiple goals of gender equity, higher fertility, and child development.

What have been the effects? Proponents typically emphasize the contribution of family policies to the goals of gender equity and child development, enabling women to combine careers and motherhood, and altering social norms regarding

■ *Claudia Olivetti is Professor of Economics, Boston College, Boston, Massachusetts, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Barbara Petrongolo is Professor of Economics, Queen Mary University London; Director of the Labour Economics Programme, Centre for Economic Policy Research (CEPR); and Research Associate, Centre for Economic Performance, London School of Economics, all in London, United Kingdom. Their email addresses are claudia.olivetti@bc.edu and b.petrongolo@qmul.ac.uk.*

gender roles. Opponents often warn that family policies may become a long-term hindrance to women's careers because of the loss of work experience and the higher costs to employers that hire women of childbearing age.

Understanding the causal impact of family policies on gender outcomes has faced two main challenges. First, family legislation is complex. Parental leave can vary in length, job protection, income support, and availability to either parent. The rules and costs governing preschool education and child care vary considerably across countries. Some countries have enabled direct family transfers and tax allowances to low-income working parents, differing in rules and magnitudes. Recent and increasingly common mandates include a legal right to part-time work and flexible working time. A further complication is that policies should not be analyzed in isolation. If a nation passes longer parental leave entitlements, the effects will be determined in part by benefit coverage during leave and the cost and availability of childcare services once leave entitlements expire.

A second challenge is determining cause-and-effect relationships. For example, the evolution of social norms towards more egalitarian gender roles may induce both family legislation and higher female labor force participation. An empirical approach that attributes the entire increase in female participation to the passage of the legislation will overstate its impact.

Existing research has tackled these issues at both the country and the individual levels. The country-level approach captures the impacts of policies based on between- and within-country variation in intervention, exploiting internationally consistent data on a variety of labor market outcomes. This approach has the advantage of considering an array of policy interventions and interdependencies among them, as well as general equilibrium effects of the policies. But such measurement is invariably coarse and the identification of the causal impacts of interest can be problematic. Because we will show some estimates based on country-level data, we will need to emphasize these limitations throughout our discussion.

The micro-level approach evaluates the causal impact of specific policies within a country by combining rich microdata with variation from natural experiments, such as the lengthening of leave policy or the provision of paid leave. The approach generally considers just one policy intervention at a time, but detailed characterization of the institutional environment can allow for more meaningful comparisons.

We draw lessons here from existing work and our own analysis on the effects of parental leave and other interventions aimed at aiding families. The outcomes of interest are female employment, gender gaps in earnings, and fertility. We begin with a discussion of the historical introduction of family policies ever since the end of the nineteenth century and then turn to the details regarding family policies currently in effect across high-income nations. We sketch a framework concerning the effects of family policy to motivate our country- and micro-level evidence on the impact of family policies on gender outcomes. Most estimates range from negligible to a small positive impact. But the verdict is far more positive for the beneficial impact of spending on early education and child care.

Historical Background

While all developed countries now have in place some form of parental leave policy and family transfers, the path to policy adoption has differed widely across countries, in terms of both its timing and political rationale. In the mid-19th century, early efforts to regulate working conditions in industrialized (or industrializing) countries often encompassed special provisions for female work. Britain and Switzerland pioneered this movement by introducing specific restrictions on female work shifts since the 1840s. Starting around mid-19th century, virtually all US states gradually adopted legislation on maximum weekly hours for women, and most western economies restricted the employment of married women in general or in specific professions. Later in the century, Germany, Sweden, Austria, Belgium, the Netherlands, Denmark, and Switzerland introduced explicit regulations for (mostly unpaid) maternity leave, followed by France, United Kingdom, Italy, Spain, and Greece in the early 20th century.[1]

The emphasis in early legislation was mostly phrased in terms of protecting physically weaker workers from extreme working conditions, and concerns for the health of mothers and children typically led to bans on female employment within a few weeks of birth. Mandated leave was only sporadically accompanied by job protection or income support. Unions often latched onto such special provisions for women in order to lobby for a shorter workweek for men (Goldin 1988). In 1919, the International Labour Organization advocated maternal rights to 12 weeks' leave from work around the time of birth, combined with job protection and partial income support. While maternal leave was ratified in most member countries, job and income protections did not become the norm until much later in the 20th century.[2]

In the 1950s, the design of family policies across Europe emphasized traditional gender roles, and explicitly protected women in their capacities as wives and mothers. During World War II, women in countries with high rates of male military mobilization filled jobs in male-dominated sectors like manufacturing, transportation, and military industry. Despite these developments—or sometimes as a response to them—family policy legislation in some European countries often seemed designed to reaffirm women's household roles. For example, some countries extended job leave rights without granting job protection (Ruhm 1998, and references therein), which can be interpreted as encouraging women to take leave, while raising uncertainty about the ability to return to work in a similar position.

The late 1960s and 1970s brought important changes in maternity leave provisions and set the basis for a wider selection of modern family policies. The sharp rise in female labor market participation generated greater demands for maternity

---

[1] See Wilkander, Kessler-Harris, and Lewis (1995).

[2] In Appendix Table A1, available online with this paper at http://e-jep.org, we report a summary of early legislation based on a comparative study published by the US Department of Labor Children's Bureau (Harris 1919).

leave provisions as a way to reconcile careers and motherhood. Countries that had adopted maternity leave earlier often extended these provisions substantially, while other countries like Canada and Australia introduced such provisions. Most high-income countries combined leave periods with job protection and increased income support during employment breaks. Sweden was the first country to introduce explicit paternity leave rights in 1974, allowing mother and father to share six months of parental leave. Other European countries started to supplement "maternity leave," available to mothers around the time of childbirth, with "parental leave," available to both parents during a child's early years (as reported in the OECD Family Database "PF 2.5 Annex: detail of change in parental leave by country"). These changes, together with the decline in the manufacturing sector and the weakening of trade unions, contributed to eroding the male breadwinner model in most high-income countries.

The United States notably lagged behind these general trends. Back in 1919, the Children's Bureau published a comparative study of "Maternity Benefit System in Certain Foreign Countries" (Harris 1919), which stated that the report was commissioned "in the hope that the information might prove useful to the people of one of the few great countries which as yet have no system of State or national assistance in maternity—the United States." Despite having in place equal pay legislation since 1963, and maternity leave legislation in a few states starting in the late 1960s, the only maternity provision adopted at the federal level until the 1990s was the Pregnancy Discrimination Act of 1978, prohibiting unequal treatment of pregnant women. Parental leave rights were introduced in the United States with the passage of the Family and Medical Leave Act of 1993, allowing eligible employees up to 12 weeks of unpaid leave for pregnancy and newborn care. A major selling point emphasized by its proponents was its claimed beneficial impact—beyond mothers' welfare and careers—on child development, reduced abortions, and men's access to leave (Anthony 2008). Opposition was driven to a large extent by its perceived costs to employers. Currently 25 states, most notably California, have more generous parental leave provisions than the federal law.

In 1996, the EU Directive on Parental Leave ratified rights to at least three months of parental leave for childcare purposes, over and above maternity leave rights, seeking to "facilitate the reconciliation of parental and professional responsibilities for working parents." The EU Directive also encouraged member states to limit transferability of each parent's rights across parents, so as to achieve a more equal participation of parents in child care. The 2010 Parental Leave Directive further extended leave rights to four months. Pronatal motives often underpinned the recent waves of parental leave and other family policy reforms, aiming to remedy the the problem of birth rates falling below replacement levels and demographic aging in several European countries (Raute 2015).

While maternity or parental leave has historically been the most important dimension of family policies, the introduction of leave rights has often been followed, with long and varying lags, by other family-friendly policies such as public or subsidized child care, workplace practices such as part-time work or flexible

working time, and in-work benefits for parents. The rationale behind these policies was often to encourage fertility while limiting the career penalty of motherhood.

Family Policies in High-Income Countries

At present, all high-income industrialized countries have in place paid maternity leave rights (with the exception of the United States where this is unpaid), and provide some support, in cash or kind, for child care. Table 1 provides a snapshot of some key family policies in a recent cross-section of developed economies, including the United States, Canada, Australia, Japan, and 11 large European countries. All indicators reported are obtained from the OECD Family Database and Social Expenditure Database and refer to the latest available year, between 2011 and 2015.[3]

Countries are organized in decreasing order of duration of job-protected leave provisions for mothers, which is reported in column 1. This includes maternity leave and the maximum job-protected parental leave available to mothers for home care of children, whether or not income support is also included. For simplicity we will refer to this variable as "parental leave." The median parental leave is about 60 weeks, with very wide variation across countries, summarized by a standard deviation of about one year. Germany, France, Spain, and Finland have leave entitlements above three years, followed by Norway and Sweden with around 20 months of entitlement. At the other extreme, the United States has 12 weeks of parental leave. While this figure refers to federal entitlements, 25 US states have expanded in some way or another upon federal legislation. Interestingly, cross-country variation in parental leave rights is much wider than in other labor market institutions such as the unemployment benefit replacement ratio and the tax wedge—and, as we discuss later, wider than in gender employment outcomes.[4]

Variation in maternity leave provisions around the time of childbirth, shown in column 2, is modest in comparison, with most countries ranging between 14 and 22 weeks. As shown in column 3, on average about one-third of this must be taken before birth. The bans that some countries have on working during late pregnancy are likely a vestige of early legislation, from a time when a larger share of jobs, like many manufacturing jobs of the past, were physically strenuous.

---

[3] Table A2, available in the online Appendix to this paper at http://e-jep.org, reports this information for a larger sample of 30 OECD countries.

[4] The cross-country coefficient of variation in the length of parental leave is between two and three times the coefficient of variation in the unemployment benefit replacement ratio or the tax wedge (indicators from OECD Benefit and Wages data, http://www.oecd.org/els/benefits-and-wages-statistics.htm, and Tax Wedge data, https://data.oecd.org/tax/tax-wedge.htm, 2014). While leave entitlements are a good predictor of the percentage of employed women who are on leave during the first year after birth, representing a measure of their take-up rate, the corresponding correlation (measured on 18 countries for which data on take-up is available) is only 0.44.

Table 1

Cross-country Variation in Family-Friendly Policies

	Maximum job-protected leave for mothers (weeks) (1)	Total maternity leave (weeks) (2)	Pre-birth leave (% maternity leave) (3)	Total paid leave available to mothers (weeks) (4)	Average payment rate for mothers (% of average, 2014, national earnings) (5)	Total paid leave available to father (% total paid leave for both parents) (6)	Early childhood education and care (% GDP) (7)	Accumulate days off and vary start/end of daily work (% companies) (8)
Spain	166	16	63	16	100	12	0.6	34.07
France	162	16	38	42/110[a]	44.7	40/33[a]	1.2	54.29
Germany	162	14	43	58	73.4	13	0.5	62.00
Finland	161.03	17.5	29	161.03	26.5	5	1.1	86.05
Norway	91	13	23	91	50.0	10	1.2	—
Sweden	85	15.6	45	60	63.4	14	1.6	74.18
United Kingdom	70	52	21	39	31.3	5	1.1	46.83
Greece	60.33	43	19	43	53.9	1	0.1	20.60
Japan	58	14	43	58	61.6	47	0.4	—
Australia	52	6	100	18	42.0	10	0.6	—
Canada	52	17	47	52	52.6	0	0.2	—
Denmark	50	18	22	50	54.1	4	2.0	76.91
Italy	47.7	21.7	18	47.7	52.7	0	0.6	39.96
Netherlands	42	16	38	16	100	2	0.9	66.48
United States	**12**	**0**	**0**	**0**	**0**	**0**	**0.4**	—

Sources: Columns (1) to (5): OECD, Family Database, Public policies for families and children (PF), indicator "PF2.5 Trends in parental leave policies since 1970." Column (6): OECD, Family Database, Public policies for families and children (PF), indicator "PF2.1 Key characteristics of parental leave systems." Column (7): OECD, Social Expenditure database "PF3.1 Public spending on childcare and early education, in percentage of GDP" (December 2014). Column (8): OECD Family database—The labour market position of families (LMF), indicator "LMF2.4 Family-friendly workplace practices based on European Company Survey, 2013." See http://www.oecd.org/els/family/database.htm and www.oecd.org/social/expenditure.htm.

Notes: Column (1): Maximum number of weeks of employment-protected parental leave available to mothers, regardless of income support. This is the sum of weeks of maternity leave, parental leave, and home care of children. Column (2): Total number of weeks of maternity leave available to employed women, regardless of income support. These are defined as employment-protected leaves of absence for employed women around the time of childbirth, or adoption (in some countries). Column (3): Percentage of total weeks of maternity leave that a woman is allowed to take before the expected date of childbirth. Column (4): Total number of weeks for which a mother can receive payments from the benefit attached to or associated with parental leave, regardless of the period of employment protection. Column (5): Number of weeks of paid leave reserved for the exclusive use of the father, divided by the sum of weeks of paid leave for the exclusive use of the father and weeks of paid leave available to mothers. (6): The "average payment rate" is the proportion of previous earnings replaced by the benefit over the length of the paid leave entitlement for a person earning 100% of average national (2014) earnings. If this covers more than one period of leave at two different payment rates, a weighted average is calculated based on the length of each period. Column (7): Government spending in early childhood education and care measured as the sum of benefits in cash and kind, in percentage of GDP (see online Appendix for details). Column (8): Percentage of companies in the European Company Survey (administered by the European Foundation for the Improvement of Living and Working Condition, Eurofound) that report providing the possibility to accumulate hours for days off (full or half days) and to vary the start and end of daily work to at least some employees. Sample period: latest year available. This is 2015 for columns (1) to (6), 2011 in column (7), and 2013 in column (8).

[a] In columns (4) and (5), we report two statistics for France. The first number refers to families with one child, the second to families with two children.

In all countries except the United States, a substantial portion of parental leave is paid, as shown in column 4. Leave benefits are usually funded by a combination of social insurance systems and employee and employer social security contributions. The proportion of previous earnings replaced by maternity benefits is on average 52 percent, based on 2014 country-specific average earnings, as shown in column 5. According to column 6, paid leave entitlement for fathers is on average 11 percent of total paid entitlement per household. Wherever parental leave is available, entitlement is not entirely transferable between parents (for details, see OECD 2016).

Column 7 reports data on public expenditure on early childhood and educational care, in cash or kind (or, for brevity, "early childhood spending"). This is on average 0.8 percent of country-specific GDP, but up to 2 percent in Denmark, and above 1 percent in the rest of Scandinavia, the United Kingdom, and France. North American and Southern EU countries have the lowest rates of early childhood public spending. In the United States, early childhood public spending is 0.4 percent of GDP.

The final column of Table 1 reports an indicator of work time flexibility, which is only available for EU member states. On average, 56 percent of firms across EU countries offer employees the opportunity to accumulate days off and to vary the start and end of daily work. This proportion falls below 40 percent in southern Europe and rises above 75 percent in Scandinavia.

In comparing family policies in Table 1, it's important to bear in mind that the introduction of parental leave rights and family-related subsidies has often been accompanied by or has followed changes in a country's social norms and attitudes towards gender roles in the home and the market. For example, while family policy legislation in the post–World War II period reflected the role of women as primary providers of child and home care, women's movements of the late 1960s contributed to introducing the first elements of equal parental treatment in family intervention. In reality, different countries may be adopting, say, generous parental leave out of quite different motivations, and it would not be clear a priori whether this is more in line with paternalistic considerations of protecting the "weak" or goals of gender equality.[5]

To give an example, women in Denmark and Italy have very similar entitlement to parental leave around 50 weeks, with nearly identical replacement ratios. However, maternity leave extensions in Italy happened mostly before the 1960s, with long mandatory absence periods before and after birth, especially in manufacturing and agriculture, and no provisions for fathers. In Denmark, the bulk of parental leave legislation came into play after 1960, during decades of rapidly evolving social norms, and with limited substitutability between maternal and paternal leave rights. Comparable maternal leave rights are currently coexisting with relatively

---

[5]Previous work has shown that countries with more conservative gender norms exhibit lower female employment rates, higher gender gaps in college education, and disproportionately lower marriage rates for highly educated women (among others, see Fortin 2005; Bertrand, Cortes, Olivetti, and Pan 2016). However, less is known about how family policies correlate to such norms.

gender-biased norms in Italy—where, according to the European Values Survey, 70 percent of the population agree or strongly agree with the statement "Pre-school children suffer from a working mother," but with much more gender-neutral attitudes in Denmark—where only 10 percent of the population agree with that statement. In fact, cross-country evidence does not reveal any clear-cut association between the generosity of parental leave and answers to gender-related survey questions.[6] However, countries with more conservative views on men and women's roles in society tend to spend less on early childhood education and child care, and are less likely to accommodate flexible working arrangements.

Framework

Most family policies are intended to encourage female labor supply. For example, subsidized child care seeks to provide direct substitutes for maternal child care. Maternity leave seeks to enable mothers to stay attached to the labor market during temporary interruptions of employment, while retaining firm-specific or occupation-specific human capital. Similar arguments can be made for flexible or part-time work arrangements. However, extended maternity leave may have detrimental effects on female labor supply in the long run if it induces women to stay out of work for long enough periods, or repeated periods, in a way that hinders them from effectively re-entering employment on the same pre-maternity track.

Besides these first-order impacts on labor supply, family policies may feed into labor demand decisions via at least two channels. On the one hand, insofar as part of the costs of these arrangements directly or indirectly trickles down on employers, the demand for female labor (and especially for women of child-bearing age) would be negatively affected. On the other side, if family policies effectively make continuity of employment for mothers easier, and their enhanced labor market attachment is incorporated into employers' beliefs, the extent of statistical discrimination (if any) against women would be reduced, with beneficial effects on labor demand for women.

In a competitive labor market with imperfect substitution of inputs, the change in the gender wage ratio as a result of family policies is theoretically ambiguous, depending on the relative shifts in labor supply and labor demand and the context in which such shifts occur. For example, if equal pay legislation effectively prevents a fall in female wages, then policies that would raise the cost of hiring women may lead to a fall in female employment at constant wages. Similar effects are to be expected in the presence of union contracts or binding minimum wages. In most countries considered, collective wage negotiations set gender-neutral, industry-specific wage

---

[6]In Table A3 of the online Appendix, we show bilateral correlations between each policy indicator reported in Table 1 and several qualitative measures of gender norms from the World Values Survey and the European Values Study.

floors. If such floors are above equilibrium wages, most of the effect of family legislation would show up in gender differences in employment rates.

The case of in-work government benefits for lower-wage workers is different, as it implies an increase in female labor supply, at no extra cost for employers, leading to an increase in employment and a fall in wages paid by employers to such workers (unless such wages are sticky downward).

If the labor market is not perfectly competitive—for instance, due to job search frictions and imperfect labor mobility—policies that ease continuity of job-to-worker relationships could be especially valuable in allowing women to retain their match-specific search capital after childbirth. In this setting, firms would have a degree of monopsony power and workers would be paid below their marginal product, according to the wage elasticity of labor supply to the individual employer (Manning 2003). In this case, the costs of family policies may be absorbed by the wedge between the wage and the marginal product without a detrimental impact on female employment. In particular, Manning (2003, chap. 7) argues that, because women with domestic responsibilities may be relatively more restricted in commuting time, hours of work, and other nonwage attributes, their labor supply to the individual firm may be less wage-elastic than male labor supply, implying a higher wedge for women than for men.

In summary, the introduction of family policies might be expected to lead to an increase in female employment and possibly a fall in female relative wages, depending on wage elasticities of labor demand and supply. As these elasticities may vary with skill, age, and family composition of women, the effects of policy intervention may be heterogeneous along these and other dimensions. Wage effects may be mitigated or even reversed whenever continuous labor market attachment or labor market experience is highly valuable, as in the presence of search frictions, high returns to actual labor market experience, and feedback mechanisms onto employers' beliefs. On the other hand, theories of gender statistical discrimination suggest that these policies might backfire by reinforcing employers' beliefs and social norms regarding women's comparative advantage in child care and home production more generally.

Cross-Country Evidence

Given wide international variation in family policies, several papers have compared institutions and gender labor market outcomes across high-income OECD countries. In a prominent early study, Ruhm (1998) examined the effect of parental leave on female employment and wages during 1969–1993 in nine EU countries that experienced significant changes in their respective leave mandates. His analysis indicates that short periods of paid entitlement around three months lead to a 3 to 4 percent rise in female employment rates, with little effect on wages, while longer entitlements of more than nine months lead to negligible additional impact on employment but sizeable negative impacts of about 3 percent on female

wages. Such employment effects can be generated by stronger labor market attachment during job-protected leave and the right of mothers to return to pre-birth jobs, as well as entitlement effects for women who would not otherwise participate in the labor force but intend to accumulate work experience to later qualify for leave benefits. Detrimental wage effects after long periods of absence may be driven by loss of actual labor market experience, as well as by nonwage costs to firms such as disruption and replacement costs. Given the employment effects of leave, changes in wages may also result from the outward shifts in female labor supply and/or selection effects on the composition of female employment.

There are caveats to a causal interpretation of these results, duly noted by Ruhm (1998). In particular, the estimates overstate the true impact of leave rights insofar as their extension is accompanied by the implementation of other family-friendly policies, such as subsidized child care. Furthermore, female labor supply shifts may create political support for parental leave rights and lead simultaneously to both extended rights and higher female employment rates.

The general approach in Ruhm (1998) has been extended by later work to cover more recent years, a wider set of countries, and a richer set of institutions. Thévenon and Solaz (2012) broadly confirm Ruhm's findings on a cross-section of 30 countries observed between 1970 and 2010. Using data on a sample of 17 high-income OECD countries for 1990–2010, Blau and Kahn (2013) find that gender gaps in both employment and wages shrink with parental leave rights, the generosity of benefits, the right of part-time work, and equal treatment legislation (although only the effects of the latter two are statistically significant). The authors conclude that the greater expansion of these policies outside the United States is an important factor explaining why female employment growth in the United States has been weaker relative to these other OECD countries since the early 1990s. Cipollone, Patacchini, and Vallanti (2014) find evidence of heterogeneous policy effects by showing that female participation of medium- and highly-educated women is more responsive to family-oriented policies—as measured by a synthetic index encompassing parental leave, family subsidies, and elderly subsidies—than participation of less-educated women.

A few papers have exploited the staggered introduction of parental leave rights across geographies within a country. Baum (2003) focuses on the partial state-level adoption of leave rights in the United States ahead of the Family and Medical Leave Act in 1993, and fails to detect any significant impact of leave rights on employment or wages of mothers. Using a similar approach, Han, Ruhm, Waldfogel, and Washbrook (2009) detect detrimental employment effects of parental leave and welfare benefits, and positive effects of childcare spending, for single mothers and the less-skilled. Baker and Milligan (2008) find that the introduction of leave rights in Canadian provinces delays return to work of mothers shortly after birth, but eases return to the pre-birth employer.

Below we complement existing cross-country evidence by bringing together data on 30 countries that are currently in the OECD. Figure 1 summarizes evidence on female employment in these countries since the 1970s (or the 1980s wherever

Figure 1
Evolution of Female Employment Rates: 1970s to 2010s
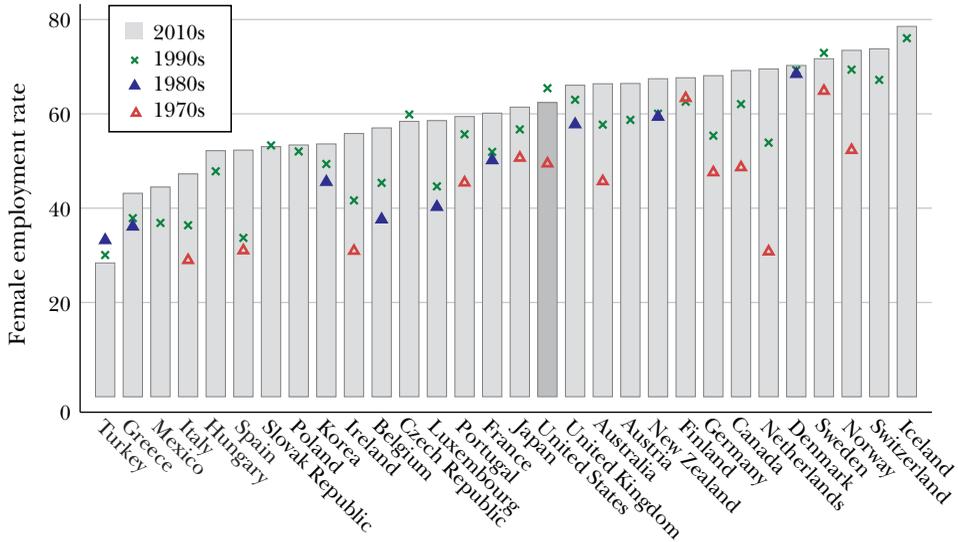
Notes: The figure reports average employment rates for women aged 25–54, by decade. The employed are defined as those who work for pay or profit for at least one hour a week, or who have a job but are temporarily not at work due to illness, leave, or industrial action. We report data since the 1970s or the earliest available decade.

earlier data are not available). The employment rate is measured as the number of individuals aged 25–54 who are employed, divided by the relevant population. Countries are ranked in ascending order of female employment in the 2010s, ranging from 28 percent in Turkey to 79 percent in Iceland. The average female employment rate in the sample is currently 60 percent, with a standard deviation of 10 percent. The US female employment rate of 62 percent is just above the sample average. Scandinavian countries rank towards the top of the chart, followed by most English-speaking countries, while southern European countries and lower-income countries rank towards the bottom. In relative terms, there is much wider variation in parental leave rights across these countries than in female employment. In most countries, female employment has increased in recent decades, from 49 percent on average in the 1980s to 60 percent in the 2010s. However, there is no evidence of narrowing differences in female employment across countries. Until the 1990s, the female employment rate in the United States was among the highest in this sample of countries, but it actually declined since then, from about 66 percent in the 1990s to about 62 percent in the 2010s, and now ranks very close to the sample median.

Table 2
Correlations between Family-Friendly Policies and Women's Outcomes
(coefficients in bold are statistically significant)

	Female Employment Rate	Gender Gap		Total Fertility Rate
		Employment	Earnings	
Maximum weeks of job-protected leave available to mothers	0.188 (0.320) 30	**−0.385** (0.036) 30	−0.134 (0.482) 30	**−0.472** (0.009) 30
Total paid leave available to mothers (weeks)	0.205 (0.278) 30	**−0.320** (0.085) 30	0.018 (0.925) 30	**−0.372** (0.043) 30
Average payment rate, mothers (% average earnings)	−0.103 (0.590) 30	0.175 (0.355) 30	−0.108 (0.570) 30	−0.134 (0.480) 30
Total paid leave available to fathers (% total paid leave)	0.151 (0.426) 30	−0.034 (0.859) 30	0.298 (0.110) 30	−0.129 (0.496) 30
Early childhood education and care (% GDP)	**0.513** (0.004) 30	**−0.466** (0.009) 30	−0.153 (0.419) 30	**0.396** (0.030) 30
Accumulate days off *and* vary start/end of daily work (% companies)	**0.735** (0.000) 21	**−0.657** (0.001) 21	0.275 (0.228) 21	0.195 (0.396) 21

Sources: OECD (2016) Employment Database, http://stats.oecd.org/Index.aspx?DataSetCode=LFS_SEXAGE_I_R. OECD, 2016, Gender wage gap (indicator). doi: 10.1787/7cee77aa-en. OECD (2016), Fertility rates (indicator). DOI: 10.1787/8272fb01-en.
Notes: For each institution/outcome combination, we report the correlation coefficient, its *p*-value (in parentheses), and the number of observations. Coefficients in bold are statistically significant at conventional levels. The employment rate by gender refers to the number of individuals aged 25–54 who are employed, divided by the relevant population. The employed are defined as those who work for pay or profit for at least one hour a week, or who have a job but are temporarily not at work due to illness, leave, or industrial action. The gender gap in employment is the male–female difference in employment rates. The wage gap is computed as the log of the gender ratio between median earnings of working age men and women. Data on earnings are for full-time employees. The total fertility rate is defined as the total number of children that would be born to each woman if she were to live to the end of her childbearing years and give birth to children in alignment with the prevailing age-specific fertility rates. See notes to Table 1 for sources and definition of family-friendly policies indicators. Sample period: all outcomes are 2010–2014 averages.

Data on labor market institutions and outcomes are brought together in Table 2, in which six family policies are considered. The four columns show four different outcomes: the female employment rate; the gender gap in employment, measured by the male–female difference in employment rates; the gender wage gap, computed as the log of the ratio between median earnings of working age men and women for full-time employees; and the total fertility rate, defined as the total number of children that would be born to each woman throughout her childbearing years. The table reports raw correlations between such average outcomes for 2010–2014, and the family policy indicators reported in Table 1.

Figure 2
Employment Gap and Maximum Length of Job-Protected Leave for Mothers
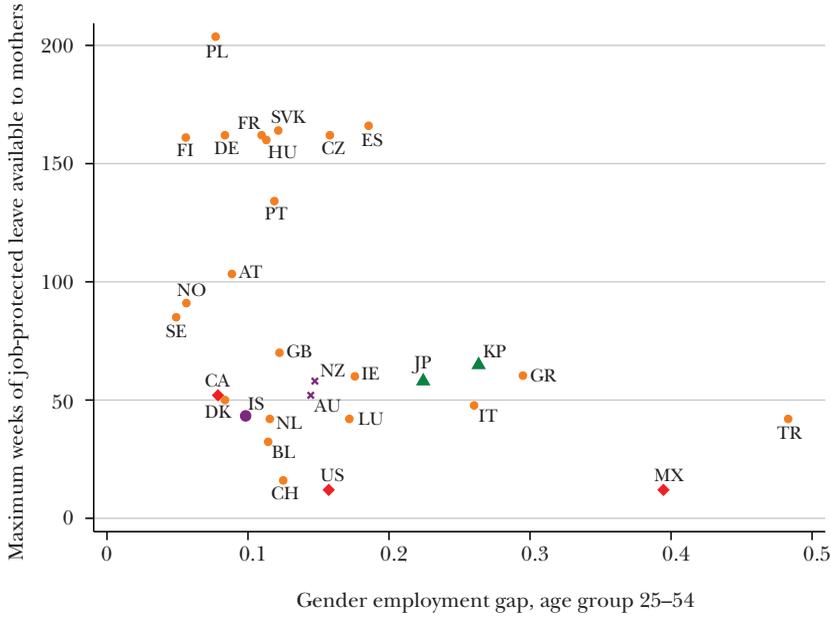
Source: OECD Employment Database, 2016, http://stats.oecd.org/Index.aspx?DataSetCode=LFS_SEXAGE_I_R. OECD, Family Database, Public policies for families and children (PF), indicator "PF2.5. Trends in parental leave policies since 1970," 2016.
Note: The figure plots the gender gap in employment rates (as defined in Table 2) against the maximum weeks of job-protected leave available to mothers for the countries in our sample.

The first two policy choices, maximum weeks of leave available to mothers and total paid leave available to mothers, have a negative, statistically significant correlation with the gender gap in employment. These correlations are overall robust, by which we mean that they are not driven by any small subset of countries in particular, as it can be grasped visually in Figure 2, which plots the gender gap in employment rates against the maximum weeks of job-protected leave available to mothers. The length and financial coverage of parental leave are also negatively correlated to the rate of fertility, but such correlations are entirely driven by the presence of the four eastern European countries in our sample, that are characterized by record low fertility and generous leave provisions. If one removes Poland, Czech Republic, Hungary, and Slovakia from this sample, the correlation between the fertility rate and the length of parental leave becomes small and statistically insignificant.

The next two policy choices—the average payment to mothers as a share of average earnings, and paid leave to fathers—are not significantly associated with any of the outcome variables. The final two policy choices, spending on early childhood care and the index of workplace flexibility, are positively correlated with female

employment and negatively correlated with the employment gap. These correlations are very strongly significant and robust to the exclusion of (small) geographic clusters of countries. Spending on early childhood care is the only one of these policies to have a positive (and robust) correlation with the fertility rate. None of the policy variables are significantly correlated with the gender earnings gap.

We have explored the various correlations further, looking at different groups of women. When we differentiate relevant outcomes across three skill groups—below secondary education, secondary education, and tertiary education—the results show that only for the less-skilled are female relative earnings higher in countries with flexible working arrangements. On the other hand, correlations with employment outcomes are consistent across the skill distribution.[7]

Of course, variation in parental leave may itself induce mechanical variation in measured employment rates, as the employment count includes individuals at work in a reference week as well as those who were temporarily absent from work, and parental leave is among named reasons for such temporary absence. Thus an individual who is not working t months after childbirth may be classified as "employed" in a country with leave entitlement longer than t, and "nonemployed" in a country with shorter entitlement, despite performing the same activity in the two contexts. Comparing ILO employment rates to employment rates obtained by classifying as "nonemployed" any individual temporarily absent from work due to parental leave, we find that female employment rates would be on average 2 percentage points lower with the second definition of employment status—ranging between 3.3 percentage points in Austria and Scandinavian countries, to virtually zero in the United States.[8]

We next look at the impact of family policies on gender outcomes exploiting their evolution over time, and controlling for country and year fixed-effects—while bearing in mind the caveats to a causal interpretation of coefficients expressed above. The results are reported in Table 3, which considers the same four outcomes as in Table 2: the female employment rate, the gender gap in employment and earnings, and the total fertility rate. For each outcome, we use two alternative specifications. The first specification (columns 1, 3, 5, and 7) only controls for the maximum job-protected leave and its square, allowing for nonlinearities in the effect of parental leave on various outcomes as in Ruhm (1998). The second specification also controls for the percentage of the total leave that is income-protected, the average replacement ratio, and the percentage of GDP that is spent on early

---

[7]For details of these additional correlations, see Table A5 of the online Appendix.

[8]Specifically, we compare male and female ILO employment rates to employment rates obtained by classifying as "nonemployed" any individual temporarily absent from work due to parental leave. We can only perform this exercise for countries covered by the Eurostat Labor Force Survey and the United States, based on the Current Population Survey. For this subsample of countries (23 out of 30), the correlations with the set of institutions in Table 2 do not change with the employment definition. In contrast to the result for women, mentioned above, excluding workers on parental leave barely alters male employment rates, with the exception of Iceland and Sweden, with relatively generous paternity leave provisions and high take-up rates. For details, see Table A4 in the online Appendix, in which countries are sorted according to the discrepancy between the two alternative employment measures for women.

Table 3
Family-Friendly Policies and Women's Outcomes

	Female employment rate		Employment gap		Earnings gap		Fertility rate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Maximum weeks of job-protected leave	0.113***	0.063**	–0.050***	0.023	–0.011	–0.210***	0.002	–0.001
	(0.019)	(0.029)	(0.018)	(0.022)	(0.033)	(0.033)	(0.001)	(0.001)
Maximum weeks squared/100	–0.078***	–0.062***	0.043***	0.012	0.016	0.108***	0.001	0.001**
	(0.010)	(0.014)	(0.010)	(0.011)	(0.016)	(0.016)	(0.001)	(0.001)
Percentage of total leave that is paid		–0.037***		0.029***		0.006		0.002***
		(0.008)		(0.006)		(0.008)		(0.000)
Average payment rate		–0.036***		0.027***		0.012		0.000
		(0.011)		(0.008)		(0.019)		(0.000)
Early childhood education and care		3.613***		–1.587***		–2.852**		0.270***
		(0.903)		(0.564)		(1.258)		(0.024)
Constant	43.955***	47.007***	41.954***	37.892***	44.709***	52.367***	2.810***	1.753***
	(1.561)	(2.016)	(1.913)	(2.497)	(0.936)	(1.144)	(0.117)	(0.057)
R^2	0.914	0.921	0.931	0.944	0.943	0.967	0.718	0.692
Mean of dependent variable	*54.8*	*55.1*	*20.6*	*21.0*	*23.4*	*23.7*	*1.9*	*1.7*
Observations	1,026	667	1,026	667	545	340	1,325	806
Sample period	1970–2014	1970–2010	1970–2014	1970–2010	1970–2013	1970–2010	1970–2014	1970–2010
Number of countries	30	22	30	22	30	22	30	22

Notes and Sources: Robust standard errors in parentheses. All specifications include country and year effects. The average payment rate is from the Max Planck Institute's Comparative Family Policy Database (Gauthier 2011). It is computed as a weighted average of payment rates for maternity leave, parental leave, and childcare leave with weights given by the length of each leave type. The cash benefits are expressed as a percentage of the average female wage in manufacturing. See notes to Tables 1 and 2 for all other variable definitions and sources. Percentage of total leave that is paid is the ratio of total paid leave available to mothers to maximum weeks of job-protected weeks (paid/unpaid) available to mothers.
***, ** and * denote significance at 1, 5, and 10 percent levels, respectively.

years education and care. While it is important to control for other family policies, as well as parental leave, because of possible correlation between their timings of introduction, one should bear in mind that this reduces our sample size by about one-third, as the average replacement ratio is not available for eight countries in our sample.[9]

---

[9] These are: Czech Republic, Hungary, Iceland, Korea, Mexico, Poland, Slovakia, and Turkey. The replacement ratio data for 1970–2010 are obtained from the Max Planck Institute for 22 countries, while OECD data on replacement ratios are only available for 2015. If one drops the replacement rate from the regressions while still controlling for the percentage of parental leave that is income-protected and early childhood spending, the estimated effect of parental results is very similar to the one obtained when parental leave is the only included regressor.

Results reported in the first two columns confirm Ruhm's (1998) findings of a nonmonotonic relationship between the duration of parental leave and female outcomes. In our case, however, this holds irrespective of whether the leave is paid, while Ruhm's analysis focuses on paid leave. Our estimates in the full specification of column 2 imply that female employment rises with job-protected parental leave up to 50 weeks and declines thereafter. Quantitatively, however, this effect is small—with a maximum 1.6 percentage point gain, corresponding to 3 percent of the sample average.[10] If the United States—with currently 12 weeks of entitlement—were to extend entitlement to 70 weeks as in Sweden, the corresponding gain in US female employment would be 1.4 percentage points. Conditional on the length of leave entitlements, female employment falls with the percentage of the total leave that is paid and the average replacement ratio, and rises with early childhood spending. Half a percentage point increase in early years spending (corresponding to roughly its standard deviation in this sample) is associated with a 1.8 percentage point increase in female employment.

Column 3 of Table 3 shows results for the difference between male and female employment rates. Because this is a gender gap, it also takes into account the effects, if any, of parental leave on male employment rates. The coefficients on parental leave rights are smaller (in absolute value) than in column 1, which is consistent with a finding that male employment responds to the length of parental leave in the same direction as female employment, but with reduced magnitude. The estimated effect of parental leave becomes not significantly different from zero in column 4, whose sample is restricted to the set of countries for which information on the replacement ratio is available.

Columns 5 and 6 show results for earnings gaps, measured as gender differences in median log annual earnings for full-time workers, and thus encompassing differences in weeks worked as well as differences in weekly wages. Wage effects may in turn be driven by both a causal impact of policy on wages of those in employment, and the changing composition of employment along wage characteristics. The specification in column 5 shows evidence of closing earnings gaps during the first eight months of parental leave, but quantitatively this effect is tiny and not significantly different from zero. The impact of parental leave on the earnings gap becomes significant and sizeable in the specification of column 6. One possible explanation for the different results is that the results of column 6 are obtained after dropping from our working sample countries for which information on the replacement ratio is unavailable, and those countries happen to have systematically lower rates of union density than the rest of the countries.[11] In other words, the

---

[10] Based on a specification of the relationship between employment and parental leave duration of the type: $y = \beta_1 w - \beta_2 w^2/100$, where y denotes employment rates and w denotes weeks of parental leave, the maximum employment effect is detected in correspondence of $w = 50\beta_1/\beta_2$ weeks, and it is equal to $25\beta_1^2/\beta_2$.

[11] Average union density in the Czech Republic, Hungary, Iceland, Korea, Mexico, Poland, Slovakia, and Turkey is 21 percent (down to 12 percent excluding Iceland), against 31 percent in the other countries in our sample.

results of column 6 are obtained on a sample of countries with a lower average incidence of binding union contracts than those of column 5, and may indicate a more sizeable wage response to policy in a context in which wages are relatively more flexible. Overall, coefficients on parental leave denote a stronger effect on earnings gaps in column 6 than on employment gaps in column 4, which implies that wage gaps are also closing for a wide range of parental leave durations.

Column 6 also shows evidence of a relatively strong effect of early years spending on closing earnings gaps, which is larger than the corresponding effect on employment gaps in column 4. By the same logic, this implies that wage gaps are predicted to shrink with childhood spending.

In columns 7 and 8, we show that the effect of parental leave on fertility is also nonmonotonic, but quantitatively this is negligible throughout, independent of the specification used, consistent with Shim's (2014) finding that fertility decisions are not that responsive to parental leave unless leave is also adequately paid. Early childhood spending has a sizeable correlation with fertility, with one extra percentage point of GDP spending associated with 0.2 extra children per woman. The results are overall consistent with Adema, Ali, and Thévenon's (2014) findings that public spending on family benefits and the duration of paid child-related leave for mothers is significantly associated with an increase in the total fertility rate.

In Table 4, we consider heterogeneous policy effects by educational attainment. The sample size is much smaller, due to more limited availability of outcome data, and to save on degrees of freedom we only report specifications based on the whole sample, which only control for parental leave entitlements. For simplicity we do not report regression results for employment gaps, and we cannot report results for fertility, as the fertility rate is only available for the overall population. Thus, the focus here is on female employment and earnings gaps. The notable result from these skill-specific regressions is that the beneficial effects of leave entitlement on female employment are mostly confined to less-skilled women, while high-skill women see their relative earnings fall as a consequence of longer entitlements.

Taken together, these findings suggest that moderate job-protected leave entitlements—up to about one year in our estimates—are associated with higher female employment while not having much correlation with fertility. However, beneficial effects are driven by the low-skill subsample, with possibly detrimental effects for the more educated. Longer and more generously paid entitlements may instead be detrimental for female employment at all skill levels. These effects are consistent with a pattern of progressive labor market detachment during long periods of absence, potentially combined with negative feedback effects on employers' beliefs about work attachment. For college-educated women, longer parental leave seems instead associated with wider earnings gaps. The returns to job-specific experience for this group is plausibly higher than for the less-skilled, and skilled women have more to lose from missed opportunities for career advancements. The one indicator that is across the board associated with more equal gender outcomes is spending in early childhood education and care. Presumably the availability of cheap substitutes

Table 4
Family-Friendly Policies and Gender Gaps by Educational Attainment

	Less than Upper Secondary		Upper Secondary, non-tertiary		Tertiary	
	Female employment rate (1)	Earnings gap (2)	Female employment rate (3)	Earnings gap (4)	Female employment rate (5)	Earnings gap (6)
Maximum weeks of job-protected leave	0.164** (0.083)	−0.112 (0.152)	0.097 (0.060)	0.062 (0.114)	−0.011 (0.046)	0.232** (0.107)
Maximum weeks squared/100	−0.171*** (0.059)	−0.257** (0.130)	−0.097* (0.054)	−0.122 (0.081)	−0.054* (0.030)	−0.138* (0.078)
Constant	47.872*** (3.274)	62.487*** (5.852)	63.132*** (2.013)	50.854*** (4.537)	79.560*** (1.736)	39.626*** (4.955)
R^2	0.946	0.840	0.956	0.883	0.921	0.758
Mean of Dependent Variable	*46.6*	*44.3*	*65.7*	*40.0*	*78.7*	*42.6*
Observations	492	300	504	300	504	300
Sample period	1997–2013	1997–2013	1997–2013	1997–2013	1997–2013	1997–2013
Number of countries	30	29	30	29	30	29

Notes and Sources: Robust standard errors in parentheses. All specifications include country and year effects. In column (1) data for Japan are only available from 1997 to 2001. Iceland is excluded from the sample in columns (2), (4), and (6). See notes to Table 1, Table 2, and online Appendix Table A2 for variable definitions and data sources.
***, ** and * denote significance at 1, 5, and 10 percent levels, respectively.

to maternal care encourages female labor supply, with positive, rather than negative, effects on the accumulation of actual experience.

Micro-level Studies

The micro-level approach aims to identify the causal impact of family policies on several labor market outcomes by focusing on specific country reforms and combining rich microdata—often social security records with detailed information on births, working hours, earnings, and leave take-up—and variation from natural experiments. In this section, we discuss available micro-level evidence on the impact of parental leave, subsidized child care, and in-work benefits.

Maternity and Parental Leave

A few decades ago, a major Austrian reform extended the duration of parental leave from one to two years for children born after July 1, 1990. Lalive and Zweimüller (2009) used the period before and after this time as a natural experiment for analyzing the effects of these changes in entitlement on fertility and mothers'

labor market outcomes. Based on variation across births on either side of policy introduction, the authors detect substantial fertility effects of leave extension, accounting for 12 additional children per 100 women. Extended leave also delays return to work of mothers, even after the benefits are exhausted, resulting in significant reductions in female employment and earnings during the first three years after birth, but only minor effects beyond three years. While fertility effects are stronger for women with below-median pre-birth earnings, the short-run reduction in earnings is larger for high-wage than for low-wage women. Later Austrian reforms of 1996 and 2000 shortened and extended, respectively, entitlement to replacement benefits, leaving job-protected leave unchanged, and Lalive, Schlosser, Steihauer, and Zweimüller (2013) estimate that longer cash benefits significantly delay return to work of mothers when leave is job-protected, but less so once job protection has expired.

Germany enacted five major expansions in maternity leave coverage between 1979 and 1993, which led to gradual and staggered extensions in job-protected leave from 2 to 36 months and in the time of receipt for cash benefits from 2 to 24 months. Schönberg and Ludsteck (2014) find that extension of coverage at short durations leads to small delays in return to work, and the extension of coverage at longer durations leads to larger delays, but it has almost no effect on employment rates and earnings for women more than three years after childbirth. However, extensions of cash benefits beyond the job protection period produce significant long-run employment and earnings losses for affected mothers, which suggest a role for job guarantees in avoiding long-lasting negative effects of benefit extensions.

Norway enacted a series of seven expansions in paid maternity leave, which nearly doubled from 18 weeks in 1977 to 35 weeks in 1992. Dahl, Løken, Mogstad, and Salvanes (2016), using a regression discontinuity approach based on timing of births, find that the paid leave expansion didn't crowd out unpaid leave, thereby substantially expanding maternal time at home (and out of the labor market) in the months following birth, without a reduction in household income. Other than delaying the return to work, the reforms did not have a discernible impact on female labor supply in the long run, fertility, marriage, divorce, or children's schooling outcomes. The authors also note that the extension in paid leave implied regressive transfers towards eligible women, who had systematically higher income than the noneligible, and hence conclude that available evidence on social benefits and redistribution may not support the case for extensions to maternity leave.

In a 2007 reform, Germany linked maternity benefits to pre-birth income for up to 14 months after birth, thus raising the financial incentives to take-up parental leave for higher-earning women. This reform appears to have significantly delayed the return to work of affected mothers during the first year after birth but enhanced female employment in the medium run (Kluve and Tamm 2013). Medium-run employment gains however seem to be limited to part-time jobs, with no discernible impact on full-time employment (Kluve and Schmitz 2014). Raute (2015)

investigates fertility effects of the 2007 German reform and finds sizeable fertility gains for women with above-median earnings and older women.[12]

While most high-income countries currently have in place leave provisions for fathers, their relatively recent introduction, as well as their more limited take-up rate, imply that the evaluation of their effects on female outcomes is still in its infancy. Available evidence shows that the introduction of one month of exclusive paternity leave in Sweden in 1995 increased fathers' time off work after birth while it did not alter fathers' share of child care as measured by their take-up rate of leave for care of sick children (Eckberg, Eriksson, and Friebel 2013). Thus, to date, there is no evidence of beneficial impacts of paternity leave rights on mothers' careers, but the recent extension of exclusive father's leave in other European countries should lead to further evaluations of their impact on mothers' labor market outcomes.

Subsidized Child Care

In several countries, parents returning to work after childbirth continue to receive state support in the form of subsidized or publicly provided child care and preschool programs. For the United States, Cascio and Schanzenbach (2013) evaluate the impact of the introduction of universal preschool for four year-olds in Oklahoma and Georgia since the 1990s, exploiting both state-level variation in policy adoption and age targets, and find only mild evidence for an impact of such programs on maternal labor supply. Despite a substantial increase in preschool enrollment among four year-olds, maternal employment gains are limited to the less-skilled and only during the first few years after the program is in place. Other studies have found that kindergarten increases maternal labor supply. Gelback (2002) uses quarter of birth as an instrument for when children enroll in kindergarten, based on the fact that many states determine kindergarten enrollment based on a child's birthdate, and finds using data from the 1980 Census that child enrollment in kindergarten substantially raises weekly hours, annual weeks, and employment rates for single mothers, but less so for married mothers. Similar qualitative effects of kindergarten subsidies are documented by Cascio (2009), who looks at staggered patterns of state-level provision of subsidies to school districts providing kindergarten in the 1960s and 1970s.

The Canadian province of Québec introduced childcare subsidies for four year-olds in 1997, combined with wider availability and high quality of service. Lefebvre and Merrigan (2008) find a sizeable impact of this scheme on maternal outcomes of Québec mothers relative to other provinces, leading to 8 percentage points higher labor force participation and 231 extra annual hours. In follow-up work, Lefebvre, Merrigan, and Verstraete (2009) and Haeck, Lefebvre, and Merrigan (2015) find that these beneficial effects for mothers' outcomes persist in the long run.

---

[12] The discussion here is not intended to be a comprehensive literature review. Some other studies include Geyer, Haan, and Wrohlich (2015) with an analysis of the employment effects of higher leave benefits in Germany and Asai (2015) for a similar study on Japan. Rossin-Slater, Ruhm, and Waldfogel (2013) find that the expansion of paid leave policy in California did not have a negative impact on employment.

Norway undertook a large-scale expansion in subsidized public child care in 1975. Havnes and Mogstad (2011) find that despite a high take-up rate, subsidies did not encourage maternal employment, as they largely crowded-out informal childcare arrangements. Similar conclusions for more recent years are drawn by Givord and Marbot (2015) for the French case, in which an average 50 percent subsidy to childcare spending introduced in 2004 only raised female participation by 1 percentage point. Bettendorf, Jongen, and Muller (2015) and Nollenberger and Rodríguez-Planas (2015) detect slightly stronger effects for the Netherlands and Spain, respectively.

Unsurprisingly, micro-level studies find relatively smaller effects of childcare subsidies in countries with relatively low childcare costs at baseline (like Norway and France) than in countries in which the cost of child care is higher (like the United States and Canada).

In-Work Benefits

By providing tax credits (mostly) to low-income workers with children, in-work benefits combine poverty alleviation with incentives to work. Brewer, Francesconi, Gregg, and Grogger (2009) and Blundell and Hoynes (2004) provide cross-national overviews on in-work benefits.

Although these tax credits are not primarily targeted to female labor force participation, eligibility and take-up rates are higher among mothers than fathers. A large literature has examined the labor supply effects of the Earned Income Tax Credit (EITC) in the United States, most notably for single mothers. Nichols and Rothstein (2016) survey this literature and highlight a consensus around the sizeable positive effects of the EITC on the labor supply of single mothers, especially for the less-skilled, with most estimates ranging between 3 and 6 extra percentage points of participation (for example, Hotz and Scholz 2003). The effects of interest are mostly identified via an expansion of the EITC for multi-children households, or via state-level changes in the generosity of benefits. The estimated effects are quite different for married women: Eissa and Hoynes (2004) find evidence of small *disincentive* effects on the participation rate of married women, consistent with the fact that the EITC raises average taxation on the secondary earner's earnings. Hotz and Scholz (2003), Nichols and Rothstein (2016), and references in those papers offer detailed discussion of the effects of the EITC on work, poverty, health, and family outcomes.

In the United Kingdom, the main in-work benefit is the Working Family Tax Credit, introduced in 1999, and its effects on the labor supply of various groups (most notably single mothers) were evaluated both via simulations based on structural models of labor supply (Blundell, Duncan, McCrae, and Meghir 2000; Brewer, Duncan, Shepard, and Suarez 2006) and difference-in-differences models based on comparisons with married or cohabiting mothers and/or single women without kids (for example, Francesconi and van der Klaauw 2007; Gregg, Harkness, and Smith 2009). The consensus from this literature is that the WFTC raised the employment rate of lone mothers by 4–5 percentage points, although the interpretation of

difference-in-differences estimates may be complicated by the existence of trending differences in the labor force behavior of treatment and control groups, and the introduction of other welfare policies that were phased in at the same time as the WFTC, including the Childcare Credit and the New Deal for Lone Parents.

Overall, the available micro-level studies find beneficial effects of in-work benefits on female employment, although these effects are typically sizeable only for single mothers.

Multiplier Effects

While legislation on family policies typically defines eligible groups, based for example on age or cohort of children and/or family income, social interactions across peers may affect policy take-up as well as labor market outcomes beyond the targeted population. Failing to recognize such interdependencies (if any) has implications for the evaluation of policy interventions, as outcomes for control groups may be contaminated by interactions with the treatment group, and a comparison of outcomes between treatment and control groups would yield a downward-biased estimate of policy impact.

Contamination from the treated to the nontreated population may result from sharing information about the characteristics and availability of a certain family policy, or learning about its effects on outcomes for one's peers. Progress in the identification of these effects has been limited by well-known problems. There is the "reflection problem" of drawing an inference about how one group affects another, so-called because when two groups are changing at the same time—like your reflection changes in the mirror when you move—figuring out how one group affects the other is problematic (Manski 1993). Similar issues arise because of correlated unobservable characteristics, endogenous group membership, and the difficulties of observing peer groups in available data (Dahl, Løken, and Mogstad 2014).

In Norway, after the introduction of one extra month of parental leave for fathers of children born after April 1, 1993, Dahl, Løken, and Mogstad (2014) sought to estimate peer effects in the take-up rate of parental leave. Before 1993, parents could share without restrictions a parental leave of given length, while the introduction of an exclusive "daddy month" was supposed to promote gender equality in the household. Taking advantage of the timing of the reform, the authors identify peer effects in the take-up rate of parental leave among both workplace and family networks using a regression discontinuity design. Their results show evidence of substantial peer effects in both networks, as co-workers and brothers of fathers induced to take-up leave by the reform became 10 and 15 percentage points more likely, respectively, to take paternity leave out of the (nonexclusive) parental quota. These effects—which are amplified over time—are interpreted as evidence of transmission of information about costs and benefits of paternal leave, eventually leading to a redefinition of social norms about parents' roles. Welteke and Wrohlich (2016) highlight similar spillover effects in the take-up rate of parental leave among mothers following the 2007 German reform discussed above. Future research should continue looking for potential peer effects and network effects in other programs.

Comparing Lessons from Macro and Micro Studies

While both the macro and micro literatures tend to find overall positive effects of subsidized child care on female employment, the discussion above illustrates that no obvious consensus emerges from the literature that has studied the labor market impact of parental leave rights and benefits. Cross-country studies, with weaker identification, point to a positive correlation with maternal employment rates, albeit this effect is limited to short or intermediate leave durations, and mostly applies to less-skilled women, with virtually no impact for the more educated. Extremely long leave durations seem instead to have inhibitive effects. On the other hand, studies on microdata tend to find that parental leave mostly delays return to work, with no discernible effects on employment rates in the long run.

Several factors may potentially explain such discrepancies. First, the beneficial impact of maternity leave may be overestimated in cross-country studies insofar as exogenous shocks to female labor force participation induce family-friendly legislation via political support and/or changes in social norms towards gender roles. Second, Ruhm (1998) does detect positive effects of short-lived parental leave (around three months), and our more recent estimates find positive effects up to one year and negative effects afterwards, but widespread extensions to leave rights in most countries have inevitably shifted the focus of later studies based on microdata towards variations in parental leave at much longer durations, up to three years. Thus, it might be possible that the availability of some job protection, relative to no protection at all, would ensure continuity of employment and discourage transitions out of the labor market, while further extensions would simply delay return to work without further gains in employment. Third, cross-country studies often provide joint perspectives on various family-oriented policies, and Blau and Kahn (2013) find that parental leave has no significant impact on female participation once other policy instruments are controlled for. On the other hand, micro-level studies that compare employment outcomes for treated and nontreated parents may underestimate the impact of policy reforms if the behavior of treated parents induces changes in the behavior of the nontreated via learning, imitation, and other spillover mechanisms.

Conclusions

What can we learn from the evolution of family policies across high-income economies? It is a complex tale in which changing economic, cultural, and political economy considerations appear to shape (and be shaped by) these policies. No obvious consensus on the labor market impact of parental leave rights and benefits emerges from the empirical literature. Although there are some exceptions, it seems a fair summary that cross-country studies tend to find more positive effects on female employment than micro-level studies for relatively short leave durations, and more negative effects for longer entitlements. Employment and earnings impacts tend to be more beneficial for the less skilled, possibly with a detrimental impact on the

earnings of high-skill women. In a nutshell, there is little compelling evidence that extended parental leave rights have an overall positive effect on female outcomes.

The policies with the strongest evidence for reducing gender disparities seem to be early childhood spending (in both cross-country and microdata) and in-work benefits (in the microdata). A potential common theme here is that making it easier to be a working mother may matter more than the length of leave or the payments that new parents receive while out of the labor force.

The United States has been an outlier in the adoption of family policies across high-income countries since the turn of the twentieth century. As Goldin and Mitchell argue in this symposium, the female labor force participation in the US has evolved into a pattern with very high rates of employment early in the life cycle, which then sharply decline with motherhood, which is being progressively delayed. The cross-country and micro-level evidence has not found an overall strong connection between maternity leave and female labor force participation. But the relatively short leave entitlements available to mothers in the United States may possibly contribute to this life-cycle pattern of delaying motherhood, with persistently low rates of labor force participation for women in their 30s and 40s.

References

Adema, Willem, Nabil Ali, and Olivier Thévenon. 2014. "Changes in Family Policies and Outcomes: Is There Convergence?" OECD Social, Employment and Migration Working Paper 157. http://dx.doi.org/10.1787/5jz13wllxgzt-en.

Anthony, Deborah J. 2008. "The Hidden Harms of the Family and Medical Leave Act: Gender Neutral versus Gender Equal." *Journal of Gender, Social Policy and the Law* 16(4): 459–501.

Asai, Yukiko. 2015. "Parental Leave Reforms and the Employment of New Mothers: Quasi-Experimental Evidence from Japan." *Labour Economics* 36: 72–83.

Baker, Michael, and Kevin Milligan. 2008. "How Does Job-Protected Maternity Leave Affect Mothers' Employment?" *Journal of Labor Economics* 26(4): 655–691.

Baum II, Charles L. 2003. "The Effect of State Maternity Leave Legislation and the 1993 Family and Medical Leave Act on Employment and Wages." *Labour Economics* 10(5): 573–96.

Blau, Francine D., and Lawrence M. Kahn. 2013. "Female Labor Supply: Why Is the United States Falling Behind?" *American Economic Review* 103(3): 251–56.

Bertrand, Marianne, Patricia Cortés, Claudia Olivetti, and Jessica Pan. 2016. "Social Norms, Labor Market Opportunities, and the Marriage Gap for Skilled Women." NBER Working Paper 22015, February.

Bettendorf, Leon J. H., Egbert L. W. Jongen, and Paul Muller. 2015. "Childcare Subsidies and Labour Supply—Evidence from a Dutch Reform." *Labour Economics* 36: 112–23.

Blundell, Richard, Alan Duncan, Julian McCrae, and Costas Meghir. 2000. "The Labour Market

Impact of the Working Families' Tax Credit." *Fiscal Studies* 21(1): 75–104.

Blundell, Richard, and Hilary W. Hoynes. 2004. "Has 'In-Work' Benefit Reform Helped the Labor Market?" In *Seeking a Premier League Economy: The Economic Effects of British Economic Reforms, 1980–2000*, edited by David Card, Richard Blundell, and Richard B. Freeman, pp. 411–60. University of Chicago Press.

Brewer, Mike, Alan Duncan, Andrew Shepard, and Maria Jose Suarez. 2006. "Did Working Families' Tax Credit Work? The Impact of In-Work Support on Labour Supply in Great Britain." *Labour Economics* 13(6): 699–720.

Brewer, Mike, Marco Francesconi, Paul Gregg, and Jeffrey Grogger. 2009. "Feature: In-Work Benefit Reform in a Cross-National Perspective—Introduction." *Economic Journal* 119(535): F1–F14.

Cascio, Elizabeth U. 2009. "Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools." *Journal of Human Resources* 44(1): 140–70.

Cascio, Elizabeth, and Diane Whitmore Schanzebach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." *Brookings Papers on Economic Activity*, no. 2 (Fall): 127–92.

Cipollone, Angela, Eleonora Patacchini, and Giovanna Vallanti. 2014. "Female Labour Market Participation in Europe: Novel Evidence on Trends and Shaping Factors." *IZA Journal of European Labor Studies* 3(1): 1–40.

Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad. 2014. "Peer Effects in Program Participation." *American Economic Review* 104(7): 2049–74.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes. 2016. "What is the Case for Paid Maternity Leave?" *Review of Economics and Statistics* 98(4): 655–70.

Eissa, Nada, and Hilary Hoynes. 2004. "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit." *Journal of Pubic Economics* 88(9–10): 1931–58.

Ekberg, John, Rickard Eriksson, and Guido Friebel. 2013. "Parental Leave—A Policy Evaluation of the Swedish 'Daddy-Month' Reform." *Journal of Public Economics* 97: 131–43.

European and World Values Surveys Four-Wave Integrated Data File, 1981–2004, v.20060423. 2006. Aggregate File Producers: Análisis Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain/Tilburg University, Tilburg, The Netherlands. Data Files Suppliers: Analisis Sociologicos Economicos y Politicos (ASEP) and JD Systems (JDS), Madrid, Spain/Tillburg University, Tillburg, The Netherlands/Zentralarchiv fur Empirische Sozialforschung (ZA), Cologne, Germany. Aggregate File Distributors: Análisis

Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain/Tillburg University, Tilburg, The Netherlands/Zentralarchiv fur Empirische Sozialforschung (ZA) Cologne, Germany. http://www.europeanvaluesstudy.eu/page/integrated-values-surveys-1981-2008.html.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2015. *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0.* [Machine-readable database]. Minneapolis: University of Minnesota. http://doi.org/10.18128/D030.V4.0.

Fortin, Nicole M. 2005. "Gender Role Attitudes and the Labour-Market Outcomes of Women across OECD Countries." *Oxford Review of Economic Policy* 21(3): 416–38.

Francesconi, Marco, and Wilbert van der Klaauw. 2007. "The Socioeconomic Consequences of 'In-work' Benefit Reform for British Lone Mothers." *Journal of Human Resources* 42(1): 1–31.

Gauthier, Anne. H. 2011. Comparative Family Policy Database, Version 3 [computer file]. Netherlands Interdisciplinary Demographic Institute and Max Planck Institute for Demographic Research (distributors). http://www.demogr.mpg.de/cgi-bin/databases/fampoldb/index.plx.

Gelbach, Jonah B. 2002. "Public Schooling for Young Children and Maternal Labor Supply." *American Economic Review* 91(1): 307–322.

Geyer, Johannes, Peter Haan, and Katharina Wrohlich. 2015. "The Effects of Family Policy on Maternal Labor Supply: Combining Evidence from a Structural Model and a Quasi-Experimental Approach." *Labour Economics* 36: 84–98.

Givord, Pauline, and Claire Marbot. 2015. "Does the Cost of Child Care Affect Female Labor Market Participation? An Evaluation of a French Reform of Childcare Subsidies." *Labour Economics* 36: 99–111.

Goldin, Claudia. 1988. "Maximum Hours Legislation and Female Employment: A Reassessment." *Journal of Political Economy* 96(1): 189–205.

Gordon B. Dahl, Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes. 2016. "What is the Case for Paid Maternity Leave?" *Review of Economics and Statistics* 98(4): 655–70.

Gregg, Paul, Susan Harkness, and Sarah Smith. 2009. "Welfare Reform and Lone Parents in the UK." *Economic Journal* 119(535): F38–65.

Haeck, Catherine, Pierre Lefebvre, and Philip Merrigan. 2015. "Canadian Evidence on Ten Years of Universal Preschool Policies: The Good and the Bad." *Labour Economics* 36: 137–57.

Han, Wen-Jui, Christopher Ruhm, Jane Waldfogel, and Elizabeth Washbrook. 2009. "Public Policies and Women Employment after

Childbearing." NBER Working Paper 14660.

Harris, Henry J. 1919. "Maternity Benefit System in Certain Foreign Countries." US Department of Labor, Children's Bureau. Legal Series No. 3, Bureau Publication No. 57, Washington, Government Printing Office.

Havnes, Tarjei, and Magne Mogstad. 2011. "Money for Nothing? Universal Child Care and Maternal Employment." *Journal of Public Economics* 95(11): 1455–65.

Hotz, V. Joseph, and John K. Scholz. 2003. "The Earned Income Tax Credit." In *Means Tested Transfer Programs in the United States*, edited by Robert A. Moffitt, 141–97. University of Chicago Press.

Kluve, Jochen, and Sebastian Schmitz. 2014. "Social Norms and Mothers' Labor Market Attachment: The Medium-run Effects of Parental Benefits." Ruhr Economic Papers 481.

Kluve, Jochen, and Marcus Tamm. 2013. "Parental Leave Regulations, Mothers' Labor Force Attachment and Fathers' Childcare Involvement: Evidence from a Natural Experiment." *Journal of Population Economics* 26(3): 983–1005.

Lalive, Rafael, Analía Schlosser, Andreas Steihauer, and Josef Zweimüller. 2013. "Parental Leave and Mother's Careers: The Relative Importance of Job Protection and Cash Benefits." *Review of Economic Studies* 81(1): 219–65.

Lalive, Rafael, and Josef Zweimüller. 2009. "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments." *Quarterly Journal of Economics* 24(3): 1363–1402.

Lefebvre, Pierre, and Philip Merrigan. 2008. "Child-Care Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada." *Journal of Labor Economics* 26(3): 519–48.

Lefebvre, Pierre, Philip Merrigan, and Matthieu Verstraete. 2009. "Dynamic Labour Supply Effects of Childcare Subsidies: Evidence from a Canadian Natural Experiment on Low-Fee Universal Child Care." *Labour Economics* 16(5): 490–502.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets.* Princeton University Press.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60(3): 531–42.

Nichols, Austin, and Jesse Rothstein. 2016. "The Earned Income Tax Credit." Chap. 2 in *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt. University of Chicago Press.

Nollenberger, Natalia, and Núria Rodríguez-Planas. 2015. "Full-Time Universal Childcare in a Context of Low Maternal Employment: Quasi-Experimental Evidence from Spain." *Labour Economics* 36: 124–136.

OECD. 2014. OECD Social Expenditure Database (SOCX), https://www.oecd.org/social/expenditure.htm.

OECD, Social Policy Division-Directorate of Employment, Labour and Social Affairs. 2016. "PF2.5. Trends in parental leave since 1970," and "PF2.5 Annex: Detail of Change in Parental Leave Policy, pp. 1–49" (Annex last updated 01/07/2014), in the OCED Family Database, http://www.oecd.org/els/family/database.htm.

Raute, Anna. 2015. "Can Financial Incentives Reduce the Baby Gap? Evidence from a Reform in Maternity Leave Benefits." Unpublished paper.

Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel. 2013. "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes." *Journal of Policy Analysis and Management* 32(2): 224–45.

Ruhm, Christopher J. 1998. "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *Quarterly Journal of Economics* 113(1): 285–317.

Schönberg, Uta, and Johannes Ludsteck. 2014. "Expansions in Maternity Leave Coverage and Mothers' Labor Market Outcomes after Childbirth." *Journal of Labor Economics* 32(3): 469–505.

Shim, Joyce. 2014. "Family Leave Policy and Fertility Rates in OECD Countries including East Asia." Paper presented at the Population Association of America 2014 Annual Meeting, Boston, MA. http://paa2014.princeton.edu/papers/140973.

Thévenon, Olivier, and Anne Solaz. 2012. "Labour Market Effects of Parental Leave Policies in OECD Countries." OECD Social, Employment and Migration Working Papers 141.

Welteke, Clara, and Katharina Wrohlich. 2016. "Peer Effects in Parental Leave Decisions." IZA Discussion Paper 10173.

Wilkander Ulla, Alice Kessler-Harris, and Jane Lewis, eds. 1995. *Protecting Women: Labor Legislation in Europe, the United States and Australia, 1880–1920.* University of Illinois Press.

How to Write an Effective Referee Report and Improve the Scientific Review Process

Jonathan B. Berk, Campbell R. Harvey, and David Hirshleifer

The review process for academic journals in economics has grown vastly more extensive over time. Journals demand more revisions, and papers have become bloated with numerous robustness checks and extensions (Ellison 2002b; McAfee 2010; Spiegel 2012; Harvey 2014). For example, Card and DellaVigna (2013) document that recent published papers are on average three times longer than in the 1970s—a total which excludes the many robustness checks now contained in online appendices. Even if the extra resulting revisions do on average lead to improved papers—a claim that is debatable—the cost is enormous. We argue that much of the time involved in these revisions is a waste of research effort.

The leading explanation for the expansion of the review process, as modeled by Ellison (2002a), is that social norms have evolved toward increased demands for revisions. We believe that part of the explanation for why the profession has evolved to this equilibrium is that referees feel the need to demonstrate their intelligence or industriousness to editors by identifying problems in papers. The result is that in many cases reviewers inflate minor blemishes to the status of major flaws.

■ *Jonathan B. Berk is A. P. Giannini Professor of Finance, Graduate School of Business, Stanford University, Stanford, California. Campbell R. Harvey is Professor of Finance at the Fuqua School of Business, Duke University Durham, North Carolina. David Hirshleifer is Professor of Finance and Merage Chair in Business Growth, Paul Merage School of Business, University of California-Irvine, Irvine, California. Their email addresses are jbberk@ stanford.edu, cam.harvey@duke.edu, and david.h@uci.edu.*

Another cause for concern is the level of disagreement amongst referees. As Welch (2014) documents in a study of eight prominent journals in economics and finance, in cases with multiple referees on the same paper, the probability that one referee recommends inviting a revision rather than rejecting, conditional on another referee doing so, is only marginally higher than the unconditional probability. Furthermore, the correlation of referee recommendations at a major finance conference was only 0.28. This pattern of disagreement suggests a high level of arbitrariness in the review process.[1]

To identify and highlight what is going right and what is going wrong in the reviewing process, we wrote to a sample of former editors of the *American Economic Review*, the *Journal of Political Economy*, the *Quarterly Journal of Economics, Econometrica*, the *Review of Economic Studies* and the *Journal of Financial Economics*, and asked them for their thoughts about what might improve the process. We found a rough consensus that referees for top journals in economics tend to make similar, correctable, mistakes. The italicized quotations throughout this paper are drawn from our correspondence with these editors and our own experience. Their insights are consistent with our own experiences as editors at the *Journal of Finance* and the *Review of Financial Studies*. Our objective is to highlight these mistakes and provide a roadmap for how to avoid them.

This article is not a comprehensive "how-to guide" for refereeing. Readers interested in such a guide can consult Berk, Harvey, and Hirshleifer (2015), or the essays in this journal by Hamermesh (1992, 1994). Instead our objective is to draw attention to specific shortcomings in the process that we believe can and should be improved. We begin by discussing what we see as three central current problems of refereeing in economics: 1) giving appropriate weight to the importance and innovativeness of a paper (and not being distracted that such papers may have minor flaws); 2) drawing a clear and explicit separation between comments that must be dealt with to make the paper publishable and suggestions that do not affect the publication decision and are therefore optional; and 3) living up to the implicit contract involved in a revise-and-resubmit process. We then offer thoughts on some other topics involved in refereeing: conflicts of interest, cover letters, politeness, and acting ethically.

Importance: The Hardest Decision

Perhaps the hardest part of a referee's job is assessing the *importance* of the paper, which involves judgments about whether a paper is of sufficiently broad interest and offers a sufficiently original contribution to be appropriate for the journal to which is has been submitted. Submissions exceed the available space in all A-level and most B-level journals; there are plenty of "correct" papers that do

---

[1] Similar findings are reported in a study of 823 submissions to the *Journal of Applied Psychology* by Gilliland and Cortina (1997). In an earlier study, Laband (1990) tries to measure the value added by referees by analyzing the relation between reviewer comments and citations.

not make a significant enough contribution to existing knowledge to appear in a top-tier journal. The editor needs to assess the importance of the contribution. Thus, a referee report should contain a scientifically based argument that explains the referee's assessment of the importance of the work and details the considerations that bear upon your judgment. Only an argument that is scientifically based is useful and informative to an editor in making the decision.

> *I quickly learned that there were three critical determinants of whether a paper was publishable in the Journal: (i) whether it addressed a question of sufficiently broad interest; (ii) whether it made a sufficient leap over existing literature; and (iii) whether the analysis was correct. Since most of the papers I received were "correct", my decision often boiled down to considering how well the paper fared according to the first two criteria. … I often felt that if referees more clearly understood the factors that make a paper publishable, they would organize the reports accordingly, making them more useful overall.*

The importance of a contribution can be undervalued in some cases by referees and editors. After all, papers that are more ambitious are often more likely to have loose ends, which gives referees and editors a reason to avoid taking a chance on them. Indeed, Arrow (1995) pointed out: "I think the publication selection procedure at the major journals has become methodologically more conservative, more given to preferring small wrinkles in existing analysis to genuinely new ideas. This conservative tendency also appears in the allocation of grants by government agencies and in faculty appointments and promotions."

Thus, we offer this advice to referees: Do not dismiss papers that attack larger issues merely because flaws can be found. The important question that you need to assess is whether the flaws actually invalidate the contribution. If the flaws do not rise to this level and you judge the contribution to be important enough to warrant publication, then you should recommend publication. All papers have flaws, and no amount of revision removes all uncertainties. There is always need for further research to provide deeper perspectives. Try to ask yourself the following question: Flaws and all, would I have been pleased to have written such a paper? If yes, that gives a strong hint that it should be strongly considered for publication, flaws and all.

Separating the Essential and the Suggested: The Highest-Cost Mistake

All papers have weaknesses, and catching a serious problem can be an indicator of referee quality. As a result, referees have a perverse incentive to persuade the editor that their own intellectual work is of high quality and that they are smart by pointing to minor blemishes and trumpeting them as serious problems. The distinction between these two categories may not be obvious to an editor who is not a specialist in the area of the paper: for discussion of this dynamic, see the signal-jamming model of Hirshleifer (2015). So by pointing out these supposedly

critical problems and requiring authors to address them, referees believe they are improving their reputation with the editor.

> *Some younger referees feel that they need to be overwhelmingly negative about everything in a paper in the report to the author to prove their own mettle and critical insights. They don't appreciate the need to at least demonstrate that they have read the paper and to provide some kind comments indicating some understanding of what the author is trying to do.*

Often, the ultimate outcome of such signal-jamming is an unwieldy and lengthy paper. This raises the question of why signal-jamming would be more important now than in the past. One possible reason is that growth in the profession makes it hard for an editor to have an independent assessment from personal contact of the quality of referees—even those at top schools. This increases perceived incentives to boost reputation by other means. Also, growth in the body of knowledge needed to evaluate papers makes it harder for editors to see through signal-jamming efforts (though in our experience, signal-jamming referees are often not nearly as subtle as they think they are).

Perhaps more importantly, signal-jamming pressure may cause social norms about standards for publishability to evolve over time toward increased demands for revisions, in the spirit of the model of Ellison (2002a). When referees engage in signal-jamming, more ambitious and potentially more important papers tend to lose in competition with routine papers in the review process because of a standard that effectively demands, as a prerequisite for publication, the virtual absence of minor shortcomings. The emphasis on superficial perfection over substantive importance may have the disastrous side effect of reducing the incentive to engage in innovative research.

> *I handled a paper in which the reviewer offered micromanagement of how the paper should be written, and made a specific demand for how the paper should be repositioned—in a way that highlighted a paper of the reviewer's. As for cases in which reviewers demanded numerous unimportant extensions or robustness checks—this was so common that specific examples are not especially noteworthy.*

One key step to reduce dysfunctional signal-jamming is that a referee report should divide comments into two clearly demarcated sections: 1) problems that make the paper unpublishable, which (if revision is invited) must be addressed before the paper is publishable; and 2) problems that are not essential for the publishability of the paper, which should labeled as "suggestions." From our perspective, perhaps the most common and most crucial mistake most referees make when providing a revise and resubmit recommendation is not clearly distinguishing between suggestions that are required for publication and ones that are not.

From the referee's perspective, the key difference between these two categories is the amount of justification that is required in the report. Statements about

problems that must be addressed before publication require rigorous justification; others do not. More specifically, must-address statements require a *scientifically convincing* argument for why these problems currently render the paper unpublishable. Of course, the same requirement may lead to a recommendation that the paper be rejected. A referee is *obligated* to provide a scientifically based argument for why a perceived problem renders the paper unpublishable. The argument needs to be clear and understandable to the editor (and authors). After all, the goal is for this argument to be used in the editor's decision.

It is surprising how many reviewers have a highly refined olfactory sense; several editors write about handling empirical papers in which a referee comments on the "smell" of the results. A referee hunch that theoretical or empirical results are wrong is not a scholarly argument against a paper's conclusions. Instead, as referee you must provide scientifically based reasons *why* you suspect the results are not correct.

> *A theory paper is submitted to the Journal and a senior referee recommends rejection because they "do not believe the proof". This is insufficient. While it is expected that the referee should be skeptical, the referee needs to provide a solid foundation as to why the proof is incorrect. Hunch is not sufficient.*

If there is a mistake in the proof, the appropriate recommendation is often rejection, though in some cases a mistake is fixable. A mistake can also be identified by finding a counterexample to a claim. However, if the authors demonstrate (correctly) an error in your counterexample, then their proposition stands. It is not grounds for rejection to retreat into generalized skepticism. A revise-and-resubmit for a theoretical paper can also ask the authors to explain how they reconcile their results with existing findings, or the extent to which their result may depend on nonstandard assumptions. But remember that it is possible that the results in the paper are right and it is the previous literature that got it wrong.

When it comes to empirical papers, a Bayesian should be skeptical of a conclusion that has a low prior probability—more evidence needs to be brought to bear, and greater validation of the evidence is required, to be persuasive. But a referee cannot dismiss evidence simply because he or she finds the results to be surprising. Obviously, surprising results have a greater probability of being wrong, so the appropriate response is to request confirming evidence. You can ask the authors to explain how they reconcile counterintuitive or unexpected empirical results with existing findings. You can also evaluate the robustness checks that the authors have done and consider whether it might be important to perform a limited number of other robustness checks. But in making such comments, the explicit line between what is necessary and what is suggested should be preserved. If the author satisfies these additional requirements, then the referee should leave it to subsequent research to evaluate whether existing preconceptions need to be updated based on the paper's results.

> *An empirical paper is submitted to the journal and the referee recommends rejection because "the empirical results do not pass the smell test". While there are plenty of reasons that the result may not be solid, the referee needs to ask questions or request specific robustness tests. "Smell" is an insufficient reason for rejection.*

The review process for academic journal articles should sharply focus on what is essential. This focus allows authors greater freedom to write papers in the way that they think is best. All else equal, scholars should be able to develop their ideas as they see them. On issues that are not essential for publishability, there is no presumption that authors and referees need to see eye to eye. As a referee, it is important to keep in mind that unless comments about secondary issues are carefully labeled as such, such comments are inherently coercive. Few authors have the luxury of risking rejection of a project in a top journal for the sake of skipping the nth robustness check mentioned by a referee, or of expressing their ideas in the way that the author thinks is most effective but the referee does not

The Implicit Bargain in a Revise-and-Resubmit

As a referee, it is important to keep in mind the implicit deal you are making with the author when you recommend a revise and resubmit: if the author satisfactorily addresses the issues that you have raised, you will recommend publication. When you make a revise-and-resubmit recommendation, you are actually making three statements: 1) the paper is of sufficient importance in terms of scope and findings that you believe it is suitable for the journal at hand; 2) there are problems with the paper that currently make it unpublishable in its current form; and 3) these problems are correctable.

In short, you are helping the editor provide a road map to publication. Keep in mind that the editor's road map will require heavy investment of time by the authors and usually relies heavily on referee advice. It is also an implicit contract. If the authors address these requests in a satisfactory way, then the editor will likely accept the paper. This should be the case even if other good ideas for improvement occur to the referees in the next round. Avoid putting the editor in the very unfortunate position of discovering that there are serious problems with the paper that were not noticed in the first round. It is very important that your own ideas for improvement be as refined as you can make them in the *first* round, so that the editor is positioned to offer a useful and reliable road map.

In a revise-and-resubmit recommendation, there will likely be other problems with the paper that are not severe enough to render the paper unpublishable. For such other problems, you do not as a referee need to provide detailed reasons (or a scientifically based argument) for your opinion. However, if the author chooses not to address these problems, you cannot use that lack of reaction as a basis for recommending rejection of the revised draft. In many cases, reasonable people can disagree about what should and should not go into a given paper. Ultimately, the

author's name goes on the paper, not the referee's name, so the decision on how best to write the paper is the author's. Do not hold a submission hostage because, in your (undoubtedly wise) opinion, the paper could be better written.

> *I handled a paper where the referee provided a thoughtful report which led to a revise and resubmit. The author addressed the comments and resubmitted. On the revision, the referee produced a new, long list of additional comments—each of which could have been detailed in the initial report.*

In making requests of authors, weigh the costs of the request. It is not enough that a particular request will improve the paper. The benefits must exceed the costs, so that the improvement has positive net present value. Since the author bears the costs, it is easy for a referee to make absurd demands thoughtlessly. Don't.

> *I receive a positive referee report on a paper that uses 19 years of hand-collected data. The earliest data was the most time consuming to load because the authors had to visit the archives and deal with paper documents. The referee insists that the data be expanded back to 20 years. This 20th year was not a special year but it was clearly going to be punitive for the authors to collect this extra year of data. In addition, the extra year would unlikely be influential for the results. This is an example of "make work" and as Editor it is often very difficult to separate the work that must be done from the frivolous work. In this case, it was easy and I instructed the authors to ignore the referee's comment. I never called on that referee again.*

By following these suggestions, the revise-and-resubmit process should generally take just one round before reaching a final decision on publication, and that is how referees should view the process. Of course, sometimes a second round may be required for unavoidable reasons. For example, perhaps the authors only partially addressed the necessary changes specified in the first round. In such cases, referees and editors should pause and consider the wisdom of going another round rather than just recommending that the paper be rejected. A related possibility is that in an intelligent and honest effort by authors to address the comments of the first round, unforeseen issues have unavoidably arisen. Finally, cases may arise in which a request for a second revision is based on an issue that should have been apparent on the first round. In this case, the referee has made a mistake, and the referee should admit this oversight to the author and to the editor. Such requests present the editor with a hard choice, and should be rare.

> *I get a positive recommendation from a referee with a list of items for the authors to take care of in a revision. I issue a R&R [revise and resubmit] and the paper comes back with the authors making a serious effort to revise the paper and they address each of the referee's comments. The referee declines to provide a report on the resubmission but writes to me saying that he recommends rejection because of "insufficient incremental contribution". The rejection had nothing to do with the author responses. The referee changed his*

mind with no reasoning. If there was insufficient incremental contribution, that case must be made in the initial submission.

In summary, if the author has appropriately satisfied the key requests of a previous revise-and-resubmit referee report, the paper should then be accepted. Do not invent a new set of requests—or even worse, reject capriciously.

Other Advice

Thus far we have highlighted the mistakes by referees and editors that impose the highest costs on authors. In this section, we will briefly cover a few other important lessons uncovered in our own experience as editors and in our correspondence with other editors.

Declining the Invitation

Upon receiving a referee request, decide immediately whether you will be able to complete the review within the allotted time. If not, respond immediately with suggestions for alternative reviewers. Often the editor is not a specialist in the specific area, and does not know exactly who the best match for a submission is, so thoughtful suggestions for other reviewers are valuable. Also, editors understand that peak loads can create the occasional need for a declined invitation or a request for an extended deadline—or at least greatly prefer this to having to deal with a severely late report, or a systematically derelict referee.

Agreeing to do a report at the requested deadline and then never doing the report on time and never responding to reminders or queries about the report me or the editorial office. In many of these cases, the individual did not have time to do the report but felt compelled to say they would do it and embarrassed to respond. It is much better for all involved to upfront decline a request immediately then to say yes when it won't be feasible to do the report in a timely manner. You lose a lot more face in my view from going into hiding and not responding or doing reports much later than promised, than from being upfront and saying you can't do the report at all.

Try to form an objective assessment about whether you are a good match for the submission. If you know little about the relevant literature, it might be wise to contact the editor to make sure that a mistake has not been made. The editor may have had a good reason for selecting you, and verifying this can make your job more useful and focused. For example, in a paper with both theory and empirical work, the editor might be seeking the views of the empiricist for the empirical sections and of a theorist for the theory sections.

If you have reviewed the paper for another journal, again immediately alert the editor. Some editors would prefer getting a fresh view. Let the editor make

that decision. Other editors may be satisfied that the referee can assess effectively whether the paper has improved.

You might also feel that you cannot complete the report anonymously. For example, you may have discussed the paper and feel your report would be so close to the discussion it will be obvious to the author that you are the referee. If you wish to maintain your anonymity, it is reasonable to decline to referee the paper. If you do not mind the loss of anonymity, you should still alert the editor, since nonanonymous reviewing can create agency problems. You may also provide advice to the editor in a letter without a report. Your views will be important even when they cannot be passed on to the author, though such service is not as useful to the editor as a formal report. In this case, the same rules apply. If you feel the paper is not publishable you must provide a scientifically based argument that will allow the editor to make an informed judgment.

If you are already working on a paper that overlaps substantially with the submission, you should notify the editor. As discussed further below, this is a potential conflict of interest and might be a reason to recuse yourself.

Finally, if a paper is obviously far below the bar at the journal where it has been submitted, a short (one-page) report is perfectly acceptable. It should be straightforward to provide a scientifically based justification. If it is not, you should reconsider whether the paper really is that far below the bar. In this case, impress the editor by returning your report within a week—not by spending time on a lengthy analysis of what, by hypothesis, is an obvious conclusion that would be shared by the overwhelming majority of referees. If the editor disagrees with your assessment, the editor can then turn to another referee without undue delay.

The Cover Letter

The cover letter should contain an assessment of the contribution of the paper along with very concise reasoning supporting your recommendation. The editor wants to know the positives as well as the negatives of a paper. The cover letter should be brief. It should *not* be a cut and paste of the referee report, which is a waste of the editor's time. In addition, your cover letter and recommendation to the editor should be consistent with your report. Otherwise you impose on the editor the task of explaining the discrepancy to the author. This will irritate both of them.

> *I assigned a paper by a well-known author to a particular referee. I had a very low prior on the paper. The referee submits the report which is amazingly positive using words like "pathbreaking", "important insight", "game-changer", "will change the way people think about a problem", etc. I do notice that the referee has checked off "reject" in the electronic system—which I assumed must have been a mistake. However, after reading the cover letter, I understand. The referee thinks the paper is a poor idea and is nowhere near the hurdle for the journal. I reject the paper and the author is furious and immediately appeals the decision quoting the very positive referee report. Such inconsistencies between recommendations and the report frustrate both authors and editors—and create a lot of extra work for the editors.*

The ideal cover letter should succinctly provide three types of information. First, it should include a statement of the broad interest and importance of the paper's contribution relative to existing work. Keep in mind that the editor may not be an expert in this subfield, and it is often hard to figure out the paper's main point or line of reasoning. Second, is the analysis convincing? Third, the cover letter should offer a frank assessment as to whether the paper is publishable as it stands, or whether the paper is likely to be publishable within one round of revision.

Be decisive. As a referee, you are being asked to make a recommendation: accept, revise, or reject. The reasons given for the recommendation are at least as important as its summary value. But it is much easier for the editor to understand which considerations the referee considers most important when a summary value is provided. For example, if you are recommending a revise-and-resubmit, you must give a sound reason, or the editor will likely reject the paper.

It is fine to mention reasons for uncertainty about the recommendation. If the paper is somewhat outside your area, you might suggest that a second opinion be sought, and you should provide names of candidate referees, and if possible, what specific issues the alternative referee can address that you felt were outside your area of expertise. But a request for a second opinion should not be used as a cover for indecisiveness or not wanting to spend a lot of time studying details.

Length of the Referee Report

A referee report is not a mind-dump about the paper. Polish your report with an eye to condensing. It is almost always possible with an extra revision by the reviewer to make the report both shorter and more incisive.

The main purpose of the report is to help the editor decide whether to publish. It can also be extremely valuable to have extra material to help the authors improve the paper, but only when presented in a way that does not interfere with the main purpose. So if comments about minor details are provided, they must be segregated into a clearly marked separate section that is easy for the editor to skip. Broad philosophical issues or rhetorical flourishes should be eliminated if they have no bearing on improving the paper or on determining whether the paper should be published. Finally, our suggested format of sharply separating comments which affect publishability from those which are just to improve the paper will be very helpful for the editor in making a decision efficiently and accurately.

> *Overly long referee reports are a burden on editors. Unless a referee needs to make extremely technical points, 2–3 pages should be sufficient. Going beyond this raises the likelihood of coercion/overburdening. As an editor, I view a 10-page referee report as punitive.*

We would not go so far as this editor to endorse a blanket principle that 2–3 pages is the right length for almost all papers, but brevity is valuable. Indeed, brief reports often contain more important content than lengthy ones. We have seen reports that are too minimal in both length and thought, but the more

common problem is lengthy reports that are not accompanied by correspondingly extensive insight.

A more mechanical point, but one that has a surprisingly large effect on the efficiency of the review process, is that comments should be numbered. A report consisting mainly of discursive undifferentiated paragraphs is a burdensome disservice to all involved in the review process. Insightful discussion is a plus, but not at the expense of a clear bottom line that includes a numbered list of suggested actions or problems. We suggest separate numbering for category 1 of key problems and category 2 of suggestions.

Ethics

A referee who has any conflict of interest with the manuscript must alert the editor—promptly, and *before* agreeing to accept the assignment. The editor might decide to find a new referee or might ask the original referee to complete the report. Conflicts that require alerting the editor include when an author of the paper is: a past (over the past five years), current, or planned coauthor; a current colleague; a former student or advisor; a close personal friend or family member; or a person with a financial relationship with the reviewer. Conflicts can also arise if you have current research that is competing with the research in the submitted paper. If the paper contradicts or corrects your research, you need to tell the editor. If there have been disputes between you and one of the authors in the past, alert the editor.

Manipulation by referees of recommendations and reports for the purpose of advancing or holding back publication with the goal of advancing their own work is unethical. One such manipulation is for a referee to focus on adding citations to the referee's own work.

> *I received a paper that I had a low prior on that cited the work of a potential reviewer. I assigned this person as the referee and the reviewer enthusiastically recommended a revise and resubmit. The report made reference to an extraordinary six extra citations of the reviewer's work. Essentially, the reviewer wanted to increase the cite count from two to eight in the paper. After carefully reading the paper, I rejected the manuscript.*

In addition, it is not acceptable to hold up the reviewing process with the goal of buying time for your own related research agenda.

> *I assigned a paper to a referee on a hot current topic. I sent many reminders to the referee because the report was very late. I then notice the referee posts a competing paper on SSRN. The competing paper cites the paper in review (which was also public on SSRN). I withdraw the referee request but the referee has successfully delayed the review process for the original paper. While it seems like the referee was successful, that was not the case. I alerted the Editors of the other top journals. I never saw the referee's paper in print.*

Of course, it is inevitable that a referee will sometimes get an idea based on reading a submitted paper for how to write another paper on the same broad topic. If the author has not distributed the paper publicly, and if you have already agreed to review a paper, you cannot write your own paper on the topic, as you are not in a position to reference the submitter's prior work appropriately.

In contrast, if a paper is publicly distributed, you are free to work on the same topic (just as everyone is), as long as you clearly cite the relevant paper as prior existing work. If you decide that you will be working on the same topic before completing your review, you *must* inform the editor immediately. The editor can then decide whether to remove you as referee; and whether to invite an additional, potentially more objective referee. To be ethical, do not advance your new paper in your work queue for the purpose of pre-empting the submitter's existing paper.

Do not seek to game the system by magnifying a paper's drawbacks or contributions. In addition to being unethical, such gaming can hurt your reputation. Editors often catch unethical reviewer behavior.

Finally, if you are aware that the authors have behaved unethically with respect to the submission—such as submitting essentially the same paper to multiple journals, or they have committed plagiarism—notify the editor immediately. Focus on the facts, not on expressions of contempt or outrage.

Courtesy

Be courteous in the report and focus on substance. Avoid ascribing bad intent to authors ("The authors were trying for a cheap publication," "The authors were trying to brush past literature/conflicting findings under the rug …") and focus on the substance of the paper. Do not insult the authors, or use overly emotional or accusatory language.

> *Reports that are totally negative and nasty in language undercut the critical points made [in the report]. And one can be critical and tough and still be polite in the write up. If one really wants to be nasty to the authors, do it in the cover letter to the editor and not in the report itself.*

If there are indications of intellectual dishonesty, state the facts rather than speculating on intent. If an accusation of such dishonesty is made, leave it for the cover letter to the editor, who can then decide whether it should be passed along to the authors.

Moreover, be mindful that referees are subject to the same behavioral biases as everyone else. It is a mistake to approach a paper looking for evidence that confirms your pre-existing viewpoint and discount evidence that does not. Try to avoid forming unduly favorable assessments of work that appear to be written by well-known authors, members of the referee's personal social network, or papers that cite the reviewer.

Conclusion

The peer review process that occurs under the auspices of academic journals is crucial for the advancement of research. It is a central mechanism by which the profession determines collectively, through a decentralized process, both the standards for economic research and what constitutes progress in such research. However, we believe that fundamental change in how researchers review each other's journal submissions is needed to improve the integrity, quality, and efficiency of the review process. We believe that such change in refereeing culture is possible, and that when this is widely recognized, it will happen. Such change will improve how new research is developed and communicated, and will allow scholars to reallocate time from navigating the publication process to developing innovative research.

Refereeing is a hard job. Unfortunately, just like others in the profession, the three of us are all guilty of making many of the mistakes highlighted in this article. We hope that by discussing guidelines for referees, and by pointing out some dysfunctional features of current refereeing practices, we can improve our own refereeing, and play some small part in changing the culture of the review process in economics. It may be too ambitious to aim for a world in which nobody makes the mistakes that we underscore. But to improve, we need to be cognizant of our failings. A general awareness that certain refereeing practices are barriers to the advancement of knowledge would be a very large step forward.

References

Arrow, Kenneth. 1995. "Forward." In *Rejected: Leading Economists Ponder the Publication Process,* edited by George B. Shepherd, p. vii. Sun Lakes, AZ: Thomas Horton and Daughters.

Berk, Jonathan, Campbell R. Harvey, and David A. Hirshleifer. 2015. "Preparing a Referee Report: Guidelines and Perspectives." Available at SSRN: http://ssrn.com/abstract=2547191.

Card, David, and Stefano DellaVigna. 2013. "Nine Facts about Top Journals in Economics." *Journal of Economic Literature* 51(1): 144–61.

Ellison, Glenn. 2000a. "Evolving Standards for Academic Publishing: A q - r Theory." *Journal of Political Economy* 110(5): 994–1034.

Ellison, Glenn. 2002b. "The Slowdown of the Economics Publishing Process." *Journal of Political Economy* 110(5): 947–93.

Gilliland, Stephen W., and José M. Cortina. 1997. "Reviewer and Editor Decision Making in the Journal Review Process." *Personnel Psychology* 50(2): 427–52.

Hamermesh, Daniel S. 1992. "The Young

Economist's Guide to Professional Etiquette." *Journal of Economic Perspectives* 6(1): 162–79.

Hamermesh, Daniel S. 1994. "Facts and Myths about Refereeing." *Journal of Economic Perspectives* 8(1): 153–63.

Harvey, Campbell R. 2014. "Reflections on Editing the *Journal of Finance*, 2006–2012." In *Secrets of Economics Editors,* edited by Michael Szenbert and Lall Ramrattan, 67–81. Cambridge: MIT Press.

Hirshleifer, David. 2015. "Cosmetic Surgery in the Academic Review Process." *Review of Financial Studies* 28(3): 637–49.

Laband, David N. 1990. "Is There Value-added from the Review Process in Economics? Preliminary Evidence from Authors." *Quarterly Journal of Economics* 105(2): 341–52.

McAfee, R. Preston. 2010. "Edifying Editing." *American Economist* 55(1): 1–8.

Spiegel, Matthew. 2012. "Reviewing Less—Progressing More." *Review of Financial Studies* 25(5): 1331–38.

Welch, Ivo. 2014. "Referee Recomendations." *Review of Financial Studies* 27(9): 2773–2804.

Retrospectives
Do Productive Recessions Show the Recuperative Powers of Capitalism? Schumpeter's Analysis of the Cleansing Effect

Muriel Dal Pont Legrand and Harald Hagemann

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact Joseph Persky, Professor of Economics, University of Illinois, Chicago, at jpersky@uic.edu.

Introduction

The process of "creative destruction," which Joseph A. Schumpeter (1942, p. 82) considered to be "the essential fact about capitalism" since the Industrial Revolution, involves large reallocations of labor and capital from shrinking to growing sectors of the economy. Schumpeter believed that the process of liquidation and reallocation of productive resources which took place in the recession and the depression phases of cyclical fluctuations was not only an essential and unavoidable characteristic of capitalist evolution, but also was necessary, and ultimately beneficial, for long-run development. As a result, Schumpeter has often been interpreted, especially, but not only, by modern economists (such as De Long 1990; Caballero

■ *Muriel Dal Pont Legrand is Professor of Economics at the Université Côte d'Azur, Centre National de la Recherche Scientifique (CNRS) and a Researcher at the Groupe de Recherche en Droit, Economie, Gestion (GREDEG), all in France. Harald Hagemann is Professor of Economic Theory, University of Hohenheim, Stuttgart, Germany. Their email addresses are: muriel.dalpont@gredeg.cnrs.fr and harald.hagemann@uni-hohenheim.de.*

2008) as a "liquidationist." The term does not have a precise definition, but liquidationists are usually considered in the literature as convinced that economic crises are necessary and unavoidable and thus that government noninterventionism in such crises is a sound policy. As one example of a complaint of this nature levied against Schumpeter, Paul Samuelson (2015, p. 33) wrote that Schumpeter "was such a bad depression macroeconomist. … At the prime age of 51, in the ludicrous book by several Harvard senior professors Schumpeter praised the great depression as a 'healthy catharsis' of the economic system. This was a garish 'uncreative' version of what 1942 Schumpeter later called 'creative destruction.'"

The first two sections of this paper discuss Schumpeter's views in greater detail, and suggest that categorizing him as a "liquidationist" is an oversimplification and as an unrepentant "noninterventionist" is incorrect. Although Schumpeter was certainly not a strong supporter of public interventions, he did see a role for public expenditure programs in particular circumstances. During periods of recession, Schumpeter (1934b [1989 p. 110]) believed firmly in what he described as the "recuperative powers of capitalism." However, when a depression becomes "*pathological*" (Schumpeter 1941, pp. 349–50), there could be a role for government to intervene. In order to understand the overall picture of Schumpeter's message, we will first try to explain Schumpeter's analysis of recessions, depressions, and the other stages of business cycles. We will also discuss how Schumpeter perceived the recuperative powers of capitalism, a core concept in Schumpeter's analysis that allows him to distinguish between physiological and pathological recessions.

We then investigate the similarities and differences between Schumpeter's analysis and modern contributions on these issues. In the 1990s, an active line of research examined the possibility that recessions may have a productive character along with their more obvious negative outcomes, because recessions in some way might hasten the process of reallocating economic recourses from slower-growth to faster-growth sectors. Such models were sometimes referred to as "neo-Schumpeterian," but given our analysis of Schumpeter's work, we will question whether this label is appropriate.

The Wavelike Movement of Capitalist Development

Schumpeter believed that a theory of economic development (which, as Hagemann 2008 explains, he later took to calling "economic evolution") had to be constructed as a theory of business cycles, and this belief is apparent in his early writings. His 1911 [1926] book *The Theory of Economic Development*, originally published in German, held that progress occurs in waves, and crises are considered as "turning points of economic development" (p. 425; see also Schumpeter 1910, p. 294). These waves are both an unavoidable consequence of the preceding prosperity and necessary for the process of re-equilibration and long-run development. "The Wavelike Movement of Economic Life" (1914–15) provided the title of a paper in which Schumpeter summarized for an American audience his vision of the cyclical nature

of the evolution of a capitalist economy, as described in greater length in *The Theory of Economic Development*. This paper was originally presented at a Harvard seminar in January 1914.

After working on-and-off for three decades on the subject, Schumpeter (1939) presented a three-cycle schema in *Business Cycles*. Short-term 3–5 year Kitchin cycles (drawing from Kitchin 1923) were based on variations in psychological factors and crop yields. Medium-run Juglar cycles (Juglar 1862) were based on fluctuations in levels of fixed investment, often stemming from waves of innovation. Long-run Kondratieff cycles (Kondratieff 1925 [1935]) happened every 45–60 years, and were based on major technological shifts. Schumpeter also considered including additional cycles, like the Kuznets cycle (Kuznets 1930) based on construction swings, but he ultimately focused on the first three.

Schumpeter further distinguished four phases of economic fluctuations: prosperity, recession, depression, and recovery, which include two movements away from equilibrium (prosperity and depression) and two movements towards it (recession and recovery). It may seem odd to modern economists to consider "prosperity" as a movement away from equilibrium, and recession as a movement toward equilibrium. But one can readily imagine a macroeconomy fluctuating around equilibrium in which upward movements can be categorized into returning to equilibrium through "recovery" and then going beyond equilibrium into "prosperity," while downward movements are categorized into "recession," which is a decline back to equilibrium, and "depression," which is a further decline past equilibrium. In seeking to understand Schumpeter, one needs to be cognizant of his terminology.

Schumpeter (1931a, p. 6) attributed to Clement Juglar (1862) "the discovery that … crises are nothing but turning points from prosperity into depression, and it is the alternation between prosperity and depression which is the really interesting phenomenon." Concerning Schumpeter's own approach, the decisive element in his assessment is Juglar's diagnosis of the nature of depression, namely that "'the only cause of depression is prosperity'" (Schumpeter 1954, p. 1124). (It should be mentioned that despite intensive research, nobody has found what Schumpeter called this "famous sentence" in Juglar's actual writings.) In *History of Economic Analysis* (1954, p. 1123), Schumpeter elevated the work of Juglar into the pantheon of economics, praising him time and again for identifying the cyclical character of economic development and for having been seminal in combining theoretical, statistical, and historical analysis. For a detailed comparison of the analysis of business cycles in Juglar and Schumpeter, see Dal Pont Legrand and Hagemann (2007).

In Schumpeter's vision, economic development was initiated by the introduction of innovations by pioneering entrepreneurs and fostered by bank credit, which implies that divergences from macroeconomic equilibrium should often be viewed as endogenous disturbances. His argument was that exogenous impulses, like harvests or wars, would not explain the regularity of the phenomenon and the permanent "struggle for a new equilibrium" (Schumpeter 1931a, p. 15).

Later discussions of Schumpeter's multiple cycles point out that they are conceptually related to different types of capital goods as the main causal factor.

For example, the short-term Kitchin cycles mainly involve fluctuations in inventories: medium-term Juglar cycles are more related to fluctuations in fixed capital investment; Kuznets cycles involved fluctuations in construction investment; and Kondratieff cycles involved fluctuations in basic capital goods as a result of fundamental innovations. However, Schumpeter did not want to veer from his initial vision established in *The Theory of Economic Development* (1911), and continued to adhere to a monocausal argument focused on innovations. "Innovations, their immediate and ulterior effects and the response to them by the system, are the common 'cause' of them all, although different types of innovation and different kinds of effects may play different roles in each" (Schumpeter 1939, p. 172).

In his discussion of "the secondary wave" in the chapter "The Contours of Economic Evolution," Schumpeter (1939, chap. 4, p. 146) deplores that "the element of innovation has been so much neglected by the traditional analysis of the business cycle." Innovations take the system out of equilibrium. However, "[a]ny prosperity … induces a period of liquidation which, besides eliminating firms that have become obsolete beyond the possibility of adaptation, also involves a painful process of readjustment of prices, quantities, and values as the contours of the new equilibrium system emerge" (p. 148). Schumpeter had already expressed this idea in the concluding nine theses summarizing his fundamental ideas in "On the Nature of Economic Crises," his first article on the subject in 1910 (pp. 324–5 [2005, p. 50]).

For Schumpeter (1939), the notion of balanced economic growth was self-contradictory. In his view, capitalist development is characterized by recurrent disequilibria which cannot be absorbed smoothly but only through a painful process of creative destruction. While an economy is in a recession, a mechanism is at work that "cleans out" the excesses of the boom and draws the economy towards equilibrium, whereas an economy in a depression moves away from equilibrium and opens up the possibility of "abnormal liquidation" (p. 149). However, "depressions are not simply evils, which we might attempt to suppress, but—perhaps undesirable—forms of something which has to be done, namely adjustment to previous economic change" (Schumpeter 1934b [1989, p. 115]).

Schumpeter and the Great Depression

Schumpeter viewed the severity of the Great Depression that started in 1929 as a combination of the downswing of the short-run, medium-run, and long wave cycles. He expresses this view in a number of places: for example, in a January 8, 1931, letter to the German economist Arthur Spiethoff (reprinted in Schumpeter 2000), and in his Tokyo lecture in that same month (Schumpeter 1931a; see also Schumpeter 1931b [1989, pp. 96–97]). The problem of many simultaneous waves led Schumpeter (1939) to the idea of *superposition* (pp. 212ff.), meaning the composite of three cycles of various durations: "[I]f innovations are at the root of cyclical fluctuations, these cannot be expected to form a single wavelike movement" (p. 166).

For this reason, Schumpeter did not perceive the Great Depression of 1929–1933 as unique: instead, he viewed it as similar in severity to the depressions that occurred in Britain in 1825 and in the United States in 1873, events with which he makes comparisons. He later wrote: "[D]epressions of such severity have repeatedly occurred—roughly once in fifty-five years" (Schumpeter 1942, p. 64). Schumpeter recognized that special historical circumstances can aggravate a crisis which otherwise is endogenous to the capitalist system: "that what we are faced with is never simply a depression but always a depression *moulded and made worse by forces not inherent to the working of the economic engine as such*" (Schumpeter 1934b [1989, p. 114]). Among the noneconomic causes, great wars —like the Napoleonic wars, the US Civil War, and World War I—loom large in his telling. Thus, Schumpeter (1941 [1991, p. 351]) argued: "There was nothing unprecedented in this breakdown in 1929–1932 … the intensity of the depression will be in some way proportional to the intensity of the preceding progress."

Whereas a recession is a process leading to equilibrium which Schumpeter (1941 [1991]) considered as "the normal working of the evolutionary mechanism," "the excesses of speculation and loose banking methods make the thing much worse than it otherwise would be" (1991, p. 350). These factors are at the root of the pathological processes in an overshooting boom which cause the depression, leading the economy away from an equilibrium position. In Schumpeter's view, the main function of the economic cycle in capitalist development is that movements away from equilibrium—that is, prosperity and depression—are parts of an evolutionary perspective that overall has positive effects. In the decade after the outbreak of the Great Depression, Schumpeter continued to emphasize "that in the breakdown there was promise of a harvest" (1991, p. 351). Schumpeter argued that the crisis was a precondition for a recovery.

Schumpeter (1946, p. 8) would later criticize the common indictment of economists' "alleged inability to offer a satisfactory explanation of the world crisis of 1929–32" as "without foundation." He argued that it is important "to distinguish between facts that explain why there should have been a 'depression' and facts that turned this 'depression' into a 'disaster.'" In other words, it was important to distinguish between a *normal* depression and an *abnormal* or pathological depression.

Schumpeter also spelled out a mechanism through which recession and depression periods, despite their negative concomitants, exert a selective function in the process of creative destruction. During boom periods, almost every business could make a profit, or as Schumpeter (1934b [1989, p. 113]) put it, "errors and misbehaviour should be abnormally frequent in prosperity." However, "everything that is unsound for either reason shows up when prices break and credit ceases to expand in response to decreased demand for it." Thus, he viewed economic development as intrinsically interrelated with the ebb and flow of innovations and the process of setting up new production functions. In this perspective, a normal recession in the direction of equilibrium and the process of overshooting into depression indicates the functioning of the capitalist engine in which, through a painful process, outdated and ill-adjusted firms and activities are eliminated. According

to Schumpeter, a functioning capitalism does not need a government stabilization policy. He believed that it would be better to adopt a prophylactic policy that would minimize the chance of severe economic crises by adhering to clear rules—by which he meant a gold standard for monetary stability, a productivity-oriented wage policy, and other elements—rather than to intervene in a way that would be likely to hinder this economic evolution.

However, the situation changes when the away-from-equilibrium depression becomes "abnormal" or "pathological" destroying "many things which could and would have survived without it" (Schumpeter 1939, p. 149). As decisive factors causing the downward vicious spiral or the disaster of the Great Depression, Schumpeter (1946, pp. 9–10) diagnosed the wild speculative excesses in the preceding years, the weakness of the United States banking system, and the mortgage situation due to reckless borrowing and lending. This distinction between normal and abnormal liquidation was part of Schumpeter's thinking before the Great Depression arrived. For example, it appears already at the end of the second German edition of his *The Theory of Economic Development* (1926, pp. 386–89 [English translation 1934a, pp. 254-55]).

When the depression stage becomes pathological, not so much because of the severity of the depression but when the recuperative powers do not work anymore, Schumpeter held that the abnormal overshooting of the depression would have negative consequences for long-run development, and thus was willing to advocate government intervention.[1] Schumpeter shared Keynes' position that in the depression only an expansive fiscal policy, but not an expansive monetary policy, could be an efficient remedy. "The readiness to let a budget run into a deficit in a depression by keeping up expenditures in the face of shrinking revenues is a policy which will alleviate much suffering and keep many things going which would otherwise crash" (Schumpeter 1941 [1991, p. 370]).

Schumpeter was clearly reluctant to accept the principle of government intervention, and his writings are sprinkled with warnings about the dangers of continued government borrowing. For example, in lectures given in 1941, he proposes "the ruthless principle that the budget ought to be balanced under any circumstances" (Schumpeter 1941 [1991, p. 371]). While conceding that "[a]ction on that principle makes things worse in a depression," it might nevertheless be healthier in the long run because a too large and long dose of budget deficit as a stimulant can make a country dependent like a "morphinist."

But taking Schumpeter's writings as a whole, it doesn't seem appropriate to consider him a pure noninterventionist. In 1934—that is, two years before the publication of Keynes *General Theory*—he wrote of "the pivotal importance, in crises, of government expenditure on public works, and especially on direct relief for

---

[1] For a more detailed discussion of Schumpeter's views on economic policy, see Dal Pont Legrand and Hagemann (forthcoming), where we point out the time-consistency of Schumpeter's position, which is characterized by a long-run perspective fostering economic growth.

the unemployed" (1934c, p. 239). Schumpeter (1934b [1989, p. 117]) concluded another article on the Great Depression in this way:

> Especially if a country has steadily improved its public finances during prosperity as the United States did in the decade which proceeded the present crisis, enough means are available, and other means can be procured, for an expenditure which will blot out the worst things without injury to the economic organism, provided only that action on this line is taken promptly and followed up by equally sound fiscal habits as soon as recovery gets under way.

Schumpeter clearly did not believe in trying to manage the fluctuations of normal business cycles, but in the case of abnormal depressions—that is, when the recuperative powers of capitalism are not abrogated—he did support expansionary fiscal policy. His views on economic policy were already an element of his approach in earlier contributions well before the Great Depression (Stolper and Seidl 1985).

The Modern Literature on Recession and Restructuring

In line with new growth theory models that emerged in the mid-1980s, neo-Schumpeterian growth models were developed which sought to operationalize Schumpeter's notion of creative destruction (Aghion, Akcigit, and Howitt 2015). Those models were built on explicit micro foundations and benefited from rich firm-level data, making empirical tests possible. The development of this literature involved a wide variety of contributions, including a subset of papers on how business cycles can affect resource allocation. Here, we focus only on research that deals with the effect that recessions can have on restructuring and growth.[2]

Characteristically, the studies in this modern literature often quote Schumpeter's statement that recessions "are but temporary. They are the means to reconstruct each time the economic system on a more efficient plan." A little oddly, the reference is typically indirect, often quoting Stiglitz (1993, p. 2), who originally presented this paper at the meeting of the International Joseph Schumpeter Society in Kyoto, August 1992, and who appears to be the only modern author to provide the complete reference to Schumpeter's (1934b [1989, p. 113]) comment.

Despite their invocation of the Schumpeterian spirit, these models have their own distinct orientation. For example, the modern literature no longer sees fluctuations as a succession of phases, so the distinctions Schumpeter introduced between toward-the-equilibrium recession and away-from-equilibrium depression, or between normal and pathological depressions, do not appear. These models treat recessions as demand shocks, not as the outcome of technology shocks. Moreover,

---

[2]We exclude models based on a process of creative destruction but dealing with reverse causality—that is, models examining how growth can affect fluctuations, like most of the chapters included in Aghion and Howitt (1997).

they mostly concentrate on decisions taken during recession on the rate of growth, not on decisions over the business cycle as a whole (for an exception, see Caballero and Hammour 2005).

This literature can be divided into two broad categories. In the first class of models, developed by Hall (1991), Aghion and Saint-Paul (1991, 1993, 1998a), and Saint Paul (1994), firms face an intertemporal trade-off between current production and restructuring activities that will increase future output, like research and development. Under certain theoretical conditions, this reallocation in favor of the productivity-improving activities is more likely during a recession, because production is less advantageous in bad times. However, this theory has difficulty in finding clear-cut empirical support. In early research, Bean (1990) and Gali and Hammour (1992) find evidence for the domination of the reallocation effects during recession over a pro-cyclical learning-by-doing effect. But other empirical literature appeared questioning that link. More recently, Barlevy (2007) finds that research and development spending does not rise during recessions as predicted by the pure allocation model: instead, it is pro-cyclical. In a discussion drawing on the concept of research and development externalities, Barlevy argues that entrepreneurs concentrate the implementation of new ideas during booms (p. 1132). François and Lloyd-Ellis (2009) have similar findings when they decompose the process of innovation into research and development, commercialization, and implementation. Finally, Aghion, Angeletos, Banerjee, and Manova (2005) suggest that because credit is procyclical, credit constraints are more likely to be binding during recessions, which means that productivity-improving activities may be pro-cyclical, too. To be clear, the analytical framework here is about reallocation within firms, not between them, and does not deal explicitly with entry or exit. But the overall evidence from this line of research suggests that viewing recessions as times when firms are more likely to engage in productivity-enhancing activities is not empirically well-founded.

In a second class of models, Caballero and Hammour (1994, 1996; for an overview, see Aghion and Saint Paul 1998b) consider technology as embodied in capital, skills, and work organization. This framework allows them to drop the assumption of a single representative firm. Instead, heterogeneous firms can react differently to variations in demand: "[T]hey can vary either the rate at which production units that embody new techniques are created or the rate at which outdated units are destroyed" (Caballero and Hammour 1994, p. 1351). "This process of growth through Schumpeterian creative destruction results in an on-going reallocation of factors of production from contracting production sites to expanding ones" (Caballero and Hammour 1996, p. 805). This model suggests that the effect of recession on productivity will depend on the rates of entry and exit during the business cycle. They evaluate a range of evidence, noting (for example) that during recessions, the rate of entry of new firms is lower, which limits the productivity-enhancing effect of phasing out less-efficient firms. Looking at the available evidence, Caballero (2008, p. 3) finds that "on the average, recessions depress restructuring."

To put older contributions into modern perspective is always a difficult exercise. Modern theories and empirical work are inevitably not part of Schumpeter's

earlier conceptual framework. However, modern work has specified the mechanisms that could possibly be at work during the process of creative destruction, including restructuring within firms, resource reallocation between firms, and the exit of weak firms and entry of stronger firms. One can examine these mechanisms more closely to assess the market's efficiency and the recuperative and regenerative powers of capitalism proposed by Schumpeter. Indeed, Caballero and others both elucidate how these mechanisms work and also examine the consequences of their dysfunction, a situation that Schumpeter perhaps would have considered equivalent to the collapse of the recuperative powers of capitalism.

Modern empirical work suggests that the process of creative destruction is important for economic growth: "Over the long run, the process of creative destruction accounts for over 50 percent of productivity growth" (Caballero 2008). Nevertheless, the creativity and restructuring that happens during economic upswings is likely to be more important for this process than the destructive processes unleashed by recessions, and in addition, the reallocations of creative destruction are *hindered* by recession, not accelerated. Interestingly, in his early study concerning the empirical characteristics of long waves, Kondratieff (1925 [1935], p. 111) had reached this conclusion as well. He wrote: "During the recession of the long waves, an especially large number of important discoveries and inventions in the technique of production and communication are made, which, however, are usually applied on a large scale only at the beginning of the next long upswing."

Finally, Caballero and Hammour (2005) investigated the potential social costs of recessions from a broadly Schumpeterian perspective. Caballero (2008, p. 3) described their results in this way: "While we sided with Schumpeter and others on the view that increasing the pace of restructuring of the economy is likely to be beneficial, we provided evidence that, contrary to conventional wisdom, restructuring falls rather than rises during contractions." Aside from its empirical contribution challenging the more extreme liquidationist interpretations of recessions, in the context of this paper, the summary is interesting for another reason. It reveals an unexpected paradox: while these authors see themselves as Schumpeterian when interpreting recessions as productive episodes, they do not recognize that it is precisely when they express their more balanced views about the effect that recessions can have on growth that they actually have much more in common with Schumpeter's own analysis.

References

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt. 2015. "Lessons from Schumpeterian Growth Theory." *American Economic Review* 105(5): 94–99.

Aghion, Philippe, George-Marios Angeletos, Abhijit Banerjee, and Kalina Manova. 2005. "Volatility and Growth: Credit Constraints and Productivity-Enhancing Investment." NBER Working Paper 11349.

Aghion, Philippe, and Peter Howitt. 1997. *Endogenous Growth Theory.* Cambridge, MA: MIT Press.

Aghion, Philippe, and Gilles Saint-Paul 1991. "On the Virtue of Bad Times: An Analysis of the Interaction between Fluctuations and Economic Growth." CEPR Discussion Paper 578.

Aghion, Philippe, and Gilles Saint-Paul 1993. "Uncovering Some Causal Relationships between Productivity Growth and the Structure of Economic Fluctuations: A Tentative Survey." NBER Working Paper 4603.

Aghion, Philippe, and Gilles Saint-Paul 1998a. "Virtues of Bad Times: Interaction between Productivity Growth and Economic Fluctuations." *Macroeconomic Dynamics* 2(3): 322–44.

Aghion, Philippe, and Gilles Saint-Paul. 1998b. "Uncovering Some Causal Relationships between Productivity Growth and the Structure of Economic Fluctuations: A Tentative Survey." *Labour* 12(2): 279–303.

Barlevy, Gadi. 2007. "On the Cyclicality of Research and Development." *American Economic Review* 97(4): 1131–64.

Bean, Charles R. 1990. "Endogenous Growth and the Procyclical Behavior of Productivity." *European Economic Review* 34(2–3): 355–63.

Caballero, Ricardo J. 2008. "Creative Destruction." *The New Palgrave Dictionary of Economics*, 2nd edition, edited by Steven N. Durlauf and Lawrence E. Blume.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Recessions." *American Economic Review* 84(5): 1350–68.

Caballero, Ricardo J., and Mohamad L. Hammour. 1996. "On the Timing and Efficiency of Creative Destruction." *Quarterly Journal of Economics* 111(3): 805–52.

Caballero, Ricardo J., and Mohamad L. Hammour. 2005. "The Cost of Recessions Revisited: A Reverse-Liquidationist View." *Review of Economic Studies* 72(251): 313–41.

Dal Pont Legrand, Muriel, and Harald Hagemann. 2007. "Business Cycles in Juglar and Schumpeter." *History of Economic Thought* 49(1): 1–18.

Dal Pont Legrand, Muriel, and Harald Hagemann. Forthcoming. "Business Cycles, Growth, and Economic Policy: Schumpeter and the Great Depression." *Journal of the History of Economic Thought.*

De Long, J. Bradford. 1990. "'Liquidation' Cycles: Old Fashioned Real Business Cycle Theory and the Great Depression." NBER Working Paper 3546.

Francois, Patrick, and Huw Lloyd-Ellis. 2009. "Schumpeterian Cycles with Pro-cyclical R&D." *Review of Economic Dynamics* 12(4): 567–91.

Gali, Jordi, and Mohamad Hammour. 1992. "Long-Run Effects of Business Cycles." Colombia Graduate School of Business paper 92-26.

Hagemann, Harald. 2008. "Schumpeter on Development." In *Marshall and Schumpeter on Evolution: Economic Sociology of Capitalist Development,* edited by Yuichi Shionoya and Tamotsu Nishizawa, pp. 225–42. Cheltenham, UK, and Northampton, MA: Edward Elgar.

Hall, Robert E. 1991. "Labor Demand, Labor Supply, and Employment Volatility." In *NBER Macroeconomic Annual,* edited by Olivier Blanchard and Stanley Fischer, pp. 17–47. Cambridge, MA: MIT Press.

Juglar Clément. 1862. *Des crises commerciales et de leur retour périodique en France, en Angleterre et aux Etats-Unis,* Paris, Guillaumin et Cie, libraires – Editeurs, 1862, Second Edition in 1889.

Kitchin, Joseph. 1923. "Cycles and Trends in Economic Factors." *Review of Economics and Statistics* 5(1): 10–16.

Kondratieff, Nikolaï D. 1925 [1935]. "Long Waves in Economic Life." In *Voprosy Conjunktury,* vol. I, pp. 28–79. (Translated in 1935 with some summary and omissions into English from a 1926 German version by W. A. Stolper, *Review of Economics and Statistics,* 17(6): 105–115; complete English translation 1979, *Review* (Fernand Braudel Center) 2(4): 519–62.)

Kuznets, Simon. 1930. *Secular Movements in Production and Prices: Their Nature and their Bearing upon Cyclical Fluctuations.* Boston: Houghton Mifflin.

Saint-Paul, Gilles. 1994. "Les récessions sont-elles productives?—Une approche ne termes de coût d'opportunité." *Revue Economique* 45(3): 897–904.

Samuelson, Paul A. 2015. "The Harvard-Circle." *Journal of Evolutionary Economics* 25(1): 31–36.

Schumpeter, Joseph A. 1910. "Über das Wesen der Wirtschaftskrisen." In *Zeitschrift für*

Volkswirtschaft, Sozialpolitik und Verwaltung, 19: 271–325. (2005 English translation is "On the Nature of Economic Crises," in *Business Cycle Theory: Selected Texts 1860–1939*, vol. 5, edited by M. Boianovsky, 3–50. London: Pickering & Chatto.)

Schumpeter, Joseph A. 1911. *Theorie der wirtschaftlichen Entwicklung*, 2nd edition. Munich and Leipzig: Duncker & Humblot. (German edition completely revised in 1926; English translation of the revised edition in 1934a is *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle.* Cambridge, MA: Harvard University Press.)

Schumpeter, Joseph A. 1914–15. "Die Wellenbewegung des Wirtschaftslebens." *Archiv für Sozialwissenschaft und Sozialpolitik* 39(1): 1–32.

Schumpeter, Joseph A. 1931a. "The Theory of the Business Cycle." *Keizaigaku Ronshu –Journal of Economics* 4(1): 1–32.

Schumpeter, Joseph A. 1931b. "The Present World Depression: A Tentative Diagnosis." *American Economic Review* 21(1): 179–82. (Reprinted in 1951 as *Joseph A. Schumpeter: Essays on Entrepreneurs, Innovations, Business Cycles and the Evolution of Capitalism*, edited by R. V. Clemence, Cambridge University Press; republished in 1989, with a new introduction by Richard Swedberg, pp. 96–99, Transaction Publishers.)

Schumpeter, Joseph A. 1934b. "Depressions: Can We Learn from Past Experience?" In *The Economics of the Recovery Program*, edited by D. V. Brown et al. London and New York. McGraw-Hill. (Reprinted 1951 in *Joseph A. Schumpeter: Essays on Entrepreneurs, Innovations, Business Cycles and the Evolution of Capitalism*, edited by Richard V. Clemence, Cambridge University Press; republished in 1989, with a new introduction by Richard Swedberg, pp. 108–117, Transaction Publishers.)

Schumpeter, Joseph A. 1934c. "Statement." In *Economic Reconstruction. Report of the Columbia University Commission*, Addendum, pp. 239–42. New York: Columbia University Press.

Schumpeter, Joseph A. 1939. *Business Cycles. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, 2 vols. New York: McGraw-Hill.

Schumpeter, Joseph A. 1941. "An Economic Interpretation of Our Time: The Lowell Lectures." (Reprinted 1991 in *Joseph A. Schumpeter: The Economics and Sociology of Capitalism*, edited by R. Swedberg, pp. 339–400, Princeton University Press.)

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy.* New York and London: Harper & Brothers. (2nd edition 1947.)

Schumpeter, Joseph A. 1946. "The Decade of the Twenties." *American Economic Review* 36(2): 1–10.

Schumpeter, Joseph A. 1954. *History of Economic Analysis.* London: George Allen & Unwin.

Schumpeter, Joseph A. 2000. *Briefe/Letters*, selected and edited by U. Hedtke and R. Swedberg. Tübingen: Mohr Siebeck.

Stiglitz, Joseph E. 1993. "Endogenous Growth and Cycles." NBER Working Paper 4286.

Stolper, Wolfgang F., and Christian Seidl eds. 1985. *Joseph A. Schumpeter: Aufsätze zur Wirtschaftspolitik.* Tübingen: J. C. B. Mohr (Paul Siebeck).

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of under-graduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

The United Nations Development Programme has published its *Arab Human Development Report 2016*, subtitled "Youth and the Prospects for Human Development in a Changing Reality." "Most recent statistics indicate that two-thirds of the Arab region's population is below thirty years of age, half of which falling within the 15–29-year age bracket. … The report asserts that today's generation of young people is more educated, active and connected to the outside world, and hence have a greater awareness of their realities and higher aspirations for a better future. However, young people's awareness of their capabilities and rights collides with a reality that marginalises them and blocks their pathways to express their opinions, actively participate or earn a living. As a result, instead of being a massive potential

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

for building the future, youth can become an overwhelming power for destruction. … Unemployment among youth in Arab countries is the highest in the world, 29 percent in 2013, versus 13 percent worldwide. … The region needs to create more than 60 million new jobs in the next decade to absorb the large number of workforce entrants and stabilize youth unemployment. ... The state-led development model … hobbles promising enterprises, discourages economic efficiency and deters young talents because its goal is not to promote innovation or competition, but solely to preserve access to wealth and power among a few. The result is a top-down model that is based on hand-outs, undermines individual agency and encourages short-term consumption at the expense of long-term investment in human capabilities and competitive production. ... Resource rents in the region have been channelled into lavish and conspicuous real estate projects, unproductive public sector spending and military expenditures, but the spending benefits a tiny slice of society." November 2016. http://arab-hdr.org/PreviousReports/2016/2016.aspx.

Laurence Ball, Joseph Gagnon, Patrick Honohan, and Signe Krogstrup make the case that central banks continue to have considerable power to engage in an expansionary monetary policy, should they wish to do so, in their monograph *What Else Can Central Banks Do?* "Short-term interest rates have been near zero in advanced economies since 2009, making it difficult for central banks to cut rates further and provide needed economic stimulus. There is reason to believe that this lower bound problem will be common in the decades to come. … [T]here is much that policymakers can and should do to provide further stimulus in environments with underemployment and inflation below target. We emphasise two policy options: • *Negative interest rates.* Since 2014, some central banks have pushed nominal interest rates modestly below zero, providing some increase in stimulus. Although there is a limit to how far below zero interest rates can go, it is likely that rates could go somewhat further than what has been done so far without adverse consequences. • *Quantitative easing.* Central banks, beginning with the Federal Reserve and the Bank of England, have already used this tool to mitigate the slumps in their economies since 2008. More stimulus can be provided if policymakers increase the scale of quantitative easing, and if they expand the range of assets they purchase to include risky assets such as equity. … We argue, however, that any side effects are manageable and not of a magnitude to justify timidity in using available tools to regain price stability and restore full employment." Geneva Reports on the World Economy #18, published by the International Center for Monetary and Banking Studies and the Centre for Economic Policy Research, at http://cepr.org/sites/default/files/geneva_reports/GenevaP285.pdf.

Caroline M. Hoxby delivered the 2016 Martin Feldstein Lecture at the National Bureau of Economic Research on the subject: "The Dramatic Economics of the U.S. Market for Higher Education." "We have in the United States what is arguably the world's only true *market* for higher education, as opposed to systems that are largely centrally controlled or financed. This market exhibits a strong positive correlation between students' college readiness (hereafter 'CR') and the educational resources of the institution they attend. … Strikingly, among institutions that experience strong market forces, the productivity of a dollar of educational resources is fairly

similar, even if the schools serve students with substantially different CR. On the other hand, among institutions that experience weak market forces, productivity is lower and more dispersed. … If we take the productivity results and the resources-CR correlation as manifestations of market forces, then it follows that a student with higher CR must make more productive use of any marginal dollar of educational resources than a student who is less prepared for college. This property, which economists call 'single-crossing,' has long been hypothesized to be a law of nature, at least in tertiary education. This is the first compelling evidence. Single-crossing has profound consequences for the role of higher education in income growth …" *NBER Reporter* (2016, Number 3, pp. 1–6) has published an edited version of the talk at http://www.nber.org/reporter/2016number3/2016number3.pdf. You can also view the hour-long lecture and the slides at http://www.nber.org/feldstein_lecture_2016/feldsteinlecture_2016.html.

The International Labour Organization has published "Non-Standard Employment Around the World: Understanding Challenges, Shaping Prospects." "In over 150 countries, the average use of temporary employees in registered private sector firms is 11 per cent, with about one-third of countries around this mean. … Casual employment is a prominent feature of labour markets in developing countries, and has grown in importance in industrialized countries. … In Australia, where casual employment is a specific employment category, one out of four employees is casual. … Approximately 10 per cent of the workforce in the United States have irregular and on-call work schedules, with the lowest-income workers the most affected … In countries with available data, TAW [temporary agency work] spans from 1 to over 6 per cent of wage employment. … NSE [non-standard employment], particularly when it is not voluntary, may increase workers' insecurities in different areas …" November 2016, http://www.ilo.org/global/publications/books/WCMS_534326/lang--en/index.htm.

Symposia and Special Issues

Cityscape offers a 10-paper symposium with various perspectives on gentrification. For example, in their essay "Advancing our Understanding of Gentrification," Ingrid Gould Ellen and Lei Ding (the editors of the symposium) write: "In sum, since 2000, U.S. cities have seen greater increases in the SES [socioeconomic status] index and other measures in downtown neighborhoods and an expansion of SES index increases to more cities and neighborhoods. Compositional shifts toward White, prime-age, and college-educated households—not population growth—are more characteristic of recent changes. Although lower-skilled or lower-education jobs continue to suburbanize, jobs employing college-educated workers have stopped declining or have even increased in traditional downtowns. Downtown safety and amenity values appear to have increased. A sizable number of downtown neighborhoods in big cities, however, have not seen increases in our SES index at all, and a number of peripheral neighborhoods in smaller metropolitan

areas have seen dramatic changes." In "Commentary: Causes and Consequences of Gentrification and the Future of Equitable Development Policy," Derek Hyra writes: "Gentrification, in some places, is associated with political and cultural displacement. … In certain respects changing norms may be positive in terms of counteracting norms of violence or a lack of health-producing amenities and activities, but do the new norms and incoming amenities in gentrifying neighborhoods sufficiently cater to the preferences of low-income people or do they predominately represent newcomers' tastes and preferences?" U.S. Department of Housing and Urban Development, 2016, vol. 18, no. 3, at https://www.huduser.gov/portal/periodicals/cityscpe/vol18num3/index.html.

Future of Children has published a 10-paper symposium, plus an overview essay, about the existing research on "Starting Early: Education from Prekindergarten to Third Grade." In the overview essay, "Starting Early: Introducing the Issue," Jeanne Brooks-Gunn, Lisa Markman-Pithers, and Cecilia Elena Rouse write: "[W]e believe the weight of the evidence, as reflected in the articles in this issue, indicates that high-quality pre-K programs can indeed play an important role in improving later outcomes, particularly for children from more disadvantaged families. At the same time, significant questions remain. … At the end of most evaluated [early childhood education] programs, researchers find effects on school achievement, though these effects diminish over elementary school. When program effects are large, they tend to be maintained into elementary school, though they are smaller than the initial impacts. At the same time, we see long-term effects on adult outcomes—for example, a reduction in crime or the completion of more schooling. It's puzzling that during elementary school, the achievement-test scores of children who attended prekindergarten converge with the test scores of children who did not, a phenomenon commonly called fadeout. Studies document that those who participate in a pre-K program have a significant advantage in kindergarten in terms of educational achievement. But those assigned to the control group tend to catch up in the first through third grades; in most evaluations, more than half the difference between the two groups disappears by the end of first grade." Fall 2016, http://www.futureofchildren.org. The symposium extends and expounds on a number of themes from the paper by Greg J. Duncan and Katherine Magnuson, "Investing in Preschool Programs," in the Spring 2013 issue of this journal.

The *Russell Sage Foundation Journal* has devoted an issue to the topic "A Half-Century of Change in the Lives of American Women," including a short introduction and 10 readable essays. As one example, Kim A. Weeden, Youngjoo Cha, and Mauricio Bucca write on "Long Work Hours, Part-Time Work, and Trends in the Gender Gap in Pay, the Motherhood Wage Penalty, and the Fatherhood Wage Premium": "One of the key empirical insights of this literature is that the gender gap in wages at the aggregate level is perpetuated by persistent gender differences in individual labor market behaviors: whether men and women work for pay, the occupations and industries in which they work, and the number of hours per week they work. These gender differences emerge in the context of structural changes in the distribution of jobs with particular attributes (such as expected work hours) and in the wages associated

with these attributes, resulting in complex and offsetting effects on the gender gap in wages. For example, … the diffusion of long work hours in the United States in the 1990s and 2000s, coupled with the persistent gender gap in long work hours and rising hourly compensation for long work hours, was associated with an increase in the gender gap in wages after adjusting for other wage-relevant attributes. These trends largely offset wage-equalizing shifts in women's educational attainment. A second empirical insight is that much of what appears to be a gender wage gap is better understood as a gender-specific family gap in pay or, as they are known in the economic and sociological literatures, the motherhood wage penalty and fatherhood wage premium: mothers earn less than observationally similar childless women, and fathers earn more than observationally similar childless men." August 2016, vol. 2, issue 4, at http://www.rsfjournal.org/toc/rsf/2/4. These papers complement the symposium on "Women in the Labor Market" in this issue.

Prominent Economists

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for 2016 has been awarded to Oliver Hart and Bengt Holmström "for their contributions to contract theory." Each year, the committee posts a "Popular Science Background" with a readable overview of the laureates' work. Here's a slice of the description of Holmström's work:

> This *informativeness principle* does not merely say that payments should depend on outcomes that can be affected by agents. For example, suppose the agent is a manager whose actions influence her own firm's share price, but not share prices of other firms. Does that mean that the manager's pay should depend only on her firm's share price? The answer is no. Since share prices reflect other factors in the economy—outside the manager's control—simply linking compensation to the firm's share price will reward the manager for good luck and punish her for bad luck. It is better to link the manager's pay to her firm's share price relative to those of other, similar firms (such as those in the same industry). A related result is that the harder it is to observe the manager's effort—perhaps due to many distorting factors blurring the relationship between her effort and the company's performance—the less the manager's pay should be based on performance. In industries with high risk, payment should thus be relatively more biased towards a fixed salary, while in more stable environments it should be more biased towards a performance measure.

Here's a slice of the description of Hart's work:

> The main idea is that a contract that cannot explicitly specify what the parties should do in future eventualities, must instead specify who has the right to

decide what to do when the parties cannot agree. … In complex contracting situations, allocating decision rights therefore becomes an alternative to paying for performance. … Suppose a new invention requires the use of a particular machine and a distribution channel. Who should own the machine and who should own the distribution channel—the inventor, the machine operator, or the distributor? If innovation is the activity for which it is most difficult to design a contract, which seems realistic, the answer could be that the innovator should own all the assets in one company, even though she may lack production and distribution expertise. As the innovator is the party that has to make greater non-contractible investments, she also has greater need of the future bargaining chip that property rights bring to the assets. … Should providers of public services, such as schools, hospitals, and prisons, be privately-owned or not? According to the theory, this depends on the nature of non-contractible investments.

The "Popular Science Background" and a more technical "Scientific Background" paper called "Oliver Hart and Bengt Holmström: Contract Theory" are available with other background information at https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2016.

Douglas Clement has an "Interview with David Autor," and the subheading reads: "MIT economist on tech, trade & job markets, how Chinese imports affect U.S. politics & family structure, and the Janus-faced gig economy." "Computerization changes what type of work people do—that's very clear; we see the occupational change going on. But the part that people miss is that displacement of a set of tasks or even entire job categories does not augur the end of work. … Most of the jobs we have didn't really exist in any significant number 200 years ago. As a result, work is much better. It's more interesting, it's more productive, it's safer and more rewarding. … We've adjusted to the displacement of human labor by automation along at least three margins. One is that we've just created many new and interesting things to do. … Two, more of us work, but we work fewer hours. People don't work until the day they die. They work 40- and 50-hour weeks instead of 80-hour weeks. They work five days a week instead of seven. They take vacations. So they've spread the work in a way that's constructive and leads to a better quality of life. And the other thing, of course, is that as we get wealthier, our consumption demands rise, so we create a lot of work because we choose to consume rather than just taking it all in leisure. If a worker in 2015 wanted to have a 1915 level of income, he or she could work about 17 weeks a year. But most of us choose not to. We'd rather have a bigger house and a couple of cars and whatever else." *The Region*, Federal Reserve Bank of Minneapolis, September 7, 2016, https://www.minneapolisfed.org/publications/the-region/interview-with-david-autor.

David A. Price has an "Interview" with Josh Lerner with a focus on venture capital, private equity, crowdfunding, and entrepreneurship. On fees: "An interesting thing is that fees in private equity and venture capital are remarkably sticky.

The compensation structures don't look that different in today's era of $10 billion-plus funds than they did back in an era of $10 million funds. They've come down somewhat, so instead of 2 percent committed capital, it's more likely to be 1.5 percent. But given the economies of scale of running a larger fund, it means the profits per partner can be staggering." On crowdfunding: "Moreover, when you look at attempts to create entrepreneurial finance models with crowdfunding-type flavors to them, the outcomes have not been great. For instance, there was an effort in Europe during the 1990s to create a whole series of small capitalization models where riskier young companies could list and so forth with relatively lax regulations. They ended up with a phenomenon where the bad drove out the good. All it took was a few scammers to come in and undertake 'pump and dump' schemes, and the interest in those markets declined precipitously. And I think some of the same danger lurks here." On the timing of initial public offerings: "For instance, among venture-backed firms today, the average company going public was 12 years old at the time of IPO last year. Historically, it was around four or five years old. And so you've got all these companies that are privately held sitting there raising money but staying private. They're getting funded not just by venture capitalists, but also by sovereign wealth funds and family offices and even mutual funds. *Econ Focus*, Federal Reserve Bank of Richmond, Second Quarter 2016, pp. 26–30, https://www.richmondfed.org/publications/research/econ_focus/2016/q2/interview.

Discussion Starters

Diane Whitmore Schanzenbach, Ryan Nunn, Lauren Bauer, Megan Mumford, and Audrey Breitwieser present "Seven Facts on Noncognitive Skills from Education to the Labor Market." "[T]there is a robust and growing body of evidence that noncognitive skills function similarly to cognitive skills, strongly improving labor-market outcomes. These noncognitive skills—often referred to in the economics literature as soft skills and elsewhere as social, emotional, and behavioral skills—include qualities like perseverance, conscientiousness, and self-control, as well as social skills and leadership ability. The value of these qualities in the labor market has increased over time as the mix of jobs has shifted toward positions requiring noncognitive skills. Evidence suggests that the labor-market payoffs to noncognitive skills have been increasing over time and the payoffs are particularly strong for individuals who possess both cognitive and noncognitive skills …" Hamilton Project at the Brookings Institution, October 2016, http://www.hamiltonproject.org/assets/files/seven_facts_noncognitive_skills_education_labor_market.pdf.

The IMF *Fiscal Monitor* for October 2016 is subtitled "Debt: Use it Wisely." "The global gross debt of the nonfinancial sector has more than doubled in nominal terms since the turn of the century, reaching $152 trillion in 2015. About two-thirds of this debt consists of liabilities of the private sector. Although there is no consensus about how much is too much, current debt levels, at 225 percent of world GDP, are at an all-time high. The negative implications of excessive private debt (or

what is often termed a "debt overhang") for growth and financial stability are well documented in the literature, underscoring the need for private sector deleveraging in some countries. The current low-nominal-growth environment, however, is making the adjustment very difficult …" http://www.imf.org/external/pubs/ft/fm/2016/02/pdf/fm1602.pdf.

Alain Marciano and Nathalie Moureau discuss "Museums, Property Rights, and Photographs of Works of Art: Why Reproduction Through Photograph Should be Free." From the abstract: "The law concerning the reproduction of works of art is unambiguous: the owner of the physical item does not own the right to copy and reproduce it. The copyright or right to reproduce a work of art either belongs to the artist and his/her heirs, or to everybody when the work is in the public domain. However, a large number of museums use their property rights to assume a copyright, i.e. a right to reproduce works of art. These illegal practices are the result of choosing a business model based on the desire to cross-subsidise the upstream market of the services provided to the public with the benefits obtained by monopolising the 'downstream' market of the copies or reproductions of works of art. … We argue that this strategy conflicts with the mission upheld by museums and prevents certain externalities from circulating in the society." *Review of Economic Research on Copyright Issues*, 2016, vol. 13, no. 1, pp. 1–28). Also available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2840305.

Samuel Hammond has written "Bone Marrow Mismatch: How Compensating Bone Marrow Donors Can End the Transplant Shortage and Save Lives." "The most common goal of bone marrow donation is to harvest hematopoietic (blood-forming) stem cells (HSC) for transplant in patients with life threatening diseases like leukemia and anemia. For decades, the main method of HSC extraction was aspiration: a painful procedure wherein a needle is inserted directly into the donor's bone, extracting the soft, spongy tissue from its source. Today, over 70% of HSC donations are instead extracted indirectly from the bloodstream through a safe, non-invasive technique known as apheresis, after which the cells quickly regenerate. … [W]hile the volunteer donor registry has grown in recent years, it remains beset by low follow-through rates. … [T]his paper estimates that the gap between the actual and optimal size of the bone marrow registry leads to 1318 fewer transplants and 275 deaths per year. … This report argues that legalizing compensation for HSC donors would help to close that gap by increasing registry size among target populations while improving follow-through rates of those who are called upon to donate … Internal market research by the healthcare company Hemeos indicates a payment of $2000 is sufficient to increase the follow-through rate of pledged donors to 90% …" Niskanen Center Discussion Paper, November 15, 2016, https://niskanencenter.org/wp-content/uploads/2016/11/Bone-Marrow-Mismatch-1.pdf.

CONSIDERATIONS FOR THOSE PROPOSING TOPICS AND PAPERS FOR JEP

Articles appearing in the journal are primarily solicited by the editors and associate editors. However, we do look at all unsolicited material. Due to the volume of submissions received, proposals that do not meet *JEP*'s editorial criteria will receive only a brief reply. Proposals that appear to have *JEP* potential receive more detailed feedback. Historically, about 10–15 percent of the articles appearing in our pages originate as unsolicited proposals.

Philosophy and Style

The *Journal of Economic Perspectives* attempts to fill part of the gap between refereed economics research journals and the popular press, while falling considerably closer to the former than the latter. **The focus of *JEP* articles should be on understanding the central economic ideas of a question, what is fundamentally at issue, why the question is particularly important, what the latest advances are, and what facets remain to be examined.** In every case, articles should argue for the author's point of view, explain how recent theoretical or empirical work has affected that view, and lay out the points of departure from other views.

We hope that most *JEP* articles will offer a kind of intellectual arbitrage that will be useful for every economist. For many, the articles will present insights and issues from a specialty outside the readers' usual field of work. For specialists, the articles will lead to thoughts about the questions underlying their research, which directions have been most productive, and what the key questions are.

Articles in many other economics journals are addressed to the author's peers in a subspecialty; thus, they use tools and terminology of that specialty and presume that readers know the context and general direction of the inquiry. By contrast, **this journal is aimed at all economists, including those not conversant with recent work in the subspecialty of the author.** The goal is to have articles that can be read by 90 percent or more of the AEA membership, as opposed to articles that can only be mastered with abundant time and energy. Articles should be as complex as they need to be, but not more so. Moreover, the necessary complexity should be explained in terms appropriate to an audience presumed to have an understanding of economics generally, but not a specialized knowledge of the author's methods or previous work in this area.

The *Journal of Economic Perspectives* is intended to be scholarly without relying too heavily on mathematical notation or mathematical insights. In some

cases, it will be appropriate for an author to offer a mathematical derivation of an economic relationship, but in most cases it will be more important that an author explain why a key formula makes sense and tie it to economic intuition, while leaving the actual derivation to another publication or to an appendix.

JEP does not publish book reviews or literature reviews. Highly mathematical papers, papers exploring issues specific to one non-U.S. country (like the state of agriculture in Ukraine), and papers that address an economic subspecialty in a manner inaccessible to the general AEA membership are not appropriate for the *Journal of Economic Perspectives*. Our stock in trade is original, opinionated perspectives on economic topics that are grounded in frontier scholarship. If you are not familiar with this journal, it is freely available on-line at <http://e-*JEP*.org>.

Guidelines for Preparing JEP Proposals

Almost all *JEP* articles begin life as a two- or three-page proposal crafted by the authors. If there is already an existing paper, that paper can be sent to us as a proposal for *JEP*. However, given the low chances that an unsolicited manuscript will be published in *JEP*, no one should write an unsolicited manuscript intended for the pages of *JEP*. **Indeed, we prefer to receive article proposals rather than completed manuscripts.** The following features of a proposal seek to make the initial review process as productive as possible while minimizing the time burden on prospective authors:

- Outlines should begin with a paragraph or two that precisely states the main thesis of the paper.
- After that overview, an explicit outline structure (I., II., III.) is appreciated.
- The outline should lay out the expository or factual components of the paper and indicate what evidence, models, historical examples, and so on will be used to support the main points of the paper. The more specific this information, the better.
- The outline should provide a conclusion
- Figures or tables that support the article's main points are often extremely helpful.
- The specifics of fonts, formatting, margins, and so forth do not matter at the proposal stage. (This applies for outlines and unsolicited manuscripts).
- Sample proposals for (subsequently) published *JEP* articles are available on request.
- For proposals and manuscripts whose main purpose is to present an original empirical result, please see the specific guidelines for such papers below.

The proposal provides the editors and authors an opportunity to preview the substance and flow of the article. For proposals that appear promising, the editors provide feedback on the substance, focus, and style of the proposed article. After the editors and author(s) have reached agreement on the shape of the article (which may take one or more iterations), the author(s) are given several months to submit a completed first draft by an agreed date. This draft will receive detailed comments from the editors as well as a full set of suggested edits from *JEP*'s Managing Editor. Articles may undergo more than one round of comment and revision prior to publication.

Readers are also welcome to send e-mails suggesting topics for *JEP* articles and symposia and to propose authors for these topics. If the proposed topic is a good fit for *JEP*, the *JEP* editors will work to solicit paper(s) and author(s).

Correspondence regarding possible future articles for *JEP* may be sent (electronically please) to the assistant editor, Ann Norman, at <anorman@ JEPjournal.org>. Papers and paper proposals should be sent as Word or pdf e-mail attachments.

Guidelines for Empirical Papers Submitted to JEP

The *JEP* is not primarily an outlet for original, frontier empirical contributions; that's what refereed journals are for! Nevertheless, *JEP* occasionally publishes original empirical analyses that appear uniquely suited to the journal. In considering such proposals, the editors apply the following guidelines (in addition to considering the paper's overall suitability):

1. The paper's main topic and question must not already have found fertile soil in refereed journals. *JEP* can serve as a catalyst or incubator for the refereed literature, but it is not a competitor.

2. In addition to being intriguing, the empirical findings must suggest their own explanations. If the hallmark of a weak field journal paper is the juxtaposition of strong claims with weak evidence, a *JEP* paper presenting new empirical findings will combine strong evidence with weak claims. The empirical findings must be robust and thought provoking, but their interpretation should not be portrayed as the definitive word on their subject.

3. The empirical work must meet high standards of transparency. *JEP* strives to only feature new empirical results that are apparent from a scatter plot or a simple table of means. Although *JEP* papers can occasionally include regressions, the main empirical inferences should not be regression-dependent. Findings that are not almost immediately self-evident in tabular or graphic form probably belong in a conventional refereed journal rather than in *JEP*.

The American Economic Association's Committee on Economic Education

in Conjunction with the

Journal of Economic Education

and the

Federal Reserve Bank of Denver

hosts the

Seventh Annual Conference on Teaching & Research in Economic Education

May 31 – June 2, 2017
Denver, Colorado
Westin Denver Downtown
Dinner/plenary speaker at the Federal Reserve Bank of Denver

Plenary Sessions include:

- **Dean Karlan** (Yale): Tackling Market Failures: Making the World a Better Place with Ideas from Economics 101
- **Troy Davig** (Federal Reserve Bank of Kansas City): TBD
- **Anna Maria Lusardi** (George Washington): Financial Literacy and the Importance of Financial Education

Call for papers and all information regarding the conference can be found at
www.aeaweb.org/committees/AEACEE

Compliments of the American Economic Association

Webcasts of Selected Sessions from the 2017 AEA Annual Meeting . . .

Now available on the AEA Website

January 6, 2017

"Brexit: Six Months Later"

Olivier Blanchard
Jonathan Portes
Andrew Lilico
Karl Whelan

"Nobels on Where is the World Economy Headed?"

Dominick Salvatore
Angus Deaton
Roger Myerson
Edmund Phelps
Robert Shiller
Joseph Stiglitz

AEA/AFA Joint Luncheon

Alvin Roth
Eric Budish

"Gender Agenda"

Alexander Bick
Nicola Fuchs-Schuendeln
Muriel Niederle
Patricia Cortes
Jessica Pan
Claudia Goldin
Claudia Olivetti

Richard T. Ely Lecture

Alvin Roth
Esther Duflo

January 7, 2017

"Economists as Engineers"

Alvin Roth
Paul Milgrom
Atila Abdulkadiroglu

"Economic Issues Facing the New President"

Greg Mankiw
Jason Furman
Glenn Hubbard
Alan Krueger
John Taylor

Nobel Luncheon

Alvin Roth
Christina Paxson
James Heckman
David Laibson

"Publishing and Promotion in Economics: The Curse of the Top Five"

James Heckman
George Akerlof
Angus Deaton
Drew Fudenberg
Lars Hansen

AEA Presidential Address

Alvin Roth
Robert Shiller

2017 AEA Continuing Education webcasts also available

Visit https://www.aeaweb.org/conference/webcasts/2017

The American Economic Association

MIX
Paper from
responsible sources
FSC™ C132124
FSC
www.fsc.org

AMERICAN
ECONOMIC
ASSOCIATION