The Journal of

Economic Perspectives

A journal of the
American Economic Association

Celebrating 30 Years

Spring 2016

The Journal of

Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by R. R. Donnelley Company, Jefferson City, Missouri, 65109, USA.

No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Spring 2016, Vol. 30, No. 2. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. E-reader versions are free. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 30 • Number 2 • Spring 2016

Symposium

Inequality Beyond Income

Articles

Features

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.

Journal of Economic Perspectives

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 3–28

Consumption Inequality

Orazio P. Attanasio and Luigi Pistaferri

Much of the debate over the rising levels of inequality in the United States and other developed countries is phrased in terms of income, or in terms of components of income like wages and earnings. But for economists, a basic utility function of individuals typically refers to consumption and leisure, not income.

The distinction between income and consumption could make a meaningful difference in thinking about inequality if the distribution of consumption at a given point in time is less wide than that of income, or if its evolution over time is smoother than that of income. Consumption can differ from income if consumers borrow or save, or if they receive transfers from other family members or the government in response to income shocks. The joint analysis of consumption and income inequality can be informative in several ways. It can show the presence (or lack) of such consumption-smoothing mechanisms. It can shed light on the nature of income shocks, and in particular the extent to which they should be understood as temporary (which may be easier to smooth out for consumption

■ *Orazio P. Attanasio is Jeremy Bentham Professor of Economics, University College London, London, United Kingdom, and Research Fellow, Institute for Fiscal Studies, both in London, United Kingdom. Luigi Pistaferri is Professor of Economics and Ralph Landau Senior Fellow, Stanford Institute for Economic Policy Research (SIEPR), both at Stanford University, Stanford, California. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts, and also Research Fellows, Centre for Economic Policy Research (CEPR), London, United Kingdom. Their email addresses are o.attanasio@ucl.ac.uk and pista@stanford.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.3 doi=10.1257/jep.30.2.3

purposes) or permanent. If one is interested in the effects of inequality on those in the poorest segments of society, consumption might reveal different insights than income—for example, because of different dynamics in the relative prices of goods consumed by rich and poor households. Finally, higher consumption of leisure could partly offset lower consumption of goods when it comes to overall welfare measurement.

In this essay, we begin with a discussion of the sources of consumption data, and some of the issues that arise when looking at data on consumer spending while trying to infer the economically relevant concept of consumption. We then offer our interpretation of the research that has compared trends in income inequality with trends in consumption inequality. The narrative has evolved very sharply from arguing that trends in consumption inequality are quite different from those of income inequality to concluding that they track each other closely. This change in findings has been shaped in substantial part by the adoption of strategies aimed at dealing with measurement problems in consumption data, as well as by some reinterpretation of the underlying economic forces.

We then discuss some additional aspects of consumption. We look at specific data on inequality in consumption of food, ownership of major household appliances, leisure, and persistence in consumption across generations. These comparisons suggest ways in which aggregate consumption inequality fails to tell the entire story. In the concluding section, we take stock of the evidence and summarize challenges for future work. Our main conclusion is that researchers interested in measuring inequality in well-being need to go beyond the fact that consumption is unequally distributed and realize that a full picture of the evolution in welfare requires taking a stand on quality concerns and on the value that people attach to leisure, among other things.

Consumption Data: Sources and Concepts

What Consumption Data Do We Have?

US researchers who want to study income inequality have a considerable array of data sources from which to choose. If they want household survey data on incomes, the Current Population Survey, Panel Study of Income Dynamics, Survey of Income and Program Participation, National Longitudinal Survey of Youth, and even the decennial Census (for studying long-run trends) all offer large, consistent samples and detailed information on income and its components. Researchers may also have access to administrative-level data (where measurement error issues may be less important with regard to income data), such as data from the Internal Revenue Service and data from the W-2 forms that employers use to report income paid. Overall, datasets with measures of household income resources (such as wages, earnings, and income) are more frequently available, typically have larger samples, and have more consistent variable definitions than datasets containing information on consumption (Pistaferri 2015).

In contrast, household surveys on household expenditure are rare, small, and lack a consistent longitudinal component. The Consumer Expenditure Survey (CE), the only dataset with comprehensive and detailed information on household expenditure and its components, is available on a continuous basis since 1980. It is used by the Bureau of Labor Statistics primarily to form weights placed on price changes of goods in the computation of the overall Consumer Price Index. The CE is composed of two distinct surveys. In the Interview survey, respondents are sampled every three months for a total of four quarters. In the Diary survey, respondents fill a two-week diary of their expenditures and are sampled only once. In producing aggregate means, the Bureau of Labor Statistics routinely uses and tabulates certain items from the Interview survey and others from the Diary survey.

The other dataset that is widely used by academic researchers to study consumption behavior is the Panel Study of Income Dynamics (PSID), which is available on an annual basis from 1968 to 1997, and on a biannual basis since then. The initial goal of the PSID was to study income dynamics (and poverty) between and across generations. For this reason, information on consumption was considered ancillary, and until 1997, the PSID collected information only on a few consumption items: food (at home and away from home), home rent, and (occasionally) utility payments. Starting with the 1999 wave, however, the PSID began to collect information on a larger range of items, covering about 70–90 percent of the spending collected in the Consumer Expenditure Survey. Respondents typically report spending for broad categories, with the reference period being (with some exceptions) the previous calendar year. Blundell, Pistaferri, and Saporta-Eksten (2016) show that these data track national accounts aggregates well.

Comprehensive administrative data on consumption spending are not available, but some partial sources do exist. For example, some researchers have used spending data from credit card expenditure records (as in Gross and Souleles 2002; Aydin 2015). Others have used data on spending, income, and assets for consumers using online financial aggregators such as Mint.com or Check.com (Baker 2014; Gelman et al. 2014). Finally, there is spending data from checkout scanners from the Nielsen Homescan datasets (Handbury 2014; Broda and Weinstein 2008), which refer primarily to grocery store items. While these new sources of administrative data on consumption constitute remarkable steps ahead, they are either not representative of all households or not representative of all the goods that people buy.[1]

Looking at Spending, Thinking about Consumption

Consumption can be harder to measure accurately than income, and measurement errors may be differently severe in different parts of the income distribution.

[1] For some Nordic countries, researchers have proposed using longitudinal administrative tax record information on income and wealth to create consumption using the intertemporal budget constraint, so that expenditure can be derived as income minus the change in wealth: see Browning and Leth-Petersen (2003) and De Giorgi, Frederiksen, and Pistaferri (2015) for Denmark; Koijen, Van Nieuwerburgh, and Vestmanz (2015) for Sweden; and Autor, Kostøl, and Mogstad (2015) for Norway.

Survey data like the Consumer Expenditure Survey typically report consumer spending, which may not coincide with consumption for at least four reasons. First, consumption is overstated relative to spending for those who have made durable purchases in the current period, and understated for those who made durable purchases in the past. Most surveys of household consumption have no information on the stock of existing durables owned by the household. In the Consumer Expenditure Survey, the only exception is cars, as consumers report the year, type, and make of the cars they own. However, there is no information on the resale values, which must instead be estimated. For other durables, there is some information on ownership and number of items owned, but no information on current values. As for housing, the survey contains information on imputed services, as homeowners report how much their house would rent for. Second, some consumption is received in kind, through transfers from friends, relatives, private institutions (charities or churches), or the government (in the form of in-kind or voucher-provided benefits like food stamps, school lunches, health care services through Medicaid or Medicare, rent subsidies, and so forth). Third, some consumption is produced at home using time and good inputs, like child care provided by parents or siblings. Finally, the conversion from spending to consumption requires knowledge of prices paid for the goods people consume, a requirement that is usually solved by assuming that households face the same prices. This assumption is violated in practice for many nontradeable goods (such as housing), and even for tradeable, homogenous goods due to shop-specific effects, bulk purchases, or loyalty cards. Ignoring the distinction between consumption and spending can either understate or overstate differences in well-being across individuals with different levels of spending.

Survey nonresponse and measurement errors create a different set of issues. Sabelhaus et al. (2015) study nonresponse rates in the Consumer Expenditure Survey and conclude that they have risen over time, especially among the high-income population. Sabelhaus and Groen (2000), using a variety of techniques, argue that the ratio of consumption to income for richer households is downward biased. This may affect the measurement of trends in aggregate consumption and consumption inequality, respectively. In this journal, Meyer, Mok, and Sullivan (2015) conclude that measurement error in survey data has also increased over time. In principle, there is no obvious reason to expect errors in reporting spending to be worse than errors in reporting income. On the one hand, the changing nature of spending modes and patterns may heighten reporting errors; on the other hand, the move from cash to e-commerce may facilitate the collection of administrative data on spending. Several papers discuss strategies to elicit consumption information in general-purpose surveys (for instance, Crossley and Winter 2015).

The combination of these issues makes any measure of consumption naturally problematic, no matter how much effort researchers put into making it accurate. In contrast, income measured in surveys is arguably closer to the relevant economic concept (except in cases involving businesses). However, we should also note that income is not easier to measure than consumption for every household. For low-income individuals (in both developed and developing countries) income can be complex to measure, because it includes a myriad of different sources including wages, interpersonal and government transfers, and so on. In comparison, consumption for the poorest households may be fairly straightforward. The situation is probably reversed for well-off households for which administrative income data might be accessible and reliable, while their consumption can be complex and varied and difficult to measure (Deaton and Grosh 2000).

Survey Data versus National Accounts Data on Consumption

Consumer spending data available in the Consumer Expenditure Survey appears increasingly detached from the Personal Consumption Expenditure (PCE) data collected by the Bureau of Economic Analysis (which forms the basis for the national income and product accounts data). For example, Passero, Garner, and McCully (2015) report that the ratio of total expenditures in the CE data compared to the PCE data has declined from 0.70 in 1992 to 0.58 in 2010. Of course, whenever two different methods of measuring a similar economic concept give different answers, it's a matter for concern.

Some of the discrepancy is due to the fact that the two series measure different concepts and cover different entities. The Personal Consumption Expenditure data includes the value of goods and services purchased by US resident households (including imputed rents for owner-occupied housing) and by nonprofit organizations on behalf of households (typically, employer-paid health insurance and medical care, and expenses associated with life insurance and pension plans). It also includes purchases by US government civilian and military personnel stationed abroad and US residents traveling or working abroad for one year or less. For most consumption categories, the PCE is estimated using a "commodity-flow" method. This approach computes the value of domestic output based on data from the Census of Manufactures, which looks at the value of manufacturers' shipments and inventories. Next, domestic consumption (denominated in producers' prices) is estimated by adding imports and subtracting exports and changes in inventory. Finally, the value of consumer purchases is converted from producers' prices to purchasers' prices by adding wholesale margins and taxes, transportation costs, and retail margins, and taxes. Clearly, many steps in the calculation of the PCE are also likely to contain sizeable measurement error.

Both the population coverage and the methodology for obtaining total spending is different in the Consumer Expenditure Survey (as discussed in Slesnick 1991). The Personal Consumption Expenditure data includes institutionalized individuals; the Consumer Expenditure Survey does not. The Consumer Expenditure Survey excludes spending made by US residents abroad and by nonprofit institutions on behalf of households (with the most obvious difference being the value of Medicare and Medicaid spending, which tripled in real terms between 1990 and 2014). The PCE concept includes imputed rents on owner-occupied housing, while the Consumer Expenditure Survey aggregates typically exclude them. Indeed, the discrepancy between the two measures is less dramatic when comparing items that are conceptually comparable and definitionally similar. Passero, Garner, and

McCully (2015) compare different components of consumption and conclude that "non-durables are most alike for the CE and PCE with about 93 percent of total non-durable expenditures identified as comparable within the CE and within the PCE." Their conclusion is that "focusing on comparable goods and services only, CE to PCE ratios have steadily decreased," but slightly less than when comparing unadjusted statistics. For example, the CE–PCE ratio for total comparable goods and services decreased from 84 percent for 1992 to 74 percent for 2010 (as opposed to 70 and 58 percent, respectively, in unadjusted figures). They write: "The greatest decline in CE to PCE ratios is for durables, with a decrease of 24 percentage points," from 0.82 to 0.62. The decline is smaller for services (0.95 to 0.86) and for nondurables (0.70 to 0.63).

Bee, Meyer, and Sullivan (2015) assess the performance of the Consumer Expenditure Survey on a good-by-good basis and report three findings. First, in general the Interview survey performs better than the Diary survey in matching numbers from the Personal Consumption Expenditure data for some categories. Second, the coverage ratios are excellent for some goods (food at home, rent, and utilities) and, in those areas, have not changed appreciably over time; on the other hand, the coverage ratios for other items (such as clothing or alcoholic beverages) are low and declining. Finally, some durable stocks and durable purchases appear to be reported sufficiently well (new vehicles), while the quality of others has worsened considerably (furniture). Overall, they conclude that the consumption categories that tend to be reported poorly are those that involve small and infrequent purchases, while large and regular purchases are reported sufficiently well. Moreover, they write that "based on observable characteristics, the [Consumer Expenditure Survey] appears to be fairly representative, although there is strong evidence of under-representation at the top of the income distribution and under-reporting of income and expenditures at the top." This is another reason for the growing discrepancy between the CE aggregates and the PCE from the National Accounts. If consumption growth is higher at the top of the distribution, declining survey response among the rich may easily explain the deterioration of the match between CE aggregates and the PCE.

Quantity versus Prices

Survey data like the Consumer Expenditure Survey measure expenditure, which is the product of prices and quantities. To make comparison of consumption across periods meaningful, researchers deflate expenditure using an overall measure of the cost of living, typically the Consumer Price Index (CPI). Indeed, as mentioned above, the CE is collected primarily to compute the weights for the CPI. However, average weights may not be relevant for all households. It is possible that the composition of the consumption basket varies substantially and systematically across different households, as a direct consequence of differences in access to resources as well as needs and tastes. Luxuries will be more prominent in the expenditure basket of the rich, while necessities will account for a larger share of poor households' expenditure. Health costs may be more relevant for older individuals and certain types of entertainment more relevant for younger ones. Therefore changes in relative prices can have distributional consequences.

Moreover, there may be price differences across space (or stores within a given geographical location), or across time within space (because of high-frequency sales) even for relatively homogenous goods. Because of differences in prices over space, consumers might have an incentive to search for the best deals and these incentives may vary for individuals with different values of time.

The presence of differences in prices for homogeneous goods and changes in the relative prices of goods that are more or less relevant for different groups of individuals might lead researchers to overstate or understate the level and trends of inequality. If the poor live in relatively cheaper areas (or if they shop in relatively less-expensive stores), or if the prices for the goods that they typically purchase grow less than the prices of the goods typically purchased by rich households, then inequality in consumption (that is, spending deflated by an index that accounts for household-specific price differences) will be less (and grow more slowly) than inequality in spending deflated with a common price index. There are two reasons why this may be the case. First, increasing trade with low-wage countries lowers the price of imported goods. This may reduce consumption inequality even in the absence of any change in income inequality if goods imported from low-wage countries are relatively more important in the consumption basket of low-income individuals than in the basket of high-income individuals. Moreover, the diffusion of mega-stores (such as Walmart) has likely benefited low-income individuals more than high-income individuals. These differences are partly attenuated by the consideration that better quality and the experience of shopping in certain stores have an amenity value.

Datasets where researchers can disentangle the two components of expenditure are hard to come by. The Nielsen Homescan data is one exception, but it is limited to groceries and a few other items. Other data sources (such as the ACCRA Cost of Living Index produced by the Council for Community and Economic Research) provide city-specific indexes on a few categories of interest. In the Consumer Expenditure Survey, the geographical detail is very limited due to confidentiality concerns (and in any case, it would miss information on the type of store where goods are purchased). Later in the paper, we discuss some recent work on the implications of price inequality.

Does Consumption Inequality Track Income Inequality?

Consumption Smoothing: Why and How

Is consumption inequality a better measure of changes in welfare than income inequality? For economists, using consumption inequality has theoretical appeal. The life-cycle hypothesis of Modigliani and Brumberg (1955) and the permanent income hypothesis of Friedman (1957), which constitute the workhorse theory of how people make their consumption decisions, suggest that risk-averse households prefer a smooth to a variable consumption flow. Hence, households would choose consumption to be a constant fraction of their permanent or lifetime income, not current income. Because current income can be highly volatile from one year to the next, it may give a partial snapshot of people's living standards. The extent to which households can achieve a smooth consumption flow depends on the tools they have to move resources over time and states of nature. Savings can be used to absorb certain income shocks and can be accumulated for such a purpose. Other tools for consumption smoothing may include access to credit and insurance markets, and interpersonal and government transfers.

The ability to move resources across time and states explains why consumption may not track income. Consumption may exceed current income because a consumer is borrowing (permanent income is above current income, as in the case of a medical student taking out loans in the expectation of higher future earnings) or it may be below current income because the consumer is saving (and the doctor is now repaying medical school loans). Large wealth effects can also have a considerable influence on consumption independently of income. It is then possible for the income distribution to reveal no changes in well-being even though the underlying consumption distribution is shifting in response to wealth effects. Consumption may vary from income for other reasons as well. For example, consumption falls below current earnings and wages because of taxes paid, and above them because of government transfers—a different form of consumption smoothing especially relevant for households at the bottom of the distribution. Even if *full* smoothing is not feasible, perhaps because of borrowing restrictions and imperfections in insurance and credit markets, some consumption smoothing would still occur. Recent surveys on consumption (or marginal utility) smoothing are Browning and Crossley (2001) and Attanasio and Weber (2010).

These considerations imply that how consumption (and welfare) react to changes depends on the tools available for consumption smoothing and on the nature of income changes. In general, we can think of current income as including two components: one reflects long-term or permanent factors such as the level of skills and human capital, while the other reflects temporary or transitory factors (like being out of the labor force due to human capital investments, fertility, job loss, and the like). For any given household, permanent income changes slowly over time. When it does, it is because of unpredictable events, like the way in which new technologies affect the price and quantity demanded of one's skills (for better or worse), along with unexpected promotions or demotions, a change in the local economy that affects wage levels, and other factors. MaCurdy (1982) and Abowd and Card (1989) are two representative papers in a vast labor literature that tries to decompose income (or earnings, or wages) into transitory and permanent components.

The distinction between temporary and persistent shifts in the wage or income distribution has important policy and welfare implications. Policies aimed at reducing inequality under the two scenarios are very different. In the first case, it is probably necessary to reduce inequality in the endowments of human capital, whilst in the latter it may be sufficient to improve the access to smoothing mechanisms. In theory, permanent shocks are harder to absorb and insure and are thus more likely to be reflected in substantial changes in consumption and welfare. In contrast, temporary shocks are easier to smooth through borrowing or running down accumulated assets. Hence, if all changes in income inequality were of a transitory nature, we could expect no large changes in consumption inequality.

The Evidence on Consumption and Income Inequality

The mainstream narrative on consumption inequality has evolved considerably over time, from earlier uncertainty over whether consumption was rising less than income inequality to the current belief that it has been rising just about as much as income inequality.

The first few papers in the earlier literature that looked at different dimensions of inequality in consumption were Cutler and Katz (1991, 1992) and Slesnick (1994). These papers used the Consumer Expenditure Survey data (in an era in which the measurement error issues discussed above were less well-known than they have since become). Cutler and Katz (1991) found that "changes in the distribution of consumption parallel changes in the distribution of income." In contrast, Slesnick (1994) found that consumption inequality had grown more modestly than income inequality.[2] A number of later studies (for example, Krueger and Perri 2006) found evidence similar to Slesnick (1994). In the terminology of consumption smoothing and the permanent income hypothesis, this finding can be interpreted as implying that a sizeable proportion of shocks to income were both temporary and insurable. Evidence on this comes from two sources: direct evidence on the income process, and indirect evidence coming from the response of consumption to income changes of different natures.

On the first front, researchers working on income inequality were finding strong evidence that the rise in income inequality was partly of transitory nature—which in turn implies that some portion of the rise in inequality could be more smoothed in consumption. Gottschalk and Moffitt (1994) argued that the rise in the variance of the transitory component of income (what they called "income instability") represented about one-third to one-half of the overall rise in income inequality observed in the 1980s and 1990s. (Income instability is often measured using the variance of income changes, or growth. This is because the change in income across two periods is approximately equal to the change in the transitory component if permanent income evolves along the expected path.)

The fact that a good chunk of the rise in income inequality was of a transitory nature would not matter much (in terms of separating income from consumption inequality) if consumers were unable to smooth transitory shocks. However, several

[2] The differences between the two studies arise partly from their different consumption definitions. Cutler and Katz (1991) include spending on durables (other than housing and vehicles), while Slesnick (1994) imputes service flows. Moreover, Slesnick adjusts for topcoding in some spending categories.

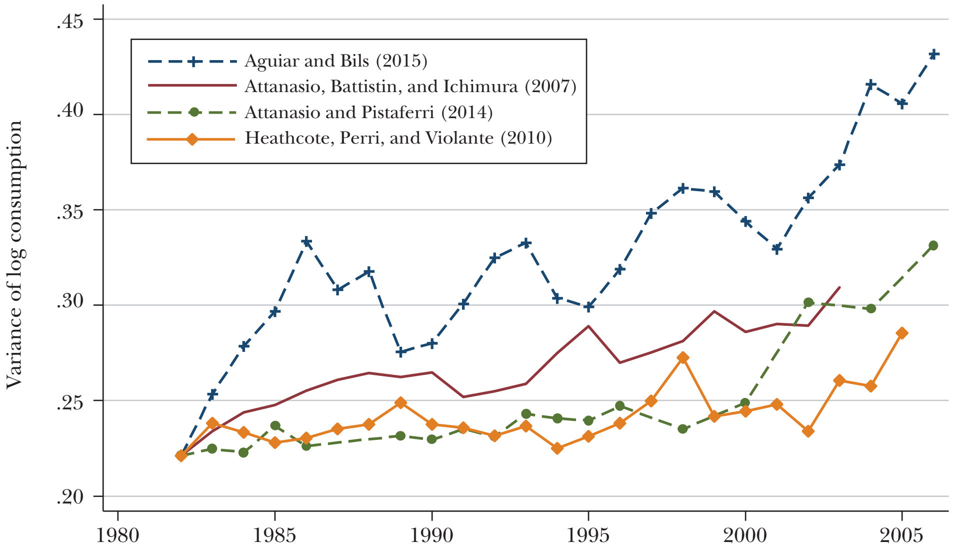
papers show that consumers are able to smooth short-run shocks (Dynarski and Gruber 1997), although less so if they have low assets or low education (Blundell, Pistaferri, and Preston 2008). Attanasio and Davis (1996) focused on the relationship between relative wages and relative consumption across different groups in the US population, where groups were defined on the basis of the year of birth of the household head and on their educational achievement. They found that long-run relative movements in wages across these groups were mirrored in relative movements in consumption. This correlation was driven by the relative movements across education groups: the increases in the return to education in terms of wages and earnings seemed to be reflected in increases in the return to education in terms of consumption. The online appendix available with this paper at http://e-jep.org contains an update of Attanasio and Davis (1996), extending the data to 2012. As in the original paper, we find that when we consider the impact on consumption of one-year wage changes, the variability of which is probably dominated by temporary fluctuations in wages that can be smoothed in some way, we do not find a significant relationship between relative changes in wages and consumption. However, when considering longer (five- and eight-year) horizons, where instead persistent wage factors are more likely to be at play, the relationship between changes in consumption inequality and income inequality becomes strongly significant.

In keeping with these two pieces of evidence, Krueger and Perri (2006) show that while in the 1980–2003 period the variance of log income increases from 0.35 to 0.57, the variance of log consumption increases only from 0.18 to 0.24. In other words, inequality rises in both income and consumption, but the rise in income inequality is much larger.

More recently, however, researchers have started to question the evidence about consumption inequality, rethinking the measurement issues that arise from considering measures of expenditure in the Consumer Expenditure Survey. The survey seems to be affected by serious nonclassical measurement error whose importance is increasing over time. One possible strategy is to focus on the components of the Consumer Expenditure Survey that appear to be measured most accurately and to use alternative datasets for other components of consumption. The challenge of course is that one would like to make statements about inequality in overall consumption, not necessarily about inequality in some components. Once one corrects for the measurement problems afflicting the Consumer Expenditure Survey, uses alternative data sources when these seem preferable, or measures consumption in alternative ways, consumption inequality seems to rise by more than previously believed, and to track income inequality closely.

A number of papers develop this view using a variety of data sources and empirical approaches (Attanasio, Battistin, and Ichimura 2007; Aguiar and Bils 2011; Attanasio and Pistaferri 2014). Figure 1 gives an overall view of the evolution of consumption inequality over time and across papers (and empirical strategies). In this figure, consumption inequality is measured by the variance of log consumption (deflated by the Consumer Price Index and expressed in per capita terms).

Figure 1
The Evolution of Consumption Inequality over Time as Measured by Different Papers



Note: Heathcote, Perri, and Violante (2010) used the Interview survey of the Consumer Expenditure Survey. Attanasio, Battistin, and Ichimura (2007) combined consumption items from the Interview survey and the Diary survey (in the attempt of picking the survey component that best measures each item). Aguiar and Bils (2015) used the Consumer Expenditure Survey but computed consumption as the difference between disposable income and active savings. In Attanasio and Pistaferri (2014), we used Panel Study of Income Dynamics consumption data available from 1999 onward, estimated an inverse demand function for food for the 1999–2009 period, and then used the estimated coefficients to predict consumption for the period before 1999 (when only food data were available). Consumption inequality is measured by the variance of log consumption (deflated by the Consumer Price Index and expressed in per capita terms).

Heathcote, Perri, and Violante (2010) used the Interview survey of the Consumer Expenditure Survey, and their findings reproduce the flat profile of consumption inequality shown by Krueger and Perri (2006). Attanasio, Battistin, and Ichimura (2007) combined consumption items from the Interview survey and the Diary survey (attempting to pick the survey component that best measures each item), with results showing a more marked increase in inequality. Aguiar and Bils (2015) used the Consumer Expenditure Survey but computed consumption as the difference between disposable income and active savings, and they find an even larger increase in inequality. Finally, in Attanasio and Pistaferri (2014), we used Panel Study of Income Dynamics consumption data available from 1999 onward, estimated an inverse demand function for food for the 1999–2009 period, and then used the estimated coefficients to predict consumption for the period before 1999 (when only food data were available). We found that inequality also increases more than the Heathcote, Perri, and Violante measure, especially in the last years of the sample period.[3]

To obtain a sense of how much consumption inequality grows relative to income inequality, and how the response depends on the methodology used to measure consumption, consider the following calculation. Over the period considered in the figure, the variance of the log of family income from the PSID (deflated by the Consumer Price Index and expressed in per capita terms) increases by 27 points (or about 20 points when using an after-tax measure available in the Consumer Expenditure Survey, as reported by Heathcote, Perri, and Violante 2010). If we take the Aguiar and Bils measure of consumption inequality shown in Figure 1 as the most credible, the variance of log consumption increases by about 18 points over the same time period. In contrast, the Heathcote, Perri, and Violante measure would suggest an increase of only about 10 points. Meyer and Sullivan (2013) show that the tracking between income and consumption inequality is stronger at the top of the distribution (as measured by the 90th–50th percentile difference) and in the 1980s and 1990s than in subsequent years. Aguiar and Bils's (2015) core exercise is actually to measure consumption inequality by looking at how high- and low-income households allocate spending to luxuries and necessities. In particular, inequality in the luxury/necessity spending ratio (scaled by the difference in demand elasticities, which can be obtained from estimation of a demand system) is shown to provide a measure of consumption inequality that is robust to measurement error in overall spending, as well as to household-specific measurement errors (for example, more severe underreporting by high-income households) and good-specific measurement errors (more severe underreporting for some goods than others). Using this alternative metric, Aguiar and Bils confirm that over the 1980–2007 period, inequality in consumption grows as much as income inequality.

The common element of the papers above is that once one makes an attempt to move away from the traditional measurement of consumption inequality using the Interview component of the Consumer Expenditure Survey, and tries to correct for the measurement problems, then the trends in consumption inequality appear much steeper than initially believed. Of course, the conclusion reached by these papers may also be premature, because the strategies adopted, while ingenious, are based on data that may have different types of measurement problems.

One aspect that seems to militate in their favor, however, is that the change in the consensus about the trends in consumption inequality has been accompanied by changes in the consensus on the evolution of income inequality, based on

[3] An important caveat is that the consumption series used in the four papers are not identical. For example, the Attanasio and Pistaferri (2014) measure is limited by the fact that the PSID collects a limited amount of information on expenditure, while the Aguiar and Bils (2015) measure, by definition, does not use any consumption information. The Heathcote, Perri, and Violante (2010) and Attanasio and Pistaferri (2014) measures include out-of-pocket spending on health and education, while the Attanasio, Battistin, and Ichimura (2007) measure excludes them. The differences between the series should thus be seen as illustrative. We normalize all series to equal the Heathcote, Perri, and Violante (2010) value in 1982 (the first year in which we observe all four series).

improved income data. Recent work using administrative data about income—which is less prone to measurement error issues than survey data—finds that most of the increase in wage and earnings variance has been structural, or of a more permanent nature (for example, DeBacker et al. 2013; Kopczuk, Saez, and Song 2010; Guvenen, Ozkam, and Song 2014). Kopczuk, Saez, and Song (2010) use Social Security data over a very long time horizon and present a formal decomposition between total, transitory, and persistent earnings variances. Using this decomposition, the rise in total variance during the period of interest is primarily driven by a rise in structural factors. In contrast, there is very little evidence of a rise in the variance of the transitory component.

These recent findings about the nature and dynamics of income inequality are consistent with the revised thinking in the dynamics of consumption inequality. If income volatility is stable, it means the variance of the transitory component has not increased. Hence, the bulk of the change in income inequality has occurred because of a rise in the variance of the permanent component. It is possible for consumption inequality to rise less than income inequality even in a setting in which income volatility is stable. This is because consumers may be able to insure even some shocks to their permanent income, at least partially. For example, the Disability Insurance program seeks to attenuate the economic cost of permanent shocks to health that result in permanent inability to work. But it is clearly more difficult to smooth changes in permanent income, and as a consequence, it is not surprising that consumption inequality rises by roughly as much as income inequality. Indeed, their rise is explained by the same forces (absent strong insurance mechanisms).

Inequality in Prices

As mentioned earlier, recent research has started to look at data on individual purchases and has documented the existence of important heterogeneity in prices of even very homogenous goods, both in different stores and within a store over short periods of time, through the use of sales and discounts. One of the first papers to document the existence of substantial heterogeneity in the prices of very homogeneous goods is Aguiar and Hurst (2007). They correlate observed prices and shopping behavior with consumer characteristics. They show that older consumers are more likely to shop longer and more frequently and, probably as a consequence, pay lower prices for similar goods. Griffith, Leibtag, Leicester, and Nevo (2008) use British scanner data to show that low-income households realize considerable savings by buying in bulk and by buying economy brands, while savings from sales, coupons, and the like are nonlinear in income—specifically, higher at the top and bottom of the income distribution. More recently, Nevo and Wong (2015) show that during the Great Recession, consumers switched to buying more on sale, using more coupons, buying more generics and larger pack sizes and these changes were larger in states that suffered larger increases in unemployment rates.

Kaplan and Menzio (2015) use scanner data on a large sample of US households covering grocery stores purchases in 54 geographical markets over the 2004–2009 period. They find that the distribution of prices is symmetric and with fatter tails than the normal distribution, and its average standard deviation is between 19 and 36 percent. They also show that, when decomposing the variability of prices of homogeneous goods into a store component, a store-good component, and a transaction component (the dispersion of prices within a store), most of the variability of prices in their sample is explained by the latter two factors. They suggest that price dispersion is more likely to be driven by intertemporal price discrimination and search frictions than differences in amenities or marginal costs across stores. This hypothesis is explored more fully in Kaplan, Menzio, Rudanko, and Trachter (2016).

The extent to which differences in prices actually paid affect the dynamics of consumption inequality, either through differences in consumer baskets or through price heterogeneity induced and sustained by frictions and retailer behavior, is an open question and one of considerable interest. The availability and use of scanner data on individual transactions can be very useful in this respect, as is the development of models that incorporate price discrimination and frictions in price-setting behavior. More broadly, the measurement of consumption and income inequality is a lively area of research. The existing work is undoubtedly subject to improvement as better data or more creative approaches to overcome measurement issues come along.

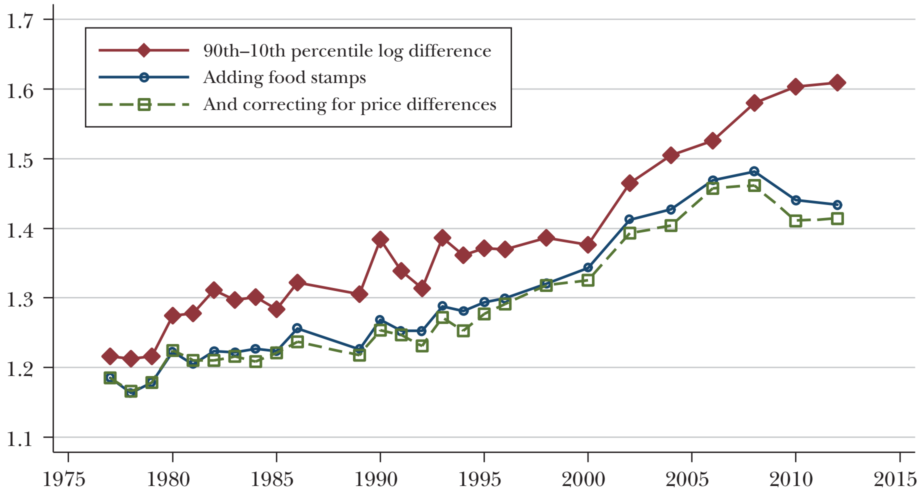
Inequality in Components of Consumption

Looking at inequality of consumption across specific components of consumption may be interesting for several reasons. First, the measurement of some components of consumption is of better quality than others, thus alleviating concerns about whether results are affected by measurement error. Second, the analysis of different groups of commodities with different income elasticities can be informative about the nature of shocks and about mechanisms for smoothing consumption. Third, changes in the patterns of expenditure on durables can be informative about the perception of future shocks, because individuals know that commodities such as furniture or cars provide services for long periods and can be sold only subject to large transaction costs. Finally, disparities in consumption necessities such as food may be more worrying from a welfare point of view than disparities in the consumption of luxuries, such as exotic vacations.

Food

In Figure 2, we use data from the Panel Study of Income Dynamics and plot the difference between the 90th and the 10th percentile of the logarithm of food consumption distribution over the 1977–2012 period. Food consumption is defined as the sum of spending on food at home, food away from home, and the monetary value of food stamps. Data are in real terms and adjusted for family composition by dividing by an OECD scale (defined as $\$1 + 0.7(n - 1) + 0.3k$, where n is the number of adults and k the number of kids). PSID food data exist before 1977, but it is only in 1977 that the Food Stamps Act established national standards of eligibility.

Figure 2
The 90th–10th Percentile Log Food Difference



Source: Authors using data from the Panel Study of Income Dynamics.
Note: This figure plots the difference between the 90th and the 10th percentiles of the logarithm of food consumption distribution from 1977 to 2012.

Our sample includes all households whose head is aged 25–85. The sample includes the poverty subsample of the Panel Study of Income Dynamics and hence sampling weights are used throughout. To emphasize the distinction between consumption and spending (which in the case of food may be particularly relevant due to government transfers), we plot the 90th–10th percentile difference both including and excluding food stamps from our definition of food consumption.

Clearly, inequality in food consumption is rising. Most of the rise is coming from a decline in spending at the bottom (not shown separately here). The difference between the top and intermediate lines shows clearly the insurance value of government transfers. In particular, during the Great Recession the availability of food stamps allowed poor households to maintain their food consumption, while spending declined substantially.

We should also note that some of the lower spending on food by the poorest households may be due to a decline in the prices of the food items that they purchase (Broda, Leibtag, and Weinstein 2009). To have some sense about the importance of price differentials, we also plot the 90th–10th percentile difference allowing the price deflator to be good-specific (that is, food at home plus food stamps, and food away from home). Correcting for price differentials has a small effect, although it is more pronounced in recent years. Because the price of food at home (a necessity consumed in large fractions by households at the bottom of the distribution) has been steadily declining relative to the price of food away from home (a luxury consumed in large fractions by households at the top of the distribution), inequality in consumption is lower when adjusting for these price differences.

The decline in spending on food consumption at the bottom of the distribution may not indicate a decline in caloric intakes. Households may spend less on food without modifying the caloric intake of the food purchases they make (as argued by Aguiar and Hurst 2005). Indeed, Singh et al. (2009) report that energy intake is not statistically different between US adults with income below the poverty line and those with income above 500 percent of the poverty line. After all, sugars and fats, which are high in calories, can be considerably less expensive than diets based on vegetables, fruits, whole grains, and lean meat.

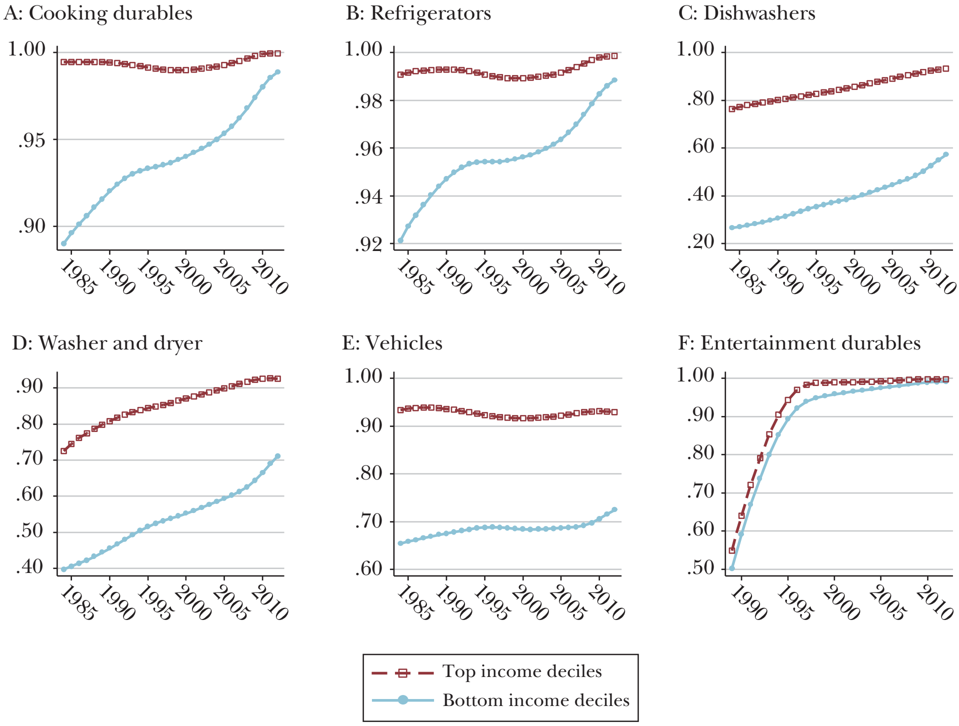
The different qualities of food raise a question: Should an assessment of inequality in food consumption be based on its monetary cost, its energy content, its healthfulness, or some other measure of quality? The US Department of Health and Human Services and the US Department of Agriculture have proposed to measure diet quality with an index known as the Healthy Eating Index (HEI). The index gives a 0–10 score to 12 food components (like Total Fruits, Whole Fruits, and so on). For some "good" components (like Total Vegetables) a higher intake means a higher score, while for some "bad" components (like Saturated Fat) the opposite is true. Wang et al. (2014) use data from the 1999–2010 National Health and Nutrition Examination Survey, and compare the HEI index for people of different socioeconomic background and education. They find that in the population at large the quality of food consumed increases monotonically over the sample period. However, individuals with low socioeconomic status (defined by those with less than high school and income below 130 percent of the poverty line, the eligibility threshold for food stamps) make no progress in terms of HEI from 1999 to 2010, while most of the improvements are concentrated among medium and high socioeconomic status groups. We want to stress that while these differences reflect changes in the "quality" of food consumed between rich and poor individuals, they are silent regarding the reasons. One possibility is that tastes for healthy food changed differently for rich and poor individuals (or that the rich were more receptive or attentive to "eating healthy" campaigns). Another possibility is that salience was similar but the higher price of healthy food or its lower availability in poor neighborhoods represent significant "barriers to entry" in healthier eating habits for poor individuals.

Durable Goods

Consumer durables reflect an element of standard of living that may not be captured by current spending (as people buy them infrequently). Hence, another way to look at consumption inequality is to see how many and what type of households own certain home appliances and durable goods. Also in this case—as we did in the discussion of food consumption—we stress that the quality of what is being consumed or purchased can matter substantially in thinking about the welfare consequences of inequality.

Figure 3

Share Owning Durables in the Top and Bottom Income Deciles



Source: Authors using data from the Consumer Expenditure Survey
Note: For different categories of durables, the figure compares ownership rates of the bottom and the top after-tax income deciles.

Given that the Panel Study of Income Dynamics has no information on durables except cars, we use the Consumer Expenditure Survey, which contains consistent series on durable goods ownership over a long period time. For some appliances like refrigerators, washing machines, and others, "availability" is probably a more appropriate term than "ownership" if such items are attached to the housing unit. Figure 3 offers a comparison of ownership rates for the bottom and the top after-tax income deciles. For some categories, we have a long series from 1984 to 2012; for others, the series starts in 1989. For most categories, there is evidence of catching up. For example, at the beginning of the time period, ownership of cooking durables (stoves or microwaves) and refrigerators is almost universal among the top 10 percent households, while the proportion of households in the bottom income decile owning such appliances is below 90 percent. For refrigerators at the start of the period, the difference is less but still noticeable. By 2012, these differences have largely disappeared. While there is convergence for these categories, the catch-up rates for dishwashers and for washers and dryers are much slower. Ownership of cars has also converged, albeit at a slower rate than food-related appliances. Finally, there is only a small difference in the fractions owning entertainment durables (TVs, sound systems, DVD players, PCs, and so on) throughout the entire period.

There are two caveats to these findings. First, a convergence in ownership does not imply convergence in the number of appliances owned. Indeed, for the durables for which this information is available (vehicles and entertainment), there is no evidence of convergence. Moreover, there may be a large quality difference between high-end and low-end appliances, but the existing data are not rich enough to measure the quality of the durables owned by socioeconomic status.

Inequality in the Consumption of Leisure

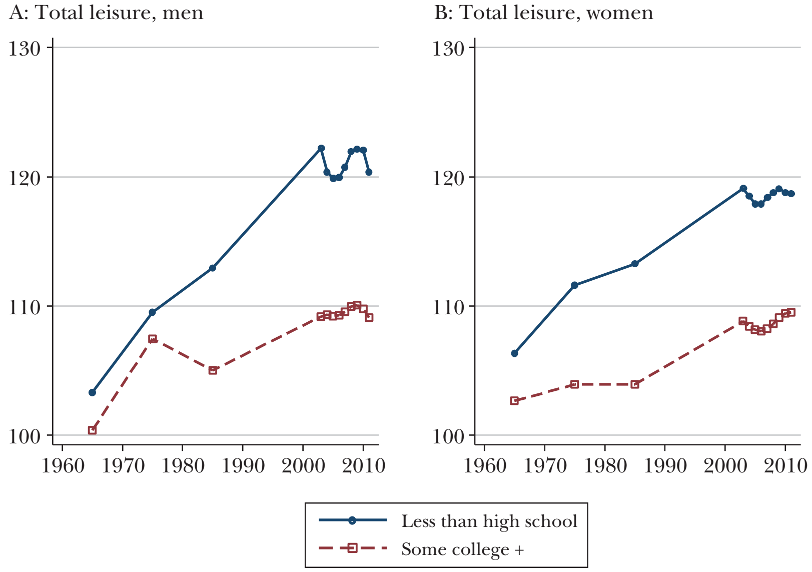
Economists traditionally write the utility function of individuals as comprising consumption and leisure. Perhaps greater inequality in the consumption of goods and services is being partially offset by greater equality in leisure time?

Measuring leisure is complex. Aguiar and Hurst (2009) have looked at trends in time use as a way of measuring leisure time. We follow a similar strategy here. In particular, we use surveys collecting information on time use over the last 50 years: the 1965–66 Americans' Use of Time, the 1975–76 Time Use in Economics and Social Accounts, the 1985 American Use of Time, and the 2003–11 integrated American Time Use Survey. The datasets do not include detailed or consistent information on income or on other measures of economic resources. Thus, we will use education as a rough measure of socioeconomic status.

To display the sharpest differences, we consider only the top and bottom education categories: individuals with less than a high school degree, and those with at least some college. In these datasets, people report the number of minutes they spend in various activities in the previous 24 hours. The main time use categories are: "work" (including time spent searching for jobs), "chores" (all household activities such as cooking, cleaning, and others), "child care," "social" (watching sports, going to movies, partying, and so on), "organizational" (for example, volunteering, religion), "personal care" (sleeping, eating, and so on), "shopping," "education," "active leisure" (sport activities, playing games, and others), and "passive leisure" (watching television, listening to radio, relaxing, and others). All figures are weighted with sampling weights (except 1965–75 where no weights were released) and expressed in hours per week.

In Figure 4, we plot trends in total leisure time—the sum of social activities, active and passive leisure, and time devoted to personal care—controlling for the day of the week the diary was filled in. This measure of leisure may not be without problems. For example, it includes time spent assisting or helping adult household members (which for some people may represent "chores"); it excludes gardening or cooking (which for some individuals may represent a form of leisure). However, excluding personal care does not affect the main trends. The less-than-high-school education group has more leisure, and in the last few decades most of the growth of leisure has happened for this group, too. Moreover, growth has been stronger for men than women.

Figure 4
Trends in Total Leisure Time

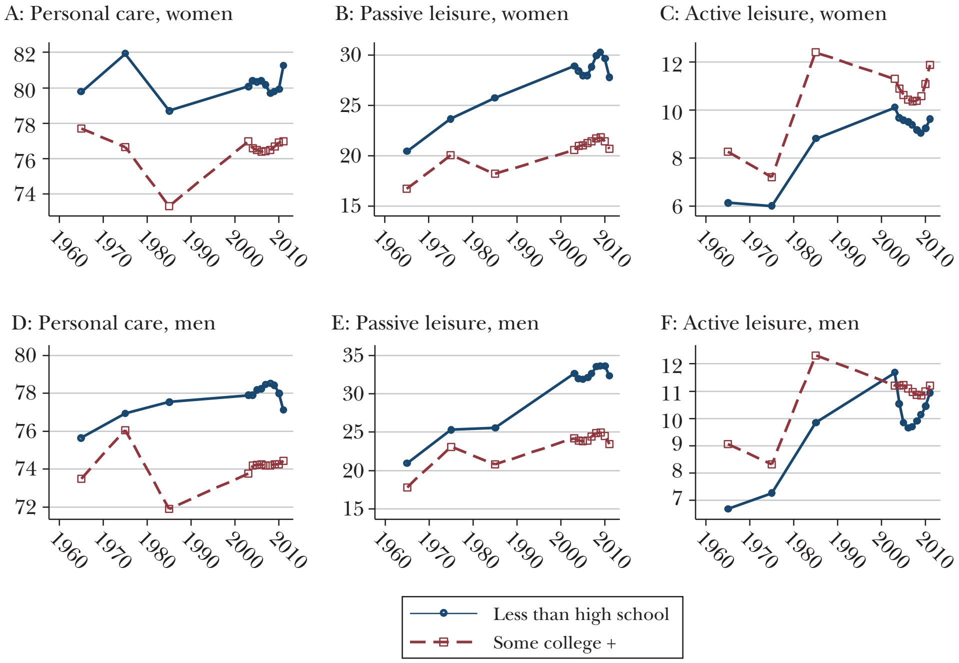


Source: Authors using data from the 1965–66 Americans' Use of Time, the 1975–76 Time Use in Economics and Social Accounts, the 1985 American Use of Time, and the 2003–11 integrated American Time Use Survey.
Note: In the figure, we plot trends in total leisure time—the sum of social activities, active and passive leisure, and time devoted to personal care.

How do we interpret these trends? It is tempting to argue that the rise in consumption inequality has been to some extent counterbalanced with increasing leisure time among the poor, implying that the increase in the inequality of well-being is less severe than what consumption data alone may suggest. But any such conclusion needs to be hedged around with cautions. Some of the increase in leisure is involuntary, due to lack of job market opportunities. Excluding recession years from the analysis gives the same broad picture of Figure 4. Moreover, if we repeat the analysis only for the employed, we find that the differences are smaller but still significant (especially in the 2000s). Of course, an analysis that conditions on employment does not solve the problem of making welfare comparisons across income groups that include leisure, as employment itself results from and is affected by a combination of supply and demand factors. It is also possible that there is substantial heterogeneity in preferences for leisure—which may also help to explain different educational choices in the first place.

What component of leisure time is driving these trends? In Figure 5, we decompose total leisure into three components: personal care, active leisure (plus social activities), and passive leisure. There are some notable trends. First, for both low-educated women and (especially) men, the increase in total leisure time visible from

Figure 5
Decomposing Total Leisure
(in hours per week)



Source: Authors using data from the 1965–66 Americans' Use of Time, the 1975–76 Time Use in Economics and Social Accounts, the 1985 American Use of Time, and the 2003–11 integrated American Time Use Survey
Note: Here we decompose total leisure into three components: personal care, active leisure (plus social activities), and passive leisure.

Figure 4 is coming primarily from an increase in time devoted to passive leisure activities. Second, time spent on active leisure and social activities also increases, resulting in greater similarity between high- and low-educated individuals. Finally, time spent on personal care is stable. It is possible that these changes represent an evolution of preferences for leisure across education (income) groups. It is also possible that increasing availability of durable goods in the lower-income groups (documented earlier) frees up time previously devoted to housework.

Consumption Mobility

While there is a popular image of the United States as a land where high rates of mobility across the income distribution are possible, in practice intergenerational income mobility has not changed much over the last 40 years (Chetty, Hendren, Kline, and Saez 2014). In fact, some European countries display more intergenerational income mobility than does the United States (Black and Deveraux 2011). There are also vast geographical differences in mobility across US regions.

Do the trends in intergenerational mobility in consumption mirror those found for income? To study this topic, we need longitudinal information on consumption that follows multiple generations. The Panel Study of Income Dynamics offers such data. In particular, for each household, it follows the children, the "splitoff" households, when they leave the parental home. A few authors have looked at the intergenerational dimension of the PSID data in the context of risk-sharing within the family (for example, Hayashi, Altonji, and Kotlikoff 1996; Attanasio, Hurst, and Pistaferri 2015).

We first construct a measure of household consumption using the Panel Study of Income Dynamics data. Specifically, we define consumption in this data as the sum of spending on food, rent, health, home insurance, utilities, car insurance, car repair, gasoline, parking and transportation, education, child care, clothing, vacation, and entertainment (with the last three categories only available since 2005). We add the monetary value of food stamps and imputed rents for homeowners and free-rent households. To obtain a measure of consumption and income, we deflate by the Consumer Price Index and as before use the OECD adult equivalence scale. This measure of consumption is available for the survey years 1999–2013. To reduce the impact of measurement error, we take moving averages across three subsequent surveys for both consumption and income.

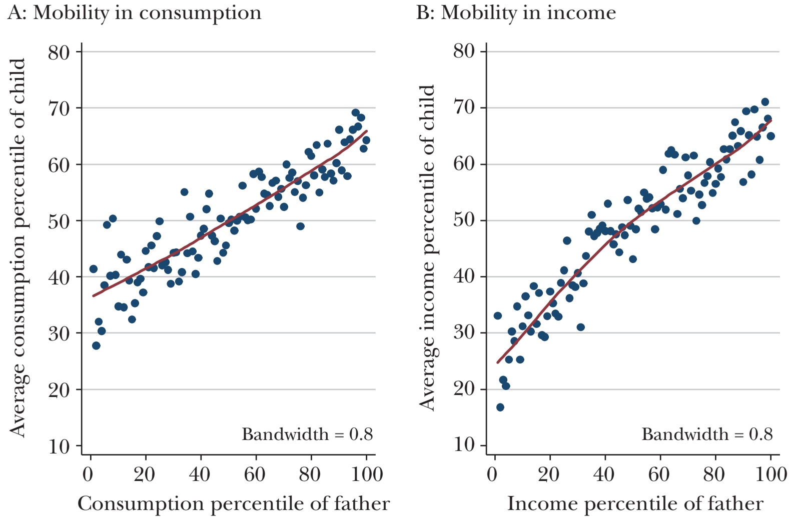
For each year in which the household is observed, we compute the percentile occupied by the household relative to the head's reference birth cohort (born in the 1900s, 1910s, and so on). We do this for the father and his children. Next, we look graphically at the relationship between the average percentile occupied by the children and the percentile occupied by the father. If there is no relationship between the ranks of parents and children, the (local) regression lines we plot separately for consumption and income in Figure 6 should be flat; on the other hand, perfect correlation between the ranks of parents and children would give a 45-degree line.[4]

We find that the slope of the local regression line for income gradient is higher than that for consumption, implying greater intergenerational mobility in consumption than income. This finding is especially true at the bottom of the distribution. Hence, as consumption is more equally distributed than income, there is also more intergenerational mobility when looking at consumption than income.[5]

[4] If we follow the suggestion of Chetty, Hendren, Kline, and Saez (2014) of conditioning on the age of parent and child, we get similar, though less-precise, results. Interestingly, the relationship we plot in the right panel of Figure 6 is remarkably similar to that reported by Chetty et al., despite the enormous differences in sample sizes.

[5] Wodon and Yitzhaki (2002) extend the traditional Sen (1973) welfare function to the dynamic case. Sen's welfare function increases with aggregate income and declines with inequality. Wodon and Yitzhaki's social welfare function increases with intergenerational mobility. Hence, social welfare is higher when considering consumption than when considering income not only because of less unequal distribution of consumption (relative to income), but also because of higher intergenerational mobility.

Figure 6
Intergenerational Mobility in Consumption and Income



Source: Authors using Panel Study of Income Dynamics data.
Note: If there is no relationship between the ranks of parents and children, the regression lines we plot separately for consumption and income should be flat; on the other hand, perfect correlation between the ranks of parents and children would give a 45-degree line. See text for details.

The explanation of the former phenomenon is in all likelihood the tendency to smooth-out income shocks whenever possible (through saving and borrowing, public programs, or informal mechanisms). As for intergenerational mobility, one can conjecture that parents transfer genetic endowments of ability (which will be reflected in both consumption and income) as well as preferences (which will be reflected primarily in consumption). However, the extent of similarity between the consumption of parents and the consumption of children also depend on the credit and insurance market frictions faced by the two generations. (Similarities can also depend on the point of the life-cycle we are observing father and child, but we neglect this complication here.) In the end, whether there is more or less intergenerational mobility in income or consumption is an empirical matter, and the data we present constitute one of the first pieces of relevant evidence in this regard.

Conclusion

The goal of this paper has been to discuss what we do and do not know about the evolution of consumption inequality in the United States, while contrasting it with trends in income inequality. There is now some cumulating evidence showing that increasing disparities in income are approximately replicated by increasing disparities in consumption. These findings suggest that a substantial portion of the nature of the shocks to income and wages that have generated the observed and well-documented increase in income inequality over the last 35 years should be viewed as permanent rather than temporary, and that households have only a limited ability to absorb such shocks for more than a short period.

While much attention in discussions of inequality has been given to the top of the income and consumption distribution, the left tail is also of considerable interest, both from a scientific and policy point of view. We have considered different components of consumption, inequality in leisure, and also the intergenerational transmission of consumption inequality. When looking at individual components, some of them show greater equality, and some raise difficult questions of how to adjust for quality changes. Ownership of major durables, which in principle raise living standards, has also been converging rapidly between low- and high-permanent-income households. While inequality in food consumption has increased, there is little evidence of growing inequality in caloric intakes—partly as a result of assistance provided by government programs (like food stamps) supplementing private spending, partly from price declines of some food items, and probably in part because low-income people spend more time searching for lower prices. The latter trend involving time use is actually more general: the consumption of leisure has increased among low-socioeconomic status individuals at a faster pace than among the higher educated.

What do we conclude about whether disparities in well-being have increased? Our opinion is that, despite the fact that some studies have suggested the opposite, inequality in the consumption of nondurables and services has increased substantially over the last few decades and has paralleled the increase in inequality in income and earnings. A consequence of this is that the increase in income inequality is reflected in an increase in inequality in welfare and well-being.

Some important caveats, however, are in order. We have provided evidence on specific goods, leisure, durable ownership, and even mobility across generations that points to relative utility gains realized by the lower-income groups. These relative gains arise as a consequence of lower prices for the goods they typically purchase, increasing availability of leisure time, increasing durable ownership, and improved consumption opportunities for their children. Obviously, assigning a value to these utility gains is hard, and we do not attempt to do this. It should also be pointed out that most of these gains have been in quantity terms, not quality terms.

This discussion suggests that if using consumption is in principle a better way than income to measure the well-being of households, a complete welfare analysis will need to go beyond looking at aggregate categories of household expenditure, and consider in addition the value that people assign to time and the quality of goods they consume, among other factors.

References

Abowd, John M., and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57(2): 411–45.

Aguiar, Mark A., and Mark Bils. 2011. "Has Consumption Inequality Mirrored Income Inequality?" NBER Working Papers 16807.

Aguiar, Mark, and Mark Bils. 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105(9): 2725–56.

Aguiar, Mark, and Erik Hurst. 2005. "Consumption vs. Expenditure." *Journal of Political Economy* 113(5): 919–48

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.

Attanasio, Orazio, Erich Battistin, and Hidehiko Ichimura. 2007. Chap. 17 in "What Really Happened to Consumption Inequality in the United States?" In: *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, edited by Ernst E. Berndt and Charles R. Hulten. National Bureau of Economic Research.

Attanasio, Orazio P., and Steven J. Davis. 1996. "Relative Wage Movements and the Distribution of Consumption." *Journal of Political Economy* 104(6): 1227–62.

Attanasio, Orazio P., Erik Hurst, and Luigi Pistaferri. 2015. "The Evolution of Income, Consumption, and Leisure Inequality in the US, 1980–2010." Chap. 4 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Attanasio Orazio, and Luigi Pistaferri. 2014. "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure." *American Economic Review* 104(5): 122–26.

Attanasio, Orazio, and Guglielmo Weber. 2010. "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy." *Journal of Economic Literature* 48(3): 693–751.

Autor, David, Andreas Ravndal Kostøl, and Magne Mogstad. 2015. "Disability Benefits, Consumption Insurance, and Household Labor Supply." https://bfi.uchicago.edu/research/working-paper/disability-benefits-consumption-insurance-and-household-labor-supply-0.

Aydin, Deniz. 2015. "The Marginal Propensity to Consume Out of Liquidity." http://stanford.edu/~daydin/DAydin_MPCL.pdf.

Baker, Scott R. 2014. "Debt and the Consumption Response to Household Income Shocks." April. http://web.stanford.edu/~srbaker/Papers/Baker_DebtConsumption.pdf.

Bee, Adam, Bruce D. Meyer, and James X. Sullivan. 2015. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Black, Sandra, and Paul Deveraux. 2011. "Recent Developments in Intergenerational Mobility." Chap. 16 in *Handbook of Labor Economics*, vol. 4, Part B, edited by Orley Ashenfelter and David Card. North Holland Press, Elsevier.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *American Economic Review* 98(5): 1887–1921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106(2): 387–435.

Broda, Christian, Ephraim Leibtag, and David E. Weinstein. 2009. "The Role of Prices in Measuring the Poor's Living Standards." *Journal of Economic Perspectives* 23(2): 77–97.

Broda, Christian, and David E. Weinstein. 2008. *Prices, Poverty and Inequality: Why Americans Are Better Off Than You Think*. Washington, DC: AEI Press.

Browning, Martin, and Thomas F. Crossley. 2001. "The Life-Cycle Model of Consumption and Saving." *Journal of Economic Perspectives* 15(3): 3–22.

Browning, Martin, and Søren Leth-Petersen. 2003. "Imputing Consumption from Income and Wealth Information." *Economic Journal* 113(488): F282–F301.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." NBER Working Paper 19843.

Crossley, Thomas F., and Joachim K. Winter. 2015. "Asking Households about Expenditures: What Have We Learned?" Chap. 1 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Cutler, David M., and Lawrence F. Katz. 1991. "Macroeconomic Performance and the Disadvantaged." *Brookings Papers on Economic Activity* no. 2, pp. 1–74.

Cutler, David M., and Lawrence F. Katz. 1992. "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980s." NBER Working Paper 3964.

Deaton, Angus, and Margaret Grosh. 2000. "Consumption." Chap. 5 in *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, Vol. 1, edited by Margaret Grosh and Paul Glewwe. Washington, DC: World Bank.

DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns." *Brookings Papers on Economic Activity*, Spring, pp. 67–142.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2015. "Consumption Network Effects." November 23. https://www.economics.utoronto.ca/index.php/index/research/downloadSeminarPaper/60199.

Dynarski, Susan, and Jonathan Gruber. 1997. "Can Families Smooth Variable Earnings?" *Brooking Papers on Economic Activity* 1: 229–305.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton University Press.

Gelman, Michael, Shachar Kariv, Matthew Shapiro, Dan Silverman, and Steven Tadelis. 2014. "Harnessing Naturally Occurring Data to Measure the Response of Spending to Income." *Science*, July 11, 345(6193): 212–15.

Gottschalk, Peter, and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity* 25(2): 217–72.

Griffith, Rachel, Ephraim Leibtag, Andrew Leicester, and Aviv Nevo. 2008. "Timing and Quantity of Consumer Purchases and the Consumer Price Index." NBER Working Paper 14433.

Gross, David B., and Nicholas S. Souleles. 2002. "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data." *Quarterly Journal of Economics* 117(1): 149–85.

Guvenen, Fatih, Serdar Ozkan, and Jae Song. 2014. "The Nature of Countercyclical Income Risk." *Journal of Political Economy* 122(3): 621–60.

Handbury, Jessie. 2014. "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities." Unpublished paper.

Hayashi, Fumio, Joseph Altonji, and Lawrence Kotlikoff. 1996. "Risk-Sharing between and within Families." *Econometrica* 64(2): 261–94.

Heathcote, Jonathan, Fabrizio Perri, and Gianluca L. Viuolante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13(1): 15–51

Kaplan, Greg, and Guido Menzio. 2015. "The Morphology of Price Dispersion." *International Economic Review* 56(4): 1165–1206.

Kaplan, Greg, Guido Menzio, Leena Rudanko, and Nicholas Trachter. 2016. "Relative Price Dispersion: Evidence and Theory." NBER Working Paper 21931.

Koijen, Ralph, Stijn Van Nieuwerburgh, and Roine Vestmanz. 2015. "Judging the Quality of Survey Data by Comparison with 'Truth' as Measured By Administrative Records: Evidence from Sweden." Chap. 11 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics* 125(1): 91–128.

Krueger, Dirk, and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies* 73(1): 163–93.

MaCurdy, Thomas E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18(1): 82–114.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4): 199–226.

Meyer, Bruce D., and James X. Sullivan. 2013. "Consumption and Income Inequality in the U.S. since the 1960s." https://www3.nd.edu/~jsulliv4/Inequality3.6.pdf.

Modigliani, Franco, and Richard Brumberg. 1955. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." Chap. 15 in *Post Keynesians Economics*, edited by Kenneth K. Kurihara. Rutgers University Press.

Nevo, Aviv, and Arlene Wong. 2015. "The Elasticity of Substitution between Time and Market Goods: Evidence from the Great Recession." http://sites.northwestern.edu/awo760/files/2015/10/Paper_Aug19_2015-y5zgke.pdf.

Passero, William, Thesia I. Garner, and Clinton McCully. 2015. "Understanding the Relationship: CE Survey and PCE." Chap. 6 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Pistaferri, Luigi. 2015. "Household Consumption: Research Questions, Measurement Issues, and Data Collection Strategies." *Journal of Economic and Social Measurement* 40(1–4): 123–49.

Sabelhaus, John, and Jeffrey A. Groen. 2000. "Can Permanent-Income Theory Explain Cross-Sectional Consumption Patterns?" *Review of Economics and Statistics* 82(3): 431–38.

Sabelhaus, John, David Johnson, Stephen Ash, David Swanson, Thesia I. Garner, John Greenlees, Steve Henderson. 2015. "Is the Consumer Expenditure Survey Representative by Income?" Chap. 8 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. University of Chicago Press.

Sen, Amartya. 1973. *On Economic Inequality*. Clarendon Press.

Singh, Gopal K., Mohammad Siahpush, Robert A. Hiatt, and Lava R. Timsina. 2011. "Dramatic Increases in Obesity and Overweight Prevalence and Body Mass Index among Ethnic-Immigrant and Social Class Groups in the United States, 1976–2008." *Journal of Community Health* 36(1): 94–110.

Singh, Rajni, Berdine R. Martin, Yvonne Hickey, Dorothy Teegarden, Wayne W. Campbell, Bruce A. Craig, Dale A. Schoeller, Deborah Anne Kerr, and Connie M. Weaver. 2009. "Comparison of Self-Reported, Measured, Metabolizable Energy Intake with Total Energy Expenditure in Overweight Teens." *American Journal of Clinical Nutrition* 89(6): 1744–50.

Slesnick, Daniel T. 1991. "The Standard of Living in the United States." *Review of Income and Wealth* 37(4): 363–86.

Slesnick, Daniel T. 1994. "Consumption, Needs and Inequality." *International Economic Review* 35(3): 677–703.

Slesnick, Daniel T. 2001. *Consumption and Social Welfare*. Cambridge Books.

Wang, Dong D., Cindy W. Leung, Yanping Li, Eric L. Ding, Stephanie E. Chiuve, Frank B. Hu, and Walter C. Willett. 2014. "Trends in Dietary Quality among Adults in the United States, 1999 through 2010." *JAMA Internal Medicine* 174(10): 1587–95.

Wodon, Quentin, and Shlomo Yitzhaki. 2002. "Inequality and Social Welfare." *A Sourcebook for Poverty Reduction Strategies*, Vol. 1: *Core Techniques and Cross-Cutting Issues*, edited by Jeni Klugman, 75–104.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 29–52

Mortality Inequality: The Good News from a County-Level Approach

Janet Currie and Hannes Schwandt

Life expectancy for the US population has shown a strong increase since 1990. The rise in life expectancy at birth holds for both men and women, as shown in Figure 1. This development has not been driven solely by improvements in life expectancy at older ages. Mortality rates for those under one year of age, for the age group 1–4, and for every five-year age group above that level, declined for both males and females between 1990 and 2010.[1] Particularly pronounced improvements in mortality occurred at younger ages, which tend to be age groups in which deaths occur predominantly among the poor.

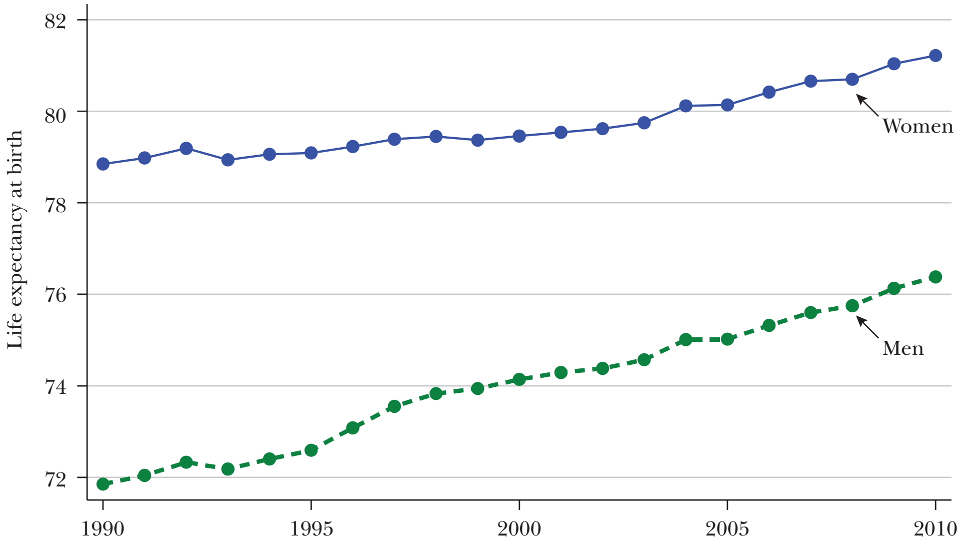
However, this overall decline in mortality rates has been accompanied by prominent recent studies highlighting that the gains have not been distributed equally (for example, Cutler, Lange, Meara, Richards-Shubik, and Ruhm 2011; Chetty et al. 2015; National Academies of Science, Engineering, and Medicine (NAS) 2015; Case and Deaton 2015). Indeed, several studies argue that when measured across educational groups and/or geographic areas, mortality gaps are not only widening, but that for some US groups, overall life expectancy is even falling (Olshansky et al. 2012; Wang, Schumacher, Levitz, Mokdad, and Murray 2013; Murray et al. 2006). It seems to have become widely accepted that inequality in life expectancy is increasing. Given that the number of years that one can expect to live is such an

[1] These one-year mortality rates are shown for 1990, 2000, and 2010 in Appendix Figure A1, available online with this paper at http://e-jep.org.

■ *Janet M. Currie is the Henry Putnam Professor of Economics and Public Affairs, Princeton University, Princeton, New Jersey. Hannes Schwandt is Assistant Professor of Economics, University of Zurich, Zurich, Switzerland. Their email addresses are jcurrie@princeton.edu and hannes.schwandt@uzh.ch.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at http://dx.doi.org/10.1257/jep.30.2.29 doi=10.1257/jep.30.2.29

Figure 1
Life Expectancy at Birth by Gender and Year



Source: Authors using data from HMD (2015).

important indicator of welfare, this finding has been heralded as yet another dimension in which overall societal inequality is increasing.

In this essay, we ask whether the distributions of life expectancy and mortality have in fact become generally more unequal. Focusing on groups of counties ranked by their poverty rates, we show that gains in life expectancy *at birth* have actually been relatively equally distributed between rich and poor areas. Analysts who have concluded that inequality in life expectancy is increasing have generally focused on life expectancy at age 40 to 50. This observation suggests that it is important to examine trends in mortality for younger and older ages separately.

Turning to an analysis of age-specific mortality rates, we show that among adults age 50 and over, mortality has declined more quickly in richer areas than in poorer ones, resulting in increased inequality in mortality. This finding is consistent with previous research on the subject. However, among children, mortality has been falling more quickly in poorer areas with the result that inequality in mortality has fallen substantially over time. This is an important result given the growing literature showing that good health in childhood predicts better health in adulthood (Currie and Rossin-Slater 2015). Hence, today's children are likely to face considerably less inequality in mortality as they age than current adults.

We also show that there have been stunning declines in mortality rates for African Americans between 1990 and 2010, especially for black men. The fact that inequality in mortality has been moving in opposite directions for the young and the old, as well as for some segments of the African-American and non-African-American populations, argues against a single driver of trends in mortality inequality, such as rising income inequality. Rather, there are likely to be multiple specific causes affecting different segments of the population.

In what follows, we first provide a brief overview of the literature on inequality in mortality. This is followed by a discussion of our methods, data, and main results. The end of this paper offers some hypotheses about causes for the results we see, including a discussion of differential smoking patterns by age and socioeconomic status. These patterns may explain a significant fraction of the increase in mortality inequality in older cohorts.

Background

Is There a Causal Relationship Between Inequality in Income and in Mortality?

It is no accident that the resurgence of interest in mortality inequality has followed growing public interest in income inequality. The two are linked in the minds of the public and many academics (Marmot et al. 1991; Wilkinson 1996). There is no doubt that lower socioeconomic status tends to be associated with higher mortality; Kitigawa and Hauser (1973) showed this relationship more than 40 years ago. However, this insight does not mean that increases in income inequality must inevitably widen differentials in mortality regardless of actual income levels or other relevant policies. Indeed, given that much of the recent increase in economic inequality is at the very top of the income distribution, it is not immediately obvious why it should result in increases in deaths for other groups.

In the academic literature, the idea that rising income inequality must necessarily lead to rising inequality in mortality has been vigorously disputed. Deaton and Paxson (2001) show that there is no necessary relationship between trends in income inequality per se and mortality trends, and that in fact, the two moved in opposite directions for much of the twentieth century. Gravelle (1998) argues empirically that places with a lot of income inequality also tend to have a lot of poverty, and that poverty and not income inequality is causally related to higher mortality.

In this journal, Smith (1999) argues that health may be driving income differences rather than the reverse, at least among adults. Similarly, Case and Paxson (2011), in their reanalysis of data from the Whitehall studies of British civil servants, show that poor health in childhood causes lower socioeconomic status in adulthood, rather than lower socioeconomic status causing poor health in adulthood. An important possible explanation may be that health trajectories are established early in childhood (Almond and Currie 2011; Smith 2007). From this perspective, mortality differentials that are seen today among middle-aged and older adults likely had their roots decades ago. Aizer and Currie (2014) show that the health of infants in the lowest socioeconomic status groups is catching up to those of higher status groups and argue that this convergence likely reflects a range of recent policies that have improved the prospects of these children.

An overall reading of this evidence suggests that it is not at all obvious how one should expect trends in mortality inequality to have evolved over the past 20 years. While income inequality has increased greatly over this period, there have been dramatic changes in access to health insurance among pregnant women and children, as well as a sea change in societal attitudes towards smoking. Fenelon and Preston (2012) place particular emphasis on smoking, estimating that about 20 percent of US mortality may be attributed to smoking, and that there are deleterious effects even on those "ever smokers" who have not smoked for many years. There have also been tremendous increases in obesity rates and addiction to prescription painkillers, as well as the rise (followed by the subsequent imperfect control of) HIV/AIDS.[2]

Complicating matters further, many of the health-related behaviors that are associated with lower socioeconomic status contemporaneously—like smoking, drinking, or overeating—do not explain differences in health at the population level. For example, Banks, Marmot, Oldfield, and Smith (2006) show that British citizens have lower morbidity than Americans even though they tend to smoke and drink more and are almost equally likely to be overweight. The question of how inequality in mortality has evolved cannot be readily inferred from the mixture of other social trends about income inequality or behavior and instead must be estimated directly.

Approaches to Measurement of Inequality of Mortality

An important point to keep in mind is that although life expectancy sounds like it measures the number of years that a particular cohort can expect to live, it is unlikely to do so. It is easiest to understand the problem with a concrete example. Suppose we are interested in the life expectancy of the cohort that is currently 20 years old. Life expectancy is computed using the assumption that when this cohort reaches age 40, it can expect to live the same number of years as the cohort that is currently 40 years old. It is easy to see that if age-specific mortality rates are changing over time, then this assumption will be false. Only in a world where mortality rates are static does life expectancy at a given age mean what most people think that it does. For this reason, we focus most of our attention in what follows on age-specific mortality rates.

A recent NAS (2015) report lays out three common approaches to measuring inequality in mortality: "One looks at differences in the mortality of populations of U.S. counties in relation to county-level economic measures. Another looks at mortality by educational attainment. A third approach looks at mortality by career earnings." One reason for the multiplicity of approaches is that each has weaknesses.

The most popular method for measuring inequality in mortality involves splitting the population either by education or by income level. Examples of studies

[2] Crime is not likely to be large factor over the time frame considered here, since during this period crime was low by historical standards. See "Gun Homincide and Violent Crime" at Pew Research Center, May 7, 2013, http://www.pewsocialtrends.org/2013/05/07/gun-crime.

looking at inequality in mortality by education level include Pappas, Queen, Hadden, and Fisher (1993), Elo and Preston (1996), Preston and Elo (1995), Olshansky et al. (2012), Meara, Richards, and Cutler (2008), Cutler et al. (2011), Montez and Berkman (2014), and Montez and Zajacova (2013). The main difficulty with this approach is that the share of the population in different educational categories has changed dramatically over time (Dowd and Hamoudi 2014; Hendi 2015; Bound, Geronimus, Rodriguez, and Waidmann 2014; Goldring, Lang, and Shubik 2015). For example, the share of white, non-Hispanic women aged 25–84 who had less than a high school degree fell by remarkable 66 percent between 1990 and 2010.[3] Moreover, it is likely that those women who would have been high school dropouts in 1990, but who now have higher levels of education, are of higher socioeconomic status and/or ability than those who remain in the high school dropout category. They might therefore have been expected to have better health in any case.

Thus, the observed decline in life expectancy among white, non-Hispanic, high school dropout women highlighted in Olshansky et al. (2012) could be mostly or entirely accounted for by changes in the composition of this group. Bound et al. (2014) address the issue by categorizing education in terms of relative rank in the overall distribution and focusing on the bottom quartile of the education distribution. They find no evidence that survival probabilities declined in the bottom quartile of the education distribution.

Another group of studies have examined mortality inequality by relative income levels (NAS 2015; Pappas et al. 1993; Waldron 2007, 2013; Bosworth and Burke 2014; Pijoan-Mas and Ríos-Rull 2014). These studies are subject to the concerns mentioned earlier about possible reverse causality—that is, the idea that economic hardship could be caused by ill health rather than vice-versa (Smith 1999, 2007). Moreover, these analyses have been limited by the fact that many potential data sources for mortality rates do not include information on income or earnings.

The Health and Retirement Study, which follows a representative sample of the US elderly population who can be linked to Social Security earnings histories, does include information on both income and mortality. Studies based on the HRS find increasing divergence in life expectancy at age 50 by income over time (for example, NAS 2015; Bosworth and Burke 2014; Pijoan-Mas and Ríos-Rull 2014). However, these analyses are constrained by the limited age ranges that are observed in different years due to the cohort structure of the data, as well as by small sample sizes; for example, the sample used in the NAS (2015) study includes 5,740 deaths, compared to the 21 million deaths analyzed in our study. These data limitations mean that strong assumptions are required to estimate and project life expectancy trends in socioeconomic subgroups, given that some subgroups have very few deaths.

The strategy we pursue here is to examine inequality by geographical areas, such as counties. One concern here is that analyses based on geographic areas are potentially subject to bias due to selective migration. For example, if the most healthy and able-bodied people tend to leave lower-income counties over time and

[3] See Appendix Table A1, available with this paper at http://e-jep.org, for details of these calculations.

migrate to higher-income counties, we might expect to see mortality increase in the poor counties and decrease in the rich counties even if in fact each individual's health remained exactly the same.

Some previous studies have taken this geographical approach (for example, Wilmoth et al. 2011; Sing and Siahpush 2006; Kulkarni et al. 2011; Wang et al. 2013; Murray et al. 2006), but concerns over migration and other changes in geographic areas can make the results tricky to interpret. For example, Wang et al. (2013) find that female life expectancy decreased by 2.68 years in the counties with the sharpest declines in life expectancy between 1985 and 2010, while it increased by 6.16 years in the counties with the largest gains in life expectancy. However, our calculations show that during this time period, the population fell 6 percent in the counties with the largest mortality declines, while the population of the top counties grew on average by 101 percent, making it extremely difficult to interpret these trends.[4]

Singh and Siahpush (2006) divide counties based on a socioeconomic index for the population in 1980 and follow these same county groups up to 2000. We follow a similar strategy here, although we reorder the counties in each Census year to insure that we are always comparing poor counties to rich ones. In practice, this refinement does not have much effect on our estimates. Our approach differs from theirs in that we examine age-specific mortality in addition to life expectancy.

In an exceptional recent paper, Chetty et al. (2015) categorize individuals by both income and location. Using tax data, they first examine mortality by percentiles of the income distribution for each age from 40 to 76. Their preferred specification uses income from two years ago in order to reduce the chance that their results are driven by reverse causality from health to inequality of income, but they show that they would obtain similar results using income from five and ten years ago. Using this information, they calculate survival curves and extrapolate them to age 90. These data are then used to calculate life expectancy at age 40 for each quartile of the income distribution, in different locations using either counties or clusters of counties that make up commuting zones. They conclude that "low-income individuals live longest in affluent cities with more educated people and higher local government expenditures. … [and] low-income individuals on both coasts experienced annual gains in life expectancy of approximately .3 years, comparable to the mean gain in the U.S. for the highest income individuals."[5] This work does show significant declines in life expectancy at age 40 among low-income individuals in some places, including Nevada, Appalachia, and southern Ohio. However, the results also suggest that some features of location boost health at least as much among the poor as among the rich.[6]

[4] These calculations are presented in Appendix Table A1, available with this paper at http://e-jep.org.

[5] Some limitations of the Chetty et al. (2015) approach include the fact that income is not observed for non-earners and race is not observed while (as we will show) there have been dramatic improvements in mortality among African Americans.

[6] Costa and Kahn (2015) provide a historical example of how improvements in a location's health environment reduce mortality among the poor more than among the rich by studying citywide clean water interventions and drug availability in New York City and Philadelphia in the early twentieth century.

Hence, the Chetty et al. (2015) study suggests an additional reason that we should be interested in the analysis of inequalities in mortality across geographical areas: There may be features of particular areas (for example, air pollution) that affect everyone living in a particular location. There may also be spillovers from rich to poor (or vice versa) within areas. For instance, if the rich insist on excellent parks and hospitals, then to the extent that the poor are able to live in the same locations, they may also benefit from these resources.[7]

Mortality Rates: Measurement Issues

An intrinsic problem in empirical work with US mortality rates is that the numerator and denominator come from different datasets, with somewhat different and changing measures of key concepts. Death statistics come from the Vital Statistics mortality data, which are collected by each local county registrar-recorder and eventually forwarded to the national government. However, population estimates come from the decennial Census and the American Community Survey. Debate continues as to the quality of the reporting in these sources (Arias, Schauman, Eschbach, Sorlie, and Bucklund 2008). Information on education, Hispanic ethnicity, race, and occupation of the decedent is supposed to be completed by the funeral director on the basis of information provided by an informant (or in the absence of an informant, based on observation). More than one race can be entered on the death certificates, but only the first-mentioned race is recorded in the Vital Statistics files.[8] In 1990, about 8 percent of certificates were missing education—a proportion that fell to 1 percent by 2010. If the missing data are concentrated in lower-education subgroups, then excluding observations with missing values will bias estimated trends in mortality by education.

Changes in the measures over time present a vexing question for the analysis of trends in mortality. For example, the Census now allows each respondent to report more than one race. Similarly, since 2003, an increasing number of states changed from reporting education by years completed to reporting educational degrees as stated on death certificates, while the Census and American Community Survey data continue to report education in years completed.

An important change regarding Hispanics occurred recently in the American Community Survey: In 2008, the wording in the questionnaire changed from "Hispanic" to "Hispanic origin." According to the Census Bureau (undated), this wording change "likely identified Hispanics—mostly native-born—who would not have been captured before."[9] If there is an increasing tendency for people to identify themselves as Hispanic in the American Community Survey, while no changes

[7] A somewhat obvious point about any analysis of inequality between counties is that such an analysis neglects inequality within-county, which may nevertheless be important. Thus, between-county inequality in mortality is only part of the story, albeit an important part.

[8] While there is evidence of a slight general underreporting of Hispanic origin in death certificates (Murphy, Xu, and Kochanek 2013), no systematic changes have occurred over time.

[9] Figure A8 shows that among US-born adults the fraction identifying as Hispanic sharply increased after 2008.

in race reporting occur in the Vital Statistics data, the mortality rate of Hispanics will mechanically decrease and the rate for non-Hispanics will increase, with the impact being larger for Hispanics than for non-Hispanics because the latter group is much larger.

A related issue that could also have a large impact on the size of the denominator is undercounting of undocumented immigrants in the US Census. To the extent that the size of the population is undercounted while the deaths are all counted, mortality rates will tend to be too high. Alternatively, to the extent that an increasing tendency to identify as Hispanic is increasing the size of the denominator used to calculate mortality rates, calculated rates may be too low. In practice, Hispanics are estimated to have the highest life expectancy at birth in the US despite large numbers of both documented and undocumented immigrants. It is not clear what rate would result if both the numerator and denominator were measured accurately and comparably (Arias, Eschbach, Schauman, Bucklund, and Sorlie 2010).

Methods and Data

In our main analysis of mortality and inequality, as in Currie and Schwandt (forthcoming), we rank counties by their poverty rates and then divide the counties into groups that each represent equal 1 percent shares (or equal 5 percent shares) of the overall US population. We do this separately for each Census year. In this way, we compare the 1 (or 5) percent of the population who lived in counties with the highest poverty levels in 1990 to the 1 (or 5) percent of the population who lived in counties with the highest poverty levels in each Census year. This approach accounts for the fact that counties may change poverty rank over time and avoids problems due to shrinking or growing counties by always looking at county bins of similar size.[10]

Our focus on mortality at the level of county groups has advantages beyond the possibility of adjusting for changes in population shares. County of residence is consistently reported both in the Vital Statistics and the Census data, unlike other proxies for socioeconomic status such as education or race. Moreover, grouping counties into equal shares of the population helps to address the problem of measurement error in mortality rates for small counties, in particular when analyzing age ranges with low mortality or racial minorities. We also look at mortality rates over a three-year period in each county, which further helps to minimize noise due to measurement error and to avoid counties reporting zero deaths. A further advantage of our approach is that several socioeconomic indicators are available at the county level. Our baseline specification ranks counties by their poverty rates, but in the online Appendix we also show results for county-level rankings in terms of the

[10] If people systematically left the poorest counties, then over time the population in a fixed group of counties would represent a smaller share of the total population. Moreover, if out-migrants were relatively healthy while the relatively unhealthy remained, then it could appear that health was declining in the poorest counties even if in fact all that was happening was selective out-migration of the healthy.

fraction of the population that are high school dropouts and in terms of median income and life expectancy.

Various issues arise when dividing up counties in this way. For example, dividing counties into groups that represent equal fractions of the population is not an exact procedure because counties at the margin will overlap the bins, making one group too large and the next group too small.[11] In practice, however, this variation in county group size is relatively small, and it is not systematically related to county-level poverty.[12]

Our analysis requires three broad categories of data: on life expectancy; on mortality rates; and about county-level characteristics including the poverty rate, median income, and the fraction of high school dropouts.

For *mortality rates,* we construct age group, gender, and race-specific three-year mortality rates at the level of county groups for the years 1990, 2000, and 2010 based on Vital Statistics mortality data and population counts from the decennial Census. In order to account for changes in the age structure within age groups (for example, the fact that within a group like "over 50" the age distribution can change over time), we age-adjust mortality rates in 2000 and 2010 using the 1990 population. This means we apply the age-specific mortality rates in 2000 and 2010 to the 1990 population, which effectively keeps the age composition within broader age groups constant over time.

The mortality data gives the month of death, which allows us to construct mortality rates based on deaths that occurred after Census Day (April 1). To be specific, the three-year mortality rate in 1990 is the ratio of all deaths that occurred in a cohort between April 1, 1990, and March 31, 1993, divided by the 1990 Census population count. We use the decedent's county of residence, which is what the Census reports, rather than the county where the death occurred.

Following Dorn (2009), we account for changes in county definitions that occurred between 1990 and 2010. Mortality rates by race are constructed using single-race definitions in the 2000 and 2010 Census. We focus on mortality rates in levels and consider there to have been a decline in inequality if the mortality rate

[11] In order to smooth the size of the county groups, we divide the five largest counties in our sample—Cook County, Illinois (which includes the city of Chicago), Los Angeles and Riverside Counties, California, Harris County, Texas (including Houston), and Maricopa County (including Phoenix), Arizona, into five smaller groups, each of identical size and with the identical mortality rates. See online Appendix Figure A2 available with this paper at http://e-jep.org for evidence that the variation in county group size is relatively small. Figure A2 also shows how median income and per capita income vary with the county group poverty ranking.

[12] Most of the poorest counties that together make up 10 percent of the US population in both 1990 and 2010 were located in the South and Southwest, together with some counties in the Midwest (in particular, in South Dakota), and in Alaska. Conversely, the counties with the lowest poverty rates that make up 10 percent of the population in both 1990 and 2010 are predominantly located in the North, with clusters in the Northeast. Thus, the geographic distribution of the counties with the highest and lowest poverty rates remained fairly stable between 1990 and 2010, and in fact, whether we readjust county groups to account for population changes or instead follow fixed sets of poor and rich counties over time, we get similar results.

in poor counties decreased more strongly in absolute terms than the mortality rate in rich counties.[13]

We calculate gender-specific life expectancy at the level of county groups based on one-year mortality rates in 19 age groups (following standard life table techniques, described for example in Chiang 1984).

Finally, county characteristics are taken from the Census (in 1990 and 2000) and from the American Community Survey (ACS) in 2010 (the ACS replaced the long form of the Census). These include: the poverty rate, median income, and the percent of respondents who are high school dropouts.

The Evolution of Inequality in Life Expectancy and Mortality

Inequality is never fully captured by any single all-inclusive measure. Thus, we slice up the data in several ways to present our findings, first looking at life expectancy at birth, then at mortality by age group, and finally at mortality by race and age. Throughout, we show separate estimates for males and females, given that there are profound gender differences in both levels and trends of mortality.

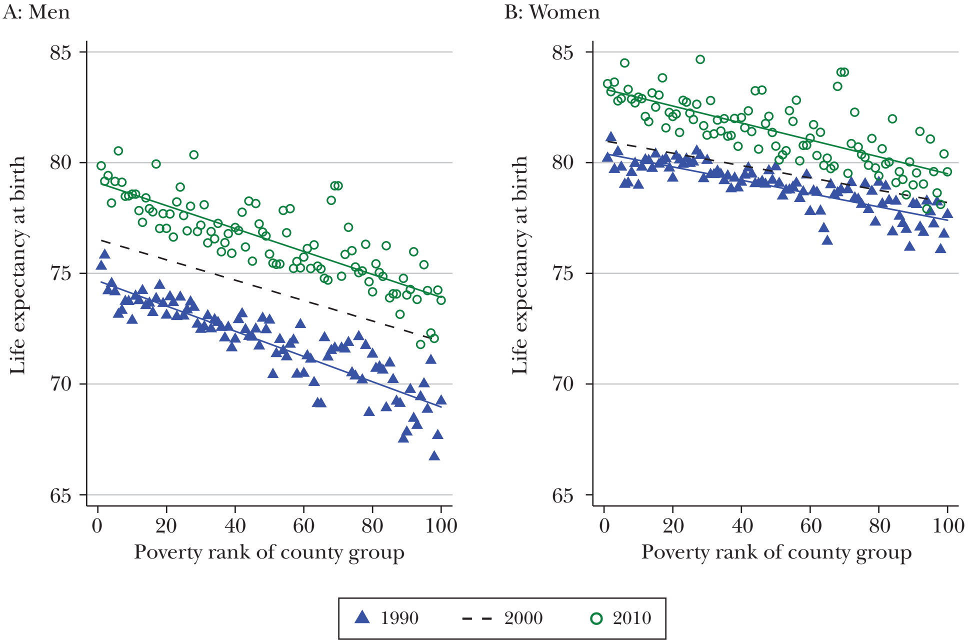
Life Expectancy at Birth

The points in Figure 2 represent measures of life expectancy at the level of county groups. On the horizontal axis, county groups are ranked from those with the lowest percentage of the population in poverty to those with the highest percentage in poverty. The size of each group represents about 1 percent of the population in the relevant year. The vertical axis shows life expectancy at birth, with the left-hand panel showing data for males and the right-hand panel for females. The triangles show the average life expectancy in each county bin in 1990, with a light best-fit regression line drawn through the points. The dashed line shows a fitted regression line for life expectancy at birth in 2000. The circles refer to the 2010 life expectancy at birth (again with a light best-fit regression line drawn through them). The negative slope of each line shows that life expectancy is lower for people in counties with higher poverty rates in each Census year. The fact that the 2010 line is consistently above the 1990 line shows that life expectancy increased in every type of county group, from those with the lowest to the highest poverty rates.

If the slope of the line becomes flatter over time, then this indicates that life expectancy is increasing more in poorer areas than in richer ones, and vice versa. For men, the shift in life expectancy over time is shown by essentially parallel lines, implying that life expectancy increased roughly equally in rich and poor counties and that inequality in life expectancy at birth neither decreased nor increased. For women, increases in life expectancy at birth have been somewhat stronger in the low-poverty county groups resulting in a steepening of the gradient

[13] Because death rates tend to be higher among the poor than among the rich, the same absolute decline in mortality represents a larger percent decline among the rich and vice versa.

Figure 2
Life Expectancy at Birth across Poverty Percentiles



Source: Authors using data from the Vital Statistics, the US Census, and the American Community Survey.
Note: Counties are ranked by their poverty rate in 1990, 2000, and 2010, and divided into groups each representing about 1 percent of the overall population. Each marker represents the life expectancy at birth in a given county group. Lines are fitted using OLS regression. For 2000, markers are omitted and only the regression line is shown. Table A2 provides magnitudes for individual life expectancy estimates and for the slopes of the fitted lines.

between 1990 and 2010, which illustrates a slight increase in inequality.[14] For women between 1990 and 2010, life expectancy at birth in the county group with the lowest poverty rate increased by three years, about one year more than in the county group with the highest poverty rate. However, changes in life expectancy at birth are positive for each county group, with an average improvement in life expectancy of about two years for the county groups with the highest poverty rates. Overall, improvements in life expectancy have been greater for men than for

[14] The online Appendix available with this paper at http://e-jep.org provides additional details. Table A2 provides numerical values. Figure A3 plots changes in life expectancy at birth between 1990 and 2010 across county groups. For men, the slope of the fitted line is 0.0043 with a standard error of 0.0044—which means that the change in the slope is not significantly different from zero. For women, the slope of the corresponding line is –0.009 with a standard error of 0.0038, indicating a small but statistically significant increase in inequality. Figure A3 also shows these changes in percent of the 1990 level. Since males in poor counties have low levels of life expectancy in 1990, the positive change in the poorest groups becomes more pronounced relative to the richer counties, implying a statistically significant decrease in inequality for males according to this measure.

women, implying a strong reduction of the gender gap (a change also visible in Figure 1).

How do these results relate to the findings of an increase in inequality in life expectancy from previous prominent studies such as Chetty et al. (2015) and NAS (2015)? One salient difference in methodology is that those studies focus on life expectancy at older ages. For example, Chetty et al. (2015) use mortality at age 40 to 63 to estimate income-specific trends in life expectancy, while NAS (2015) uses mortality at age 50 to 78, an approach that by construction does not consider developments at younger ages.[15] The next subsection investigates the potential for differences between old and young to influence trends in age-specific mortality.

Age-specific Mortality

Our data allow us to construct death rates at different ages. Figure 3 shows three-year mortality rates at the level of county groups, with counties ranked by the share of their population below the poverty line, for males and females in four different age groups.[16] In these figures, each marker shows the mortality rate for a bin representing 5 percent of the US population in the relevant year. As in the life expectancy figures, a slope that becomes steeper over time implies increasing inequality and vice versa.

Figure 3 shows dramatic reductions in mortality among children aged zero to four between 1990 and 2000, with smaller reductions between 2000 and 2010. Overall, the reductions in under-five mortality were much greater in poorer counties than in richer ones, and slightly larger for males than for females. For example, the under-five mortality rate for males fell from 4.5 per 1,000 in 1990 to 2.4 per 1,000 in the poorest counties, compared to a decline from 2.4 to 1.3 per 1,000 in the richest counties over the same period. Among children aged 5 to 19, there were large reductions in mortality for males, with more modest reductions for females (from already low levels). Once again, reductions were larger in poorer counties, implying significant reductions in mortality inequality.

Moving into young adulthood and middle age, Figure 3 shows that the different trends for males and females intensified. Males aged 20–49 experienced declines in mortality in poorer counties (though not so much in richer ones) leading to a significant decline in mortality inequality, whereas for women there was little improvement in mortality in either rich or poor county groups. This a truly remarkable stagnation in light of the significant progress in mortality reduction made in other age categories.

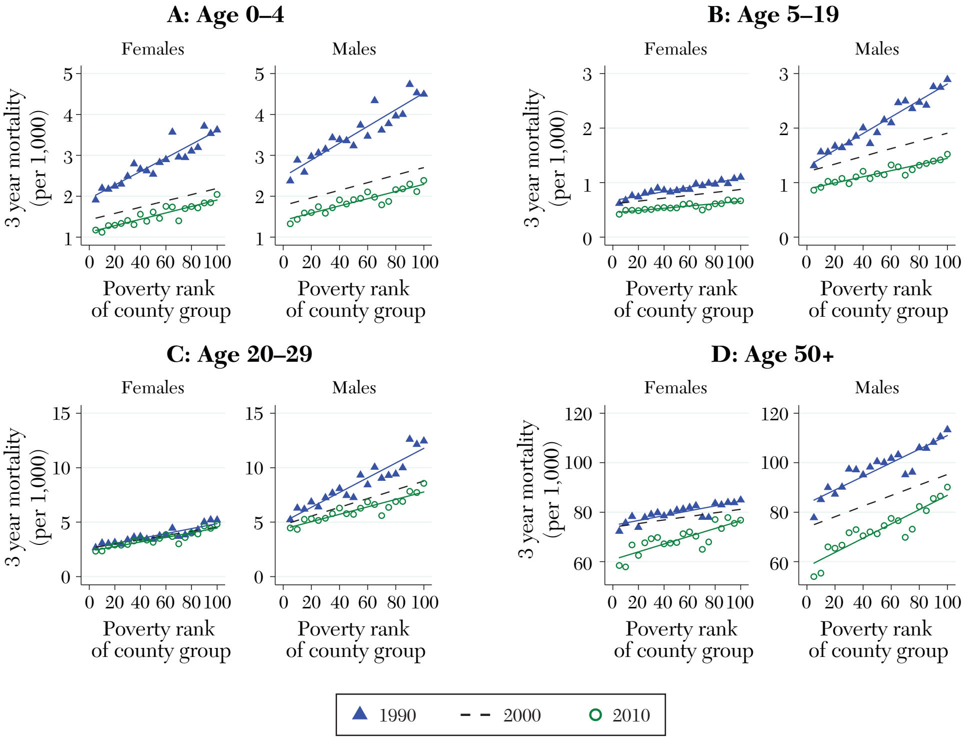
After age 50, mortality again showed large decreases over the whole 20-year period. For females, virtually all of this improvement occurred between 2000 and

[15] Online Appendix Figure A4 shows that when we use our data and method to look at life expectancy at age 50, we also find increases in inequality in life expectancy for both men and women.

[16] For an analysis by finer age groups, see Currie and Schwandt (forthcoming). Online Appendix Table A3 shows numerical values for the mortality estimates and includes tests for whether the slopes of a line drawn through the 1990 points is different from the slope of a line drawn through the 2010 points for each age group.

Figure 3

Three-Year Mortality Rates across Groups of Counties Ranked by their Poverty Rate



Source: Authors using data from the Vital Statistics, the US Census, and the American Community Survey.
Note: Three-year mortality rates for four different age groups are plotted across county groups ranked by their poverty rate. Mortality rates in 2000 and 2010 are age-adjusted using the 1990 population, that is, they account for changes in the age structure within age, gender, and county groups since 1990. Table A3 provides magnitudes for individual mortality estimates and for the slopes of the fitted lines.

2010. For men, there were larger and steadier declines in mortality. For women in this age group, gains were bigger in the richest county groups, leading to a significant increase in inequality in mortality. For men, the increase in mortality inequality is not statistically significant in the 50-plus group, though for males 65 and older, inequality in mortality is increasing significantly.[17]

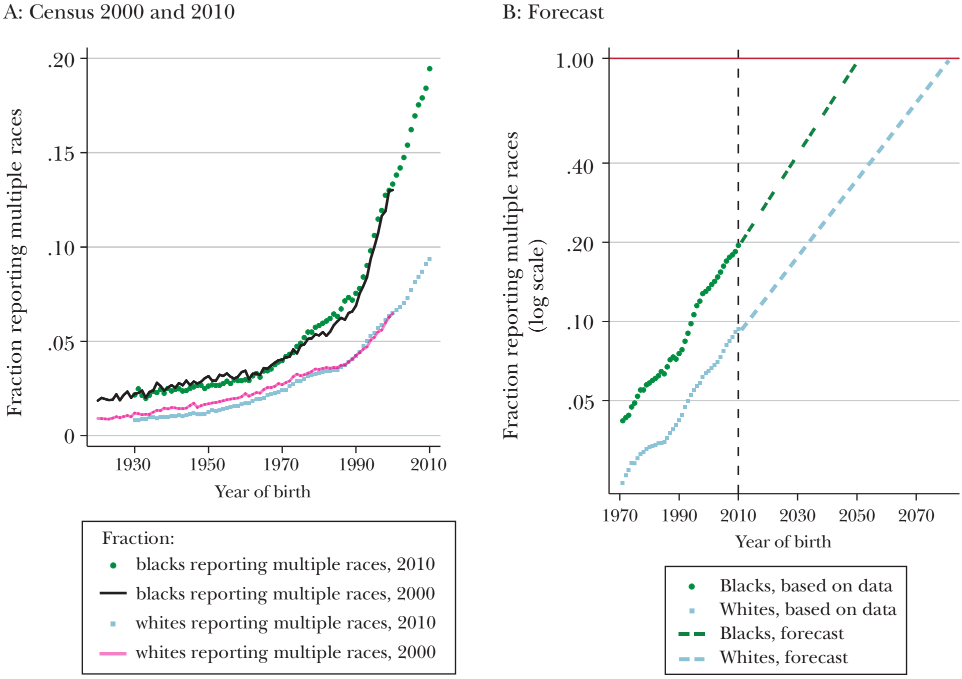
All the results in this section are robust to ordering counties using the fraction of high school dropouts, median income, or average life expectancy rather than poverty.[18]

[17] Appendix Table A3 reports mortality rates as well as tests for changes in inequality for these different age groups.

[18] For details, see online Appendix Figure A9 with this paper at http://e-jep.org. The patterns look extremely similar when counties are ranked by the fraction of high school dropouts or by life expectancy. When we sort by median income, the reductions in mortality appear to be more evenly distributed.

Figure 4

Fraction of the Black and White US Population Reporting Multiple Races



Source: Authors using data from the US Census.
Note: Figure 4A plots the fraction of people reporting multiple races among all those reporting that they are black (or white) alone or in combination, in the 2000 and 2010 Census. Figure 4B forecasts the fraction reporting multiple races among future birth cohorts. Assuming that the exponential growth continues, we fit a linear trend through the log fraction reporting multiple races for birth cohorts 1970 to 2010 in the 2010 Census and project this trend up to the 2080 birth cohort. The projected fraction reaches unity in 2051 for blacks and in 2081 for whites.

Age and Race-Specific Mortality

As discussed above, the Census now allows people to describe themselves as belonging to more than one race. Figure 4A shows a striking exponential growth in the fraction of people identifying as multiple races across birth cohorts, as reported in the 2000 and 2010 Census. While the fraction reporting multiple races is below 2.5 percent among those born in the first half of the past century, it strongly increased in more recent cohorts. For the 2010 cohort, it reached 10 percent for whites and 20 percent for African Americans. Importantly, these patterns do not reflect an age effect. The curves for 2000 and 2010 virtually match, even though the cohorts grew 10 years older between the two Censuses. As we show in panel B of Figure 4, if the observed exponential growth of multiple-race reporting continues into the future, the last single-race African-American and single-race white persons will be born in 2050 and 2080, respectively! While continuing exponential growth is a strong assumption, the patterns in Figure 4 suggest that multiple race reporting will become more important in the future. It is important to account for this development when studying trends in race-specific mortality, particularly among younger cohorts. We therefore report mortality rates based both on single and multiple race population counts.[19]

Figure 5 shows an analysis of age-specific three-year mortality rates by race.[20] Recall that only one race is reported on the death certificates, even for people who consider themselves biracial. However, in the total population data, we have counts for people who consider themselves biracial. The lines with triangles or circles are based on rates calculated using, for the denominator, people who consider themselves only white or black. For 2010, we have also added a second line, marked with squares, based on calculations in which the denominator also includes those who identify with more than one race. Of course, adding these individuals to the denominator without increasing the numerator lowers the estimated mortality rates.

Panel A shows mortality rates for children under five. What is most striking in these figures is the truly remarkable reduction in black mortality rates between 1990 and 2000, and the continuing, though smaller, decline for blacks between 2000 and 2010. In 1990, young black male children in the richest counties had mortality rates of 6.2 per 1,000, while white male children in the poorest counties had mortality rates of about 4 per 1,000. Thus, racial disparities trumped any inequality based on geographic areas. By 2010, the mortality rate for young black male children in the richest counties was still above the mortality rate for young white males in the poorest counties, but the gap had narrowed greatly. Moreover, if we use the rates calculated including people with multiple races in the denominator, the estimated black mortality rate falls even further.

Panel B shows similar figures for children aged 5 to 19. In this age group, differences between black and white females are less apparent than for those under five. However for males, there is still a very large disparity in death rates, albeit one that was greatly reduced over the 20-year period. For both black and white males, death rates fell much more in the poorest county groups. Including those with multiple races makes much less difference in these figures than in those for the children under five, though it still affects the estimated mortality rate for black males.

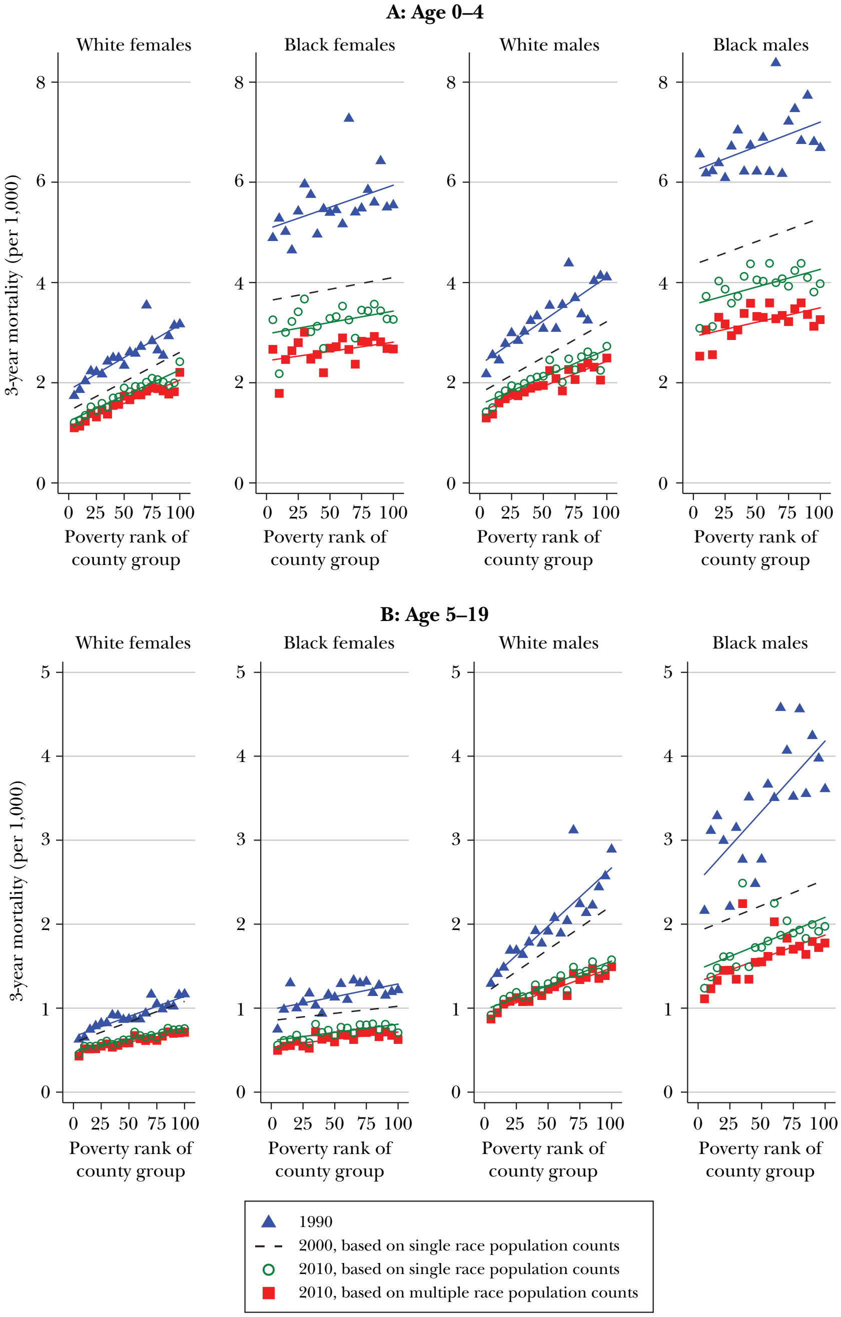
Panel C of Figure 5 focuses on people aged 20 to 49. A striking finding from this figure is the stagnation in white female mortality rates between 1990 and 2010. There is even a slight increase in the mortality in the poorest county groups. This

[19] The Census has responded to these problems by producing "bridged" estimates that attempt to allocate the entire population to one of four races (white, African American, Native American, Asian) following an imputation estimation procedure. Figure A6 provides an example of how these differences in reporting can influence the calculated death rates for those aged 20 to 24. Overall the results suggest that changes in race reporting may have important effects on estimated trends in mortality among groups where the changes in mortality are relatively small, either because mortality does not change, or because changes start from a very low baseline and are small in absolute terms.

[20] As before we rank counties by their overall poverty rate. The figure looks similar when ranking counties by race-specific poverty rates, but there seems to be a considerable sampling error for black poverty estimates in 1990, which is why we continue to use overall county poverty levels for these figures. Ranking counties the same way with the figures for both blacks and whites also facilitates comparisons.

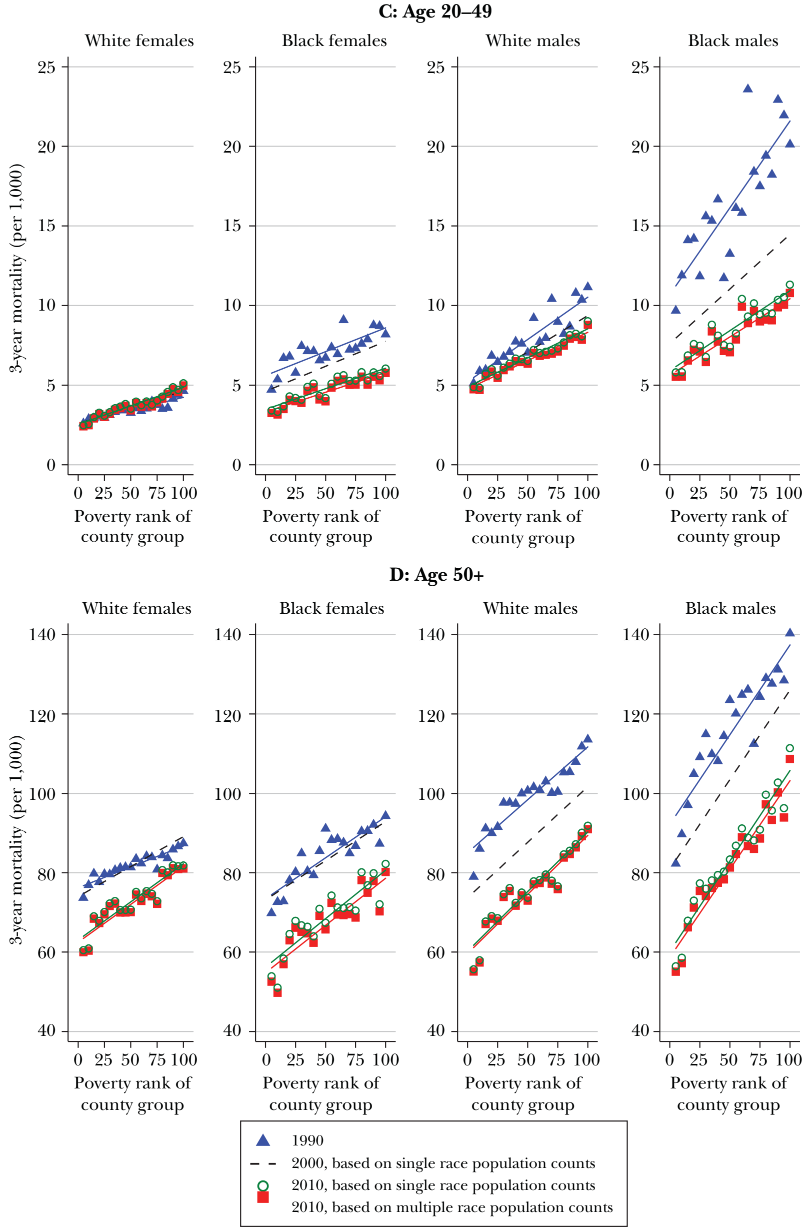
Figure 5

Three-Year White and Black Mortality Rates across Poverty Percentiles, Based on Single and Multiple Race Population Counts



(continued on next page)

Figure 5 (continued)



Source: Authors using data from the Vital Statistics, the US Census, and the American Community Survey.
Note: Three-year mortality rates for four different age groups are plotted separately for whites and African Americans across county groups ranked by their overall poverty rate. For further details see the comments below Figure 2 and in the text. Circles represent mortality rates constructed as the ratio of race-specific death counts in the Vital Statistics divided by single race population counts in the 2010 Census. The mortality rates represented by squares are based on the same death counts, but divided by population counts including multiple race reports. Mortality rates in 2000 and 2010 are age-adjusted using the 1990 population, that is they account for changes in the age structure within age, gender, race, and county groups since 1990.

result is completely consistent with those of Case and Deaton (2015), who document increases in middle-age mortality among non-Hispanic whites between 2000 and 2010.[21] Black females show reductions in mortality rates in both rich and poor counties, while white males experienced reductions only in the poorer counties, resulting in reduced mortality inequality for that group. The results for black males show, once again, huge reductions in mortality, which are greater in the poorest counties. By 2010, black males in the richest counties had considerably lower mortality than white males in the poorest counties, which had not been the case in 1990.

Results for people over 50 are shown in Panel D. Mortality fell for each of the four race and gender categories. Among females and among white males, it fell slightly more in the richest county groups, while for black males, mortality fell similarly in poor and rich county groups. Multiple race reporting appears to be a relatively insignificant issue in this age category, as one would expect given the low rate of multiple race reporting in this age range (shown earlier in Figure 4).

Important Drivers of Mortality Trends in Different Cohorts

Given that there is so much dispute about the nature of the trends in inequality in mortality rates, perhaps it is unsurprising that there is so little research seeking to establish the causes of the trends. Aizer and Currie (2014) document the fall in mortality inequality among infants and cite many possible explanations including increases in maternal education, expansions of health insurance for pregnant women, the Supplemental Nutrition Program for Women, Infants, and Children, and expansions of the Earned Income Tax Credit.

Other than our paper Currie and Schwandt (forthcoming), we are not aware of any research that has looked systematically at the causes of reductions in mortality among older children. Some possibilities include expansion of public health insurance (Brown, Kowlaski, and Lurie 2015; Cahodes, Grossman, Kleiner, and Lovenhem 2014; Currie, Decker, and Lin 2008; Miller and Wherry 2015; Wherry and Meyer 2015; Wherry, Miller, Kaestner, and Meyer 2015), other social safety net programs such as Head Start (Ludwig and Miller 2007; Hoynes, Schanzenbach, and Almond forthcoming), and reductions in pollution (Isen, Rossin-Slater, Walker forthcoming).

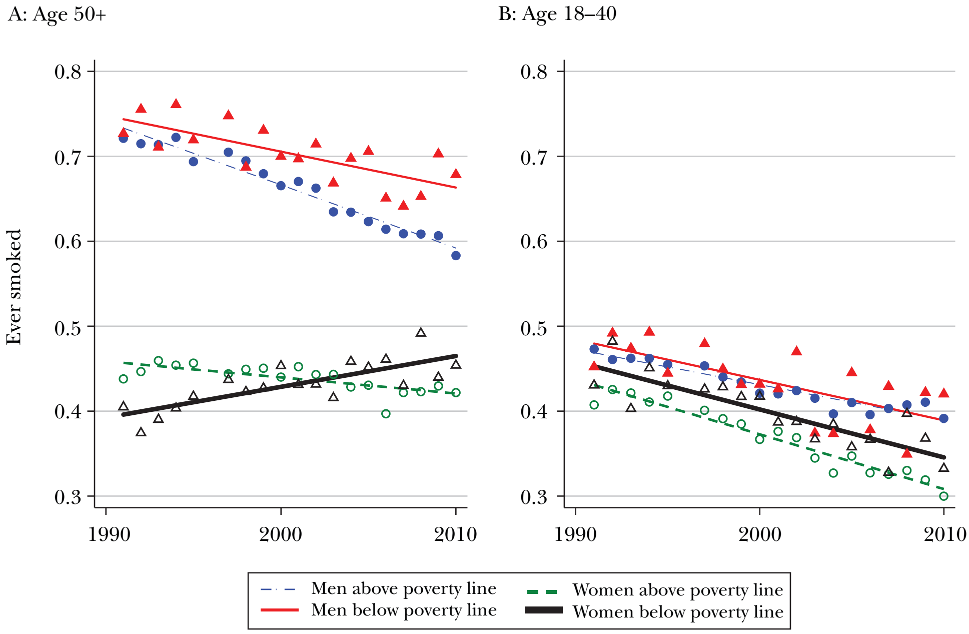
We are also unaware of research that has investigated the role of immigration in driving inequalities in mortality. To the extent that Hispanic immigrants tend to be both poorer and healthier than the average American (the so-called "Hispanic paradox"), areas that receive a lot of immigrants might see improvements in mortality differentials. One might also be more likely to see this pattern for the young than for the old, given that immigrants tend to be young.

Smoking is a major driver of spatial mortality differences among older adults in the United States; for example, Fenelon and Preston (2012) estimate that smoking

[21] In online Appendix Figure A7, we show US-wide age-specific mortality trends for non-Hispanic females and males, based on different population counts. The mortality increases in middle age between 2000 and 2010, highlighted by Case and Deaton (2015), are clearly visible across all measures, that is they are hardly affected by the way the non-Hispanic white population is counted.

Figure 6

Fraction That Ever Smoked in Old and Young Population by Poverty Status and Gender, 1990–2010



Source: Authors using data the National Health Interview Survey.
Note: Smoking rates in the overall old and young adult US population, by poverty status, are plotted from 1990 to 2010. Lines are fitted based on ordinary least squares regressions.

can explain 60 percent of the differences in age 50-plus mortality across US states. In our context, at least some of the increasing disparities that we observe in old age mortality might reflect differences in smoking take-up and cessation by socioeconomic status. De Walque (2010) shows that better-educated people stopped smoking much more quickly following the 1964 US Surgeon General's report on the dangers of smoking than less-educated people. Moreover, males started with much higher smoking rates than females, but quickly reduced their rates, while smoking continued to gain ground among less-educated women for some time after the Surgeon General's report. Cohorts in which more-educated women had already reduced smoking, while the less-educated still smoked at increasing rates, entered old age over the past two decades, implying that lifetime smoking rates between the elderly rich and the poor likely diverged during that time period.

Figure 6 shows, based on smoking histories from the National Health Interview Survey, how these patterns have continued to play out during the time period that we analyze. Among those 50 and over, men are much more likely to have ever smoked than women, but lifetime smoking rates decreased strongly between 1990 and 2010 for both rich and poor men. The decrease was somewhat stronger for rich men, which implies that the smoking gap between rich and poor men 50 and over widened during that time period. This pattern could explain why we observe strong reductions in mortality among elderly men both in rich and poor county groups, with somewhat stronger improvements among the rich.

In the cohorts of women who passed age 50 over the past two decades, smoking rates declined among rich women but increased strongly among the poor. In fact, in 1990, lifetime smoking rates were substantially lower among cohorts of poor women but by 2010, their rate had surpassed that of rich women. The smoking gap between these two groups increased by 11 percentage points during the past 20 years, almost twice as much as the parallel increase for men of 6.4 percentage points. This pattern is in line with the increasing inequality in female old-age mortality that we observe between rich and poor counties between 1990 and 2010 (note the significant steepening of the line for females age 50+ in Figure 3).

These findings suggest that at least part of the diverging mortality rates currently observed at older ages might be a temporary phenomenon driven by a strong improvement in health behavior that simply occurred with some lag among the poor. Once the later-born cohorts, which experienced strong reductions in smoking among both rich and poor, enter old age and replace these transition cohorts, smoking-induced mortality among the elderly is likely to converge to lower levels. The right panel of Figure 6 shows that the fraction who ever smoked is already much lower among adults aged 18 to 40. Moreover, the rates look quite similar regardless of poverty status for men. Among women, the poor are still more likely to smoke, but the rates are falling at roughly similar rates across all groups. When these cohorts reach old age in the coming decades, society will fully reap the benefits of the "anti-smoking dividend," resulting ceteris paribus in lower mortality and decreasing inequality in mortality at these ages.

Other factors also may have affected inequality in mortality between counties. Improvements in medical care, such as for heart disease, seem likely to have reduced health inequality as they have diffused over time, other things being equal. The gap in obesity rates between rich and poor has also been narrowing, but this development is driven by increasing obesity among the rich, which may in fact auger higher death rates for rich and poor in the future (NAS 2015). Case and Deaton (2015) highlight another factor that may be driving increased inequality in some segments of the population: the opioid epidemic. It may be possible to address these questions using the cause of death in the Vital Statistics Mortality data. However, given the issues discussed above with respect to changes in measurement, measurement error, and missing data about causes, these data are unlikely to provide a definitive answer.

Discussion and Conclusions

In contrast to many recent analyses of mortality inequality, we find improvements in overall life expectancy in both rich and poor counties. Our focus on life expectancy at birth rather than life expectancy in middle age may explain this finding. We find that inequality in mortality has fallen greatly among children. It is worth emphasizing that the reductions in mortality among African Americans, especially African-American males of all ages, are stunning and that is a major driver of the overall positive picture. This positive finding has been largely neglected in much of the discussion of overall mortality trends. Although our overall message is more positive than some earlier studies, we do find an alarming stagnation in mortality among white women aged 20 to 49. In the poorest counties mortality even increased slightly, indicating increasing inequality in mortality in that group.

It sometimes seems as if the research literature on mortality is compelled in some way to emphasize a negative message, either about a group that is doing less well or about some aspect of inequality that is rising. In contrast, this study is one of comparatively few, along with Aizer and Currie (2014) and Currie and Schwandt (forthcoming), that has emphasized improvements in life expectancy across the broad US population. Our results point to strong health improvements and decreasing inequality, particularly among the younger cohorts who will form the future adult population of the United States. Given the growing literature demonstrating a connection between health in childhood and future health (as in Currie and Rossin-Slater 2015), this improvement in health among young people in poor counties suggests that these cohorts may well be healthier and suffer less mortality inequality in the future than those who are currently middle-aged and older. In addition, much of the increase in inequality in older cohorts in the past 20 years has been driven by historical smoking patterns. Current cohorts have much lower lifetime smoking rates, which is also likely to lead to more convergence in mortality rates.

We believe that a balanced approach to the mortality evidence, which recognizes real progress as well as areas in need of improvement, is more likely to result in sensible policymaking. After all, emphasizing the negative could send the message that "nothing works," especially in the face of seemingly relentless increases in income inequality. We have emphasized considerable heterogeneity in the evolution of mortality inequality by age, gender, and race. Going forward, identifying social policies that have helped the poor and reduced mortality inequality is an important direction for future research. Similarly, understanding the reasons that some groups and age ranges have seen stagnant mortality rates will be important for mobilizing efforts to reduce inequality in mortality and improve the health of the poor.

■ *We would like to thank Anne Case, Angus Deaton, Pascaline Dupas, Ilyana Kuziemko, Ronald Lee, Kenneth Wachter, Amelie Wuppermann as well as seminar participants at Princeton, University of Zurich, ETH Zurich, Fundação Getúlio Vargas Sao Paulo, University of Bergen, University of Bonn, Berkeley, and the Chicago Fed for comments. Financial support from the Princeton Center for Translational Research on Aging (2P30AG024928) is gratefully acknowledged.*

References

Aizer, Anna, and Janet Currie. 2014. "The Intergenerational Transmission of Inequality: Maternal Disadvantage and Health at Birth." *Science,* May 23, 344(6186): 856–61.

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development Before Age Five." Chapter 15 in *Handbook of Labor Economics,* vol 4B, edited by David Card and Orley Ashenfelter. Amsterdam: Elsevier.

Arias, Elizabeth, Karl Eschbach, William S. Schauman, Eric Bucklund, and Paul D. Sorlie. 2010. "The Hispanic Mortality Advantage and Ethnic Misclassification on US Death Certificates." *American Journal of Public Health* 100(S1): S171–S177.

Arias, Elizabeth, William S. Schauman, Karl Eschbach, Paul D. Sorlie, and Eric Bucklund. 2008. "The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States." *Vital and Health Statistics* Series 2, no. 148, National Center for Health Statistics. October.

Banks, James, Michael Marmot, Zoe Oldfield, and James P. Smith. 2006. "Disease and Disadvantage in the United States and in England." *JAMA* 295(17): 2037–45.

Bosworth, Barry P., and Kathleen Burke. 2014. "Differential Mortality and Retirement Benefits in the Health and Retirement Study." Brookings Institution, Working Paper no. 2014-4.

Bound, John, Arline Geronimus, Javier Rodriguez, and Timothy Waidmann. 2014. "The Implications of Differential Trends in Mortality for Social Security Policy." University of Michigan Retirement Research Center (MRRC) Working Paper no. 2014-314.

Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie. 2015. "Medicaid as an Investment in Children: What is the Long-Term Impact on Tax Receipts?" NBER Working Paper 20835.

Cahodes, Sarah, Daniel Grossman, Samuel Kleiner, and Michael F. Lovenhem. 2014. "The Effect of Child Health Insurance Access on Schooling." NBER Working Paper 20178.

Case, Anne, and Angus Deaton. 2015 "Rising Morbidity and Mortality in Midlife among White non-Hispanic Americans in the 21s Century." *PNAS* 112(49): 15078–83.

Case, Anne, and Christina Paxson. 2011. "The Long Reach of Childhood Health and Circumstance: Evidence from the Whitehall II Study." *Economic Journal* 121(554): F183–F204.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2015. "The Relationship between Life Expectancy and Income in the United States 2001–2014." Unpublished paper.

Chiang, Chin Long. 1984. *The Life Table and Its Applications.* Malabar, FL: Robert E. Krieger Publishing.

Costa, Dora L., and Matthew E. Kahn. 2015. "Declining Mortality Inequality within Cities during the Health Transition." *American Economic Review* 105(5): 564–69.

Currie, Janet, Sandra Decker, and Wanchuan Lin. 2008. "Has Public Health Insurance for Older Children Reduced Disparities in Access to Care and Health Outcomes?" *Journal of Health Economics* 27(6): 1567–1581.

Currie, Janet, and Maya Rossin-Slater. 2015. "Early-Life Origins of Life-Cycle Well-Being: Research and Policy Implications." *Journal of Policy Analysis and Management* 34(1): 208–42.

Currie, Janet, and Hannes Schwandt. Forthcoming. "Inequality in Mortality Between Rich and Poor U.S. Counties Decreased among the Young while Increasing for Older Adults, 1990–2010." *Science.*

Currie, Janet, and Mark Stabile. 2003. "Socioeconomic Status and Child Health: Why is the Relationship Stronger for Older Children?" *American Economic Review* 93(5): 1813–23.

Cutler, David M., Fabian Lange, Ellen Meara, Seth Richards-Shubik, and Christopher J. Ruhm. 2011. "Rising Educational Gradients in Mortality: The Role of Behavioral Risk Factors." *Journal of Health Economics* 30(6): 1174–87.

Deaton, Angus, and Christina H. Paxson. 2001. "Mortality, Education, Income, and Inequality among American Cohorts." In *Themes in the Economics of Aging,* edited by David A. Wise, 129–165. University of Chicago Press.

de Walque, Damien. 2010. "Education, Information, and Smoking Decisions: Evidence from Smoking Histories in the United States, 1940–2000." *Journal of Human Resources* 45(3): 682–717.

Dorn, David. 2009. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." Dissertation no. 3613, University of St. Gallen, September 2009.

Dowd, Jennifer B., and Amar Hamoudi. 2014. "Is Life Expectancy Really Falling for Groups of Low Socio-Economic Status? Lagged Selection Bias and Artefactual Trends in Mortality." *International Journal of Epidemiology* 43(4): 983–88.

Elo, Irma T., and Samuel H. Preston. 1996. "Educational Differentials in Mortality: United

States, 1979–1985." *Social Science & Medicine* 42(1): 47–57.

Fenelon, Andrew, and Samuel H. Preston. 2012. "Estimating Smoking-Attributable Mortality in the United States." *Demography* 49(3): 797–818.

Goldring, Thomas, Fabian Lange, and Seth Richards-Shubik. 2015. "Testing for Changes in the SES-Mortality Gradient When the Distribution of Education Changes Too." NBER Working Paper 20993.

Gravelle, Hugh. 1998. "How Much of the Relation Between Population Mortality and Unequal Distribution of Income is a Statistical Artifact?" *British Medical Journal* 316(7128): 382–85.

Hendi, Arun S. 2015. "Trends in U.S. Life Expectancy Gradients: The Role of Changing Educational Composition." *International Journal of Epidemiology* 44(3): 946–55.

HMD. 2015. *Human Mortality Database.* University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). http://www.mortality.org/.

Hoynes, Hilary W., Diane Whitmore Schanzenbach, and Douglas Almond. Forthcoming. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review.*

Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. Forthcoming. "Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970." *Journal of Political Economy.*

Kitagawa, Evelyn M., and Philip M. Hauser. 1973. "Differential Mortality in the United States: A Study in Socioeconomic Epidemiology." Cambridge, MA: Harvard University Press.

Kulkarni, Sandeep C., Alison Levin-Rector, Majid Ezzati, and Christopher J. L. Murray. 2011. "Falling Behind: Life Expectancy in US Counties from 2000 to 2007 in an International Context." *Population Health Metrics* Vol. 9, Article 16.

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159–208.

Marmot, Michael G., Stephen Stansfeld, Chandra Patel, Fiona North, J. Head, Ian White, Eric Brunner, and Amanda Feeny, and George Davey Smith. 1991. "Health Inequalities among British Civil Servants: The Whitehall II Study." *Lancet* 337(8754): 1387–93.

Meara, Ellen R., Seth Richards, and David M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981–2000." *Health Affairs* 27(2): 350–60.

Miller, Sarah, and Laura R. Wherry. 2015. "The Long-Term Health Effects of Early Life Medicaid Coverage." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2466691.

Montez, Jennifer Karas, and Lisa F. Berkman. 2014. "Trends in the Educational Gradient of Mortality among US Adults Aged 45 to 84 Years: Bringing Regional Context into the Explanation." *American Journal of Public Health* 104(1): e82–90.

Montez, Jennifer Karas, and Anna Zajacova. 2013. "Explaining the Widening Education Gap in Mortality among U.S. White Women." *Journal of Health and Social Behavior* 54(2): 166–82.

Murphy, Sherry L., Jiaquan Xu, and Kenneth D. Kochanek. 2013. "Deaths: Final Data for 2010." *National Vital Statistics Reports,* May 8, 61(4).

Murray, Christopher J. L., Sandeep C. Kulkarni, Catherine Michaud, Niels Tomijima, Maria T. Bulzacchelli, Terrell J. Iandiorio, and Majid Ezzati. 2006. "Eight Americas: Investigating Mortality Disparities across Races, Counties, and Race-Counties in the United States." *PLoS Med* 3(9): e260.

National Academies of Sciences, Engineering, and Medicine (NAS). 2015. *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses.* Committee on the Long-Run Macroeconomic Effects of the Aging U.S. Population-Phase II; Committee on Population, Division of Behavioral and Social Sciences and Education; Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences. Washington, DC: The National Academies Press.

Olshansky, Jay S., Toni Antonucci, Lisa Berkman, Robert H. Binstock, Axel Boersch-Supan, John T. Cacioppo, Bruce A. Carnes, Laura L. Carstensen, Linda P. Friend, Dana P. Goldman, James Jackson, Martin Kohli, John Rother, Yuhui Zheng, and John Rowe. 2012. "Differences in Life Expectancy Due to Race and Educational Differences are Widening, and Many May Not Catch Up." *Health Affairs* 31(8): 1803–13.

Pappas, Gregory, Susan Queen, Wilbur Hadden, and Gail Fisher. 1993. "The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986." *New England Journal of Medicine* 329(2): 103–109.

Pijoan-Mas, Josep, and José-Víctor Ríos-Rull. 2014. "Heterogeneity in Expected Longevities." *Demography* 51(6): 2075–2102.

Preston, Samuel H., and Irma T. Elo. 1995. "Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health* 7(4): 476–96.

Singh, Gopal K., and Mohammad Siahpush. 2006. "Widening Socioeconomic Inequalities in US Life Expectancy, 1980–2000." *International Journal of Epidemiology* 35(4): 969–79.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status." *Journal of Economic Perspectives* 13(2): 145–66.

Smith, James P. 2007. "The Impact of Socioeconomic Status on Health over the Life-Course." *Journal of Human Resources* 42(4): 739–64.

US Census Bureau. Undated. "The American Community Survey (ACS) Mail Questionnaire from 2005 to 2008." http://www2.census.gov/programs-surveys/acs/methodology/questionnaires/SQuestChanges05to08.pdf.

Waldron, Hilary. 2007. "Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by Socioeconomic Status." *Social Security Bulletin* 67(3).

Waldron, Hilary. 2013. "Mortality Differentials by Lifetime Earnings Decile: Implications for Evaluations of Proposed Social Security Law Changes." *Social Security Bulletin* 73(1).

Wang, Haidong, Austin E. Schumacher, Carly E. Levitz, Ali H. Mokdad, and Christopher J. L. Murray. 2013. "Left Behind: Widening Disparities for Males and Females in US County Life Expectancy, 1985–2010." *Population Health Metrics* 11: 8.

Wherry, Laura R., and Bruce D. Meyer. 2015. "Saving Teens: Using a Policy Discontinuity to Estimate the Effects of Medicaid Eligibility." *Journal of Human Resources,* published ahead of print November 30.

Wherry, Laura R., Sarah Miller, Robert Kaestner, and Bruce D. Meyer. 2015. "Childhood Medicaid Coverage and Later Life Health Care Utilization." NBER Working Paper 20929.

Wilkinson, Richard G. 1996. *Unhealthy Societies: The Afflictions of Inequality.* New York: Routledge.

Wilmoth, John R., Carl Boe, and Magali Barbieri. 2011. "Geographic Differences in Life Expectancy at Age 50 in the United States Compared with Other High-Income Countries," edited by Eileen M. Crimmins, Samuel H. Preston, and Barney Cohen. In *International Differences in Mortality at Older Ages: Dimensions and Sources*, 337–72. Washington, DC: National Academies Press.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 53–78

Health Insurance and Income Inequality

Robert Kaestner and Darren Lubotsky

Most analyses of economic inequality have focused on wage rates, earnings, or incomes. Wages or earnings are the appropriate measures to study the changes in the return to skills, the structure of the labor market, and shifts in the demand and supply for different types of labor (for example, Murphy and Topel 2016). Income is a broader measure of living standards and is therefore more useful when studying how government taxes and transfers affect inequality and control over real resources. However, health insurance and other in-kind forms of compensation and government benefits are typically not included in measures of income and analyses of inequality. This omission is important: for example, health expenditures in 2014 accounted for over 17 percent of GDP, and nearly 70 percent of these expenditures were made by public or private health insurance plans.[1] Given the large and growing cost of health care in the United States and the presence of large government health insurance programs such as Medicaid and Medicare, it is crucial to understand how health insurance and related public policies contribute to measured economic well-being and inequality.

[1] Figures are from the National Health Expenditure Accounts at Centers for Medicare and Medicaid Services: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html.

■ *Robert Kaestner is Professor of Economics and Darren Lubotsky is Associate Professor of Economics, both at the University of Illinois at Chicago, Chicago, Illinois. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are kaestner@uic.edu and lubotsky@uic.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.53 doi=10.1257/jep.30.2.53

Consider Medicaid, which is targeted at low-income families and accounts for about 16 percent of national health expenditures. Medicaid spending on a family of three (say, a mother and her two children) was, on average, $9,125 in 2011 (Young, Rudowitz, Rouhani, and Garfield 2015), which represents approximately 80 percent of mean income of families in the bottom quintile of the income distribution in that year.[2] As these figures indicate, Medicaid has the potential to significantly raise well-being among low-income families and reduce inequality. Medicare, which is targeted at the elderly, accounts for about 20 percent of national health expenditures. In 2014, spending per enrollee was $11,400, which is approximately one-third of median family income among the elderly. Given that the elderly tend to have lower incomes than the nonelderly, Medicare will tend to reduce inequality (Cooper and Gould 2013).

Government tax policy regarding health insurance also influences inequality because employer contributions for employer-provided health insurance are not subject to income or payroll taxes. In 2016, this tax expenditure is estimated to be about $348 billion (US Office of Management and Budget 2015, Table 14.1). In contrast to Medicaid and Medicare, however, the tax deductibility of employer-provided health insurance will increase inequality because the tax exclusion is larger for higher-income persons, who are more likely to have employer-provided insurance and face higher marginal tax rates. Consider a family with combined income of about $135,000 (approximately the 90th percentile of family income in 2013), which faces a combined federal, state, and payroll tax rate of 40 percent and receives employer-provided health insurance that costs $16,000 per year. If the premium were taxed as income, the family would owe an additional $6,400 in taxes. The tax benefit for employer-provided insurance for high-income families in this example is thus about two-thirds of the size of Medicaid expenditures, although it obviously represents a far smaller share of family income.

The importance of integrating health policies into an analysis of inequality is underscored by the well-known inequality in health across the income distribution. Based on our calculations from the Medical Expenditure Panel Survey (MEPS) data, 25 percent of those in the bottom quintile of the family income distribution reported being in poor or fair health in 2012, while only 7 percent of those in the top quintile of family income report the same.[3] This pattern of those with low incomes being in worse health extends to many other measures of health including mortality (Deaton and Paxson 2004; Cutler, Deaton, and Lleras-Muney 2006).

[2] Medicaid spending is from Kaiser Family Foundation, "Medicaid Spending per Full-Benefit Enrollee," http://kff.org/medicaid/state-indicator/medicaid-spending-per-full-benefit-enrollee/. Measures of the distribution of income are from the US Census Bureau "Historical Income Tables: Income Inequality: https://www.census.gov/hhes/www/income/data/historical/inequality/.

[3] The calculations are derived from the 2012 full year Medical Expenditure Panel Survey for people age 18 and older. Family income is total family income (of all persons in family) and is adjusted for family size by dividing by the square root of family size. Respondents are asked to rate their health as being Excellent, Very Good, Good, Fair, or Poor. We define poor health as responses of "Fair" or "Poor," which is standard in the literature.

Low-income families face a relatively greater burden of disease, which has direct effects on their well-being and makes health care spending particularly important to their well-being.

Our paper assesses the effect on inequality of the primary government programs that affect health insurance. First, we begin with descriptions of the principal government transfers and tax expenditures related to health insurance and their changes over time. We highlight how these programs may affect measurement of income inequality. Second, we review the small literature that has included health care in the analysis of income inequality. We augment this review with empirical analyses that illustrate the effects on inequality of various components of health care policy. Third, we discuss conceptual and empirical issues that arise when trying to integrate health care into the analysis of income inequality, including the potential behavioral responses and distortionary effects of government health care transfers and tax policies. Finally, we discuss the implications of our analysis for future research and policy.

Our analyses lead us to the following conclusions. First, including the value of Medicare and Medicaid in income reduces the ratio of the 90th to the 10th percentile (the so-called 90–10 ratio) of the after-tax income distribution in 1995 by about 24 percent and in 2012 by about 30 percent. These programs clearly have the effect of reducing inequality, broadly defined to include more than income. Second, adding the value of employer-provided insurance to income raises measured inequality. On net, including both publicly-provided and employer-provided insurance in a more comprehensive measure of income results in a downward revision in measured inequality at a point in time and reduces the growth in inequality that has occurred over the last 20 years. Third, the tax exclusion for employer-provided insurance has modest effects on income inequality. Taxing employer contributions to employee health insurance plans would reduce the 90–10 ratio in 2012 by approximately 4 percent. Fourth, behavioral responses to public health insurance programs such as reductions in labor supply to qualify for Medicaid are unlikely to alter these conclusions unless they were much larger than estimates in the literature suggest. Finally, income inequality is not the same as inequality in well-being or utility. Policies could improve the health or overall well-being of the poor and thereby lower inequality more broadly construed, but still lead some measures of inequality to rise.

An Overview of Government Transfers and Tax Expenditures for Health Care

Government health policy affects inequality primarily through the in-kind provision of health insurance and through the tax treatment of employer-provided health insurance and out-of-pocket expenses. As noted earlier, the amount of in-kind transfers associated with government health insurance policies are large and are not equally distributed throughout the income distribution. Thus, in-kind transfers related to health care may have a substantial effect on the level of inequality, and changes over time in the generosity and extent of these government transfers—most recently, in the aftermath of the Patient Protection and Affordable Care Act in 2010—may have affected the trend in inequality.

Medicare and Medicaid are the primary publicly provided health insurance programs. Medicare is a federal health insurance program that covers nearly all Americans aged 65 and older, as well as some people under 65 who are disabled. It is financed through a 2.9 percent payroll tax, plus an additional 0.9 percent tax on incomes above a threshold (in 2015, the threshold is $200,000 for Single or Head of Household taxpayers and is $250,000 for Married Couples filing jointly).[4] In fiscal year 2014, total Medicare spending was $619 billion or about $11,400 per beneficiary (Centers for Medicare and Medicaid Services 2014a).

Medicaid and the State Children's Health Insurance Program (CHIP) are health insurance programs for low-income persons. Most Medicaid beneficiaries are children and mothers, but also included are a significant number of elderly (who also have dual eligibility for Medicare). Medicaid and CHIP are jointly financed by the federal government and the states through general revenue sources. Federal law sets minimum standards for eligibility and covered services, while states have flexibility to cover additional people or services. In fiscal year 2014, total Medicaid and CHIP expenditures were $476 billion and covered nearly 60 million people.

To provide some sense of the potential effects of Medicaid and Medicare on inequality, Table 1 shows the share of individuals in each decile of family income in 1995, 2004, and 2012 that participated in these programs during the year.[5] The data come from the 1996, 2005, and 2013 Annual Social and Economic Supplement of the Current Population Survey, which collects information on respondents' income and program participation during the prior calendar year. Individuals are assigned to one of ten deciles of "adjusted" family income, which is total family income divided by the square root of family size (following Burkhauser, Larrimore, and Simon 2012).

Two patterns are evident from the figures in Table 1. First, those in the lower half of the income distribution are more likely to receive government-financed health insurance, particularly Medicaid. For example, between 35 and 45 percent of those in the first deciles are covered by Medicaid. Thus, adding the value of this insurance to income would tend to flatten the income distribution. Indeed, the potential distributional impact of Medicaid is likely significantly understated in the figures because of the known and substantial underreporting of Medicaid participation in the Current Population Survey (Call, Davern, Klerman, and Lynch 2013; Meyer, Mok, and

[4] We do not discuss the effect on inequality of the tax policies that finance government health programs. For an example of a study that integrates an analysis of spending and tax policy, see McClellan and Skinner (2006).

[5] We begin our analysis with data from 1996 since this is the first year where the Current Population Survey identifies whether a respondent has employer-sponsored coverage in their own name or that of a family member. We chose 2012 (2013 CPS) as the endpoint because the health insurance questions in the Current Population Survey were redesigned in 2014.

Table 1
Medicare and Medicaid Participation Rates by Decile of Adjusted Family Income

	Medicare participation rate (%)			*Medicaid participation rate (%)*		
Decile of adjusted family income	*1995*	*2004*	*2012*	*1995*	*2004*	*2012*
Bottom decile	7.2	9.8	8.4	44.8	35.7	41.6
2	23.3	23.8	21.6	22.7	23.4	31.6
3	21.4	23.1	24.8	9.2	13.2	17.7
4	18.5	18.4	22.0	4.6	7.6	11.0
5	14.5	13.8	16.9	2.8	3.9	6.5
6	10.7	10.1	13.7	1.5	2.6	3.8
7	8.7	8.7	12.5	0.8	1.5	2.2
8	7.3	6.9	10.5	0.6	1.0	1.5
9	6.2	6.8	9.3	0.6	0.6	0.8
Top decile	6.6	6.5	9.3	0.5	0.4	0.8

Source: Data are from the 1996, 2005, and 2013 Annual Social and Economic Supplement to the Current Population Survey in which respondents report sources of health insurance coverage in the prior year.
Notes: The table shows the fraction of families in each decile in which at least one member participates in Medicare or Medicaid. Adjusted family income is total family income divided by the square root of family size.

Sullivan 2015).[6] Second, there has been some growth in Medicaid and Medicare participation in 2012 relative to earlier years. If the value of these benefits is taken into account in calculating "income" broadly understood as ability to consume, then the growth in participation in these two programs will tend to moderate growth in income inequality.

Recently, the Patient Protection and Affordable Care Act of 2010 has had a major impact on health insurance coverage and its impact has not been uniform throughout the income distribution. The Affordable Care Act created income-based subsidies for the purchase of individual health insurance on state or the federal health insurance "marketplaces" for persons with incomes up to 400 percent of the federal poverty level. The law also allowed states to expand Medicaid to all adults with incomes below 138 percent of the federal poverty level, and approximately half the states did so.

The subsidies in the newly created health insurance "marketplaces" and the expansion of Medicaid were fully implemented in 2014, and as a result, a substantially greater proportion of people in the lower part of the income distribution are

[6] The measurement error is also revealed by the presence of a surprising proportion of persons in the upper deciles of the income distribution that have Medicaid. Other surveys such as Survey of Income and Program Participation (SIPP) also underreport Medicaid participation.

Table 2
Health Insurance Coverage Rates of the Nonelderly in 2012 and 2014 by Decile of Adjusted Family Income

	Fraction uninsured			*Fraction on Medicaid*			*Fraction with private insurance*		
Decile of adjusted family income	*2012*	*2014*	*Δ2014–2012*	*2012*	*2014*	*Δ2014–2012*	*2012*	*2014*	*Δ2014–2012*
Bottom decile	39.3%	31.1%	−8.2	43.1%	47.6%	4.5	22.7%	26.5%	3.8
2	37.8	28.3	−9.5	34.4	38.9	4.5	34.3	40.0	5.7
3	30.9	23.3	−7.6	21.6	23.9	2.3	54.8	60.4	5.6
4	22.5	16.5	−6.0	15.3	17.1	1.8	69.2	73.6	4.4
5	15.5	11.8	−3.7	12.3	13.1	0.8	79.1	81.8	2.7
6	11.1	8.4	−2.7	9.9	11.1	1.2	85.3	87.0	1.7
7	7.8	6.1	−1.7	8.7	8.9	0.2	89.7	90.8	1.1
8	5.6	4.3	−1.3	7.4	7.7	0.3	92.6	93.4	0.8
9	4.1	3.3	−0.8	6.5	6.7	0.2	94.8	95.1	0.3
Top decile	2.8	2.3	−0.5	5.1	5.2	0.1	96.4	96.6	0.2

Source: Data are from the 2012 and 2014 one percent samples of the American Community Survey in which respondents report their current sources of insurance coverage. The sample is adults ages 22 to 64.
Notes: Medicaid category includes all public insurance programs. Adjusted family income is total family income divided by the square root of family size. Row totals do not add to 100 percent because people report more than one type of health insurance coverage. Figures in Tables 1 and 2 differ because of differences in survey design and sample.

now covered by health insurance. Table 2 shows the changes between 2012 and 2014 in health insurance coverage by decile of family income.[7] It is clear that the proportion of people without health insurance decreased between 2012 and 2014, and the decrease was largest for those in the lowest income decile and smallest for those in the highest income decile. Among those in the three lowest deciles of family income, the proportion of people without health insurance declined by between 8 and 10 percentage points, and among those in the three highest deciles of family income, the proportion of people without health insurance decreased by only 1 percentage point. While it is unlikely that the entire change in insurance coverage shown in Table 2 is due to the Patient Protection and Affordable Care Act of 2010, it is reasonable to attribute a substantial portion of the change to that law. The relatively large increase in health insurance in the lowest income deciles was due to a relatively large increase in Medicaid coverage, which was likely almost all due to the Affordable Care Act, as the recovering economy would have

[7] Data are from the 2012 and 2014 one percent samples of the American Community Survey in which respondents report their current sources of insurance coverage. Note that the ACS was begun after 2000 and did not collect information on health insurance coverage prior to 2008. The ACS is arguably preferable to the Current Population Survey (CPS) for measuring health insurance coverage because of the survey design and consistency of the survey question over time. The CPS redesigned its health insurance questions in 2014.

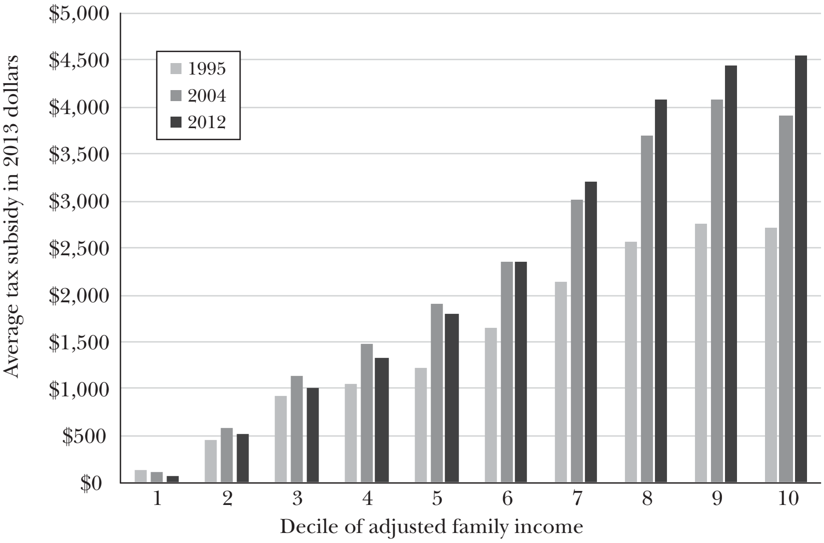
dampened Medicaid enrollment in the absence of the 2010 law. Increases in private insurance coverage between 2012 and 2014 were also concentrated in the lowest deciles, which is consistent with the income-based subsidies available in the health "marketplaces, "but some of this increase may be due to an improving economy.

The implication of the figures in Table 2 is that the Patient Protection and Affordable Care Act of 2010 has reduced inequality. The increase in Medicaid coverage in 2014 relative to 2012 is highly concentrated in the lower deciles of the income distribution. The increase in private insurance coverage is also concentrated in the lower half of the income distribution. Notably, according the Centers for Medicare and Medicaid Services, 85 percent of the people who obtained private insurance through the health insurance "marketplaces" received some amount of subsidy (in-kind benefit) and the probability of receiving a subsidy and the amount of the subsidy decreases with income (Centers for Medicare and Medicaid Services 2015). Thus, the government transfer associated with the increase in private insurance coverage is likely to be even more concentrated in the lower deciles of the income distribution than the observed increase in private insurance coverage. Moreover, the reduction in uninsured and increase in other types of health insurance coverage are expected to grow over time because of greater awareness of the Affordable Care Act, increased tax penalties for people who do not obtain health insurance, and the expansion of Medicaid in states that occurred post-2014 (for example, Pennsylvania and Indiana).

Government tax policy for health care affects inequality because employer-provided health insurance is generally not taxed as individual income at the federal or state level, or through the payroll tax. To the extent that individuals, rather than firms, capture the incidence, the tax exclusion provides a benefit that increases with the cost of insurance and with the workers' marginal tax rate (and hence total income). This tax exclusion is estimated to cost the US Treasury about $216 billion in foregone income taxes and $132 billion in foregone payroll taxes in fiscal year 2016 (US Office of Management and Budget 2015, Table 14.1).

The tax treatment of employer-provided health insurance has potentially large effects on inequality. Employer-sponsored coverage is more prevalent among higher-income families. Approximately 80 percent of those in the top three income deciles have employer-provided insurance (either in their own name or as a dependent of someone else's policy), whereas only 10 to 30 percent of those in the lowest three income deciles have this type of insurance. In addition, our estimated measure of average household premiums for employer-provided health insurance is much higher in the top income deciles. Premiums for employer-provided health insurance are not reported in the Current Population Survey. Therefore, we impute premiums by single-digit industry and firm size using data from the Medical Expenditure Panel Survey–Insurance Component. By this metric, the average value of employer-provided health insurance premiums for those in the top two income deciles exceeds $10,000, while the average value for those in the bottom two deciles (including values of zero for many in the bottom deciles who do not have employer-provided health insurance) is less than $1,000. Thus, including

Figure 1
Average Tax Subsidy for Employer-Provided Health Insurance by Decile of Adjusted Family Income



Source: Authors' calculations from the 1996, 2005, and 2013 Annual Social and Economic Supplement to the Current Population Survey.
Notes: Adjusted family income is total family income divided by the square root of family size. The calculation of the subsidy is described in the text. Calculations include zeros for families not covered by employer-provided insurance. Insurance coverage, family income, and the tax subsidy refer to the year prior to each survey.

the value of employer-sponsored health insurance as income will tend to increase measured inequality. Its inclusion may also exacerbate the growth in measured inequality since premiums have been rising over time and coverage among lower-income families has been falling over time.

The tax treatment of employer-sponsored insurance will also increase income inequality since marginal tax rates and premiums rise with income. Figure 1 shows our estimated tax subsidy for employer-sponsored insurance per family, by year and decile of adjusted family income and including zeros for families without employer-provided insurance. To calculate these figures, we used the TAXSIM program of the National Bureau of Economic Research.[8] We estimate the total tax bill for families in the 1996, 2005, and 2013 Current Population Survey, first ignoring any employer-sponsored health insurance and then including imputed health insurance premiums as wage income. The difference between these is our estimate of the

[8] TAXSIM version 9.3. For details, see Feenberg and Coutts (1993) and TAXSIM related files at the NBER: http://www.nber.org/taxsim/.

subsidy for employer-provided health insurance.[9] The tax subsidy grows with income and has been growing over time. For example, the tax subsidy among families in the bottom three deciles is under $1,000 (in 2013 dollars). The subsidy in the top three deciles was about $2,500 in 1995 (in 2013 dollars) and was $4,000 to $4,500 in 2012.

Several other ways in which the government subsidized healthcare spending—much smaller in size than Medicare, Medicaid, and the tax exemption for employer-provided health insurance—are worth mentioning. For example, individual spending on medical care is deductible from income if it exceeds a threshold level of adjusted gross income (currently 10 percent of income for people under age 65). People can also make tax-deductible contributions to a medical savings account, which can be used for medical expenses after retirement. In 2016, deductibility of out-of-pocket expenses cost the US Treasury about $7.6 billion, and the deductibility of medical savings accounts cost another $5.6 billion (US Office of Management and Budget 2015, Table 14.1).

The government also spends money on public health, for example, the Centers for Disease Control (CDC) and National Institutes of Health (NIH). These and other types of public health spending may affect inequality through their direct effect on population health (Miller, Roehrig, Hughes–Cromwick, and Lake 2008). However, government spending on what may be called public health is relatively small with estimates in the 1 to 2 percent range of all government health care spending (Miller et al. 2008).

Accounting for Taxes and Transfers in Measured Inequality

A natural place to begin an assessment of how health policies affect inequality is to adjust family income for in-kind health insurance benefits and for tax subsidies. In this section, we review the small number of studies that have done this and also provide some new estimates. The existing papers assess how including the value of government-provided health insurance and employer-provided health insurance in income affected inequality, both at a point in time and with respect to changes over time.

Pierce (2001) was one of the first papers to extend the inequality literature to include "nonwage" aspects of income by incorporating employee fringe benefits into the analysis. His focus was on explaining inequality in the returns to market work, rather than inequality in well-being more broadly defined, and thus did not incorporate publicly provided insurance into his analysis. Health insurance is one of the most important fringe benefits accounting for approximately 5 percent of total compensation for civilian employees in 1997. Because health insurance coverage rises with employee earnings, as mentioned earlier, Pierce showed that the inclusion of employer-provided health insurance widens compensation

[9] When we include employer-provided health insurance as taxable income, we compute the Earned Income Tax Credit and Child Tax Credit based on wage income alone.

inequality slightly at a point in time (Pierce 2001; see also Chung 2003). In addition, employee health insurance has grown as a share of compensation: by 2015, health insurance costs accounted for 8.4 percent of compensation (Bureau of Labor Statistics 2016). This growth would have caused both a rise in inequality in total compensation and a larger discrepancy between inequality in wages and inequality in total compensation.

Pierce (2001) also showed how changes in fringe benefits affected changes in inequality in hourly compensation between 1982 and 1996, exacerbating the increase in inequality that occurred between 1982 and 1996 by 15 percent. Pierce did not separately identify the effect of health insurance, although he did report that the incidence of employer-provided health insurance was declining over the period. Health care costs and insurance premiums were also rising during this period (General Accounting Office 1997). Given that most of the decline in health insurance coverage was concentrated at the low-end of the earnings distribution, changes in health insurance benefits likely increased the change in inequality during the period. Chung (2003) used different data (Current Population Study and Chamber of Commerce Employee Benefits Survey) than Pierce (2001) and reached similar conclusions.

In a series of papers, Burkhauser and colleagues provided assessments of integrating employer-provided health insurance, Medicaid, and Medicare into the analysis of inequality (Burkhauser and Simon 2010; Burkhauser, Larrimore, and Simon 2012, 2013). An important measurement issue highlighted in these studies is how to value Medicare and Medicaid benefits (Burkhauser, Larrimore, and Simon 2013). The US Census Bureau has developed an approach that estimates the "fungible" value of Medicare and Medicaid, which measures the resources from a family's budget freed up by Medicare or Medicaid that the family can use for other purposes. This definition presumes that no money is freed up if the family has insufficient income to cover their basic food and housing costs, and thus the fungible value is set to zero.[10] This definition seems problematic if the purpose is to assess inequality in resources; after all, Medicaid benefits do make a family better off in terms of consumption, even if all nonmedical spending by that family remains the same. Therefore, Burkhauser et al. (2012, 2013) impute the value of Medicaid benefits as the per person Medicaid expenditures by state, year, and age group. The value of Medicare benefits is set equal to Medicare expenditures per person by state and year. They impute a value to employer-provided insurance using average premium information from the Medical Expenditure Panel Survey–Insurance Component, by state, year, and firm size.

[10] Specifically, the fungible value is an imputed value of medical expenditures for households who have sufficient income to cover their basic food and housing costs and enough income left over to purchase insurance on their own. The fungible value is zero for families who do not have sufficient income to cover to their food and housing costs and is prorated for families who have enough to cover their food and housing needs but not enough to cover the full cost of insurance. See "Calculating Fungible Values: Medicare, Medicaid" at the US Census Bureau website: https://www.census.gov/cps/data/fungible.html.

The Burkhauser et al. (2012, 2013) studies yield several important results. First, the overall effect of including the value of employee health insurance and Medicaid/Medicare reduces the amount of inequality at a point in time. In Burkhauser et al. (2013), for example, the ratio of the 90th to 10th percentiles of the income distribution is 8.2 in 1995 when pre-tax (post-transfer) income is the measure of income, but this ratio declines to 6.0, or by 27 percent, when the value of employee health insurance and Medicaid/Medicare are included in the income measure. Second, accounting for employee health insurance benefits and government healthcare programs had modest effects on overall changes in inequality over time. Burkhauser et al. (2012) reported that the Gini coefficient for family income increased by 13 percent between 1979 and 2007 when employer-provided health insurance and Medicaid/Medicare benefits were ignored, while including these benefits resulted in the Gini coefficient increasing by 10 percent. In contrast, accounting for the value of health insurance had a much larger effect on income inequality between the top and bottom income quintiles. Between 1979 and 2007, mean income in the top income quintile grew 3.3 times faster than in the bottom quintile when the value of health insurance is ignored, but including the value of health insurance reduces this figure to 2.0. A contemporaneous study of inequality by the Congressional Budget Office (2011) reported similar results.

Burkhauser et al. (2013) also assessed how the Patient Protection and Affordable Care Act of 2010, which expanded Medicaid to more low-income adults (in states that agreed to participate in the expansion) and also provided subsidies to families to purchase health insurance through exchanges, would affect income inequality. Aaron and Burtless (2014) conducted a similar analysis using similar data from the Medical Expenditure Panel Survey and methods. As shown in Table 2, the Affordable Care Act expanded insurance coverage and provided monetary benefits mostly to those in the lower half of the income distribution with benefits increasing as income decreases. Aaron and Burtless (2014) concluded that the largest effects of the Affordable Care Act will be to increase full income (income plus in-kind benefit of health insurance) for the 15th to 30th percentiles of the income distribution. This prediction is consistent with the data in Table 2. Thus, Medicaid expansions and insurance premium subsidies from the Affordable Care Act will reduce income inequality generally, but particularly between the top income categories and the first to third deciles.

In two recent papers, Meyer and Sullivan (2003, 2012) have argued that inequality in consumption is preferable to income as a way to measure inequality in wellbeing, primarily because consumption is conceptually a better way to measure long-run, permanent resources than income, which has a substantial transitory component. Meyer and Sullivan (2012) compare and contrast the level and trends in inequality using income- and consumption-based approaches, and their results are consistent with the other studies. Accounting for noncash transfers including, but not limited to, the (fungible) value of public and private health insurance reduces the level of inequality at a point in time, but has little effect on changes in inequality over time, whether measured using income- or consumption-based measures of inequality.

To illustrate more clearly the effect of including employer-provided health insurance benefits in the analysis of inequality, we conducted an analysis similar to those in Burkhauser et al. (2013) using data from the 1996, 2005, and 2013 Current Population Surveys (which report income and insurance coverage for the prior years). Table 3 present the results, showing the 10th, 50th, and 90th percentiles of various measures of family income adjusted for family size, and the ratio of the 90th to the 10th, the 50th to the 10th, and the 90th to the 50th percentiles of adjusted family income. For all measures of family income, we assign each household member their family income and then compute the percentiles across all people. Inequality is rising between 1995 and 2012. This is true whether income is measured pre-tax or post-tax. For example, the 90–10 ratio of the distribution of pre-tax adjusted family income rose from 9.7 to 10.3 to 11.7 across the three years, an increase of 21 percent. The 90–10 ratio of post-tax income rose from 6.6 to 7.9 (20 percent). The 50–10 and 90–50 ratios in post-tax incomes each rose by 10 percent from 1995 to 2012.

Row 3 of Table 3 presents measures of inequality in which we added to family income an imputed value of Medicare or Medicaid. We follow Burkhauser et al. (2012, 2013) and assign to each recipient average Medicare expenditures by year and state and average Medicaid expenditures by age, year, and state.[11] The 90–10 ratio of after-tax family income in 1995 falls from 6.6 to 5.0, or by 24 percent, when our imputed value of Medicare and Medicaid is added to family income. Including these benefits has a slightly larger effect on measured inequality in 2012. The 90–10 ratio in 2012 falls from 7.9 to 5.6, a fall of 29 percent. Not surprisingly, all of the effects of including Medicare and Medicaid on inequality are within the lower half of the income distribution.

Rows 4 and 5 of Table 3 demonstrate what happens to the measures of income inequality if we add employer contributions for employer-provided health insurance to family income.[12] When employer-provided insurance is not taxed as income (as is currently the case), overall inequality increases when we add employer-contributions to family income. Comparing rows 2 and 4, in 2012, the 90–10 ratio in after-tax family income rises from 7.9 when employer-provided health insurance is ignored in the calculation of family income to 8.2 when employer contributions are added to

[11] Data on Medicare expenditures are from the State Health Expenditures files at the Centers for Medicare and Medicaid Services, which contain data from 1991 to 2009. We impute Medicare expenditures for respondents in the 2013 CPS (which contains program participation for 2012) using Medicaid data from 2009, adjusted for the overall growth in Medicare spending from 2009 to 2012. Data on Medicaid expenditures are from Medicaid Statistical Information System (MSIS). The MSIS has expenditures by state and age groups from 1999 to 2012, though many states are missing from the 2012 data. For respondents in the 1996 CPS (which reports insurance for 1995), we impute Medicaid expenditures from 1999 and discount the figure back to 1995 using the overall change in Medicaid expenditures between 1995 and 1999. For respondents in the 2005 CPS, we impute Medicaid expenditures from 2004. For respondents in the 2013 CPS, we impute Medicaid expenditures from 2011 and adjust the figures for the overall change in Medicaid spending between 2011 and 2012. Details are available upon request.

[12] Like total premiums, we impute employer contributions by single-digit industry and firm size using data from the Medical Expenditure Panel Survey–Insurance Component.

Table 3

Measures of Points in the Distribution of Family Income and Income Inequality in 1996, 2005, and 2012

(*in 2013 dollars*)

	10th percentile	*50th pecentile*	*90th percentile*	*90/10*	*50/10*	*90/50*
Panel A: 1995						
1. Pre-tax family income	$8,369	$32,654	$80,976	9.7	3.9	2.5
2. After-tax family income	$8,219	$25,364	$54,492	6.6	3.1	2.1
3. After-tax income plus imputed value of Medicare and Medicaid	$11,078	$27,179	$55,894	5.0	2.5	2.1
After-tax family income plus employer-provided health insurance (EPHI) premiums						
4. where EPHI is not taxed	$8,472	$27,545	$57,807	6.8	3.3	2.1
5. where EPHI is taxed	$8,399	$26,490	$55,904	6.7	3.2	2.1
After-tax family income plus EPHI premiums and imputed value of Medicaid and Medicare						
6. where EPHI is not taxed	$11,637	$29,354	$59,121	5.1	2.5	2.0
7. where EPHI is taxed	$11,341	$28,337	$57,162	5.0	2.5	2.0
Panel B: 2004						
1. Pre-tax family income	$8,829	$35,222	$90,726	10.3	4.0	2.6
2. After-tax family income	$8,879	$28,297	$63,339	7.1	3.2	2.2
3. After-tax income plus imputed value of Medicare and Medicaid	$12,082	$30,682	$65,092	5.4	2.5	2.1
After-tax family income plus employer-provided health insurance (EPHI) premiums						
4. where EPHI is not taxed	$9,366	$31,748	$68,262	7.3	3.4	2.2
5. where EPHI is taxed	$9,293	$30,127	$65,625	7.1	3.2	2.2
After-tax family income plus EPHI premiums and imputed value of Medicaid and Medicare						
6. where EPHI is not taxed	$13,071	$33,879	$69,709	5.3	2.6	2.1
7. where EPHI is taxed	$12,814	$32,332	$67,234	5.2	2.5	2.1
Panel C: 2012						
1. Pre-tax family income	$7,610	$33,122	$89,401	11.7	4.4	2.7
2. After-tax family income	$8,100	$27,549	$64,189	7.9	3.4	2.3
3. After-tax income plus imputed value of Medicare and Medicaid	$11,941	$30,973	$66,911	5.6	2.6	2.2
After-tax family income plus employer-provided health insurance (EPHI) premiums						
4. where EPHI is not taxed	$8,502	$31,169	$70,101	8.2	3.7	2.2
5. where EPHI is taxed	$8,460	$29,560	$66,919	7.9	3.5	2.3
After-tax family income plus EPHI premiums and imputed value of Medicaid and Medicare						
6. where EPHI is not taxed	$12,624	$34,478	$72,307	5.7	2.7	2.1
7. where EPHI is taxed	$12,478	$32,849	$69,354	5.6	2.6	2.1

Source: Authors calculations' from the 1996, 2005, and 2013 Annual Social and Economic Supplement of the Current Population Survey, as described in the text. Survey respondents report sources of health insurance coverage in the prior year.

Notes: The table shows the 10th, 50th, and 90th percentiles of various measures of family income adjusted for family size, and the ratio of the 90th to the 10th, the 50th to the 10th, and the 90th to the 50th percentiles of adjusted family income. All incomes are expressed in 2013 dollars.

income but are not taxed. The effect of including employer contributions towards insurance in 1995 and 2004 is the same order of magnitude as the effect in 2012. The increase is entirely driven by increased inequality in the 50–10 ratio, where the disparity in employer-provided health insurance is larger.

The tax subsidy for employer-provided health insurance increases income inequality modestly. Comparing rows 4 and 5, in 2012, the 90–10 ratio would fall from 8.2 to 7.9 if employer-provided insurance were taxed, a decline of about 4 percent.[13] This effect comes from declines in inequality in the lower half of the distribution (the 50–10 ratio). Taxing employer-provided health insurance has slightly smaller effects in 1996 and 2005.

Finally, in Rows 6 and 7 we present measures of inequality in post-tax family income that include the imputed values of Medicare, and Medicaid, and employer-contributions to health insurance. The net effect of including all three sources of insurance is to reduce measured inequality in after-tax income. The percentage change in inequality between rows 2 and 6 is 23 percent in 1995, 25 percent in 2004, and 28 percent in 2012. Moreover, the growth in after-tax income inequality is moderately smaller when health insurance is included in income. Inequality in after-tax income (row 2) increased by 20 percent between 1995 and 2012, while inequality inclusive of public and private insurance (row 6) increased by 13 percent. A comparison of the final two rows of Table 3 indicates that taxing employer-provided health insurance would reduce the 90–10 ratio by about 4 percent in 2012 and would have slowed the growth in equality from 1995 to 2012 by 2 percent.

To summarize, our reading of the evidence is that adjusting for public and private health insurance has a considerable effect on inequality at a point in time. Incorporating Medicare and Medicaid tends to flatten the distribution of income because the benefits of Medicaid, and to a lesser extent Medicare, accrue largely to those at the bottom of the resource distribution. Moreover, these programs have lessened the growth in inequality over time. The Patient Protection and Affordable Care Act of 2010 strengthened this effect by expanding Medicaid and providing subsidies for those with incomes up to 400 percent of the federal poverty line. Our conclusion is similar to that from the literature on poverty, which shows that accounting for noncash transfers has a substantial poverty-reducing effect (Burtless and Smeeding 2001; Haveman, Blank, Moffitt, Smeeding, and Wallace 2015). Accounting for employer-provided health insurance tends to increase measured inequality since coverage and total premiums tend to rise with income, and focusing on the tax subsidy for employer-provided insurance further exacerbates measured inequality.

[13] In these calculations, we assume that both the employer and employee portions are taxed at the federal and state level, including payroll taxes for Social Security and Medicare. We assume that the Earned Income Tax Credit and Child Tax Credit are not affected by the tax treatment of employer-provided health insurance.

Conceptual and Empirical Issues

Valuing Government- and Employer-Provided Health Insurance

In this section, we highlight some significant issues for assessing the effect of health policy on inequality. Perhaps the most important issue is how to value government and employer-provided insurance for the purposes of assessing inequality, an issue we briefly discussed above. Suppose that a person on Medicaid receives insurance that would have cost $6,000 if purchased through other means (such as the individual market or through an employer-sponsored plan). Does providing Medicaid to this person have the same effect as increasing the person's income by $6,000? In one view, the answer is "yes," because Medicaid allows the person to spend $6,000 less on health insurance and thus spend $6,000 more on other goods. The Census Bureau's "estimated fungible value" is based on this reasoning. On the other hand, in the absence of the publicly provided health insurance, the person might have chosen to spend less on insurance, or have gone without insurance, depending on their risk of illness, income, and preferences. In that case, assigning the full cost of Medicaid to such a family overstates the degree to which their access to other resources that they value has changed.

Indeed, Finkelstein, Hendren, and Luttmer (2015) estimate that Medicaid enrollees in Oregon value a $1 of Medicaid benefits at between $0.20 and $0.40. That is, they would be unwilling to enroll in Medicaid if they had to pay a premium equal to the government's cost of providing such insurance. One primary reason for this finding is that most people who are enrolled in Medicaid pay only 20 to 30 percent of the cost of care when uninsured, which greatly reduces the out-of-pocket savings and insurance value of Medicaid.[14] Therefore, they would be unwilling to pay the government's cost of care, or to buy insurance on the private market, which would be even costlier than the government's cost of providing care.

To illustrate how the approach to valuing Medicare and Medicaid benefits that is used by the Census Bureau affects the income distribution, in Table 4 we show the average fungible values (fungible values are determined by the US Census Bureau) and our imputed values of Medicaid and Medicare from the 2013 Current Population Survey by decile of family income. In the calculations, we included zeros for families that do not participate in these programs. As noted earlier, the Census Bureau's estimated fungible value of Medicare and Medicaid is zero for families whose income is below that required to provide their own food and shelter. A notable feature of Table 4 is the very low value of the fungible Medicaid benefits in the first two deciles of the income distribution, even though participation is high among these families. This fact explains why using the fungible values has a far

[14] The 20 to 30 percent figure does not include potential costs associated with unpaid bills such as higher costs of borrowing or restricted access to credit. Evidence in Finkelstein et al. (2012) and Dobkin, Finkelstein, Kluender, and Notowidigdo (2016) support this approach of ignoring potential future consequences of unpaid bills. Conversely, Mazumder and Miller (forthcoming) reported that gaining health insurance was associated with improved credit scores after four years.

Table 4

Fungible and Imputed Values of Medicare and Medicaid, by Decile of Family Income, 2012

Decile of adjusted family income	Medicare		Medicaid	
	Fungible value	*Imputed value*	*Fungible value*	*Imputed value*
Bottom decile	$25	$2,054	$70	$3,560
2	$624	$4,846	$785	$3,049
3	$2,750	$5,569	$1,099	$1,638
4	$4,191	$5,019	$893	$1,026
5	$3,713	$4,010	$633	$683
6	$3,150	$3,355	$358	$424
7	$2,896	$3,083	$217	$269
8	$2,551	$2,705	$157	$195
9	$2,212	$2,363	$81	$107
Top decile	$2,174	$2,335	$71	$107

Source: Authors' calculations from the 2013 Annual Social and Economic Supplement of the Current Population Survey.
Notes: Adjusted family income is total family income divided by the square root of family size. Average values include zeros for nonparticipants and are expressed in 2012 dollars. The US Census Bureau determines fungible values. For our imputed values of Medicare or Medicaid, we assign to each recipient average Medicare expenditures by year and state and average Medicaid expenditures by age, year, and state. See footnote 11 for further details.

smaller effect on measured income inequality compared to using, for example, a measure based on average medical expenditures. For assessing the effect of health insurance on well-being, an expenditure-based measure is probably better than the fungible value, although in light of the findings in Finkelstein et al. (2015), the use of the total Medicaid cost/expenditure likely overstates the value of Medicaid. The problem highlighted by Finkelstein et al. (2015) is less likely to be present for Medicare because the elderly would likely pay a larger proportion of medical costs out-of-pocket when uninsured than Medicaid recipients.

A similar issue arises in valuing employer-provided health insurance. In the canonical model of compensating wage differentials (Rosen 1986; Currie and Madrian 1999), employees' trade off lower wages for employer-provided health insurance. A full accounting of employee compensation should, therefore, include the value of employer-provided insurance. Yet the value of the health insurance to the employee, as measured by what insurance the employee would have purchased in the individual market in the absence of employer-provided coverage, will generally differ from employer's cost of insurance. Employers' cost reflects the average medical care use over all employees. Employees who place a higher value on the insurance, either because they use more medical care or because they have stronger preferences for insurance, will earn rents that are not properly valued when the average health insurance premium is assigned to all persons in a firm (or type of firm).

In general, people's demand for health insurance depends on several factors, but most importantly, on their risk of illness, income, and preferences (such as the extent of risk aversion). Low-income people have a greater prevalence of illness, which would cause them to have relatively strong demand for health insurance and place a relatively higher value on insurance relative to the cost to provide it. On the other hand, low-income persons have a lower willingness to pay for insurance (a higher marginal cost of consumption), which tends to reduce their demand vis-à-vis higher income persons. Similarly, older people are sicker and use more medical care; indeed, the ratio of health care spending between older and younger persons can be a factor of three or four.[15] Accordingly, older people have a relatively high demand for health insurance.

The heterogeneity in the demand for health insurance, and thus the value of health insurance to the consumer, suggests that using the average expenditures in Medicaid and Medicare, or average premium for employer-provided health insurance as the value of insurance can result in large errors with respect to the person's actual valuation of those benefits. This point can be illustrated using the example of a person over age 65 who is working and who has employer-provided insurance. The average health insurance premium in the firm may be $6,000 for an individual (Kaiser Family Foundation and Health Research Educational Trust 2013). If that person were to retire, the average Medicare expenditure would be over $11,000 per person. Neither value is likely to be correct as a value for what that health insurance is worth to the consumer. The 65-year old probably has a higher value of insurance than the average employee in the firm and a lower value of insurance than the average Medicare recipient. To the extent that the heterogeneity of demands for insurance, for example, by age, differs across the income distribution, then these errors in valuation may mask important effects of health insurance on inequality. For example, young people are overrepresented in the lower part of the income distribution and they also have relatively low willingness to pay for health insurance because they are healthier and perhaps less risk averse. Thus, assigning them the average value of employer-sponsored health insurance premium represents an overstatement of the value that a young person places on the benefit.

Burtless and Svaton (2010) address this issue, at least in part. Instead of assigning an individual an average value for Medicaid, Medicare, or private health insurance, they calculated health care expenditures for each individual and added these expenditures (net of out-of-pocket payments) to income. The value of health insurance depends on the amount of expected health care expenditures and not the actual amount of expenditures in any one period. Therefore, using actual expenditures may seem incorrect as a measure of the value of insurance, but if individuals are aggregated into groups such as by income, then the average for

[15] This multiple is based on a Fact Sheet from the Centers for Medicare and Medicaid Services, "U.S. Personal Health Care Spending by Age and Gender: 2010 Highlights," available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/2010AgeandGenderHighlights.pdf.

the group will equal the expected expenditure. Adjusting for these differences in health expenditures raised income in the bottom tenth of the income distribution by 65 percent in the population under 65, and by 130 percent in the 65-and-over population. In contrast, adding health care expenditures to the incomes of the top decile would raise income by 2 percent for nonelderly and 6 percent for the elderly. These results strengthen the conclusion that adding the value of health insurance to income reduces measured inequality.

A related issue is that the market price of health insurance—for example, employer-paid health insurance premiums—will overstate the average willingness to pay for such benefits because moral hazard increases the quantity of medical care use beyond the point where the marginal benefit equals the marginal cost. As studies have shown (for example, Newhouse and the Insurance Experiment Group 1993; Finkelstein et al. 2012), the use of medical care and presumably the overstatement of willingness to pay increases with the generosity of health insurance—that is, with the extent of moral hazard. If the generosity of insurance differs across the income distribution, then this measurement error could affect the level of inequality. In fact, we know that the generosity of private health insurance has declined over time, particularly in the recent period, as more private plans have raised copayment rates and deductibles (Kaiser Family Foundation and Health Research Educational Trust 2013). Also, public insurance (like Medicaid) is usually more generous than private health insurance because of low copayment rates and subsidized (Medicare) or no (Medicaid) premium requirements. Therefore, private health insurance premiums may better reflect a willingness-to-pay than Medicaid expenditures. This disparity may be growing over time because of the increasing use of high deductible plans in private insurance.

To summarize, current methods to value health insurance benefits for analyzing its effect on inequality are imperfect and there is room for improvement. Consider the different methods described for valuing Medicaid. For some families, the government assigns zero value because the family has no discretionary income and therefore providing this family with Medicaid does not increase consumption of nonmedical goods. This approach seems incorrect because in these instances the family's total consumption (inclusive of health care expenditures) increases. However, evidence in Finkelstein, Hendren, and Luttmer (2015) suggest that it is not too far off the truth, and arguably closer to the correct value than either using the average government expenditure for that person or the amount of medical expenditures incurred by that person. Perhaps a reasonable approach, which is driven primarily by practical considerations, would be to use 50 percent of health care expenditures as the value of Medicaid benefits.

The second major issue in properly valuing health insurance benefits is the heterogeneity in demand for health care and health insurance. The demand for health care and health insurance is correlated with age, illness, and other demographic factors. Therefore, using any average value of health insurance will result in errors, and the magnitude of the errors will be correlated with income because of the correlation between age, illness, and other factors and income. Therefore, it is arguably preferable to use health care expenditures for each person. Calculating health care expenditures within an income group (say, an income decile) may be preferable because it will account for heterogeneity of demands for insurance by age, income, and other factors within that group and produce the average/expected expenditure, which in a well-functioning insurance market expenditures should equal health insurance premiums. Of course, there are market failures in health insurance markets that likely raise expenditures above the optimal amount. One crude approach, which we mainly justify on practical grounds, would be to use 80 percent of health care expenditures, assuming that 20 percent of expenditures are above the optimal amount due to market failures such as moral hazard (for example, Newhouse and the Insurance Experiment Group 1993).

Incorporating Behavioral Responses to Public Policies

Government programs and tax policies can lead to behavioral responses, which in turn may reduce or exacerbate measured income inequality. Consider Medicaid, and the possibility that it will affect labor supply (Baicker et al. 2013; Garrett and Kaestner 2014; Kaestner, Garrett, Gangopadhyaya, and Fleming 2016). Suppose a Medicaid expansion induces some people to drop out of the labor force (because they no longer must need to be employed to receive health insurance), which causes their incomes to decline. The decline in income could conceivably even be greater than the value of Medicaid benefit (however measured). A comparison of inequality before and after the Medicaid expansion would indicate that inequality in wage income rose and may indicate that inequality in the sum of wage income and an imputed value of Medicaid also rose. The labor supply response leads to an increase in measured inequality, even though the program itself clearly transfers resources to lower-income individuals.

This example highlights that analyses of inequality that focus on wage, incomes, and even on incomes plus transfers do not fully measure inequality of well-being because such measures ignore the value of time spent not working, which includes leisure and productive activities such as child-rearing that make life better for people. In the example above, the Medicaid expansion made recipients better off in terms of welfare, even if their income declined. The amount of time spent (not) working has changed over time and the change has been different for low-wage and high-wage workers; low-wage workers have increasingly spent less time working whereas high-wage workers have increasingly spent more time working (Aguiar and Hurst 2007; Kuhn and Lozano 2008; Mishel 2013). Accounting for these changes in the difference in time spent not working would suggest inequality has increased less than that suggested by changes in income.

One practical approach to this problem is to compute inequality using different assumptions about such behavioral responses. For example, Baicker et al. (2013) reported that, in the Oregon randomized control trial, obtaining Medicaid coverage was associated with a 3 percent decrease in employment and a 3 percent, or $200, decrease in earnings in 2008 (which was not statistically different from zero). This was a study of childless adults, which are a small part of the Medicaid population; and the estimates may differ across demographic groups. But if this response were incorporated into the analysis of income inequality, we would deduct $200 from the value of Medicaid before adding Medicaid benefits to income. This adjustment would have a minimal effect on inequality because $200 is a small fraction of the cost of Medicaid per person. However, this approach still ignores the improvement in well-being among people who choose not to work in response to receiving Medicaid.

On the other hand, Garthwaite, Gross, and Notowidigdo (2014) reported that between 63 and 90 percent of people who lost Medicaid coverage in Tennessee as a result of a policy change in 2005 gained employment. If we assume that earnings in Tennessee (which were not reported in the study) were about the same as in the Oregon sample, $6,500 per year, then the labor supply response in Garthwaite et al. (2014) would imply that we should deduct between $4,095 and $5,850 from the value of Medicaid to account for the reduction in labor market earnings. A labor supply response of this magnitude would imply a very small net value of Medicaid, and so accounting for Medicaid benefits would not reduce measured inequality (and may even increase it). It is important to note that the relatively large labor supply effect reported by Garthwaite et al. (2014) is an outlier in the literature.[16]

Similar considerations would apply to Medicare, which may alter labor supply among the elderly (Madrian 2005). In the case of Medicare, there is evidence that it reduces labor supply, but modestly (French and Jones 2011; Blau and Gilleskie 2008). Thus, adding in the value of Medicare benefits to elderly income would modestly overstate the impact of this benefit on inequality of income plus transfers. It is unlikely that this behavioral response would significantly affect inequality, however, because the Medicare population is a small part of the total population, and those whose labor supply decisions would be affected are an even smaller part of the population when one considers the number who would not work regardless of the availability of Medicare. In addition, Medicare benefits are found throughout the income distribution.

Finally, the tax deductibility of employer-sponsored health insurance may bring forth a labor supply response because it may lower the effective wage.[17] Workers may respond to the lower wage by working more or less depending on income and substitution effects. Here too, it would be necessary to account for changes in time spent not working to assess fully how this change will affect inequality.

Including the Value of Health

Health is also an important source of well-being, and like leisure, it has not been considered fully in analyses of inequality. This omission is significant because

[16] A related issue is that publicly provided health insurance may crowd out employer-provided insurance (Cutler and Gruber 1996). Our calculations assume that the public and employer-provided insurance are equally valuable to the employee.

[17] Since employer-provided health insurance is partly a fixed cost, taxing employer contributions for employer-sponsored health insurance may not change the after-tax wage rate appreciably. However, it could change the generosity and cost of employer-provided insurance and therefore change the mix of compensation firms and employees choose.

health is one of the most valued sources of wellbeing and consumers have a very high willingness to pay for health (Murphy and Topel 2006). Moreover, poor health may reduce both the value of leisure and of consumption.

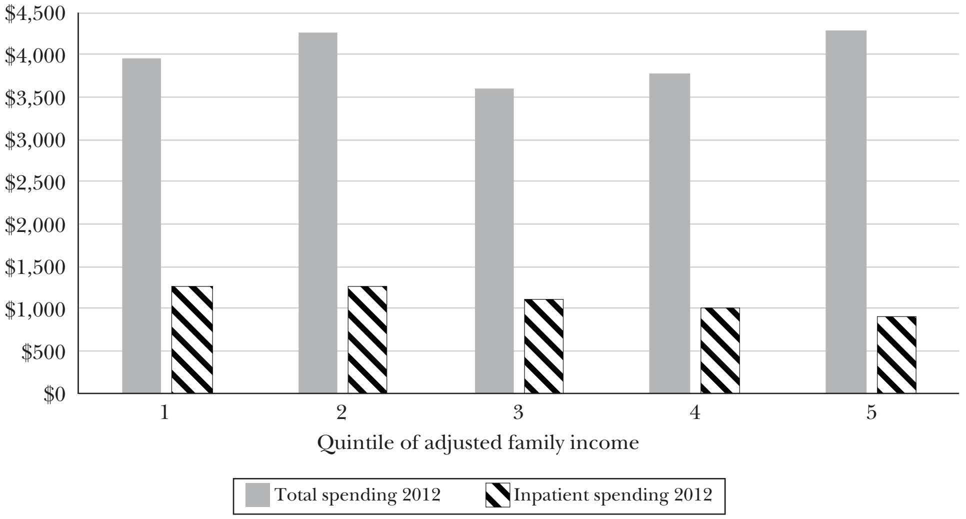
As noted earlier, health is strongly and positively correlated with income. Perhaps the most compelling evidence of this association is found in a recent National Academy of Sciences (2015) report, which reported that life expectancy of 50-year-old men in the top income quintile in 2010 was nearly 50 percent higher, or 12.7 years longer, than life expectancy of same aged men in the lowest income quintile. The analogous figure in 1980 was 20 percent (5.1 years). This is a large increase in the income–mortality gradient over a 30-year period. Thus, including the value of good health, an important source of well-being, in an analysis of inequality would likely significantly increase measured inequality and the growth in inequality because of the association between income and health that is independent of health insurance or spending on medical care.

The inclusion of health insurance benefits in analyses of inequality partly addresses this issue to the extent that health insurance reflects medical spending, and medical spending is arguably related to health. However, the association between health and spending on health care is relatively weak (Newhouse, and the Insurance Experiment Group 1993; Fisher et al. 2003; Baicker et al. 2013). A substantial part of healthcare spending reflects spending after illness that (often inadequately) restores a person to health. Figure 2 illustrates this point and shows total medical care spending and total inpatient spending for people age 18 and older in the full-year Medical Expenditure Panel Survey data for 2012. Total spending on health care is not strongly correlated with income, despite the strong correlation between health and income. However, there is a negative correlation between inpatient spending, which is usually for serious illness, and income. In the bottom quintile, inpatient spending is approximately $1,300 (32 percent of total spending) whereas in the top quintile, inpatient spending is approximately $900 (21 percent of total spending).

Conclusions

In contrast to analyses of inequality in the return to work, analyses of inequality of well-being more broadly defined should incorporate in-kind forms of compensation and government benefits. We have focused on how the inclusion of Medicaid, Medicare, and employer-provided health insurance influences measures of income inequality. While there is some debate about how to value Medicare and Medicaid benefits for the purpose of assessing how those programs influence inequality, our estimates and those in Burkhauser et al. (2013) indicate that measured inequality is about 25 to 30 percent smaller if the average cost of these programs are added to recipients' incomes. Incorporating employer-provided health insurance modestly increases measured inequality because the coverage rates and marginal tax rates rise with income. On net, however, including the value of both private and public

Figure 2
Total and Inpatient Spending on Medical Care in 2012 by Family Income



Source: Authors' calculations from the 2012 Medical Expenditure Panel Survey.
Notes: Adjusted family income is total family income divided by the square root of family size. Calculations include zeros for families with no medical care spending.

health insurance in measured income reduces inequality at a point in time and reduces the growth in inequality over the last 20 years. If employer-provided insurance were taxed as income, the 90–10 ratio would be about 4 percent smaller than it currently is. Though the effect of taxing employer-provided insurance on the 90–10 ratio may seem small compared to the effects of Medicare and Medicaid, it is important to bear in mind that the 90–10 ratio responds more to a fixed change in income among the poor than the rich.

Our discussion has highlighted some open research questions. First, assigning a value of health insurance is difficult. Most past work, including ours, has used a measure of average expenditures by firms and government. Some work has used individual expenditures (Burtless and Svaton 2010). It is arguably better to use individual expenditures, because expenditures are able to account for the heterogeneity of demands that characterize willingness to pay for health insurance. Recent work by Finklestein et al. (2015) suggests that these methods may overestimate the value of Medicaid to recipients. There is a similar debate in the literature on the trade-off between wages and employer-provided insurance and whether a dollar of insurance is worth a dollar of income. Despite these concerns over how to value health insurance, it seems fairly clear that our broad conclusions—that Medicare and Medicaid reduce inequality of well-being, and that incorporating both private and public insurance reduces inequality on net—are unlikely to be altered by better estimates of the value of insurance.

Second, inequality in income or access to resources is not the same thing as inequality in well-being. An analysis of inequality in well-being would include income, leisure and other activities not classified as work, and health, among other things. This point has been highlighted by the research of Meyer and Sullivan (2003, 2012), who focus on inequality in consumption. Consumption still does not include the value of time not working and health. There are a growing number of studies on the broader effects of public health insurance programs. Card, Dobkin, and Maestas (2009), for example, estimate that access to Medicare at age 65 is associated with a nearly one-percentage point decline in seven-day mortality among people admitted to an Emergency Department with "nondeferrable" medical conditions such as heart attacks. Finklestein et al. (2012) present evidence from Oregon that Medicaid lessens financial stress and improves mental health. Though this literature is small and not conclusive, these studies suggest that the effects of Medicare and Medicaid on inequality may be larger than that suggested simply by looking at their effects on income.

While government-subsidized health insurance significantly reduces income inequality and is an important source of well-being for the poor, the poor still are significantly disadvantaged in terms of health. Indeed, inequality in health between the top and bottom parts of the income distribution is large (National Academy of Sciences 2015). While there are many factors underlying the relationship between income and health, we close with a point made by Deaton (2002), among others, that policies that aim to improve individuals' earnings capacity will have the dual effect of both reducing income inequality and inequalities in health.

■ *We thank Gordon Hanson, Enrico Moretti, and Timothy Taylor for helpful comments. We thank Caitlyn Fleming for excellent research assistance. Any errors are our own.*

References

Aaron, Henry J., and Gary Burtless. 2014. "Potential Effects of the Affordable Care Act on Income Inequality." Research Paper, January 27, The Brookings Institution. http://www.brookings.edu/research/papers/2014/01/potential-effects-affordable-care-act-income-inequality-aaron-burtless.

Aguiar, Mark, and Erik Hurst. 2007. "Life-Cycle Prices and Production." *American Economic Review* 97(5): 1533–59.

Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein, for the Oregon Health Study Group. 2013. "The Oregon Experiment—Effects of Medicaid on Clinical

Outcomes." *New England Journal of Medicine* 368(18): 1713–22.

Blau, David M., and Donna B. Gilleskie. 2008. "The Role of Retiree Health Insurance in the Employment Behavior of Older Men." *International Economic Review* 49(2): 475–514.

Bureau of Labor Statistics. 2016. "Employer Costs for Employee Compensation—December 2015." BLS News Release, March 10. http://www.bls.gov/news.release/pdf/ecec.pdf.

Burkhauser, Richard V., Jeff Larrimore, and Kosali I. Simon. 2012. "A 'Second Opinion' on the Economic Health of the American Middle Class." *National Tax Journal* 65(1): 7–32.

Burkhauser, Richard V., Jeff Larrimore, and Kosali Simon. 2013. "Measuring the Impact of Valuing Health Insurance on Levels and Trends in Inequality and How the Affordable Care Act of 2010 Could Affect Them." *Contemporary Economic Policy* 31(4): 779–94.

Burkhauser, Richard V., and Koasali I. Simon. 2010. "Measuring the Impact of Health Insurance on Levels and Trends in Inequality." NBER Working Paper 15811.

Burtless, Gary, and Timothy M. Smeeding. 2001. "The Level, Trend, and Composition of Poverty." In *Understanding Poverty*, edited by Sheldon H. Danziger and Robert H. Haveman, 27–68. Cambridge, MA: Harvard University Press.

Burtless, Gary, and Pavel Svaton. 2010. "Health Care, Health Insurance, and the Distribution of American Incomes." *Forum for Health Economics & Policy* 13(1): 1–41.

Call, Kathleen T., Michael E. Davern, Jacob A. Klerman, and Victoria Lynch. 2013. "Comparing Errors in Medicaid Reporting across Surveys: Evidence to Date." *Health Services Research* 48(2, Part 1): 652–64.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics* 124(2): 597–636.

Centers for Medicare and Medicaid Services. 2014a. "U.S. Personal Health Care Spending by Age and Gender. 2014 Highlights." US Department of Health and Human Services. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html. Click on "Highlights."

Centers for Medicare and Medicaid Services. 2014b. "2014 CMS Statistics." US Department of Health and Human Services. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/Downloads/CMS_Stats_2014_final.pdf.

Centers for Medicare and Medicaid Services. 2015. "March 31, 2015 Effectuated Enrollment Snapshot." Fact Sheet, June 2. https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-06-02.html.

Chung, Wankyo. 2003. "Fringe Benefits and Inequality in the Labor Market." *Economic Inquiry* 41(3): 517–29.

Congressional Budget Office. 2011. "Trends in the Distribution of Household Income Between 1979 and 2007." Report, October.

Cooper, David, and Elise Gould. 2013. "Financial Security of Elderly Americans at Risk." *Economic Policy Institute,* Briefing Paper #362.

Currie Janet, and Brigitte C. Madrian. 1999. "Health, Health Insurance and the Labor Market." In *Handbook of Labor Economics* vol. 3C, pp. 3309–3416. Elsevier.

Cutler, David M., Angus Deaton, and Adriana Lleras-Muney. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives* 20(3): 97–120.

Cutler, David, and Jonathan Gruber. 1996. "Does Public Health Insurance Crowd-out Private Insurance?" *Quarterly Journal of Economics* 111(2): 391–430.

Deaton, Angus. 2002. "Policy Implications of the Gradient of Health and Wealth." *Health Affairs* 21(2): 13–30.

Deaton, Angus, and Christina Paxson. 2004. "Mortality, Income, and Income Inequality over Time in Britain and the United States." Chap. 6 in *Perspectives on the Economics of Aging*, edited by David Wise. National Bureau of Economic Research Conference Report. University of Chicago Press.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. 2016. "The Economic Consequences of Hospital Admissions." http://economics.mit.edu/files/10958.

Fisher, Elliott, S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Étoile L. Pinder. 2003. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138(4): 288–98.

Feenberg, Daniel Richard, and Elizabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12(1): 189–94.

Finkelstein, Amy, Nathaniel Hendren, and Erzo F. P. Luttmer. 2015. "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment." NBER Working Paper 21308.

Finkelstein, Amy, Sarah Taubman, Bill Wright. Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, and Katherine Baicker. 2012. "The Oregon Health Insurance

Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3): 1057–1106.

French, Eric, and John Bailey Jones. 2011. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." *Econometrica* 79(3): 693–732.

Garrett, Bowen, and Robert Kaestner. 2014. "The Best Evidence Suggests the Effects of the ACA on Employment Will Be Small." Brief, April. Urban Institute.

Garthwaite, Craig, Tal Gross, and Matthew Notodowidigdo. 2014. "Public Health Insurance, Labor Supply, and Employment Lock." *Quarterly Journal of Economics* 129(2): 653–96.

General Accounting Office. 1997. "Private Health Insurance: Continued Erosion of Coverage Linked to Cost Pressures." GAO/HEHS-97-122.

Haveman, Robert, Rebecca Blank, Robert Moffitt, Timothy Smeeding, and Geoffrey Wallace. 2015. "The War on Poverty: Measurement, Trends, and Policy." *Journal of Policy Analysis and Management* 34(3): 593–638.

Kaestner, Robert, Bowen Garrett, Anuj Gangopadhyaya, and Caitlyn Fleming. 2016. "Effects of Medicaid Expansions on Health Insurance Coverage and Labor Supply." NBER Working Paper 21836.

Kaiser Family Foundation. (Continually updated.) "Medicaid Spending Per Full-Benefit Enrollee," http://kff.org/medicaid/state-indicator/medicaid-spending-per-full-benefit-enrollee/.

Kaiser Family Foundation, and Health Research Educational Trust. 2013. *Employer Health Benefits: 2013 Annual Survey.*

Kuhn, Peter, and Fernando Lozano. 2008. "The Expanding Workweek? Understanding Trends in Long Work Hours among U.S. Men, 1979–2006." *Journal of Labor Economics* 26(2): 311–43.

Madrian, Brigitte C. 2005. "The U.S. Health Care System and Labor Markets." In *Wanting It All: The Challenge of Reforming the U.S. Health Care System,* 137–63. Research Conference Series no. 50, Federal Reserve Bank of Boston.

Mazumder, Bhaskkar, and Miller, Sarah. Forthcoming. "The Effects of Massachusetts Health Reform on Household Financial Distress." *American Economic Journal: Economic Policy.*

McClellan, Mark, and Jonathan Skinner. 2006. "The Incidence of Medicare." *Journal of Public Economics* 90(1–2): 257–76.

Meyer, Bruce D., and James X. Sullivan. 2003. "Measuring the Well-Being of the Poor Using Income and Consumption." *Journal of Human Resources* 38(Supplement): 1180–1220.

Meyer, Bruce, and James X. Sullivan. 2012. "Consumption and Income Poverty in the United States." In *The Oxford Handbook of the Economics of Poverty*, edited by Philip N. Jefferson, pp. 49–74. Oxford University Press.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4): 199–226.

Miller, George, Charles Roehrig, Paul Hughes-Cromwick, and Craig Lake. 2008. "Quantifying National Spending on Wellness and Prevention." *Advances in Health Economics and Health Services Research,* Vol. 19: *Beyond Health Insurance: Public Policy to Improve Health*, edited by Loren Helmchen, Robert Kaestner, and Anthony Lo Sasso, 1–24. JAI Press.

Mishel, Lawrence. 2013. "Vast Majority of Wage Earners Are Working Harder, And for Not Much More: Trends in U.S. Work Hours and Wages over 1979–2007." *Economic Policy Institute,* Issue Brief #348, January 30. http://www.epi.org/files/2013/ib348-trends-us-work-hours-wages-1979-2007.pdf.

Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *Journal of Political Economy* 114(5): 871–904.

Murphy, Kevin M., and Robert H. Topel. 2016. "Human Capital Investment, Inequality and Economic Growth." NBER Working Paper 21841.

National Academies of Sciences, Engineering, and Medicine, Committee on the Long-Run Macroeconomic Effects of the Aging U.S. Population–Phase II. 2015. *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses.* Washington, DC: National Academies Press.

Newhouse, Joseph P., and the Insurance Experiment Group. 1993. *Free for All? Lessons from the Rand Health Insurance Experiment.* Cambridge: Harvard University Press.

Pierce, Brooks. 2001. "Compensation Inequality." *Quarterly Journal of Economics* 116(4): 1493–1525.

Rosen, Sherwin. 1986. "The Theory of Equalizing Differences." *Handbook of Labor Economics,* vol. 1, edited by Orley C. Ashenfelter, and Richard Layard, 641–92. Elsevier.

US Census Bureau. No date. *Historical Income Tables: Income Inequality*. https://www.census.gov/hhes/www/income/data/historical/inequality/.

US Census Bureau. No date. "Calculating Fungible Values: Medicare, Medicaid." https://www.census.gov/cps/data/fungible.html.

US Office of Management and Budget. 2015. *Analytical Perspectives, Budget of the United States Government, Fiscal Year 2016.* Available at https://www.whitehouse.gov/omb/budget/Analytical_Perspectives.

Young, Katherine, Robin Rudowitz, Saman Rouhani, and Rachel Garfield. 2015. *Medicaid Per Enrollee Spending: Variation Across States.* Kaiser Family Foundation.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 79–102

Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing

Shelly Lundberg, Robert A. Pollak, and Jenna Stearns

In 1950, the family arrangements of college graduates and high school graduates were very similar. Men and women married early and most remained married. About 70 percent of 30–44 year-old female college graduates and 80 percent of female high school graduates were currently married in 1950. By 2010, women's marriage rates had fallen and the educational gradient had reversed: 69 percent of college graduate women were married, compared to 56 percent of those with a high school degree. Births to unmarried women were uncommon in 1950, but as marriage rates fell, nonmarital childbearing increased. In 1980, 5 percent of births to college graduates were to unmarried mothers, compared to 24 percent for high school graduates. By 2013, nonmarital childbearing among college graduates had risen to 11 percent, compared to 58 percent for high school graduates (Manning, Brown, and Stykes 2015). Not surprisingly, the divergence in the family arrangements of female college graduates and high school graduates is paralleled by a similar divergence in those of men. In 1950, about 85 percent of 30–44 year-old men

■ *Shelly Lundberg is the Leonard Broom Professor of Demography, University of California, Santa Barbara, Santa Barbara, California. She is also a Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany, and Adjunct Professor of Economics, University of Bergen, Bergen, Norway. Robert A. Pollak is the Hernreich Distinguished Professor of Economics, Washington University in St. Louis, St. Louis, Missouri. He is also Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany. Jenna Stearns is a PhD Candidate in Economics, University of California, Santa Barbara, Santa Barbara, California. Their emails are slundberg@ucsb.edu, pollak@wustl.edu, and stearns@umail.ucsb.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at

http://dx.doi.org/10.1257/jep.30.2.79 doi=10.1257/jep.30.2.79

were currently married at all levels of education. In 2010 only 70 percent of male college graduates and 53 percent of high school graduates were married.

Popular discussions of changes in American families over the past 60 years have revolved around the "retreat from marriage." Concern has focused on increasing levels of nonmarital childbearing, as well as falling marriage rates that stem from both increases in the age at first marriage and greater marital instability. Often lost in these discussions is the fact that the decline of marriage has coincided with a rise in cohabitation. Many "single" Americans now live with a domestic partner and a substantial fraction of "single" mothers are cohabiting, often with the child's father. The share of women who have ever cohabited has nearly doubled over the past 25 years, and the majority of nonmarital births now occur to cohabiting rather than to unpartnered mothers at all levels of education. The emergence of cohabitation as an alternative to marriage has been a key feature of the post–World War II transformation of the American family.

These changes in the patterns and trajectories of family structure have a strong socioeconomic gradient. The important divide is between college graduates and others: individuals who have attended college but do not have a four-year degree have family patterns and trajectories that are very similar to those of high school graduates. Compared with college graduates, less-educated women are more likely to enter into cohabiting partnerships early and bear children while cohabiting, are less likely to transition quickly into marriage, and have much higher divorce rates. For this group, rising rates of cohabitation and nonmarital childbearing contribute to family histories of relatively unstable relationships and frequent changes in family structure (Cherlin 2009).

We begin with a brief review of the basic facts about changes in family structure over recent decades and then explore two broad sets of explanations for the emergence of the socioeconomic gradient in marriage, divorce, cohabitation, and childbearing. The first emphasizes the diminished economic prospects of less-educated men. Rising relative wages of women have reduced the returns to specialization and exchange within marriage at all levels of education, but sociologists have focused on a shortage of "marriageable" men at the bottom of the earnings distribution as a primary cause of rising family inequality. It is unlikely, however, that men in the middle of the earnings distribution cannot contribute enough to the household to generate a positive marital surplus. For the "marriageable men" theory to explain declining marriage rates more broadly, traditional gender norms that dictate the husband should be the primary breadwinner are required. The reduced marital surplus resulting from violating these gender norms may cause some middle-earning men to become "unmarriageable." If these norms are stronger or more prevalent among those with less education, then they can, together with rising relative wages of women, cause a socioeconomic gradient in marriage.

The second set of explanations for the socioeconomic gradient emphasizes educational differences in demands for marital commitment. When marriage was based on traditionally specialized gender roles, marriage and the commitment it implies protected the interests of wives who stayed home, reared children, and failed to accumulate market-relevant human capital. As technological changes in the home and workplace reduced the gains from specialization, the value of commitment decreased. Cohabitation, with lower exit costs than marriage, allows individuals to realize many of the gains from co-residence with less commitment. We argue that college-graduate parents continue to use marriage as a commitment device to facilitate intensive joint investments in their children. For less-educated, lower-income couples for whom such investments are less desirable or less feasible, commitment and hence marriage has less value relative to cohabitation.

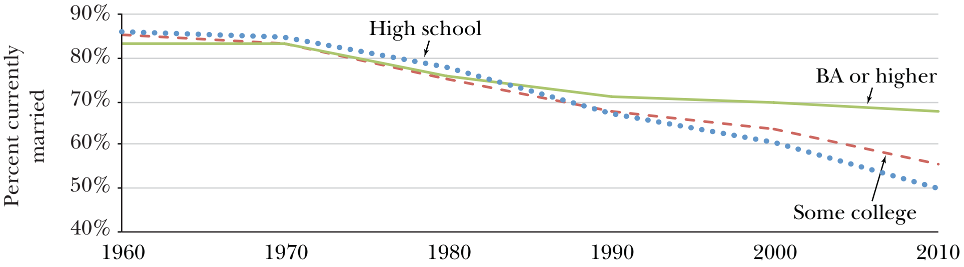
These changes in the demand for long-term commitment, and the resulting socioeconomic divergence in family structure, have important implications for children and parents. Cohabiting relationships are much less stable than marriages, so the increase in childbearing within cohabiting unions among the less-educated means that their children are more likely to experience instability in living arrangements, household income, and parental presence. This instability has been linked to adverse child outcomes, though the magnitude of the causal link is uncertain. Compared with women two generations earlier, women with low levels of education today find themselves with greater independence and control over their lives, but also at an increased risk of poverty. Less-educated men find themselves both unburdened and unmoored by weakened responsibilities of marriage and fatherhood. The new socioeconomic gradient in family structure appears to be a "mechanism" in the reproduction of inequality across generations, being both influenced by rising inequality and a potential contributor to future inequality (McLanahan and Percheski 2008).

The Uneven Retreat from Marriage

The general contours of the post–World War II transformation of American family life are well-known. The age at which men and women first marry reached historic lows during the 1950s. Between 1956 and 2013, the median age at first marriage rose by over six years for both men and women, increasing from 21.1 to 27.5 years for women and from 22.5 to 29.2 years for men.[1] Societal anxiety focused not on delay in the age of first marriage, but instead on two other changes that became apparent in the 1970s: rising rates of nonmarital childbearing and an abrupt increase in the divorce rate. The proportion of births to unmarried women rose from 5 percent in 1960 to 32 percent in 1995, and has remained at about 40 percent in recent years (Child Trends Data Bank 2015). The prevalence of divorce, which had been rising gradually in the United States since the late nineteenth century, suddenly doubled between the mid-1960s and mid-1970s (in this journal, Stevenson and Wolfers 2007).

[1] Although same-sex marriage has become more prevalent in recent years and is now legal in all 50 states, in this paper we focus on heterosexual marriages.

Figure 1
Percent of Population Aged 33–44 Currently Married, 1960–2010



Sources: 1960–2000 US Census; 2010 American Community Survey.

These changes in marriage, divorce, and nonmarital childbearing have differed by socioeconomic status (Lundberg and Pollak 2014, 2015). While the fraction of Americans currently married has declined substantially since 1960 at all levels of education, the decline is especially pronounced among the less-educated.[2] Figure 1 shows the changing share of individuals aged 30–44 currently married, by educational attainment.[3] Though differences in marriage rates by education were small in 1960, by 2010 marriage rates among college graduates were 12 and 17 percentage points higher than marriage rates for those with some college and high school graduates, respectively.

Although about 90 percent of men and women eventually marry, and the share of men and women who have ever married by middle age is similar across education groups,[4] the marriages of college graduates are much more stable. As shown in Figure 2, the trends in the share of the population aged 30–44 who are currently divorced are almost identical for the some college and high school groups, but roughly 40 percent lower for college graduates. Some of the education gap is explained by differences in age at first marriage, but the probability of divorce at

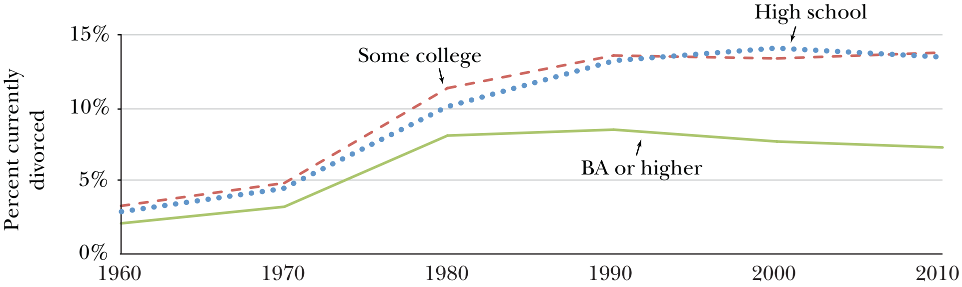
[2] This paper focuses only on individuals with at least a high school degree or equivalent. We exclude the less than high school group for two reasons. First, the composition of this group has changed substantially over time as low-skill immigration has increased. In recent decades, those without a high school degree are disproportionately Hispanic, immigrants, and noncitizens. In 1960, the share of immigrants was roughly constant across education groups. In 2010, however, over half of all 30–44 year-olds with a less than high school education were immigrants, while only about 20 percent of those with more education were immigrants. Second, the less than high school group now comprises a relatively small share of the population. In 2010, there were 15–20 million 30–44 year-olds in each of the high school graduate, some college, and college graduate groups. In contrast, only about 6 million 30–44 year-olds had less than a high school education.

[3] We use this age category because marriage rates for individuals under age 30 are strongly influenced by changes and educational differences in age at first marriage.

[4] Black men and women with a high school education or less provide an exception: they are substantially less likely to ever marry than black men and women with more education (Isen and Stevenson 2011).

Figure 2

Percent of Population Aged 33–44 Currently Divorced, 1960–2010



Sources: 1960–2000 US Census; 2010 American Community Survey.

given marriage durations is also substantially lower for college graduates than for those with some college or a high school degree.[5]

The first panel in Table 1 shows the cumulative effect of these differences on the marital histories of the late baby-boomers, using data from the National Longitudinal Survey of Youth 1979. By age 46, nearly half of the high school and some college groups who ever married have been divorced, but nearly 70 percent of the college graduates are still in their first marriage.

Focusing on these trends in the formation and dissolution of marriages ignores another important change: the rise in cohabitation. Cohabitation has become a very common domestic arrangement in the United States. The share of women who have ever cohabited has nearly doubled over the past 25 years, and today the majority of women aged 19 to 44 have been in a cohabiting relationship at some point in their lives (Manning 2013). Over 27 percent of all couples currently living together are in nonmarital unions (based on our calculations from the 2007–2013 Current Population Survey data).

Tracking changes in cohabitation over time is difficult because high-quality, population-representative data on unmarried couples is available only for recent cohorts.[6] Most estimates of cohabitation for earlier cohorts are based on inferences

[5] The probability of divorce within 20 years of marriage is 15 and 7 percentage points lower for college graduates than for those with some college or high school degrees, respectively. For white men, the probability of divorce is 19 percentage points lower for college graduates than for both the some college and high school group (Isen and Stevenson 2011).

[6] Direct measures of cohabitation are available in the 1990, 2000, and 2010 Censuses, but only if the relationship involves the head of household. The Current Population Survey from 1995–2006 and the American Community Survey also identify only cohabiting unions involving the head of household, not of other couples in the household. Kennedy and Fitch (2012) find this method misses 18 percent of cohabiting unions, so these surveys substantially underestimate cohabitation rates. After 2006, the Current Population Survey identifies all cohabiting unions. Direct questions about unmarried partners have recently been added to the Survey of Income and Program Participation and American Community Survey, as well as to several longitudinal data sources.

Table 1
Family Outcomes by Education

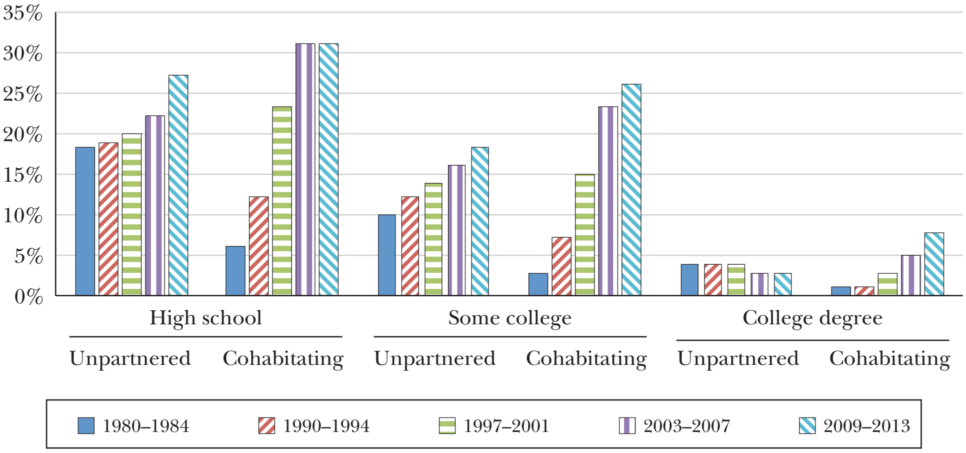
	High school graduate, no college	*Some college or associate's degree*	*College degree or higher*
National Longitudinal Survey of Youth, 1979: Marriage outcomes by age 46, birth cohorts 1957–1964 (Aughinbaugh et al. 2013)			
Percent ever married	87.0	87.1	89.0
Among those who married:			
Percent ever divorced	49.1	48.5	29.8
Percent still in first marriage	48.6	48.9	69.0
National Longitudinal Study of Adolescent to Adult Health: Family structure by age 28–32, birth cohorts 1976–1984 (authors' tabulation)			
Percent currently married	45.0	45.8	48.2
Percent currently cohabiting	21.5	19.1	14.2
Percent 2+ co-residential unions	42.1	39.5	19.3
Percent unmarried mother	32.2	26.7	8.4
Percent unpartnered mother	17.8	16.4	5.8

Sources: **For panel 1:** Aughinbaugh, Robles, and Sun (2013). **For panel 2:** Authors' tabulation from National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

from household composition and are unreliable. However, improved estimates from the Census in Fitch, Goeken, and Ruggles (2005) indicate that there was very little (reported) cohabitation prior to 1970. Data on cohabitation may be inherently flawed because it is a state that is difficult to define. Couples often enter joint living arrangements gradually (often part-time and while maintaining separate residences) and without clear expectations (Manning and Smock 2005). Sociologists and demographers have studied the causes and implications of rising cohabitation rates since the early 1980s, while economists have generally ignored cohabitation and continued to focus on the dichotomy of married versus unmarried.

Much of the retreat from marriage appears to have been a shift into cohabitation because the age at which young couples establish their first household has remained relatively constant. For cohorts born in the 1960s and 1970s, the average age at first union (married or cohabiting) stabilized at the pre-Baby Boom level of around 22.5 for women (Bailey, Guldi, and Hershbein 2014). The share of births to unmarried mothers has doubled since 1980, but most of this increase has come

Figure 3
The Share of Births to Unpartnered and Cohabiting Mothers under Age 40 by Educational Attainment in Different Periods



Source: Manning, Brown, and Stykes (2015).
Notes: For each educational group, the bars show the share of births to unpartnered and to cohabiting mothers under age 40, with the remaining share being to married mothers within that educational group. For instance, for high school women in 2009–2013, 27 percent of births are to unpartnered mothers, 31 percent are to cohabiting mothers, and the rest are to married mothers.

from a tripling in the share of births to mothers who are cohabiting rather than unpartnered (Manning, Brown, and Stykes 2015).

The strong education gradient apparent in marriage and divorce also holds for cohabitation and nonmarital childbearing. The second panel of Table 1 shows that, for a recent cohort of young adults, the marriage, cohabitation, and childbearing patterns of individuals with just a high school degree are very similar to those with some college, but starkly different from college graduates. The less-educated are less likely to be partnered, a higher fraction of their partnerships are nonmarital, and their unions are much less stable. A young woman without a college degree is approximately five times more likely than a college graduate to be a cohabiting or an unpartnered mother. Although nonmarital childbearing has increased substantially across the whole education spectrum since 1980, the rates among college graduates have remained relatively low, as shown in Figure 3. In contrast, the share of nonmarital births to both high school graduates and women with some college have increased sharply since 1980, with most of this increase driven by the higher incidence of births within cohabiting unions.

The divide in nonmarital childbearing between college graduates and those with some college within each race and ethnic group is large while the overall rates differ substantially. For example, for non-Hispanic white college graduates, the rate of nonmarital childbearing is 5.9 percent while the rate for those with some college is 30.0 percent. For Hispanics, the corresponding rates are 17.4 percent and 45.3 percent; for blacks, 32.0 percent and 68.7 percent (Lundberg and Pollak 2014). Thus, the differences by education are not simply reflections of racial and ethnic differences in educational attainment.

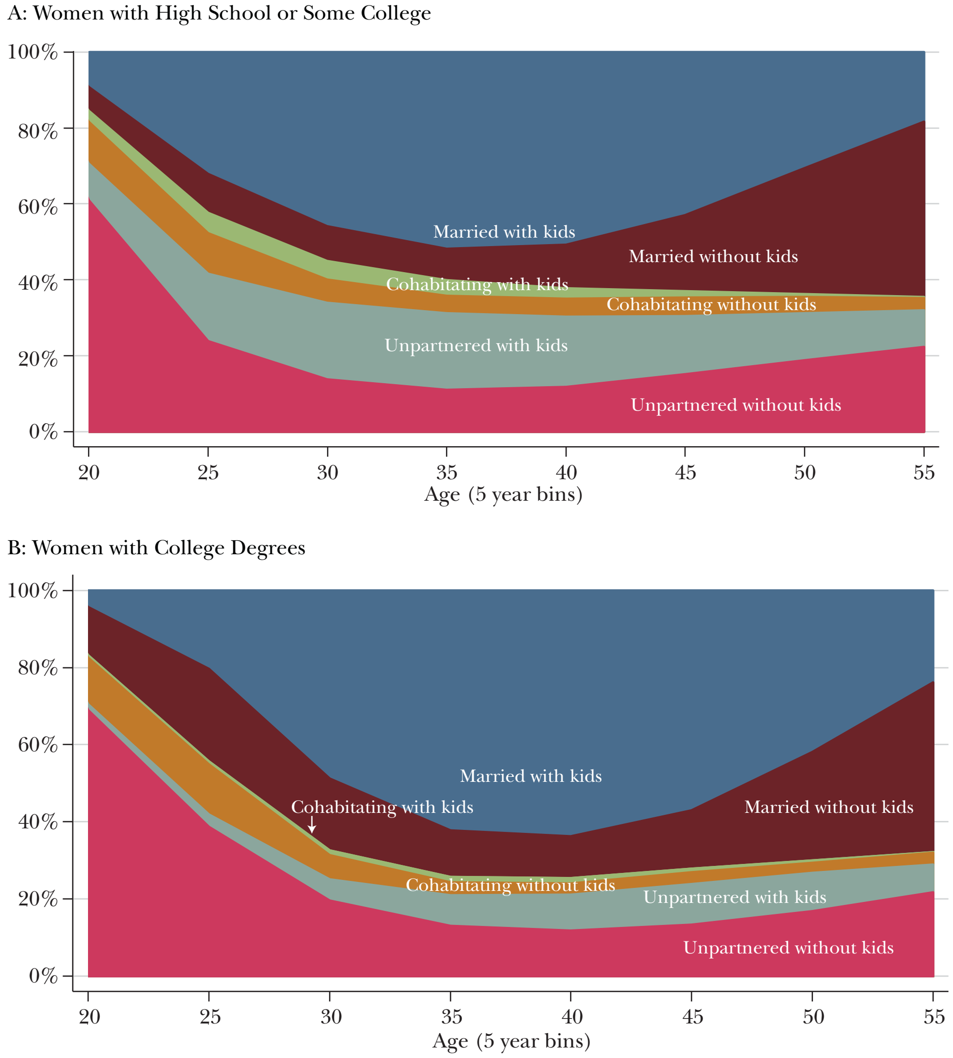
Figure 4A and B present a life-cycle perspective on how women's marriage, cohabitation, and childbearing patterns differ by educational attainment. Women with college degrees are substantially more likely to be in a union after their mid-20s than are women with lower levels of education. Conditional on being in a union—whether marital or cohabiting—college graduates are also more likely to be married than cohabiting. The differences in union status by educational attainment are even larger among women with children in the household. Only 2.4 percent of college graduate women under age 40 with children are cohabiting, compared with 8 percent of less-educated women. There is also a distinct educational gradient in the living arrangements of unmarried mothers. Unmarried mothers with college degrees are much more likely to be cohabiting rather than living alone or with relatives, compared with those with less education (Manning, Brown, and Stykes 2015).

Cohabitating unions tend to be much less stable than marriages for all education groups. The median duration of cohabitations is somewhat longer for the less-educated (22–24 months) than for college graduates (17 months), but is extremely short compared to marriage. The first premarital cohabitation spell is equally likely to dissolve within three years for all education groups, but college graduates are significantly more likely to transition into marriage and less likely to remain cohabiting for more than three years (Copen, Daniels, and Mosher 2013).

The role of cohabitation, as well as its prevalence, differs across education groups. For women who are college graduates, childbearing during cohabitation is relatively rare, and when it does occur, cohabiting unions are likely to transition quickly into marriage. Among those with less education, however, the rise of cohabitation has delayed marriage but not childbearing. The probability of a pregnancy within one year of beginning a first premarital cohabitation is 5 percent for college graduate women, 18 percent for women with some college, and 24 percent for high school graduates. Women who are college graduates and become pregnant while cohabiting are twice as likely to marry within a year as those with some college (Copen, Daniels, and Mosher 2013). In sum, traditional patterns of marital childbearing have been much more persistent among highly-educated Americans, while the decoupling of marriage and childbearing is much more prevalent among those without college degrees. For college graduates, increased cohabitation is part of a pattern of delayed marriage and childbearing to accommodate an extended period of education, facilitated by improved birth control and changes in social norms concerning premarital sex (Goldin and Katz 2002). For others, cohabitation appears to be a more direct substitute for marriage.

These differences in the role of cohabitation have important implications for the living arrangements of children whose mothers have different levels of education. Because cohabitation tends to be transitory regardless of whether there are children, children in cohabiting households are at greater risk of instability in

Figure 4
Household Type by Age of Woman and Educational Status



Source: 2007–2013 Current Population Survey.
Note: For a given age, the vertical height of each category represents the share of women in that household type.

living arrangements, parent figures, and household income. Subsequent marriages or cohabitations are often with a partner other than the father of the children, and complex families with multiple-partner fertility are common among those who bear children while cohabiting. Less than half of children in cohabiting households are living with both biological parents, compared with over 90 percent of those in married couple households (Payne 2013). Finally, fathers with some college or only a high school degree are 25 percent less likely to live with all their children than are college-graduate fathers (Guzzo and Payne 2014). These differences in family stability and paternal presence are associated with important differences in investments in children and child outcomes.

Changing Gender Roles, Marital Surplus, and Investments in Children

The economic model of marriage developed by Gary Becker in the 1970s reflected and rationalized the dominant family paradigm of the post–World War II era. Marriage was, for most, a lifetime contract between a man and a woman in which he provided income from market work and she contributed home-based cooking, cleaning, and childcare. Divorce was costly and infrequent, and both "living in sin" and the production of "illegitimate" children were stigmatized. The expected gains from marriage stemmed largely from household production—economies of scale and the returns to specialization and exchange. Subsequent economic models recognized gains from joint consumption of household public goods such as housing and children (Lam 1988). In all of these models, individuals considering marriage are assumed to compare the expected utility of this particular marriage (which depends, in turn, on expectations about the division of the surplus from marriage between the spouses) relative to the expected utility of remaining unmarried and, perhaps, continuing to search for a better spouse. Cohabitation, in these early models, is not explicitly considered as an alternative to marriage or remaining single.

Becker (1981) used the specialization-and-exchange model to explain the declining prevalence and stability of marriage in the later 1960s and into the 1970s. The proximate cause was the fall in marital surplus associated with reduced specialization, while the underlying cause was the change in the economic opportunities of women. Changes in production technology and the structure of demand increased the productivity of female workers more than male workers, increasing women's relative earnings and employment opportunities (Galor and Weil 1996). The declining gender wage gap reduced the potential gains from a sexual division of labor in the household and, as women's long-term attachment to the labor force strengthened, investments in education reinforced these changes. The opportunity costs of rearing children increased as female wages rose and the likelihood of divorce increased. In response to these changing incentives, fertility fell by half from 1960 to 1980, further reducing the returns to a couple from having one person stay home. The past 60 years have witnessed a substantial convergence in the economic lives of married men and women, and specialization in hours of market and household work has decreased dramatically (Aguiar and Hurst 2007; Lundberg and Pollak 2007 in this journal).

Cohabitation versus Marriage

The emergence of cohabitation as a widely acceptable alternative to marriage, which was in its early stages when Becker published *A Treatise on the Family* in 1981, changes the calculus of the marriage decision. Many of the gains to marriage recognized in economic models can be realized by any couple who agree to coordinate production and share consumption within a joint household. What, then, distinguishes marriage from cohabitation in an economically meaningful way?

Economic models of marriage and cohabitation have emphasized one key difference: the costs of dissolution are much higher for marriage than for cohabitation (Brien, Lillard, and Stern 2006; Matouschek and Rasul 2008). Ending a marriage involves legal formalities to divide property and debt and, if there are children, to establish custody, visitation, and support arrangements. Divorce became less costly as states adopted no-fault or unilateral divorce laws starting in the 1970s, but divorce remains a complicated, uncertain, and often expensive process in both time and money. Unlike marriages, cohabiting unions can be ended simply and quickly outside of the legal system. Finally, the cultural significance of marriage makes divorce more socially (and possibly psychologically) traumatic to individuals. Based on their ethnographic work, Edin and Kefalas (2005) conclude that fear of divorce is an important reason the unmarried mothers they study prefer cohabitation to marriage.

When a marriage dissolves, marital property is divided between ex-spouses; when a cohabiting union dissolves, there is no analogue of marital property—assets and liabilities remain with the ex-partner who holds legal title to the asset or is legally responsible for the debt. In several states, couples that enter civil unions or domestic partnerships receive some of the benefits of marriage, and most states recognize explicit contracts between cohabitants. But few cohabiting couples make written contracts, the terms and even the existence of oral contracts are often difficult to prove, and court rulings about the enforceability of such contracts are inconsistent (Bowman 2004, 2010). Common-law marriage, which requires that couples hold themselves out as married, has all but disappeared with the increasing social acceptability of cohabitation and has been abolished by statute in most states (Waggoner 2015). On the other hand, the laws governing child custody and child support have changed substantially over the last few decades, lessening the distinction between marriage and cohabitation in terms of parental rights and obligations. The distinction between legitimate and illegitimate children has virtually disappeared, so that if paternity has been established, at least in theory, child custody issues following the dissolution of a cohabiting union or a marriage are not very different.[7]

The higher cost of dissolving a marriage, relative to cohabitation, affects both the selection of couples into marriage and the level of couple-specific investments within the marriage. In traditional marriages, investments in skills that are specific to the domestic sphere, and thus to some extent marriage-specific, can generate a family version of the hold-up problem. The traditional gender division of labor that limits the market experience and skills of women requires the expectation of a

[7] One remaining difference is paternity establishment: when a married woman gives birth, the law presumes that her husband is the father of her child. Edlund (2013) emphasizes the role of paternity presumption and its implications for custody.

lifetime commitment because marital dissolution will impose heavy costs on women who have been domestic specialists. A marital regime that imposes high exit costs—legal, social, and economic—allows marriage to function as a commitment device that fosters cooperation between partners and encourages marriage-specific investments. The "divorce revolution"—the shift to no-fault or unilateral divorce—which decreased marriage exit costs and reduced the value of marriage as a commitment device appears to have played a part in reducing joint household investments (Stevenson 2007).[8]

For modern, less-specialized couples, many of the gains from marriage or cohabitation are likely to be based on shared consumption of household public goods and the pleasures of shared leisure rather than on a division of labor between household production and market work (Stevenson and Wolfers 2007). These consumption-based benefits require less couple-specific investment and therefore demand less intertemporal commitment. Cohabitation facilitates joint consumption in a lower-commitment partnership, and thus provides an attractive alternative to marriage in a society without distinct male and female spheres. Couples will sort between marriage and cohabitation depending on their demand for commitment. Not surprisingly, cohabiting partnerships tend to be substantially less specialized than marital partnerships (Gemici and Laufer 2014; Parker and Wang 2013).

Declining marital surplus has been a proximate cause of the retreat from marriage. However, the underlying forces that led to a reduced demand for long-term commitment—decreased gender specialization and a shift from production-based to consumption-based marital surplus—appear to apply to all couples regardless of education. What remains to be explained is why we have seen a large increase in nonmarital childbearing and in marital instability among low- and medium-education groups while traditional patterns of post-marital childbearing and relatively stable marriages have persisted among college graduates.

Rising Inequality, Marriageable Men, and Gender Norms

Sociologists, demographers, and family historians link the socioeconomic divergence in marriage and divorce to increasing economic inequality over the past few decades and, in particular, to the deteriorating employment and earnings prospects of less-educated men. In this view, the maintenance of the traditional family, with childbearing and childrearing within stable marriages, depends on the earnings capacity of the male partner.

An extensive literature has documented a strong empirical relationship between men's long-term economic prospects and career maturity and their transitions into marriage (Oppenheimer, Kalmijn, and Lim 1997). "Marriageable" men are those who have demonstrated their ability to be good (enough) providers for

[8] Matouschek and Rasul (2008) show that couples who married after the passage of unilateral divorce laws were positively selected and less likely to divorce. This is consistent with a model in which the principal role of the marriage contract is to act as a commitment device.

a family. The idea of marriageable men has deep historical and cultural roots. Delayed marriage was a hallmark of the "European Marriage Pattern" before the Industrial Revolution (Hajnal 1965; Wrigley 2014). Marriage required young men to be economically independent, and so couples waited to marry until the man inherited the family farm rather than forming a multigeneration household with his or her parents. Marriage ages fell in Europe as well-paying industrial jobs for young men became more prevalent (Fitch and Ruggles 2000). In the United States, age at first marriage fell to historically low levels during the optimistic and prosperous post–World War II era. Drawing upon this historical record, Ruggles (2015) attributes recent changes in family structure to the deteriorating economic prospects of men. Female wages have been rising relative to male wages at all education levels over the last few decades, but the decline in the gender earnings gap at lower levels of education is largely due to the decline in the real earnings of noncollege men (Autor and Wasserman 2013).

An economic version of the marriageable men hypothesis can explain the retreat from marriage among the severely disadvantaged. Ethnographic work in severely disadvantaged communities suggests than some men's economic prospects are so dire, due to a combination of low skills, labor market discrimination, criminal records, and substance abuse, that they are unable to make a positive contribution to a household (Edin and Nelson 2013). But a purely economic version of the marriageable men hypothesis cannot explain the falling marriage rate among men and women with some college. To explain the broad retreat from marriage in terms of the shortage of marriageable men requires a powerful role for norms defining gender roles.

As the prevalence of couples in which the wife earns more than the husband increased, studies of the relationships between relative earnings, relationship stability, and household behavior proliferated in sociology (Brines 1994; Atkinson, Greenstein, and Lang 2005; Cooke 2006).[9] A common theme in this literature is that marriages in which a wife earns more than her husband violate a norm that the husband should be the primary breadwinner. The evidence for this conclusion includes a higher probability of divorce and a higher prevalence of domestic violence in such households. Bertrand, Kamenica, and Pan (2015) invoke the stress of breaking with "gender identity" norms to motivate the apparent effects of relative spousal earnings on marriage prevalence, women's labor supply, and relationship stability. A reduction in the value of marriage when the wife earns more than the husband—as a result of violating these gender identity norms—may be more pronounced for lower-education households because traditional gender norms tend to be strongly decreasing with education (Davis and Greenstein 2009).

Becker's specialization-and-exchange theory of marriage also suggests that couples have most to gain from marriage and marital specialization when the gender wage gap is large. Using the ratio of female/male mean full-time earnings as

[9] In 2013, 38 percent of wives with positive earnings earned more than their husbands (US Bureau of Labor Statistics 2013).

Table 2
Mean Annual Wage Earnings of Full-time Workers by Education
(*2010 dollars*)

	1960	*1970*	*1980*	*1990*	*2000*	*2010*
Full-time male workers						
High school	43,333	54,129	52,010	46,223	45,950	40,967
Some college	49,382	62,332	55,842	54,579	56,039	50,501
College or more	60,094	80,490	72,553	81,366	92,226	89,187
Full-time female workers						
High school	23,653	28,598	28,983	30,059	31,755	30,288
Some college	26,078	32,762	32,853	36,398	39,160	37,413
College or more	33,898	44,169	41,389	50,973	59,133	60,902
Ratio of mean female/male earnings						
High school	0.546	0.528	0.557	0.650	0.691	0.739
Some college	0.528	0.526	0.588	0.667	0.699	0.741
College or more	0.564	0.549	0.570	0.626	0.641	0.683

Sources: Authors' calculations based on the 1960–2000 Census and 2010 American Community Survey. The sample is restricted to full-time workers (who usually work at least 35 hours per week in 1980–2010 or worked at least 35 hours last week in 1960–1970) ages 25–54. Earnings are measured by annual wage and salary income converted to 2010 dollars. Although income in the Census data is top-coded, the majority of workers whose income exceeds the top code are college educated. This issue only affects a small share of workers, but if anything, the difference between the college-educated and some college groups is slightly understated as a result.

an imperfect measure of the gender wage gap, it appears that the potential surplus from specialization within marriage may have declined less in recent decades for college graduates than for those with less education. Table 2 shows that the ratio of female to male real earnings for full-time workers has risen since 1960 for all education groups. In 1960, the gender earnings ratio was similar across education levels: 56 percent for workers with a college degree or higher, 53 percent for those with some college, and 54 percent for those with just a high school degree. After 1980, however, as the real earnings of less-educated men began to fall, these earnings ratios diverged. In 2010 the ratio of female/male earnings was 68 percent for those with a college degree, compared with 74 percent for both those with some college and those with only a high school degree. There is little evidence, however, that college graduate couples are more specialized than less-educated couples. Among married couple households in which both partners have the same level of education, there is no clear difference in the ratio of usual hours of market work between husbands and wives in the high school, some college, and college graduate groups (based on our calculations from the 1980–2000 Census and the 2010 American Community Survey).

An alternative hypothesis to explain the retreat from marriage as the gender wage gap fell is that contracting problems prevent couples from realizing potential marital surpluses: that is, problems in renegotiating the allocation of marital surplus may dissuade couples from marrying or lead them to divorce when they are unable to respond effectively to shocks. For contracting problems to explain the socioeconomic gradient in the retreat from marriage, these barriers must be more severe for less-educated couples than for college graduates. Contracting problems can reflect difficulties in negotiating mutually acceptable divisions of marital surplus with imperfect information (Peters 1986) or in making binding agreements to implement those divisions (Lundberg and Pollak 2003). For example, it may be difficult for couples to make credible commitments to share childcare and other household work when the wage gap shrinks in the face of peer pressures that support more traditional gender roles (Sevilla-Sanz 2005). Also, all marriages will face stochastic shocks to income, health, or affections, and reallocation within the marriage may require relationship skills that may be positively associated with education.

The Healthy Marriage Initiative, a set of federal marriage-promotion programs, was initially funded in 2003 based on the belief that low-income couples lack the relationship skills required to overcome the challenges they face as they deal with parenthood and economic hardship. Randomized treatment evaluations of these programs found them to be ineffective (Wood, McConnell, Moore, Clarkwest, and Hsueh 2012), which could mean either that the skills gap, if it exists, is not of central importance or that it is not much affected by the specific policy intervention. The former explanation is supported by Lundberg (2015), who found that although there are pronounced educational differences in measures of traits such as self-efficacy and emotional stability, these differences fail to explain any significant fraction of the education differences in relationship instability and lone motherhood among young Americans.

Diverging Investments in Children

Marital surplus from specialization and exchange has declined for all education groups as the gender wage gap has decreased, making it unlikely that this is the primary source of the socioeconomic divergence in marriage behavior that we have documented. Other sources of gains to marriage, including the returns to joint consumption of public goods and investments in children, may also play important roles. If the returns to these consumption-based sources of marital surplus are increasing in income (Becker 1974), then the marriage rates of college graduates may be stabilized by the "income effect" of their rising wages offsetting the "price effect" of the narrowing gender wage gap (Moffitt 2000).

There are good reasons to think that children are key to the socioeconomic differences in marriage behavior. The most important difference between the family histories of college graduates and others is not whether they marry—after all, the vast majority of men and women at all levels of education eventually marry—but rather the timing and duration of marriage and its relation to childbearing and childrearing. Less-educated mothers are substantially more likely than college graduates to give birth while in cohabiting relationships and, given the short average duration of these relationships, are more likely to rear their children alone or with a subsequent partner.

An alternative explanation for the uneven retreat from marriage that offers a better rationale for the decoupling of marriage and childbearing by parents who are not college graduates focuses on differing strategies for investments in children. We suggest that, for college graduates, marriage has become a commitment device that supports intensive joint investments in children. Marriage, because it is more costly to exit than cohabitation, can act as a commitment device for the cooperative joint project of raising economically successful children (Lundberg and Pollak 2014, 2015). Increased returns to human capital and, hence, to intense child investments, may have kept marital surplus high for college graduates, who are more likely to make these investments. Because long-term commitment facilitates this joint investment, college graduates marry late and delay having children until marriage.

Intensive investments in children, signaled by higher childcare time and by growing expenditures on children, are concentrated among college graduates. As with marital and childbearing patterns, in terms of investment patterns, Americans with some college look more similar to high school graduates than to four-year college graduates. Mothers with some college who have children under age 13 spend 30 minutes less per day in primary childcare than mothers with college degrees, and there is no difference in primary childcare time between the some college and high school groups (based on our calculations from the American Time Use Survey, 2003–2014). High- and low-educated parents may also make different types of investments in their children. Ethnographic evidence indicates that the parental aspirations and goals of poor and working class parents tend to be focused on safety and survival, rather than achievement (Lareau 2003; Edin and Kefalas 2005). Because the ethnographic literature has focused on poor and working class families, the extent to which these parental aspirations extend to the some college group is an open question.

Why might the incentives to invest in children have diverged across education groups? Rising returns to human capital have been a hallmark of the recent increases in income inequality, but an upward shift in the returns to human capital should increase investment by all parents. Indeed, parents in all education groups have increased time with children. Parents differ, however, in their resources and their capabilities. Parental academic skills will increase the productivity of their time with children. College graduate parents also appear to possess better information about how children learn and to engage with them in more developmentally appropriate ways (Kalil, Ryan, and Corey 2012). While the effect of parental productivity on time allocated to child investments is theoretically indeterminate, abundant empirical evidence indicates that childcare time increases with education (in this journal, see Guryan, Hurst, and Kearney 2008). These advantages will be reinforced by dynamic complementarities in the production of children's skills (Heckman 2000; Todd and Wolpin 2007; Aizer and Cunha 2012). If "skill begets skill," then later parental investments and formal schooling will be more productive for children who have early cognitive and health advantages. This implies that the expected returns to child investments by parents with limited resources and uncertain futures may be lower than for more educated parents with greater and more secure investment capabilities.

Implications of Growing Family Inequality for Children, Women, and Men

The relative instability of marriage and the prevalence of nonmarital childbearing among those with less education compared with college graduates have implications for the well-being of men and women, and for the transmission of resources and capabilities across generations. Because the diverging patterns of partnering and parenting across education groups reflect changes in the incentives to invest in children and in relationship-specific capital, it would be inappropriate to treat these outcomes as though they were the effects of family change itself rather than of the underlying economic and social forces that have transformed American families.

Children: Diverging Destinies

The children of women without college degrees are substantially more likely than the children of college graduates to be born to an unpartnered or cohabiting mother, to experience a change (or multiple changes) in the presence of a father or father figure in the household, and to grow up in a complex family with step- and half-siblings. The net result will be a childhood with, on average, greater instability and more limited father involvement than the children of college graduates. These trends have contributed to what McLanahan (2004) calls the "diverging destinies" of children in advantaged and less-advantaged families, with those at the top benefiting from access to the time and money of two highly educated parents while many at the bottom do not.

The enormous literature on the association between family structure and outcomes for children documents strong and consistent correlations between child outcomes such as educational attainment, crime, and mental health, and family structure indicators such as years with an unpartnered parent (McLanahan and Sandefur 1994) and family transitions (Fomby and Cherlin 2007). Parental cohabitation (as opposed to marriage) is also strongly associated with adverse outcomes for children and adolescents (Brown 2004). All such studies are of course plagued by selectivity, since unobserved parental characteristics are likely to be powerful determinants of family structure, family transitions, and child outcomes. Not surprisingly, adding controls for observed parental characteristics reduces the association between marriage and children's outcomes (Ribar 2004). Alternative identification strategies, such as sibling fixed effects and instrumental variables (for example, using parental death as an instrument for parental absence), generally show smaller but still significant effects of family structure and family transitions on child outcomes (McLanahan, Tach, and Schneider 2013). None of these studies, however, completely escape challenges to identification.

The instability of family structure also complicates estimating effects on child outcomes. Although it is convenient (or often necessary, in the absence of lifetime longitudinal data) to focus on family arrangements at a single point in the life-cycle or over a short duration—for example, whether the parents are married when the child is born or whether the child lives with both biological parents at age 15—this focus misses the "turbulence" that Cherlin (2009) identifies as a key feature of American families. While popular discussions often focus on "single parent families," such families are typically in transition between one marriage or cohabitation and another: only a small fraction of children spend their entire childhoods in single parent families (Björklund, Ginther, and Sundström 2007). This instability implies the need for a longitudinal rather than a cross-section perspective and emphasis on family structure trajectories rather than family structure measured at a point in time. Analyses of the "window problem" in studies of child outcomes have shown that single year and short duration window variables measuring childhood circumstances, including family structure and transitions, are poor proxies for childhood experience (Wolfe, Haveman, Ginther, and An 1996).

Given the identification challenges, the size and nature of any causal effects of family structure or family transitions on child outcomes remain very uncertain. It is difficult, if not impossible, to distinguish the effects of parental cohabitation on children from the high rates of parent figure transitions with which it is associated, or the unobserved characteristics of parents who have chosen not to marry. Also, the evidence does not unanimously favor the two-parent family. For example, using an estimation strategy that includes child fixed-effects, Aughinbaugh, Pierret, and Rothstein (2005) do not find significant effects of mothers' marital transitions on children's cognitive and socio-emotional development, and Brown (2006) finds that transitions from a lone-mother family into a cohabiting stepfamily are associated with negative effects on adolescent well-being. Ginther and Pollak (2004) and Gennetian (2005) find that educational outcomes for both stepchildren and biological children in blended families are similar to outcomes in lone-parent families.

Women: Independent and At Risk

Increased family instability has increased the burden of childrearing for women without college degrees relative to women with college degrees. Poverty rates for women with high school diplomas and those with some college are much higher than the poverty rates of college graduates, and some of this difference is due to the greater likelihood that less-educated women are unpartnered and rearing children. Unsurprisingly, poverty rates are substantially higher for unmarried women with children at all levels of education than for married women with children.[10] The vast majority of children living with one parent (87 percent) reside with the mother (Payne 2013).

On the other hand, as cohabitation, nonmarital childrearing, and divorce become more acceptable, women have increased freedom to reject marriages to

[10] The poverty rate is very low (1.9 percent) for college-educated women who are married with children, and 4.1 and 9.4 percentage points higher for married mothers with some college or high school degrees, respectively. Poverty rates are much higher among unmarried mothers with children: 10.8 percent of those with college degrees live below the poverty line compared to 24.6 percent of unmarried mothers with some college and 30.5 percent of those with high school degrees. These statistics are calculated by the authors from the American Community Survey 2012 five-year sample.

men with whom they have cohabited or who have fathered their children, and to exit relationships that are unrewarding or dangerous. One effect of the divorce revolution, which reduced the cost of exiting marriage, was a significant decrease in female suicide and domestic violence (Stevenson and Wolfers 2006). Although on average unmarried women are less economically well-off than married women, an important positive consequence of the retreat from marriage may be a reduction in the prevalence of relationships that are unsatisfying or harmful.

Men: Unburdened and Unmoored

There are large differences between the behavior of married men, cohabiting men, and unpartnered men, whether measured cross-sectionally or longitudinally. Transitions into both marriage and cohabitation are associated with decreases in men's risky behavior, such as binge drinking and drug use, but the decreases associated with marriage are larger and more consistent than those associated with cohabitation (Duncan, Wilkerson, and England 2006). After they marry, men work more hours and earn higher wages. Akerlof (1998) concludes that the impact of marriage is causal and that delayed marriage, the demise of the "shotgun marriage" when an unexpected pregnancy occurs, and men's reduced responsibility for, and co-residence with, children are responsible for a rise in social pathology. He argues that the transition into marriage is a rite of passage associated with a change in responsibilities that alters men's preferences, resulting in an increase in time spent in home-oriented activities. An alternative causal explanation for an abrupt change in men's behavior upon marriage is that it is part of the marital contract with their wives. If social and economic changes have reduced the value of marriage to noncollege graduates, these changes may also be responsible for a further causal, and generally deleterious, effect on men's behavior.

Finally, there is a great deal of concern among demographers and gerontologists about the fate of elderly men without wives or doting children. Data on intergenerational transfers support the hypothesis that aging fathers who did not consistently co-reside with their children as they grew up receive less support from their adult children. Fathers who never married or are divorced from their children's mothers are less likely to receive time and money transfers from children, but the same is not true for never-married or divorced mothers (Pezzin, Pollak, and Schone 2008; Astone, Peters, and Gelatt 2015; Wiemers, Seltzer, Schoeni, Hotz, and Bianchi 2015). An increasing concentration of isolated elderly men among those with low lifetime income presents challenges for social welfare policy in an aging society.

Conclusion

American family arrangements have become more diverse and more transitory in the past 60 years. Some changes have occurred broadly across the entire population, while others have a distinct socioeconomic gradient. As age at first marriage has risen, premarital cohabitation has become a common experience for men and women at all levels of education. Divorce rates remain much higher for all groups than they were before the divorce revolution as well. In other dimensions, however, college graduates have retained more traditional patterns of marriage and parenting than have men and women with less education. Childbearing in cohabiting unions has risen much more dramatically among high school graduates and those with some college, and their marital and cohabiting unions are less stable. This means that children of less-educated parents are more likely to grow up without both biological parents in the household and to experience instability in family structure. Increasing inequality in the stability of family arrangements has paralleled rising inequality in wages and earnings, and has contributed to inequality in household income.

To what extent is emerging family inequality a consequence of the well-documented increase in wage and earnings inequality? The declining gender wage gap has reduced marital surplus from specialization and exchange for individuals at all levels of education. This gap has decreased more for the high school and some college groups, in part because of the decline or stagnation in the real earnings of less-educated men, though there is little evidence that marital specialization is decreasing in education. If, in addition, less-educated individuals are more likely to face contracting problems or rigid gender norms that restrict men to the role of primary breadwinner, then the fall in the gender wage gap may explain part of the uneven retreat from marriage. However, this explanation does not account for differences in the timing of marriage in relation to childbearing across education groups.

We propose a new explanation, one that attributes the socioeconomic gradient in the timing of marriage and childbearing to diverging incentives to make intensive investments in children. If there are dynamic complementarities between early and later investments in children, high-resource men and women may respond to rising returns to human capital by using marriage as a commitment device that supports childrearing as a joint investment project. The uncertain economic prospects of the less-educated may discourage them from doing so.

Does growing family inequality in this generation contribute to economic inequality in the next? Credible estimates of the causal impacts of family structure patterns and trajectories on child outcomes still elude researchers, though most of the literature supports a negative relationship between family instability and child well-being. There is considerable evidence, however, that the divergence in child investments between high- and low-resource families is likely to exacerbate future inequality.

■ *We thank Janice Compton, Gordon Hanson, Stephanie Heger, Enrico Moretti, Dick Startz, Timothy Taylor, and Anne Winkler for helpful suggestions and insightful comments.*

References

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.

Aizer, Anna, and Flávio Cunha. 2012. "The Production of Human Capital: Endowments, Investments and Fertility." NBER Working Paper 18429.

Akerlof, George A. 1998. "Men without Children." *Economic Journal* 108(447): 287–309.

Astone, Nan M., H. Elizabeth Peters, and Julia Gelatt. 2015. "Family Structure and Intergenerational Transfers." Presented at the 2015 Population Association of America 2015 Annual Meetings.

Atkinson, Maxine P., Theodore N. Greenstein, and Molly Monahan Lang. 2005. "For Women, Breadwinning Can Be Dangerous: Gendered Resource Theory and Wife Abuse." *Journal of Marriage and Family* 67(5): 1137–48.

Aughinbaugh, Alison, Charles R. Pierret, and Donna S. Rothstein. 2005. "The Impact of Family Structure Transitions on Youth Achievement: Evidence from the Children of the NLSY79." *Demography* 42(3): 447–68.

Aughinbaugh, Alison, Omar Robles, and Hugette Sun. 2013. "Marriage and Divorce: Patterns by Gender, Race, and Educational Attainment." *Monthly Labor Review* 136(1). http://www.bls.gov/opub/mlr/2013/article/marriage-and-divorce-patterns-by-gender-race-and-educational-attainment.htm.

Autor, David, and Melanie Wasserman. 2013. "Wayward Sons: The Emerging Gender Gap in Labor Markets and Education." *Third Way Report*, March 20.

Bailey, Martha J., Melanie E. Guldi, and Brad J. Hershbein. 2014. "Is There a Case for a 'Second Demographic Transition'? Three Distinctive Features of the Post-1960 US Fertility Decline." In *Human Capital in History: The American Record*, edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 273–312. University of Chicago Press.

Becker, Gary S. 1974. "A Theory of Marriage." In *Economics of the Family: Marriage, Children, and Human Capital*, edited by Theodore W. Shultz, 299–351. University of Chicago Press.

Becker, Gary S. 1981. *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130(2): 571–614.

Björklund, Anders, Donna K. Ginther, and Marianne Sundström. 2007. "Family Structure and Child Outcomes in the USA and Sweden." *Journal of Population Economics* 20(1): 183–201.

Bowman, Cynthia Grant. 2004. "Legal Treatment of Cohabitation in the United States." *Law & Policy* 26(1): 119–51.

Bowman, Cynthia Grant. 2010. *Unmarried Couples, Law, and Public Policy.* Oxford University Press.

Brien, Michael J., Lee A. Lillard, and Steven Stern. 2006. "Cohabitation, Marriage, and Divorce in a Model of Match Quality." *International Economic Review* 47(2): 451–94.

Brines, Julie. 1994. "Economic Dependency, Gender, and the Division of Labor at Home." *American Journal of Sociology* 100(3): 652–88.

Brown, Susan L. 2004. "Family Structure and Child Well-being: The Significance of Parental Cohabitation." *Journal of Marriage and Family* 66(2): 351–67.

Brown, Susan L. 2006. "Family Structure Transitions and Adolescent Well-Being." *Demography* 43(3): 447–61.

Cherlin, Andrew J. 2009. *The Marriage-Go-Round: The State of Marriage and the Family in America Today.* New York: Alfred A. Knopf.

Child Trends Data Bank. 2015. *Births to Unmarried Women.* March 2015. Retrieved from http://www.childtrends.org/wp-content/uploads/2015/03/75_Births_to_Unmarried_Women.pdf on 7/9/2015.

Cooke, Lynn Prince. 2006. "'Doing' Gender in Context: Household Bargaining and Risk of Divorce in Germany and the United States." *American Journal of Sociology* 112(2): 442–72.

Copen, Casey E., Kimberly Daniels, and William D. Mosher. 2013. "First Premarital Cohabitation in the United States: 2006–2010 National Survey of Family Growth." *National Health Statistics Reports* 64(64): 1–15.

Davis, Shannon N., and Theodore N. Greenstein. 2009. "Gender Ideology: Components, Predictors, and Consequences." *Annual Review of Sociology* 35: 87–105.

Duncan, Greg J., Bessie Wilkerson, and Paula England. 2006. "Cleaning Up Their Act: The Effects of Marriage and Cohabitation on Licit and Illicit Drug Use." *Demography* 43(4): 691–710.

Edin, Kathryn, and Maria Kefalas. 2005. *Promises I Can Keep: Why Poor Women Put Motherhood Before Marriage.* Berkeley, CA: University of California Press.

Edin, Kathryn, and Timothy J. Nelson. 2013. *Doing the Best I Can: Fatherhood in the Inner City.* Berkeley, CA: University of California Press.

Edlund, Lena. 2013. "The Role of Paternity Presumption and Custodial Rights for Understanding Marriage Patterns." *Economica* 80(320): 650–69.

Fitch, Catherine, Ron Goeken, and Steven Ruggles. 2005. "The Rise of Cohabitation in the United States: New Historical Estimates." Minnesota Population Center, Working Paper 3.

Fitch, Catherine A., and Steven Ruggles. 2000. "Historical Trends in Marriage Formation: The United States 1850–1990." In *The Ties That Bind: Perspectives on Marriage and Cohabitation,* edited by Linda J. Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson, and Arland Thornton, 59–88. New York: Aldine de Gruyter.

Fomby, Paula, and Andrew J. Cherlin. 2007. "Family Instability and Child Well-Being." *American Sociological Review* 72(2): 181–204.

Galor, Oded, and David N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3): 374–87.

Gemici, Ahu, and Steven Laufer. 2014. "Marriage and Cohabitation." April. https://goo.gl/C6UYI4.

Gennetian, Lisa A. 2005. "One or Two Parents? Half or Step Siblings? The Effect of Family Structure on Young Children's Achievement." *Journal of Population Economics* 18(3): 415–36.

Ginther, Donna K., and Robert A. Pollak. 2004. "Family Structure and Children's Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions." *Demography* 41(4): 671–96.

Goldin, Claudia, and Lawrence F. Katz. 2002. "The Power of the Pill: Contraceptives and Women's Career and Marriage Decision." *Journal of Political Economy* 110(4): 730–70.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* 22(3): 23–46.

Guzzo, Karen Benjamin, and Krista K. Payne. 2014. "Living Arrangements of Fathers and Their Children." Family Profiles FP-14-06, National Center for Family & Marriage Research.

Hajnal, John. 1965. "European Marriage Patterns in Perspective." In *Population in History: Essays in Historical Demography,* edited by D. V. Glass and D. E. C. Eversley, 101–43. London: Arnold.

Heckman, James J. 2000. "Policies to Foster Human Capital." *Research in Economics* 54(1): 3–56.

Isen, Adam, and Betsy Stevenson. 2011. "Women's Education and Family Behavior: Trends in Marriage, Divorce, and Fertility." In *Demography and the Economy,* edited by John B. Shoven, 107–40. University of Chicago Press.

Kalil, Ariel, Rebecca Ryan, and Michael Corey. 2012. "Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children." *Demography* 49(4): 1361–83.

Kennedy, Sheela, and Catherine A. Fitch. 2012. "Measuring Cohabitation and Family Structure in the United States: Assessing the Impact of New Data from the Current Population Survey." *Demography* 49(4): 1479–98.

Lam, David. 1988. "Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications." *Journal of Human Resources* 23(4): 462–87.

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life.* Berkeley, CA: University of California Press.

Lundberg, Shelly. 2015. "Skill Disparities and Unequal Family Outcomes." *Research in Labor Economics,* Vol. 41*: Gender Convergence in the Labor Market,* edited by Solomon Polochek, Konstantinos Tatsiramos, and Klaus F. Zimmermann, 177–212.

Lundberg, Shelly, and Robert A. Pollak. 2003. "Efficiency in Marriage." *Review of Economics of the Household* 1(3): 153–67.

Lundberg, Shelly, and Robert A. Pollak. 2007. "The American Family and Family Economics." *Journal of Economic Perspectives* 21(2): 3–26.

Lundberg, Shelly, and Robert A. Pollak. 2014. "The Uneven Retreat from Marriage in the U.S., 1950–2010." In *Human Capital and History: The American Record,* edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 241–72. University of Chicago Press.

Lundberg, Shelly, and Robert A. Pollak. 2015. "The Evolving Role of Marriage: 1950–2010." *Future of Children* 25(2).

Manning, Wendy D. 2013. "Trends in Cohabitation: Over Twenty Years of Change, 1987–2010." Family Profiles FP-13-12. National Center for Family & Marriage Research.

Manning, Wendy D., Susan L. Brown, and Bart Stykes. 2015. "Trends in Births to Single and Cohabiting Mothers, 1980–2013." Family Profiles FP-15-03, National Center for Family & Marriage Research.

Manning, Wendy D., and Pamela J. Smock. 2005. "Measuring and Modeling Cohabitation: New Perspectives from Qualitative Data." *Journal of Marriage and Family* 67(4): 989–1002.

Matouschek, Niko, and Imran Rasul. 2008. "The Economics of the Marriage Contract: Theories and Evidence." *Journal of Law and Economics* 51(1): 59–110.

McLanahan, Sara. 2004. "Diverging Destinies: How Children Are Faring under the Second Demographic Transition." *Demography* 41(4): 607–27.

McLanahan, Sara, and Christine Percheski. 2008. "Family Structure and the Reproduction of Inequalities." *Annual Review of Sociology* 34: 257–76.

McLanahan, Sara, and Gary Sandefur. 1994. *Growing Up with a Single Parent: What Hurts, What Helps.* Cambridge, MA: Harvard University Press.

McLanahan, Sara, Laura Tach, and Daniel Schneider. 2013. "The Causal Effects of Father Absence." *Annual Review of Sociology* 39: 399–427.

Moffitt, Robert A. 2000. "Female Wages, Male Wages, and the Economic Model of Marriage: The Basic Evidence." In *The Ties That Bind: Perspectives on Marriage and Cohabitation,* edited by Linda J. Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson, and Arland Thornton, 320–42. New York: Aldine de Gruyter.

Oppenheimer, Valerie Kincade, Matthijs Kalmijn, and Nelson Lim. 1997. "Men's Career Development and Marriage Timing during a Period of Rising Inequality." *Demography* 34(3): 311–30.

Parker, Kim, and Wendy Wang. 2013. "Modern Parenthood: Roles of Moms and Dads Converge as They Balance Work and Family." Report, March 14. Pew Research Center.

Payne, Krista K. 2013. "Children's Family Structure, 2013." Family Profiles FP-13-19, National Center for Family & Marriage Research.

Peters, H. Elizabeth. 1986. "Marriage and Divorce: Informational Constraints and Private Contracting." *American Economic Review* 76(3): 437–54.

Pezzin, Liliana E., Robert A. Pollak, and Barbara Steinberg Schone. 2008. "Parental Marital Disruption, Family Type, and Transfers to Disabled Elderly Parents." *Journal of Gerontology: Social Sciences* 63(6): S349–S358.

Ribar, David C. 2004. "What Do Social Scientists Know about the Benefits of Marriage? A Review of Quantitative Methodologies." IZA Discussion Paper 998.

Sevilla-Sanz, Almudena. 2005. "Social Effects, Household Time Allocation, and the Decline in Union Formation." Working Paper 20015-07, Congressional Budget Office, Washington DC.

Steven Ruggles. 2015. "Patriarchy, Power, and Pay: The Transformation of American Families, 1800–2015." *Demography* 52(6): 1797–1823.

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. 2006. "Bargaining in the Shadow of the Law: Divorce Laws and Family Distress." *Quarterly Journal of Economics* 121(1): 267–88.

Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and their Driving Forces." *Journal of Economic Perspectives* 21(2): 27–52.

Todd, Petra E., and Kenneth I. Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1(1): 91–136.

US Bureau of Labor Statistics. 2013. "Wives Who Earn More than Their Husbands, 1987–2013." Labor Force Statistics from the Current Population Survey. Retrieved from http://www.bls.gov/cps/wives_earn_more.htm on 8/21/2015.

Waggoner, Lawrence W. 2015. "With Marriage on the Decline and Cohabitation on the Rise, What about Marital Rights for Unmarried Partners?" University of Michigan Public Law Research Paper 477.

Wiemers, Emily, Judith A. Seltzer, Robert Schoeni, V. Joseph Hotz, and Suzanne M. Bianchi. 2015. "The Generational Structures of U.S. Families and Their Intergenerational Transfers." Presented at the Population Association of America 2015 Annual Meeting, San Diego, CA.

Wolfe, Barbara, Robert Haveman, Donna Ginther, and Chong Bum An. 1996. "The 'Window Problem' in Studies of Children's Attainments: A Methodological Exploration." *Journal of the American Statistical Association* 91(435): 970–82.

Wood, Robert G., Sheena McConnell, Quinn Moore, Andrew Clarkwest, and JoAnn Hsueh. 2012. "The Effects of Building Strong Families: A Healthy Marriage and Relationship Skills Education Program for Unmarried Parents." *Journal of Policy Analysis and Management* 31(2): 228–52.

Wrigley, E. A. 2014. "European Marriage Patterns and Their Implications: John Hajnal's Essay and Historical Demography during the Last Half-Century." In *Population, Welfare and Economic Change in Britain 1290–1834,* edited by Chris Briggs, P. M. Kitson, and S. J. Thompson, 15–41. Woodbridge: The Boydell Press.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 103–126

Crime, the Criminal Justice System, and Socioeconomic Inequality

Magnus Lofstrom and Steven Raphael

After peaking in the early 1990s, official measures of violent and property crime rates have dropped to levels not seen since the 1960s. Proportional declines in the most serious offenses have been particularly pronounced. For example, murders/manslaughter per 100,000 declined by more than half, from 9.8 in 1991 to 4.5 in 2014, the lowest recorded murder rate since 1960. The violent crime rate overall fell by approximately half over this period, while overall property crime rates fell by nearly 50 percent. Juxtaposed against this declining crime rate has been an enormous and unprecedented expansion in US correctional populations. Between 1980 and 2013, the prison incarceration rate increased nearly 3.5 times, the jail incarceration rate increased by nearly three times, while the community correction supervision rate (numbers of people on probation or parole per 100,000) increased by 2.6 times. By 2013, roughly 3 percent of the adult population in the United States was under some form of criminal justice supervision. During this time period, the United States transitioned from a nation with an incarceration rate slightly higher than that of western European nations to the nation with the highest incarceration rate in the world.

These two coinciding trends present a provocative contrast, illustrating the conflicting manner in which changes in crime and punishment over the past few decades have impacted socioeconomic inequality in the United States. As we will

■ *Magnus Lofstrom is a Senior Fellow, Public Policy Institute of California, San Francisco, California. Steven Raphael is Professor of Public Policy, Goldman School of Public Policy, University of California, Berkeley, California. Their email addresses are lofstrom@ppic.org and stevenraphael@berkeley.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.103 doi=10.1257/jep.30.2.103

show, crime rates declined the most in poorer and more minority cities, and within cities in the poorest neighborhoods. In other words, the benefits of the crime decline have been progressively distributed. By contrast, the social costs created by the unprecedented expansion in correctional populations have been regressively distributed, with poor, disproportionately minority males (African-American males in particular) being most directly impacted and poor minority families (again African-American families in particular) disproportionately bearing the collateral social costs of the stiffening of US sentencing policy.

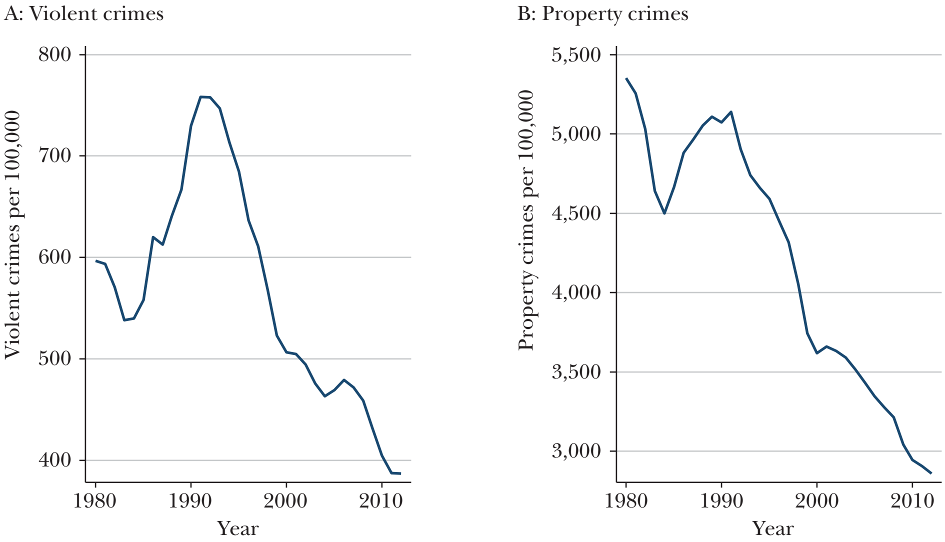
There is an ongoing debate on the extent to which the rise in incarceration and the extended reach of the criminal justice system drove recent declines in crime. There is fairly strong evidence for the United States and other nations that incarceration can have sizeable effects in reducing crime, operating largely through physical incapacitation. These effects, however, diminish with scale. Expanding the use of incarceration along the intensive margin of longer sentences results in the incarceration of individuals into advanced ages when the propensity to offend declines, while expanding along the extensive margin will lead to the incarceration of individuals who are less criminally active. There is less evidence that the more extensive use of probation, and the increased propensity of courts to levy fines and fees on those convicted of serious and less-serious criminal offenses, have contributed to crime declines.

In what follows, we document that poor and minority communities have disproportionately experienced both the decline in crime and the increase in criminal justice sanctioning. We argue that the coincidence of these two trends do not necessarily mean that one has caused the other. In particular, the crime decline commencing in the early 1990s is observed in other countries that have not greatly expanded the scope and reach of their criminal justice systems. Moreover, while increases in incarceration during the 1980s likely suppressed peak crime rates in the early 1990s, the decline in crime since that time corresponds to a period of rapid growth in incarceration levels for which there is little evidence of an appreciable impact on crime. Finally, the recent experiences of several states with reducing incarceration suggest that the contribution of higher incarceration rates to crime abatement is limited at current levels. The experience of California where the state was forced by a federal court to reduce its incarceration rate to 1990 levels is particularly instructive. In the conclusion, we argue that public policy can and should pursue ongoing reductions in crime and in the inequality of crime victimization, while simultaneously seeking to reduce the inequality of criminal justice sanctioning.

Inequality in Criminal Victimization

There are two principal sources of crime data in the United States. First, the Uniform Crime Reports (UCR) provide counts of crimes known to the police by month and crime type. Second, the National Crime Victimization Survey (NCVS) provides crime rate estimates based on an annual household survey conducted by

Figure 1
Trends in US Property Crime and Violent Crime Rates, 1980 through 2012
(per 100,000)



Source: Authors using data from the Uniform Crime Reports.
Note:s The figures present rates of property crime and violent crime, both expressed as the number of incidents per 100,000 people. Property crime rates are the number of burglaries, motor vehicle thefts, and larceny thefts per 100,000 residents. The violent crime rate is the sum of murders, rapes, robberies, and aggravated assaults per 100,000 residents.

the US Census Bureau. Crime rates tabulated from the NCVS tend to be higher than those tabulated from the UCR due to underreporting of crimes to the police. In addition, there can be notable differences in trend estimates from these two data sources. For example, during the 1970s and 1980s, the NCVS shows overall decreases in crime while crime rates as measured by the UCR increase substantially.[1] In recent years, trends in these two data sources tend to align. Both suggest an increase in serious violent crime during the 1980s. As we will soon see, both show very pronounced declines in crime and victimization since the early 1990s. The simultaneous analysis of these two data sources permits a more complete picture of how crime/victimization risk varies with socioeconomic characteristics.

Figure 1 presents rates of property crime and violent crime, both expressed as the number of incidents per 100,000 people, from the Uniform Crime Reports for the period 1980 through 2013. Property crime rates are the number of burglaries, motor vehicle thefts, and larceny thefts per 100,000 residents. The violent crime

[1] As participation of the nation's thousands of police departments in the collection of crime statistics improved during the 20th century and as victim reporting increases, aggregate crime rates rose (Boggess and Bound 1997).

rate is the sum of murders, rapes, robberies, and aggravated assaults per 100,000 residents. These seven felony crimes (often referred to as the FBI's "part 1" felony offenses) provide the standard categorization of serious offenses in the United States. Both series show peaks in 1991. While there are some doubts about whether the UCR-measured increase in property crime during the late 1980s captures an actual rise in crime or trends in crime reporting by both victims and agencies participating in the UCR (Boggess and Bound 1997), the increase in serious violent crime—and homicide in particular—beginning around 1985 is a well-documented fact (Blumstein and Rosenfeld 1998). Figure 1 also reveals very large declines in crime rates from 1991 on. Violent crime rates dropped by more than half over this period, while property crime rates fell by nearly half. As we will see shortly, victimization statistics reveal very similar overall patterns.

One cannot use the historical data from the Uniform Crime Reports to explore in a direct way how crime rates as experienced by different socioeconomic groups have changed over time. For the most part, the data are summarized at the law enforcement agency level with little micro-level information on specific criminal incidents.[2] However, police agencies tend to correspond geographically with incorporated cities, and cities vary considerably with respect to average socioeconomic and demographic characteristics. Hence, one can assess the incidence of the crime decline by examining the relationship between changes in crime rates across cities with different socioeconomic characteristics.

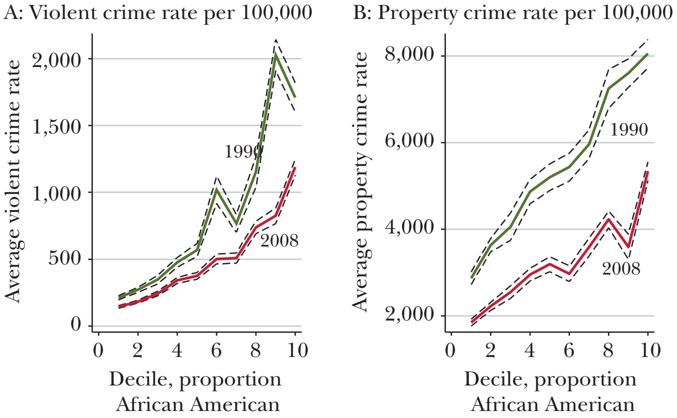
Figures 2 and 3 present such an analysis for the years 1990 and 2008 based on the data set produced by Kneebone and Raphael (2011). The data describe crime rates and demographic characteristics for the roughly 5,400 cities located within the nation's 100 largest US metropolitan areas. Figure 2 groups the cities covered by these data into deciles by the proportion of city residents that are African-American in 2000, and displays the average violent and property crime rates for each group for 1990 and 2008 (decile breaks are tabulated weighting by total 2000 city population). Figure 3 presents comparable results when cities are stratified by deciles according to the proportion of city residents that are below the poverty line (again weighted by the 2000 total population).

Property crime rates and violent crime rates are notably higher in cities where a higher proportion of residents are African-American in both years. In addition, crime rates are declining across all deciles. However, the figure reveals larger absolute drops in cities with proportionally larger African-American population. While the ratio of crime rates in decile 10 to crime rates in decile 1 actually increase slightly between 1990 and 2008, the absolute differences in crime rates narrow considerably. For example, the average violent crime rate in tenth-decile cities in 1990 exceeded

[2] The Uniform Crime Reports is slowly shifting towards the National Incident Based Reporting System (NIBRS), which includes detailed micro-level information on specific criminal incidents. As of 2012, agencies covering roughly 30 percent of the US population report criminal incidents through the NIBRS. See 2012 "NIBRS Participation by State," Federal Bureau of Investigation, https://www.fbi.gov/about-us/cjis/ucr/nibrs/2012/resources/nibrs-participation-by-state (a pdf file), accessed on September 18, 2015.

Figure 2

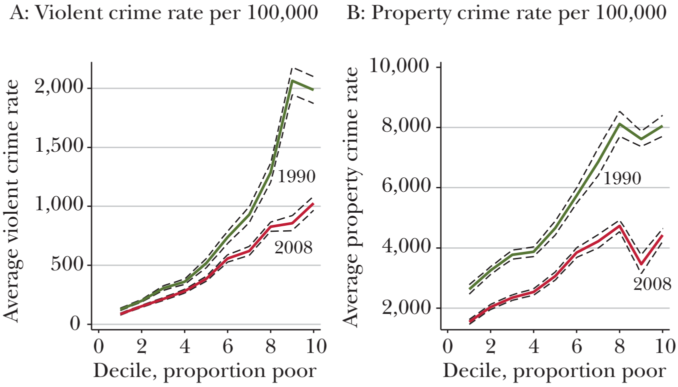
Relationship between Violent and Property Crime and Deciles of the Distribution of the Proportion of City Residents African-American in 2000



Source: Authors using data from Uniform Crime Reports, and US Census Bureau's decennial census and American Community Surveys. The data covers roughly 5,400 cities located within the nation's 100 largest US metropolitan areas.
Notes: Cities covered by these data are grouped into deciles by the proportion of city residents that are African-American in 2000, and the figure displays the average violent and property crime rates for each group for 1990 and 2008 (decile breaks are tabulated weighting by total 2000 city population). The dotted lines above and below the solid lines represent the 95 percent confidence interval.

Figure 3

Relationship between Violent and Property Crime and Deciles of the Distribution of the 2000 Proportion of City Residents in Poverty



Source: See Figure 2.
Notes: Cities covered by these data are grouped into deciles according to the proportion of city residents that are below the poverty line in 2000, and the figure displays the average violent and property crime rates for each group for 1990 and 2008. (Decile breaks are tabulated weighting by total 2000 city population). The dotted lines above and below the solid lines represent the 95 percent confidence interval.

that of first-decile cities by 1,498 per 100,000. By 2008, this difference shrinks to 1,045 per 100,000. The comparable differences for property crime rates are 5,179 per 100,000 in 1990 and 3,495 per 100,000 in 2008.

Figure 3 reveals comparable changes in the relationship between city-level poverty rates and city-level crime rates. Again, in both years, crime rates are appreciably higher in cities in which a higher share of the population was below the poverty line, and the drop in crime over time is apparent. However, the inequality between cities with the highest and lower poverty rates narrows considerably over this 18-year period. Here we observe a narrowing of both the ratio of crime rates as well as the absolute difference. Expressed as a ratio, the 1990 violent crime rate among the cities in the top poverty decile was 15.8 times the rate for the cities in the lowest poverty decile. By 2008, the ratio falls to 11.9. When expressed in levels, in 1990 the violent crime rate in the cities in the upper decile for poverty rates exceeds the violent crime rate in cities in the lowest decile for poverty rates by 1,860 incidents per 100,000. By 2008, the absolute difference in violent crime rates shrinks to 941 per 100,000. We see comparable narrowing in the differences between poorer and less-poor cities in property crime rates.

Within cities, crime tends to be geographically concentrated in poorer neighborhoods with proportionally larger minority populations. In recent years, many police departments have begun to make public geo-coded incident-level data permitting analysis of more granular geographic inequality in crime rates. However, official crimes and clearances collected under the Uniform Crime Reports are only summarized at the agency level. We are not aware of a national data source that provides geographically disaggregated crime data for the time period corresponding to the great crime decline.[3] Cohen and Gorr (2006), however, have assembled data on crime counts by census tract for the cities of Pittsburgh, Pennsylvania, and Rochester, New York, covering the period 1990 through 2001. Here we employ the data for Pittsburgh to assess whether the cross-city difference in the crime declines observed in Figures 2 and 3 are also evident at a more geographically disaggregated level. Over this time period, total violent crimes reported to the Pittsburgh police decline by 34 percent while total property crimes reported fall by 44 percent. Over this time period, Pittsburgh's population declines by roughly 10 percent. Hence, the period in time corresponds to an appreciable decline in crime rates.

Table 1 summarizes our analysis. In panel 1A, we split census tracts in Pittsburgh into quintiles according to the proportion of the census tract African-American in 2000. For each census tract, we use the data from Cohen and Gorr (2006) to measure the absolute change in violent and property offenses between 1990 and 2001. The first column presents the ratio of the change in violent crime summed across the census tracts in the given quintile to the overall change in violent crime (negative values indicate that crime increased in the respective quintile). The next

[3] The National Neighborhood Crime Study (ICPSR 27501) does provide tract-level information on crime rates merged to census demographic data for 91 cities located within 64 metropolitan areas for the year 2000. See http://www.icpsr.umich.edu/icpsrweb/RCMD/studies/27501.

Table 1
Distribution of the Decline in Crime in Pittsburgh, Pennsylvania, between 1990 and 2001 by Census Tract Racial Composition and Poverty Rates

Panel A: Distribution by tracts stratified by quintiles according to the proportion of neighborhood residents that are African American in 2000

Quintile of proportion black	*Ratio of total violent crime decline in quintile to overall decline in the city*	*Ratio of total property crime decline in quintile to overall decline in the city*	*Proportion of 2000 resident population*	*Proportion black*	*Proportion poor*
Q1	–0.01	0.07	0.17	0.05	0.15
Q2	0.03	0.17	0.18	0.10	0.16
Q3	0.12	0.18	0.24	0.17	0.19
Q4	0.31	0.23	0.19	0.47	0.29
Q5	0.54	0.34	0.23	0.78	0.32

Panel B: Distribution by tracts stratified by quintiles according to the proportion of neighborhood residents that are poor in 2000

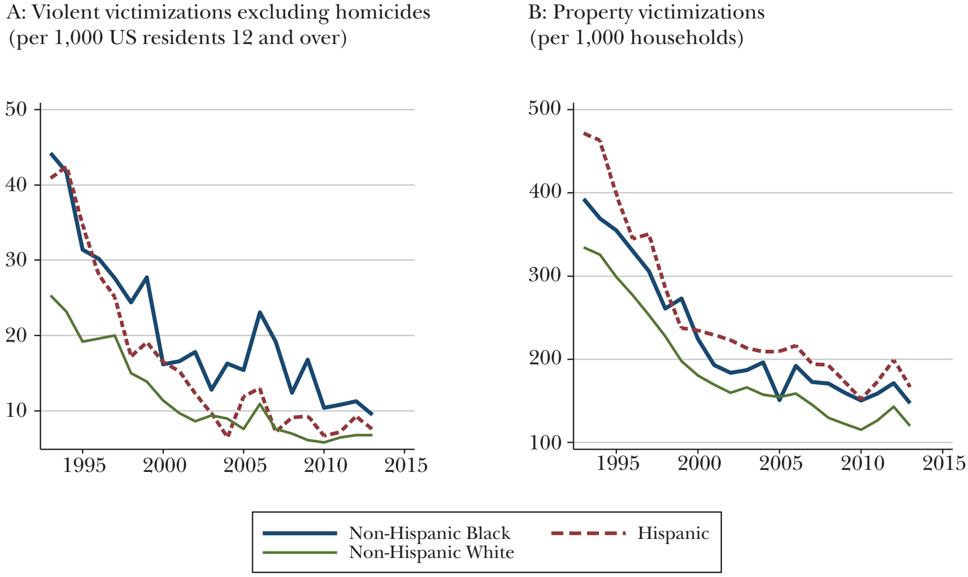
Quintile of proportion poor	*Ratio of total violent crime decline in quintile to overall decline in the city*	*Ratio of total property crime decline in quintile to overall decline in the city*	*Proportion of 2000 resident population*	*Proportion black*	*Proportion poor*
Q1	0.07	0.14	0.21	0.08	0.06
Q2	–0.01	0.12	0.23	0.17	0.12
Q3	0.04	0.14	0.20	0.23	0.18
Q4	0.30	0.29	0.19	0.42	0.27
Q5	0.60	0.31	0.17	0.67	0.47

Source: Authors using population data from the 2000 Census summary tape files 1 and 3A, and crime data from Cohen and Gorr (2006).

column presents similar tabulations for property crime. The third column shows the proportion of the city's population in 2000 in each tract group while the final two columns show the proportion African-American and the proportion poor in each group. Panel 1B produces a similar analysis where census tracts are stratified into quintiles according to the proportion of tract residents that are below the poverty line in the 2000 census.

The table reveals a very geographically concentrated crime decline within the city of Pittsburgh. The decline in violent crime in the 20 percent of tracts with the highest proportion black amounts to 54 percent of the overall decline in violent crime citywide. These tracts account for 23 percent of the city's population, have an average proportion black among tract residents of 0.78 and an average proportion poor of 0.32. Similarly, the decline in violent crime in the poorest quintile of tracts amounts to 60 percent of the citywide decline in violent crime incidents, despite these tracts being home to only 17 percent of the city's population. The

Figure 4
Violent Victimization and Property Victimizations by Race/Ethnicity, 1993 through 2013



Source: Authors using data from the National Crime Victimization Statistics.
Notes: Property crime, rates are expressed per 1,000 households according to the race of the household head. The serious violent victimization rate is measured per 1,000 residents 12 years of age and over and excludes homicide.

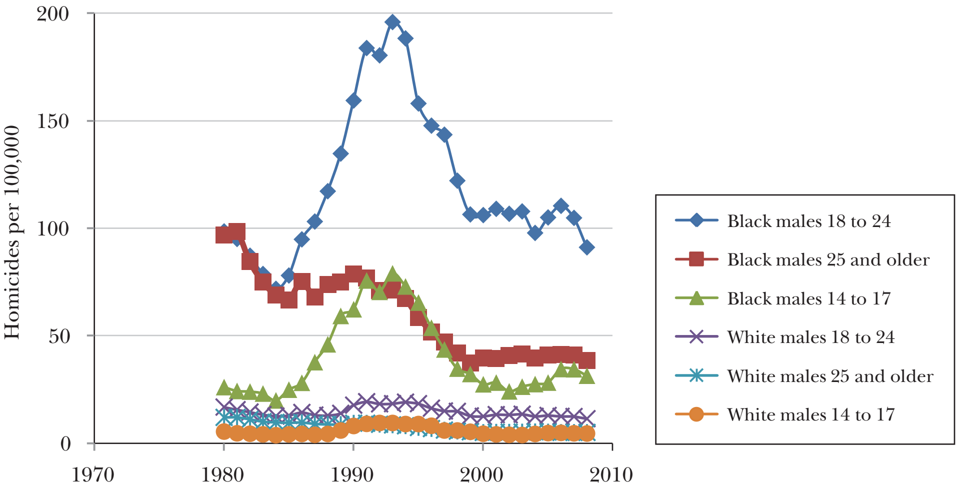
violent crime declines are considerably smaller in tracts with smaller proportions African-American and smaller proportion poor. Interestingly, while the decline in property crime is also skewed towards poorer and more minority census tracts, the geographic incidence of this decline is more even across the city's neighborhoods.[4]

Ideally, we would like to analyze trends in individual or household-level victimization rates for different income groups, such as quintiles of the household income distribution. Unfortunately, the income variable in the National Crime Victimization Statistics only reports household income ranges. Moreover, the nominal values of this variable are coded similarly over time, making it quite difficult to consistently subdivide the income distribution as nominal income increases with inflation. However, we can use the NCVS data on race/ethnicity to assess whether the cross-city patterns that we have documented (as well as the within-city patterns for Pittsburgh) are consistent with interpersonal victimization differentials.

Figure 4 presents the violent and property crime victimization rates by race/ethnicity for the period 1993 through 2013. For property crime, rates are

[4] The crime decline since the early 1990s has also considerably narrowed the difference in crime rates between the national central cities and suburbs (Kneebone and Raphael 2011).

Figure 5
Male Homicide Rates by Race and Broad Age Group
(per 100,000)



Source: Authors using data from US Bureau of Justice Statistics.

expressed per 1,000 households according to the race of the household head. The serious violent victimization rate is measured per 1,000 residents 12 years of age and over and excludes homicide. The patterns in Figure 4 confirm our cross-city and limited within-city analysis. Victimization rates decline sharply for all race/ethnic groups. However, the absolute and relative declines are largest for African-Americans and Hispanics. Given the average income differentials and differences in poverty rates between whites, Hispanics, and African-Americans, these results strongly indicate that lower-income households experienced disproportionately large reductions in criminal victimization since the beginning of the crime decline.

The most pronounced disparities in criminal victimization in the United States are found in homicide rates. There are enormous interracial disparities in homicide, with very strong interactions between gender and age. While these disparities are evident in all years with recorded data (O'Flaherty and Sethi 2010), they change drastically over time with shocks to drug markets and broader trends in crime rates. In 2008, the black homicide rate of 19.6 per 100,000 was nearly six times the white homicide rate (3.3 per 100,000). In 1991, at the peak of the run-up in black homicide rates beginning in 1986, the black homicide rate of 39.4 per 100,000 was over seven times that of the white homicide rate of 5.6 per 100,000. Figure 5 shows homicide rates for white and black males for three age groupings: 14 to 17 years of age, 18 to 24 years of age, and 25 and older. Several notable patterns emerge. First, homicide rates for black males 18 to 24 years of age are extraordinarily high in all years, reaching nearly 200 per 100,000 in the early 1990s and then declining to 91 per 100,000 in 2008. Second, homicide rates in all ranges for black males exceed homicide rates for white males. Third, black male homicide rates have fallen dramatically since the early 1990s, falling by half for males 18 and over and by over 60 percent for 14 to 17 year-old black males. White male homicide rates also dropped by roughly 40 to 50 percent, but from a much lower base.

Despite the decline in homicide rates since 1991, the homicide rates currently experienced by black males in the United States remain stunningly high. Understanding and addressing the high homicide rates for African-American males constitutes one of the most important criminal justice problems faced by the United States.

Inequality in the Incidence of the Direct and Indirect Costs of Punishment

The operation of the US criminal justice system is costly. For example, Anderson (2012) estimates annual US criminal justice expenditures circa 2010 of roughly $113 billion on police, $81 billion on corrections, $76 billion in expenditure by various federal agencies, and $84 billion devoted to combating drug trafficking. Beyond expenditures, criminal justice enforcement imposes costs on those convicted of crimes, their family members, and their communities. Some of these social costs are the direct and intended result of punishment, while others are indirect and unintended. For example, the forced removal from noninstitutionalized society associated with incarceration or the restrictions on liberties associated with a probation term are the direct and intended consequences.[5] The material deprivation of family members associated with losing an adult earner represent costs that are indirect and unintended. The prevalence and magnitude of these hard-to-measure direct and indirect social costs have increased and in an unequal manner over the past four decades. This has disproportionately affected poor minority communities, and in particular African-American men.

Before proceeding, a few institutional definitions regarding US corrections practices are in order. "Prisons" generally house those who are convicted of felonies and sentenced to serve at least one year. "Jails" house individuals awaiting arraignment and or trial, or those who are sentenced to relatively short incarceration spells. Many who are convicted of both felonies and misdemeanors are sentenced to probation in lieu of incarceration, or sometimes in combination with a short

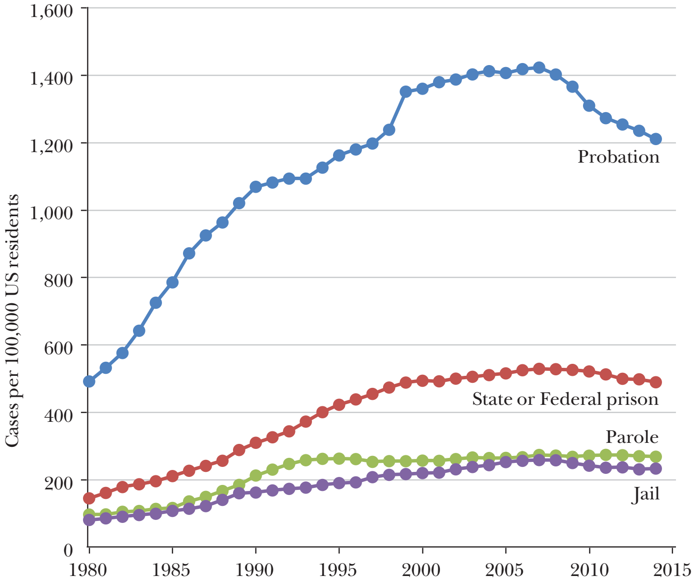
[5] To be clear, the fact that consequences are intended does not imply that they are socially optimal. Assessing the optimality of a given set of sentencing practices requires the articulation of a set of normative principals governing sentencing policy. For example, the recent National Academies of Science study of US incarceration rates articulated a normative framework for sentencing reform based on proportionality, procedural fairness, recognition of the possibilities for redemption and other deontological criteria stressing human rights and common citizenship (Travis, Western, and Redburn 2014). Alternatively, one may specify a more consequentialist position whereby the value of crime prevention net of enforcement costs defines the objective. In both frameworks, the direct effects of punishment associated with a given set of sentencing practices can be either too little or excessive.

jail sentence. Individuals on probation are monitored in the community with the degrees of stringency often depending on risk assessments conducted by local probation departments. Probation may be revoked for noncompliance with the conditions of probation, a legal action that can result in a jail or prison term. Finally, most prison releases are conditional from prison to a parole term. Paroled individuals are technically still under the custody of the state department of corrections and can have their conditional release revoked for technical reasons (missing appointments, leaving the county, testing positive for drugs) or for committing new crimes (Petersilia 1998). Because each US state has its own penal code, sentencing structure, and department of corrections, sentencing and correctional practices vary appreciably across states. Within states, each county has its own district attorney, sheriff's department (in charge of county jails), and probation departments, with practices varying widely across counties within states. In addition, there is a federal prison system for those who commit federal felony offenses and a separate system for processing federal cases.

Starting in the 1970s, federal and state jurisdictions across the United States toughened sentencing practices. Several underlying patterns are key to understanding the effects of these changes in practice. First, for the most serious crimes, we do not observe an increase in arrest rates or the number of arrests per crime. The one exception concerns drug arrests, where there is a pronounced increase in drug arrest rates starting in the mid-1980s that has been sustained through the present. Second, conditional on being arrested for a crime, the likelihood of being admitted to prison has increased for all offenses, especially those offenses for which the likelihood of being admitted to prison conditional on an arrest was low in years past (Neal and Rick 2015; Raphael and Stoll 2013; Travis, Western, and Redburn 2014). Third, effective sentence lengths (that is, the ultimate time served) within crime categories have gotten longer (Raphael and Stoll 2009, 2013; Neal and Rick 2015; Travis, Western, and Redburn 2014). This is especially true for the most serious crimes with a high likelihood of being admitted to prison upon conviction, such as murder, robbery, or rape/sexual assault. However, it is also observed for less-serious crimes. Finally, nearly all of the growth since 1980 can be explained by tougher sentences involving both more frequent use of prisons to punish felony offenses, as well as longer expected time served either conditional on conviction (Neal and Rick 2015) or conditional on prison admission (Raphael and Stoll 2013).

These changes have greatly expanded the reach of the criminal justice system, such that the proportion of American residents involved with the criminal justice system has reached historic highs. Figure 6 presents time series for various correctional populations, all expressed per 100,000 US residents. Between 1980 and 2013, all rates increased dramatically. The probation caseload is the largest in any given year, with roughly 500 per 100,000 on probation in 1980, expanding to over 1,200 per 100,000 in 2013. Between 1980 and 2007, the number of inmates in a state or federal prison increased from 145 per 100,000 to 530 per 100,000. In recent years, the prison incarceration rate has receded somewhat, in large part due to policy reforms in California, the state with the second-largest prison system in the country following the federal system. Jail incarceration rates have increased

Figure 6
Correctional Populations per 100,000 US Residents, 1980 to 2013



Source: Authors using data from Bureau of Justice Statistics, Annual Probation Survey, Annual Parole, Survey, Annual Survey of Jails, Census of Jail Inmates, and National Prisoner Statistics Program, 1980–2014.

with prison incarceration rates, from 81 inmates per 100,000 in 1980 to a peak of 259 inmates in 2007, before receding slightly to a rate of 237 in 2013. Over the time period depicted in the figure, the percent of the adult population under some form of active criminal justice supervision nearly tripled from 1.1 to 3 percent.

The population of adults involved with the criminal justice system is highly skewed towards specific demographic and socioeconomics groups (Raphael and Stoll 2013). For example, men account for roughly 93 percent of state and federal prison inmates and 88 percent of local jail inmates. Prison and jail inmates have very low levels of formal educational attainment, with 66 percent of state prisoners, 56 percent of federal prisoners, and 55 percent of local jail inmates having less than a high school degree. African-Americans are heavily overrepresented among the incarcerated, accounting for 43 percent of state prisoners, 46 percent of federal prisoners, and approximately 50 percent of jail inmates, while they are 13 percent of the US population as a whole. Hispanics are also overrepresented relative to their proportion of the general population, though to a lesser degree. Finally, most of the incarcerated are in prime working age ranges for men, ranging from their late 20s to their early 40s.

Whether measured at a point in time or as a cumulative life risk, incarceration, probation, and parole are common experiences in many minority communities. Tabulations from the 2010 American Community Survey indicate that roughly 11 percent of black men between 26 and 40 are residing in institutionalized group quarters on any given day. Narrowing the focus to black male high school dropouts in high-incarceration age ranges yields institutionalization rates of nearly one-third (Raphael 2005). The Bureau of Justice Statistics estimates that nearly one-third of black males born in 2001 will serve prison time at some point in their lives. The comparable figure for Hispanic men is 17 percent (Bonczar 2003). Using data from the Panel Study of Income Dynamics, Petit and Western (2004) estimate that for African-American men born between 1965 and 1969, 20.5 percent had been to prison by 1999. The comparable figures were 30.2 percent for black men without a college degree and approximately 59 percent for black men without a high school degree. We do not have comparable estimates for the proportions who have ever served a jail spell, been convicted of a felony or misdemeanor, been arrested, or been sentenced to probation, but such tabulations would undoubtedly reveal additional racial and ethnic disparities.

A great deal of research effort has been devoted to exploring many of the "collateral consequences" of the expansion of correctional populations; that is, the unintended consequences of punishment on convicted offenders, their families, and their communities more broadly. Collateral consequence studies have addressed the effects of criminal justice involvement on employment prospects (Grogger 1995; Holzer, Raphael, and Stoll 2006; Pager 2003; Western 2006; Mueller-Smith 2015), health outcomes (Johnson and Raphael 2009; Schnittker, Massoglia, and Uggen 2011), family budgets (Johnson 2009; Comfort 2007; Braman 2004), problem behaviors and depression among children of the incarcerated (Wakefield and Wildeman 2013), and political participation and civic engagement (Uggen and Manza 2002; Lerman and Weaver 2014), to name a few areas of inquiry. Several studies find evidence of perverse effects of incarceration spells on future criminal activity (Aizer and Doyle 2015; Mueller-Smith 2015; Nagin, Cullen, and Johnson 2009) as well as adverse effects of harsher conditions of confinement (Lerman 2013) and poor rehabilitation incentives for the incarcerated (Kuziemko 2013) on criminal recidivism.

The increasing prevalence of fines and fees imposed on those convicted of crimes raises an issue of how an intended consequence of these can lead to an array of unintended consequences. This issue recently received much attention with the release of an investigative report by the US Department of Justice Civil Rights Division (2015) analyzing the practices of the City of Ferguson, Missouri, in the wake of the shooting death of Michael Brown by a Ferguson police officer in August 2014. The report noted the aggressive use of fines and fees imposed for minor crimes, with this revenue accounting for roughly one-fifth of the city's general fund sources. The city of Ferguson is part of a broader trend. Courts and in some instances municipalities may impose a series of legal financial obligations on those convicted of crimes. These charges take many forms, including fees for the expense of jail incarceration, fees imposed on indigent defendants for the provision of a public defender, fees and surcharges for court cost reimbursements as well as for probation supervision, fines levied at sentencing for punishment, and restitution awards that compensate specific victims or that contribute to specific victim compensation funds (Bannon, Negrecha, and Diller 2010). Arrearages are common among individuals convicted of felony as well as misdemeanor offenses, with substantial heterogeneity in practices across US counties. While there is little information on cumulative outstanding legal financial obligation, we estimate that in 2012, fine and forfeiture revenue accruing to local, county, and state governments amounted to $15.3 billion.[6]

The use of fines and fees has increased in recent years. The best work on this topic is presented in Harris, Evans, and Beckett (2010).[7] In an analysis of nationally representative sentencing records and inmate surveys, the authors document an increase between 1991 and 2004 in the proportion of convicted felons with fines imposed at sentencing from 0.11 to 0.34. In addition, the proportion with outstanding restitution orders increases from 0.11 to 0.25. The authors also find that for convicted felons sentenced to jail rather than prison, or probation rather than prison or jail, the incidence of fines imposed at sentencing increases nearly threefold.

The authors also analyze administrative data on sentences imposed by Washington state superior courts in the first two months of 2004, a period of time where roughly 3,000 felony sentences were handed down. In addition to estimating mean and median monetary sanctions for these sentences, the authors randomly selected 500 individuals and cumulated lifetime monetary sanctions (including those imposed through juvenile courts) through the year 2008. The monetary sanctions exhibit great variability within offense category, and tend to be largest for drug felonies. This analysis revealed that many who are convicted of felony offenses carry substantial arrearages, and pay them off very slowly. They estimate that the median outstanding debt amounts to roughly half the likely annual earnings of the individuals impacted, while the mean balance is equal to a full year of potential earnings.

Money is fungible. When fines and fees are imposed as part of a criminal prosecution, at least some of the financial burden will devolve on to the household of the person involved with the criminal justice system. When someone who is involved in the criminal justice system has reduced employment prospects, some of those financial costs will again be borne by others in their household. We have said nothing about the family resources devoted to replenishing inmate commissary accounts, the devotion of household resources to prison phone calls, time devoted to visiting family members, and the other manners by which a family member's involvement with the criminal justice system may tax a household's resources. To our knowledge, aggregate data on such costs do not exist.

[6] This estimate is based on our tabulations from the 2012 Census of Governments: State and Local Finances revenue category U30, "Fines and Forfeits."

[7] See also Beckett and Harris (2011) and Harris, Evans, and Beckett (2011). Nagin (2008) provides a thoughtful discussion of the potential role of fines and fees in the US criminal justice system as an alternative sanction to incarceration, with attention to the implementation details, coordination requirements, and ethical tradeoffs. In addition, Ruback and Bergstrom (2006) provide a review of research on fines, fees, and restitutions and a discussion of the more systematic use of fines in western European countries.

The Criminal Justice Expansion and the Decline in Crime

We have documented unprecedented shifts in both crime and punishment. Crime rates have declined considerably since the early 1990s, and in a manner such that the benefits of this decline are quite progressively distributed. On the other hand, criminal sanctioning has become considerably more severe, with the direct and indirect impacts of this increased severity being regressively distributed. The juxtaposition of these two trends raises questions concerning what is driving the decline in crime and whether current punishment practices are necessary for maintaining currently low crime rates.

What caused the decline in US crime rates starting around 1991? There are a myriad of theories, but no smoking-gun explanation for these phenomenal changes. One body of research has focused on US time-series and cross-state evidence, both on changes in criminal justice policies and also on demographic and other factors that could have affected crime rates. However, a complicating factor is that many other western high-income countries with drastically different criminal justice systems have experienced a fall in crime rates since the 1990s, which suggests that discussions of cause and effect focused on distinctively American crime-enforcement policies and social events may be missing some important causal factors.

In the US-focused literature on the decline of rates of crime, among the many explanations that have been offered and evaluated by researchers are the general aging of the population (Levitt 2004; Baumer and Wolff 2014), a delayed effect of the legalization of abortion (Donohue and Levitt 2001, 2004; Foote and Goetz 2005), lower blood-lead levels among successive birth cohorts associated with the removal of lead from gasoline and paint (Nevin 2000, 2007; Reyes 2015), technological innovations that have made it more difficult to steal, especially locking systems in new cars (Farrell, Tilley, and Tseloni 2014), higher police staffing levels (Chalfin and McCrary 2013), innovative policing strategies (Braga and Bond 2008; Weisburd, Telep, Hinkle, and Eek 2010; Zimring 2007), an increase in the deployment of private security guards (Cook and MacDonald 2010, 2011) the waning of the crack cocaine epidemic (Fryer, Heaton, Levitt, and Murphy 2013), and the enormous rise of US incarceration rates (Levitt 1996; Liedke, Piehl, and Useem 2006; Raphael and Stoll 2013; Lofstrom and Raphael 2016). In an earlier assessment of the contribution of these factors in this journal, Levitt (2004) argues that nearly all of the US crime decline since 1991 can be explained by four factors: the legalization of abortion, the waning of the crack epidemic, the rise in the US incarceration rate, and the increase in police staffing levels.

All of these hypothesized factors remain active areas of research. Here, we will focus in particular on the possible linkage from incarceration to crime. As noted in the introduction, those who benefit most from the reduction in crime and those who are most likely to be incarcerated both come from the poorest communities in the country. Thus, the question arises as to the extent to which these communities face a tradeoff between lower crime rates and higher incarceration rates.

Before discussing specific research on the relationship between incarceration and crime, it is intriguing to note that other high-income countries have experienced a similar fall in crime rates without much change in their criminal justice enforcement or incarceration patterns. Zimring (2006, 2007) has noted the remarkable similarities between crime trends in the United States and Canada. Canada's property crime rate peaks in 1991 at 6,160 incidents per 100,000 before declining to 2,342 in 2013. Canada's violent crime rate peaked at 1,084 incidents per 100,000 before declining to a rate of 766 in 2013.[8] However, Canada's overall incarceration rate exhibits comparatively little variation. The incarceration rate inclusive of pre-trial detainees (referred to as those on remand) in 2013 stood at 139 per 100,000,[9] slightly higher than years past, but slightly less than one-fifth the comparable rate for the United States in 2013.

Tonry (2014) and Farrell, Tilly, and Tseloni (2014) provide further comparisons to mostly western European nations. While the timing of crime peaks and declines differ somewhat across countries, they observe substantial declines in violent crime and lethal violence in particular throughout Western Europe, with the timing of the declines in the United Kingdom most similar to crime trends in the United States. Taking a longer historical view, Eisner (2001, 2008, 2014) argues that criminal violence and lethal violence in particular have declined considerably and almost continuously since the thirteenth century AD. From this very long-run perspective, the increase in violent crime throughout the western world beginning in the mid-1960s appears to be an aberration from a longer-term historical trend, with the downward trend resuming in the 1990s (Eisner 2008). Incarceration rates in Western European countries are more in line with Canadian rates and are a fraction of the incarceration rates in the United States.

The comparable declines in crime in other nations raise questions regarding deeper forces in western societies that are tending towards lower offending levels and cast some doubt on the claims that the specific criminal justice policy choices made in the United States are the key to explaining the crime declines. That being said, there is considerable heterogeneity across US states and cities in criminal justice practices and changes therein, as well as ample and sometimes discrete policy variation in many national settings that permit well-identified study of the determinants of crime rates within nations. Moreover, there are important differences in either timing and/or magnitude of the US crime decline compared with the declines observed in other countries, suggesting that while the US experience may reflect broader trends in criminality worldwide, there are factors that are specific to the United States or to other specific countries that certainly merit consideration.

[8] See "Canada's Crime Rate: Two Decades of Decline" at Statistics Canada, http://www.statcan.gc.ca/pub/11-630-x/11-630-x2015001-eng.htm, accessed on February 22, 2016. The similarities between US and Canadian homicide trends figure prominently in interpretation of some time-series research pertaining to the deterrent effect of capital punishment, as in Donohue and Wolfers (2005).

[9] See Statistics Canada, http://www.statcan.gc.ca/pub/85-002-x/2015001/article/14163/c-g/desc/desc01-eng.htm, accessed on February 22, 2016.

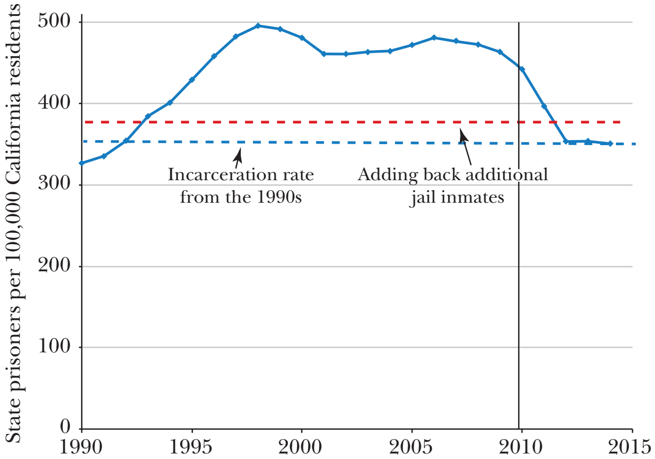
Levitt's (2004) review of the crime decline attributes one-third of the decline to increases in incarceration during the 1990s. This assessment was based largely on research studying the incarceration–crime relationship using data spanning the late 1970s, 1980s, and very early 1990s (specifically, the estimates in Levitt 1996). Since the publication of Levitt's (2004) assessment, there have been several quasi-experimental studies of the prison–crime relationship exploiting large, discrete, and policy-induced changes in incarceration rates in the US and elsewhere. There have also been advances in panel data estimates that explore the possibility of diminishing marginal effectiveness of incarceration as a crime-fighting tool. This research demonstrates that at relatively low incarceration rates, exogenous shocks to incarceration levels tend to have fairly large effects on crime, mostly through criminal incapacitation. However, this research also shows very small effects of changes in incarceration rates on crime when the incarceration rate is high—and evidence that diminishing effectiveness sets in at relatively low levels of incarceration.

For example, recent studies of policy shocks in European countries (Barbarino and Mastrobuoni 2014; Buonanno and Raphael 2013; Vollard 2012) show fairly large incapacitation effects in national settings with incarceration rates roughly one-sixth that of the United States. However, even in these very low-incarceration national settings, evidence of diminishing effectiveness is apparent. For example, Buonanno and Raphael (2013) find large reverse incapacitation effects of a mass Italian prison release in 2006 on felony offending, on the order of 13 to 18 reported felony offenses for each prison year not served. However, the effects are much smaller in Italian provinces with high pre-shock incarceration rates and larger in provinces with lower rates, with "high" incarceration provinces in Italy having combined post- and pre-trial incarceration rates that are generally below 200 per 100,000 population. Vollard (2012) finds that the application of a Dutch sentence enhancement for habitual offenders (those convicted of a new crime with ten or more prior felony convictions) netted considerably less-active offenders in Dutch municipalities that dipped further into the pool of local suspects in applying the sentencing enhancements.

In Lofstrom and Raphael (2016), we look at a recent policy shock to California. In October 2011, the state implemented sentencing reforms under pressure from a federal court order that greatly limited the use of prison for technical parole violations and defined a class of less-serious offenders to be diverted from prison sentences to locally imposed sanctions. Within one year, the state's prison population declined by nearly 28,000 inmates (roughly 13 percent), with an offsetting increase in the jail population of approximately 8,000 inmates. The reform reduced the state's incarceration rate (combining prison and jail together) to levels not seen since the early 1990s, effectively wiping away most of the prison growth coinciding with state's decline in crime, as illustrated in Figure 7.

The sharp decline in incarceration in 2011/2012 had very small effects on the state's crime rates. Crime trends in California have been comparable to those of the nation, with reported property crime rates peaking in 1991 and violent crime rates peaking in 1992. California crime rates decline considerably through 2010—the last full year before the sentencing reforms. There is a slight uptick in violent crime

Figure 7
California's Prison Incarceration Rates, 1990 through 2014

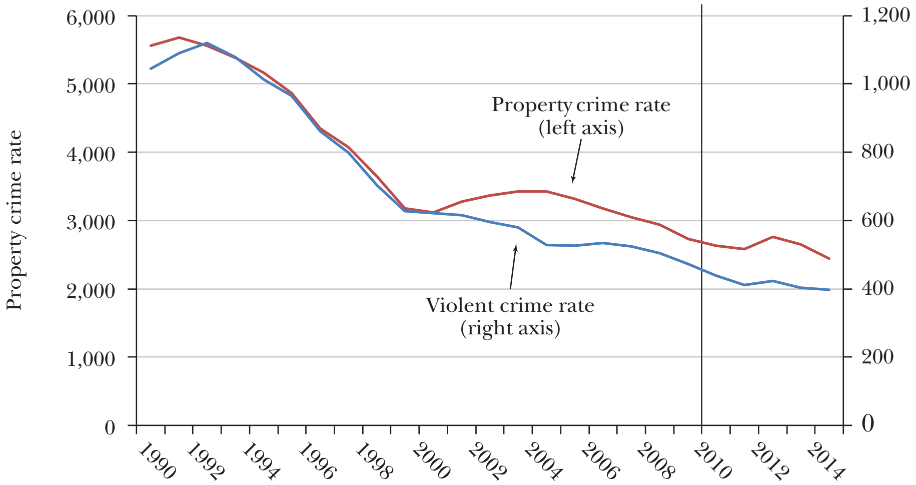


Source: Authors using data from the National Prisoner Statistics 1990 through 2014 and the California Jail Profile Survey.
Notes: The lower dashed line illustrates how the post-realignment incarceration level compares to incarceration rates from the 1990s. The upper dashed line displays the recent incarceration rate accounting for the transfer of population to local jails. 2010 is the last full year before the sentencing reforms.

in 2012, though this small uptick also occurred in states with comparable crime trends to California. Moreover, counties that experienced a larger reduction in their county-specific incarceration rates as a result of this reform did not experience relative increases in violent crime. In contrast, there is a more notable uptick in property crime above and beyond what is observed for a comparison group of states, and it is larger in counties disproportionately impacted by the reform. However, the effect is small. We estimate that California sentencing reform lead to 1.2 additional property felonies per prison year not served, with the effect almost entirely concentrated on auto theft. Notably, both property and violent crime rates in California remain at historical lows, as shown in Figure 8, far below the crime rates of the early 1990s, despite complete reversal of the state's incarceration growth since the early 1990s.

These findings for California are in line with recent panel data studies of the prison–crime effect. Liedke, Piehl, and Useem (2006) provide the earliest explicit attempt to assess whether the crime prevention effect of incarceration diminishes with scale. Using data from 1972–2000, in state-level panel data regressions that allow for interaction effects of changes in incarceration rates with the incarceration level, the authors find that the effect of incarceration on crime diminishes rapidly with scale, approaches zero somewhere between an incarceration level of 300 and 400 per 100,000, and possibly even turns positive at incarceration rates above that level. Theoretically, the use of incarceration can increase crime rates to the extent that the experience of incarceration is "criminogenic," or criminality enhancing.

Figure 8
California's Violent Crime Rate and Property Crime Rate
(per 100,000 residents)



Source: Authors using data from the Federal Bureau of Investigation Uniform Crime Reports.
Note: 2010 is the last full year before the sentencing reforms.

This outcome could occur if incarcerated individuals acculturate to criminal norms, learn how to be a better criminal while serving time, experience erosion of human capital valued in legitimate employment, or become accustomed and perhaps undeterred by the prospect of future prison spells.[10]

Johnson and Raphael (2012) provide further evidence of diminishing marginal effectiveness of incarceration. The authors use an instrument for incarceration based on the difference between a state's current incarceration rate and the state's steady-state incarceration rate implied by observable admissions and release rates. The authors derive an empirical prediction regarding the impact of this difference on next-year's change in incarceration based on a theoretical model of the relationship between crime and incarceration, and derive the conditions under which the transitory disparity between the actual and steady state incarceration rate provides a valid instrument for one-year lead changes in the actual incarceration rate. The authors then analyze state-level panel data for two time periods: 1978 to 1990 and 1991 to 2004. The former period is characterized by a relatively low incarceration

[10] Mueller-Smith (2015) and Aizer and Doyle (2015) both find evidence of net criminogenic effects of incarceration for adults sentenced in Harris County, Texas, and juveniles sentenced in Cook County, Illinois, respectively. Both identify exogenous variation in detention exploiting random assignment to judges and interjudge variation in sentencing severity. Nagin, Cullen, and Jonson (2009) offer a literature review of research assessing the effects of prior prison time on future offending.

rate (186 per 100,000) while the latter period is characterized by a much a higher incarceration rate (396 per 100,000). For the early period, an additional prison year served is estimated to prevent roughly 2.5 felony violent offenses and 11.4 felony property offenses, figures consistent with the crime–prison elasticities reported in Levitt's (1996) seminal study of the effect of prisoner overcrowding lawsuits. However, the comparable figures for crimes prevented per prison year served for the period 1991 through 2004 are 0.3 violent felony offenses and 2.7 felony property offenses. Raphael and Stoll (2013) reproduce this analysis with updated data for three time periods: 1977 through 1988, 1989 through 1999, and 2000 through 2010, with corresponding weighted-average state incarceration rates of 171, 349, and 449. This reanalysis find very small prison–crime effects for the latter two time periods (effectively zero for violent crime), but fairly large effects for the earliest time period, strongly suggestive of diminishing returns to scale.

These state-level panel data studies can be used to tabulate the contribution of expanded prison populations to declines in crime since the early 1990s. The estimates in Raphael and Stoll (2013) suggest that at most 7 percent of the decline in property crime since 1990 can be attributed to incarceration growth and none of the decline in violent crime. The larger estimates for the 1980s, however, suggest that had the prison population not been expanding between 1975 and 1989, the property and violent crime peaks in the early 1990s would have been roughly one-third higher.

These studies suggest that in drawing conclusions about how changes in incarceration rates will affect crime, one must keep the context of the study in mind. The collective clemency in Italy is obviously different from California's sentencing reforms, which were focused on limiting the use of prison for technical parole violations and less-serious crimes. In turn, the changes in California were different from the policy change of enhancing sentences for career criminals in Netherlands. It can't be assumed that levels or changes in incarceration rates or sentencing practices in one country will have similar effects in other countries with different institutions and history. In addition, changes in incarceration seems to have diminishing returns on crime, and thus it seems reasonable to argue that the rise in incarceration through the 1970s and into the 1980s may have had a substantial effect in reducing US crime rates, while simultaneously arguing that much of the growth in US incarceration rates since 1990 appears to have had little impact on crime.

Conclusion

The burdens of criminal victimization and criminal justice enforcement have changed drastically in the United States over the past three decades. Crime rates have fallen to historical lows since the early 1990s, with much larger absolute declines in relatively poor and minority communities. At the same time, the reach of the criminal justice system has greatly expanded. This predates the decline in crime by nearly a decade and a half, with prison incarceration rates and other correctional population departing from historical levels in the mid-1970s. However, this expansion accelerates in the early 1990s. In recent years, correctional populations have receded somewhat, due to selective reforms in a handful of states. However, incarceration rates, probation and parole populations, and the population of former prisoners and convicted felons among the noninstitutionalized remains at historical highs. Similar to the incidence of victimization, the distribution across demographic groups of criminal justice involvement is highly skewed towards low-income households, less-educated men, and African Americans. The great expansion in the scope and intensity of criminal sanctions has been born disproportionately by these groups.

It is certainly the case that on average criminal justice supervision of various severities deters and incapacitates and that the increases in incarceration through the early 1990s suppressed crime rates at the peak, perhaps considerably. However, the vast expansions occurring during the 1990s and during the first few years of the new century have bought little in terms of crime reduction but imposed substantial costs on the sanctioned, their families, and their communities.

Many of the same low-income predominantly African American communities have disproportionately experienced both the welcome reduction in inequality for crime victims and the less-welcome rise in inequality due to changes in criminal justice sanctioning. While it is tempting to consider whether these two changes in inequality can be weighed and balanced against each other, it seems to us that this temptation should be resisted on both theoretical and practical grounds. On theoretical grounds, the case for reducing inequality of any type is always rooted in claims about fairness and justice. In some situations, several different claims about inequality can be combined into a single scale—for example, when such claims can be monetized or measured in terms of income. But the inequality of the suffering of crime victims is fundamentally different from the inequality of disproportionate criminal justice sanctioning, and cannot be compared on the same scale. In practical terms, while higher rates of incarceration and other criminal justice sanctions may have had some effect in reducing crime back in the 1970s and through the 1980s, there is little evidence to believe that the higher rates have caused the reduction in crime in the last two decades. Thus, it is reasonable to pursue multiple policy goals, both seeking additional reductions in crime and in the continuing inequality of crime victimization and simultaneously seeking to reduce inequality of criminal justice sanctioning. If such policies are carried out sensibly, both kinds of inequality can be reduced without a meaningful tradeoff arising between them.

References

Aizer, Anna, and Joe Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130(2): 759–803.

Anderson, David A. 2012. "The Cost of Crime." *Foundations and Trends in Microeconomics* 7(3): 209–265.

Bannon, Alicia, Mitali Negrecha, and Rebekah Diller. 2010. *Criminal Justice Debt: A Barrier to Reentry*. Brennan Center for Justice, New York University School of Law.

Barbarino, Alessandro, and Giovanni Mastrobuoni. 2014. "The Incapacitation Effect of Incarceration: Evidence from Several Italian Collective Pardons." *American Economic Journal: Economic Policy* 6(1): 1–37.

Baumer, Eric P., and Kevin T. Wolff. 2014. "The Breadth and Causes of Contemporary Cross-National Homicide Trends." In *Crime and Justice* 43: 231–288.

Beckett, Katherine, and Alexes Harris. 2011. "On Cash and Conviction: Monetary Sanctions as Misguided Policy." *Criminology and Public Policy* 10(3): 505–37.

Blumstein, Alfred, and Richard Rosenfeld. 1998. "Explaining Recent Trends in Homicide Rate." *Journal of Criminal Law and Criminology* 88(4): 1175–1216.

Boggess, Scott, and John Bound. 1997. "Did Criminal Activity Increase during the 1980s? Comparisons across Data Sources." *Social Science Quarterly* 78(3): 725–39.

Bonczar, Thomas P. 2003. *Prevalence of Imprisonment in the U.S. Population, 1974–2001.* Bureau of Justice Statistics Special Report, NCJ 197976, U.S. Department of Justice.

Braga, Anthony A., and Brenda J. Bond. 2008. "Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial." *Criminology* 46(3): 577–606.

Braman, Donald. 2004. *Doing Time on the Outside: Incarceration and Family Life in Urban America.* Ann Arbor: University of Michigan Press.

Buonanno, Paolo, and Steven Raphael. 2013. "Incarceration and Incapacitation: Evidence from the 2006 Italian Collective Pardon." *American Economic Review* 103(6): 2437–65.

Chalfin, Aaron, and Justin McCrary. 2013. "The Effect of Police on Crime: New Evidence from U.S. Cities, 1960–2010." NBER Working Paper 18815.

Cohen, Jacqueline, and Wilpen L. Gorr. 2006. "Development of Crime Forecasting and Mapping Systems for Use by Police in Pittsburgh, Pennsylvania, and Rochester, New York, 1990–2001." ICPSR 4545.

Comfort, Megan. 2007. "Punishment Beyond the Legal Offender." *Annual Review of Law and Social Science* 3: 271–96.

Cook, Philip J., and John Mac Donald. 2010. "The Role of Private Action in Controlling Crime." In Cook, *Controlling Crime: Strategies and Payoffs,* edited by J. Philip J. Cook, Jens Ludwig, and Justin McCrary, 331–63. University of Chicago Press.

Cook, Philip J., and John MacDonald. 2011. "Public Safety through Private Action: An Economic Assessment of BIDS." *Economic Journal* 121(552): 445–62.

Donohue, John J. III, and Steven D. Levitt. 2001. "The Impact of Legalized Abortion on Crime." *Quarterly Journal of Economics* 116(2): 379–420.

Donohue, John J. III, and Steven D. Levitt. 2004. "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce." *Journal of Human Resources* 39(1): 29–49.

Donohue, John J., and Justin Wolfers. 2005. "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." *Stanford Law Review* 58(3): 791–845.

Eisner, Manuel. 2001. "Modernization, Self-Control, and Lethal Violence: The Long-Term Dynamics of European Homicide Rates in Theoretical Perspective." *British Journal of Criminology* 41(4): 618–38.

Eisner, Manuel. 2008. "Modernity Strikes Back? A Historical Perspective on the Latest Increase in Interpersonal Violence (1960–1990)." *International Journal of Conflict and Violence* 2(2): 289–316.

Eisner, Manuel. 2014. "From Swords to Words: Does Macro-Level Change in Self-Control Predict Long-Term Variation in Homicide Trends." In *Crime and Justice,* vol. 43, edited by Michael Tonry, 65–134. University of Chicago Press.

Farrell, Graham, Nick Tilley, and Andromachi Tseloni. 2014. "Why the Crime Drop?" In *Crime and Justice* 43: 421–90.

Foote, Christopher L., and Christopher F. Goetz. 2008. "The Impact of Legalized Abortion on Crime: Comment." *Quarterly Journal of Economics* 123(1): 407–23.

Fryer, Roland G., Paul S. Heaton, Steven D. Levitt, and Kevin M. Murphy. 2013. "Measuring Crack Cocaine and Its Impact." *Economic Inquiry* 51(3): 1651–81.

Grogger, Jeffrey. 1995. "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics* 110(1): 51–71.

Harris, Alexes, Heather Evans, and Katherine Beckett. 2010. "Drawing Blood from Stones: Legal Debt and Social Inequality in the Contemporary United States." *American Journal of Sociology* 115(6): 1753–99.

Harris, Alexes, Heather Evans, and Katherine Beckett. 2011. "Courtesy Stigma and Monetary Sanctions: Toward a Socio-Cultural Theory of Punishment." *American Sociological Review* 76(2): 234–64.

Holzer, Harry J., Steven Raphael, and Michael A. Stoll. 2006. "Perceived Criminality, Criminal Background Checks and the Racial Hiring Practices of Employers." *Journal of Law and Economics* 49(2): 451–80.

Johnson, Rucker C. 2009. "Ever-Increasing Levels of Parental Incarceration and the Consequences for Children." In *Do Prisons Make Us Safer? The Benefits and Costs of the Prison Boom,* edited by Steven Raphael and Michael A. Stoll, 177–206. New York, NY: Russell Sage Foundation.

Johnson, Rucker, and Steven Raphael. 2009. "The Effect of Male Incarceration Dynamics on AIDS Infection Rates among African-American Women and Men." *Journal of Law and Economics* 52(2): 251–93.

Johnson, Rucker, and Steven Raphael. 2012. "How Much Crime Reduction Does the Marginal Prisoner Buy?" *Journal of Law and Economics* 55(2): 275–310.

Kneebone, Elizabeth, and Steven Raphael. 2011. "City and Suburban Crime Trends in Metropolitan America." Metropolitan Opportunity Series, no. 18, Brookings Institution. May 26.

Kuziemko, Ilyana. 2013. "How Should Inmates Be Released from Prison? An Assessment of Parole versus Fixed-Sentence Regimes." *Quarterly Journal of Economics* 128(1): 371–424.

Lerman, Amy E. 2013. *The Modern Prison Paradox: Politics, Punishment, and Social Community.* Cambridge University Press.

Lerman, Amy E., and Vesla M. Weaver. 2014. *Arresting Citizenship: The Democratic Consequences of American Crime Control.* University of Chicago Press.

Levitt, Steven D. 1996. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation." *Quarterly Journal of Economics* 111(2): 319–52.

Levitt, Steven D. 2004. "Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not." *Journal of Economic Perspectives* 18(1): 163–90.

Liedke, Raymond V., Anne Morrison Piehl, and Bert Useem. 2006. "The Crime-Control Effect of Incarceration: Does Scale Matter?" *Criminology and Public Policy* 5(2): 245–75.

Lofstrom, Magnus, and Steven Raphael. 2016. "Incarceration and Crime: Evidence from California's Public Safety Realignment Reform." *ANNALS of the American Academy of Political and Social Sciences* 664(1): 196–220.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." Working paper, Department of Economics, University of Michigan.

Nagin, Daniel S. 2008. "Thoughts on the Broader Implications of 'The Miracle of the Cells.'" *Criminology and Public Policy* 7(1): 37–42.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson. 2009. "Imprisonment and Reoffending." *Crime and Justice* 38: 115–200.

Neal, Derek A., and Armin Rick. *Forthcoming.* **"The Prison Boom and Sentencing Policy."** *Journal of Legal Studies.*

Nevin, Rick. 2000. "How Lead Exposure Relates to Temporal Changes in IQ, Violent Crime, and Unwed Pregnancy." *Environmental Research* 83(1): 1–22.

Nevin, Rick. 2007. "Understanding International Crime Trends: The Legacy of Preschool Lead Exposure." *Environmental Research* 104(3): 315–36.

O'Flaherty, Brendan, and Rajiv Sethi. 2010. "Homicide in Black and White." *Journal of Urban Economics* 68(3): 215–30.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5): 937–75.

Petersilia, Joan. 1998. *Community Corrections: Probation, Parole, and Intermediate Sanctions.* Oxford University Press.

Pettit, Becky, and Bruce Western. 2004. "Mass Imprisonment and the Life Course: Race and Class Inequality in U.S. Incarceration." *American Sociological Review* 69(2): 151–69.

Raphael, Steven. 2005. "The Socioeconomic Status of Black Males: The Increasing Importance of Incarceration." In *Public Policy and the Income Distribution,* edited by Alan Auerbach, David Card, and John Quigley, 319–58. New York, NY: Russell Sage Foundation.

Raphael, Steven. 2011. "Improving Employment Prospects for Former Prison Inmates: Challenges and Policy." In *Controlling Crime: Strategies and Tradeoffs,* edited by Phillip J. Cook, Jens Ludwig, and Justin McCrary, 521–72. University of Chicago Press.

Raphael, Steven, and Michael A. Stoll. 2009. "Why Are So Many Americans in Prison?" In *Do Prisons Make Us Safer? The Benefits and Costs of the Prison Boom,* edited by Steven Raphael and Michael Stoll, 27–72, New York, NY: Russell Sage Foundation.

Raphael, Steven, and Michael A. Stoll. 2013. *Why Are So Many Americans in Prison?* New York, NY: Russell Sage Foundation.

Reyes, Jessica Wolpaw. 2015. "Lead Exposure and Behavior: Effects on Aggression and Risky Behavior among Children and Adolescents." *Economic Inquiry* 53(3): 1580–1605.

Ruback, R. Barry, and Mark H. Bergstrom. 2006. "Economic Sanctions in Criminal Justice: Purposes, Effects, and Implications." *Criminal Justice and Behavior* 33(2): 242–73.

Schnittker, Jason, Michael Massoglia, and Christopher Uggen. 2011. "Incarceration and the Health of the African-American Community."

Du Bois Review 8: 133–41.

Tonry, Michael. 2014. "Why Crime Rates Are Falling throughout the Western World." In *Crime and Justice* 43: 1–63.

Travis, Jeremy, Bruce Western, and Steve Redburn, editors. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences.* Washington, DC: National Academies Press.

Uggen, Christopher, and Jeff Manza. 2002. "Democratic Contraction? Political Consequences of Felon Disenfranchisement in the United States." *American Sociological Review* 67(6): 777–803.

US Department of Justice, Civil Rights Division. 2015. *Investigation of the Ferguson Police Department.* March 4. Washington, DC: Department of Justice Civil Rights Division.

Vollaard, Ben. 2012. "Preventing Crime through Selective Incapacitation." *Economic Journal* 123(567): 262–84.

Wakefield, Sara, and Christopher Wildeman. 2013. *Children of the Prison Boom: Mass Incarceration and the Future of American Inequality.* Oxford University Press.

Weisburd, David, Cody W. Telep, Joshua C. Hinkle, and John E. Eck. 2010. "Is Problem-Oriented Policing Effective in Reducing Crime and Disorder? Findings from a Campbell Systematic Review." *Criminology and Public Policy* 9(1): 139–72.

Western, Bruce. 2006. *Punishment and Inequality in America.* New York, NY: Russell Sage Foundation.

Zimring, Franklin E. 2006. "The Necessity and Value of Transnational Comparative Study: Some Preaching from a Recent Convert." *Criminology and Public Policy* 5(4): 615–22.

Zimring, Franklin E. 2007. *The Great American Crime Decline.* Oxford University Press.

Net Neutrality: A Fast Lane to Understanding the Trade-offs

Shane Greenstein, Martin Peitz, and Tommaso Valletti

Imagine you want to watch a movie online in the United States. For example, you might be subscribing to Comcast, the country's largest cable and broadband provider, and watching a movie delivered to you by Netflix, the giant television and movie-streaming service. What would happen if Comcast asked Netflix to pay for faster and more reliable access to its subscribers? Would Netflix be likely to agree to this request? Would Netflix charge you more for the movie? Would Comcast raise its broadband subscription fee for this improved service? If such a deal was struck, in what ways would consumer or producer welfare change? Such a deal between Comcast and Netflix actually happened in 2014. It is an example of prioritization, a network management practice that is part of the more general debate on "net neutrality."

The term "network neutrality" was introduced in a widely cited law article by Wu (2003). The article discusses whether an internet service provider should be required to treat all data from all content providers in the same way, and generally argues that net neutrality is good for the internet. For example, if a

■ *Shane Greenstein is MBA Class of 1957 Professor of Business Administration, Harvard Business School, Harvard University, Cambridge, Massachusetts. Martin Peitz is Professor of Economics, University of Mannheim, and also Director of the Mannheim Centre for Competition and Innovation (MaCCI), both in Mannheim, Germany. Tommaso Valletti is Professor of Economics, Imperial College Business School, London, United Kingdom; Professor of Economics, University of Rome Tor Vergata, Rome, Italy; and a Fellow at the Centre for Economic Policy Research (CEPR), London, United Kingdom. Their email addresses are sgreenstein@hbs.edu, martin.peitz@googlemail.com, and t.valletti@imperial.ac.uk.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.127 doi=10.1257/jep.30.2.127

net neutrality requirement does not exist, internet service providers might in some cases "throttle" certain content, slowing down delivery of that content—or even blocking it—so that it would not cause congestion and hinder other kinds of service. Internet service providers also might sign contracts to provide preferential treatment to the services of some content providers, giving their data a "fast lane" to users, so that other traffic would receive a "slow lane" during moments of congestion and arrive later.

The last decade has seen a strident public debate about net neutrality. Every developed country has had a different regulatory experience with this topic (for examples, see Marcus, Nooren, Cave, and Carter 2011; Marcus 2014). In the United States, the topic has fallen under the purview of the Federal Communications Commission, whose attempts to write rules have generated heated arguments, opposing votes along party lines, and repeated court review. In Europe, some individual states, such as the Netherlands, have introduced their own pro-neutrality legislation. A first piece of explicit Europe-wide neutrality legislation was proposed by the European Parliament in April 2014, and passed in modified form in October 2015 by the European Parliament. The policy debate remains alive today.

However, exactly what is at stake in net neutrality policy debates can be unclear. Considerable discussion has ensued among legal scholars about the precise meaning and issues surrounding net neutrality, and no consensus has emerged (for example, Yoo 2005; Sidak 2006; Lee and Wu 2009; Zittrain 2008; van Schewick 2010; Frischmann 2012). Stakeholders also take vastly different positions. Many data carriers say that differentiating charges and treatment of data will allow them to manage congestion efficiently and to provide an ongoing incentive to invest in faster service and innovate. Many content providers argue that without net neutrality rules, they will have a harder time reaching end users, reducing the benefits that end users receive from their internet connection. They also argue that the absence of net neutrality rules will deter them from innovating.

This article will provide a guide to the literature analyzing the economic trade-offs shaping policy choices. Our principal contribution is to identify the economic dimensions of this debate, and show that many questions can be informed by simple economic models of the market for internet services. This framework is useful both in teaching students and in informing the public about the economics of these important policy matters.

We begin by discussing the features of the modern internet. We introduce the key players, with a focus on internet service providers, content providers, and customers. We then summarize the insights of some models of the treatment of internet traffic. The economic literature has focused on two definitions of net neutrality. The most basic definition of net neutrality is to prohibit payments from content providers to internet service providers; this situation we refer to as a one-sided pricing model, in contrast with a two-sided pricing model in which such payments are permitted. Net neutrality may also be defined as prohibiting prioritization of traffic, with or without compensation. The research program then is to explore how a net neutrality rule would alter the distribution of rents and the efficiency of outcomes.

The economic literature examining net neutrality is young, and it would be rash to conclude that researchers have spotted all the key economic trade-offs. So throughout we identify some of the important open questions in this topic. Moreover, as we survey the literature, we highlight one particular theme: There is little support for the bold and simplistic claims of the most vociferous supporters and detractors of net neutrality. The economic consequences of such policies depend crucially on the precise policy choice and how it is implemented. The consequences further depend on how long-run economic trade-offs play out; for some of them, there is relevant experience in other industries to draw upon, but for others there is no experience and no consensus forecast.

Public net neutrality debates tend to range widely, and this too causes confusion. Our discussion will remain focused on policy for (un)equal treatment of traffic from different content providers by the "last-mile" internet service provider to which an end user subscribes. In this article, "net neutrality" *does not* encompass debates about consumer protection, like what a firm means when it advertises "unlimited" service. It also does not cover the extent to which the policies of internet service providers affect freedom of speech, privacy, or security.

Internet Structure and the Net Neutrality Debate

The modern commercial internet grew after many firms and users voluntarily adopted a set of practices for "inter-networking"—that is, transferring data between geographically dispersed local area networks and computer clients operated by different organizations. The commercial internet began to provide many revenue-generating services in the early to mid-1990s, and the network grew as many more firms and users began to participate. As of this writing, this network supports services to over three billion users (as reported at http://www.internetlivestats.com/internet-users/), and continues to grow worldwide.

Three facets of the network shape the net neutrality policy debate and arise in any economic model of this setting: complementarity between inputs provided by different firms; the direction and size of the flow of traffic and the flow of payments; and potential market power by some firms—in particular, internet service providers. We discuss these in turn.

Complementarity among inputs is almost synonymous with how the internet works, because what defines the modern internet is that it sends data from many locations to many locations. A broadband connection without access to any content is as useless as an online application without any broadband connectivity. An end user needs both. Here we see the three main players that we will study in this article: internet service providers such as Comcast, Verizon, or Vodafone; content providers such as Amazon, Facebook, Google, Netflix, Skype; and end users. A device, such as a laptop or a smartphone, is also needed, but we will ignore their (largely competitive) supply conditions.

Most economic models take for granted that the technical issues with complementarity have been solved. That is because all firms involved with moving data on the internet use the same nonproprietary “protocols,” which are standardized software commands that organize the procedures for moving data between routers, computers, and the various physical layers of the network. One design for protocols acts as the standard for today’s network, a protocol known as *TCP/IP*, which stands for Transmission Control Protocol/Internet Protocol.[1]

A major source of confusion for the economics arises from the many uses of the internet. Four types of different uses employ essentially the same internet processes: 1) static web browsing and e-mail, which tend to employ low bandwidth and can tolerate some delay; 2) video downloading, which can employ high bandwidth and can tolerate some delay; 3) voice-over IP, video-talk, video streaming, and multi-player gaming, which tend to employ high bandwidth and whose quality declines with delay; and 4) peer-to-peer applications, which tend to use high bandwidth and can tolerate delay, but can impose delay on others (Ou 2008).

Over time, the growth of the latter three applications has changed the scale and flow of data traffic on the internet, and this brought the treatment of traffic to the fore. Electronic mail dominated the volumes of traffic over the internet in the early 1990s, and tended to support nearly symmetric data flows from all locations to all locations. Though electronic mail has grown, email and web browsing made up only one-sixteenth of household internet traffic in 2014, while video made up just under two-thirds (Cisco 2015, Tables 10-13). In most developed countries, traffic related to web browsing became the majority of traffic sometime in the mid to late 1990s, and peer-to-peer traffic became the majority in the middle of the 2000s. In the last few years, streaming traffic for video applications makes up the largest fraction of traffic. On the modern internet, the majority of streaming traffic is unidirectional—mainly from content providers to users.

Most economic models of the internet overlook the details about how firms coordinate the movement of data. A first common arrangement moves data from a content provider over “backbone lines” and then to local broadband data carriers—either broadband internet service providers using DSL or cable, or mobile broadband providers. This step requires coordinated investments between content providers, backbone providers, and internet service providers, particularly at the points of interconnection between them. A second common arrangement moves traffic to servers located geographically close to the users of a broadband internet service provider. Independent third parties called content delivery networks, such as Akamai and Limelight, operate and maintain these servers. A content delivery network charges a content provider for hosting their content on servers close to the “last-mile” internet service provider, so that packets arrive at this service provider ahead of other packets.

[1] TCP/IP defines the “headers” or labels at the beginning and end of a “packet” of information. Each packet is of limited size, and as part of initial processing, larger messages are divided into several packets. Those headers inform a computer processor how to reassemble the packets, reproducing what had been sent. Vint Cerf and Robert Kahn wrote the first version of TCP/IP, and over time a large community of researchers and practitioners improved it to accommodate large-scale deployment. Useful starting points for interested readers would include Abbate (1999), Leiner et al. (2003), and Waldrop (2001). The transition from the noncommercial to the commercial Internet is explained in Greenstein (2015).

A third arrangement involves direct contracts between content providers and internet service providers, where the content providers "co-locate" their own servers inside the network of an internet service provider. Sometimes these direct contracts involve no payments; sometimes they do. Only content providers with popular content and the largest volumes of traffic choose this last option, suggesting that scale is an important consideration. In all three cases, questions often arise about who pays for investment to raise capacity for carrying data traffic.

The fact that most households have a very limited choice of broadband internet service providers adds an additional element to the policy concerns about these arrangements, motivating questions about an internet service provider's use of its market power vis-à-vis end users and content providers. As part of its 2015 Broadband Progress report for the United States, for example, the Federal Communications Commission found that a limited percentage of US households had access to a provider of broadband at 25 Mbps or more, and 20 percent had no access (Singleton 2015). Thus, an internet service provider (ISP) may enjoy a strong position in contractual negotiations with content providers, as it provides exclusive access to consumers who seek high-bandwidth services.

Have internet service providers sought to take advantage of their market status? Several recent controversies have made this question especially salient.

The Case of Bit-Torrent and Comcast. Bit-Torrent is a content provider focused on peer-to-peer file-sharing, including sharing of large files. Claims of Comcast interference with Bit-Torrent traffic had been circulating for many months, but the issue came to the forefront in 2007 when the Associated Press published a report concluding that Comcast was "throttling" Bit-Torrent traffic (Svensson 2007). Comcast claimed that it was not blocking peer-to-peer traffic, but only practicing "reasonable network management" to ensure quality service for all its subscribers. The Federal Communications Commission (2008) issued an order to Comcast for not having a "protocol-agnostic" policy—that is, a policy that applied to all content providers, not just one. Comcast altered its policies for restricting users who consume too much bandwidth (not specifically Bit-Torrent traffic), and sued the Federal Communications Commission over the scope and application of its legal authority.

The Dispute and Deal Between Netflix and Large Internet Service Providers. To accommodate growth in its streaming service, in late 2010 Netflix moved away from using Akamai's servers as its primary content delivery firm. Instead, it began streaming data through another content delivery firm, Limelight, and a backbone firm, Level3, and eventually another, Cogent. Netflix also began a program to co-locate its own servers inside ISPs. Some small ISPs agreed, with no money changing hands, but several large ISPs asked for payment for actions that reduced delays delivering data to households, such as upgrades to equipment, and for transporting data to servers with propitious locations. Users at these large ISPs began to experience delays around the middle of 2013 (with some public dispute about when this started, and why). In February 2014, Netflix and Comcast came to a deal—terms not publically disclosed—in which Netflix paid Comcast to co-locate servers inside Comcast's network. A little later, Netflix announced a similar deal with other large ISPs. Some commentators said these events illustrate a business-as-usual environment in which firms end up negotiating who will pay for certain investments (for example, Rayburn 2014). However, soon after these deals, Reid Hastings (2014), the chief executive officer of Netflix, seemed to display buyer's regret, publically raising alarms about the bargaining power of large broadband ISPs.

Data Caps and Their Exceptions. In many countries, internet service providers have adopted tiered pricing structures, in which higher bandwidth (and thus speed) comes at higher monthly expense to a household. In addition, some ISPs have adopted limits on total data usage for all users with particular contracts, which are known as "data caps." The levels and practices vary across providers, and participants hold distinct views about the consequences of these practices (Open Internet Advisory Committee 2013). Some ISPs also have adopted policies that count traffic over the public internet against the cap, but not traffic for video-on-demand using the ISP's proprietary or affiliated services. For instance, T-Mobile, a cellular provider in the United States, announced in November 2015 that it would exempt some video services such as ESPN, Netflix, and HBO from its data caps (but not others like Facebook and YouTube, for instance); however, to do so T-Mobile will stream the videos at lower quality, via a plan called "BingeOn" to which all customers are automatically opted in. Among the unresolved policy issues is whether these practices represent efficiency gains, or whether they unfairly tip the competitive landscape, raise the cost of rival services, and provide a cause for regulatory intervention.

Zero-rating Platforms. Facebook launched Free Basics in 2014, a Facebook-sponsored program that gives people in the developing world free access to cellular data for certain online services—including Facebook and WhatsApp (which belongs to Facebook). In 2015, Free Basics was available in 36 different countries, but has been temporarily banned in India while the Telecom Regulatory Authority of India sifts through public comments and explores whether the program violates the principles of net neutrality. The economics are similar to those of data caps, with the added twist that a content provider is visible as advertising and managing this program. While a "free" service is clearly good for increasing digital penetration, especially in developing countries, the biggest objection to Facebook's initiative is that it offers only a select few services chosen and controlled by Facebook, so that the platform could end up acting as gate-keeper limiting access to certain websites.

Paying for Faster Service. Orange, the largest French internet service provider, announced in 2013 that it made Google pay to deliver its YouTube traffic, though no exact figure was disclosed. The French telecommunications regulator also investigated in the same year complaints of Orange throttling against YouTube, but found no evidence of discriminatory action.

A key difference between the US and European regulatory situations is the market structure for internet service providers. European networks tend to have less concentration of broadband internet service providers, as well as less vertical integration between ISPs and content: for example, so far there is no merger in Europe equivalent to the merger of Comcast and NBC-Universal. In addition, most dominant internet content firms are based in the United States, which adds noise to every dispute concerning these issues in Europe.

While each of these events generated discussion with many legal aspects, important aspects of these debates also lend themselves to economic analysis. This is where we concentrate the bulk of our attention.

Basic Economic Analysis of Net Neutrality

A Neutrality (of Net Neutrality) Result

Let us revisit the example at the outset of this article. What happens if Comcast is allowed to charge Netflix every time that we watch a movie? The first, possibly surprising, answer that we give is this: when Comcast charges Netflix, *nothing* happens. A simple model shows how this result can arise.

Imagine an end user pays p to subscribe to Comcast and the subscription fee f to Netflix. Denote by t the "termination fee" that Netflix is asked to pay to Comcast (it is called a "termination fee" because it is the fee for bringing the content to the terminal point, that is, to the user). In this setting, Comcast and Netflix make take-it-or-leave-it offers to users and offer two services that are perfect complements. We are thus elaborating on the old issue of pricing with complementary goods, as already analyzed by Cournot (1838), enriched by side payments between the two firms offering those goods. Because broadband and content are perfect complements, the demand for both depends on the total price that an end user will have to pay. Let $q(p+f)$ denote this demand. The profits of Comcast and of Netflix are respectively given by

$$\pi_{ISP} = (p + t - c_{ISP})q$$

$$\pi_{CP} = (f - t - c_{CP})q,$$

where c_{ISP} and c_{CP} denote the per-subscriber cost to Comcast and Netflix, respectively. The two firms are free to set whatever price or subscription fee they want to for final users.

In this setting, it turns out that the payment t that flows between Comcast and Netflix does not affect margins for the firms, but only how they are earned. Remember that users face only a single total price, and that one particular level of this particular price will maximize overall joint profit for the two firms. Now imagine that Comcast tries to increase its profits by raising the termination fee t that it charges to Netflix. Netflix reacts by raising the subscription fee f that it charges to consumers to cover this change. Comcast will then react by lowering its price p so that the combined price-plus-fee charged to consumers remains at the profit-maximizing level. Indeed, as can be shown more rigorously, in this case the two firms have a symmetric position and will end up splitting the profits in half.[2] Ultimately, alterations in the termination fee change neither the bills of end users nor the profits of Comcast and Netflix. Regulation of the termination fee t would have no real economic consequences. The intuition for this is ultimately simple: There is "one price too many" in this setting, as both the internet service provider and the content provider can charge the user directly, through p and f, which means that any changes in termination fees can be easily offset.

Toward a More Complex World

This first result is not meant to end the discussion; on the contrary, the first result actually tells us that the simple model is missing several elements central to important economics trade-offs. At least four have received attention: 1) some content providers do not charge users directly, but get their money from advertising (like Facebook and Google); 2) there is considerable heterogeneity among both users and content providers, whereas the simple model deals with one content provider and one representative user; 3) we completely bypassed the "fast versus slow lane" issue, but congestion, quality of service, and network investments matter, as does investment by content providers; and 4) in some markets, multiple internet service providers compete for end users. Trade-offs are going to arise because, by introducing externalities, asymmetric information, and several dimensions of heterogeneity, often there will rather be "one or several prices too few" instead of "one price too many." To identify these trade-offs, we turn to a richer setting, although still very streamlined, that describes the internet ecosystem.

We begin with definitions for "one-sided pricing" and "two-sided pricing." To illustrate, we focus on how a single internet service provider interacts with two content providers and two end users. In two-sided pricing, the ISP can charge a subscription fee p to users and a termination fee t to content providers for delivering their content. By contrast, in one-sided pricing, the ISP can only charge a subscription fee to users. A regulatory restriction can rule out two-sided pricing.

[2] Under standard regularity conditions of the demand function, equilibrium prices are given as the solution to the first-order conditions

$$(1)\ \frac{d\pi_{ISP}}{dp} = (p + t - c_{ISP})q' + q = 0$$

$$(2)\ \frac{d\pi_{CP}}{df} = (f - t - c_{CP})q' + q = 0.$$

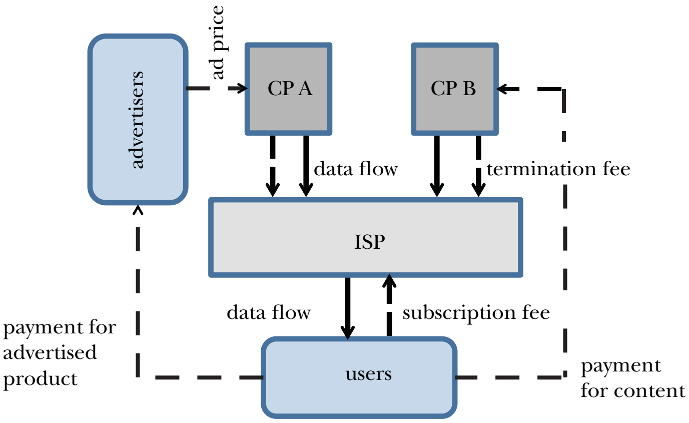
By adding (1) and (2) we get

$$(3)\ (f + p - c_{ISP} - c_{CP})q' + 2q = 0,$$

which pins down uniquely the total price $p + f$ that the end user faces, together with the respective quantity q end user buys. Notice that this total price is *independent* of the per-subscriber transfer t between Comcast and Netflix. Moreover, rewriting the two first-order conditions (1) and (2), we obtain that Comcast and Netflix have the same margins,

$$p + t - c_{ISP} = f - t - c_{CP} = -q/q'.$$

Figure 1
The Interaction between Internet Service Providers, Content Providers (CPs), and Users



Source: Authors.

In this framing, net neutrality can be thought of as a requirement that the ISP provide the same service to all content providers and users, while charging a fee *only* to users.

The first-best social welfare outcome most likely involves having all content providers and end users "on board," which is to say that welfare is maximized when all content providers can contact all end users. This outcome is realistic when thinking about the internet, with its large network externalities and low marginal costs. It immediately follows that any situation which restricts participation either on the user side or on the content provider side will be inefficient, because of the decline in the size of network externalities. (This intuition also can support regulatory mandates that internet service providers must provide access to all legal content on the internet.)

Figure 1 illustrates the general situation. The internet service provider is shown as a platform that sits between users and content providers. Of course, this model is a simplification, in the sense that a more detailed diagram would show the data from the content provider moving to its own ISP, then to the "backbone" firms of the public internet, then to the end-user's ISP, and then on to the end users. The fuller picture could also include the use of content delivery networks. But this simplification helps in understanding the first-order effects. In this diagram, the dashed arrows represent the (typical) direction of payment flows; clearly, regulatory intervention may alter such flows. The content providers in this model can have differing business models. In this diagram, for instance, content provider *B* makes revenues from selling its content to end users, while content provider *A* does not charge end users, but makes revenues from advertising.

We now consider different simplified scenarios, each of which is meant to capture a different economic mechanism. The only monetary component present in every model we consider is the subscription fee p that is paid by end users to the internet service provider.

Price Structure and Rent-shifting between Content Providers and the Internet Service Provider

We first focus on the scenario that corresponds to the situation in which content providers are of type A, as depicted in Figure 1. Ad-financing of the content provider is the case, for instance, for Facebook, Google Search, YouTube, Twitter, and many other content providers who distribute "free" content. In this case, it turns out that net neutrality can affect the distribution of rents, and may therefore not be neutral in the long run, when content providers and internet service providers have to make investment decisions. Relatedly, this example supports the economic intuition that regulating termination directly shapes margins and profits for ISPs and content firms.

Consider two content providers, denoted as 1 and 2, who provide different value to end users, and both have the same ability to generate profits R per user from advertising. If the two end users have identical tastes, the user valuation of the service from an internet service provider depends on which content is available. Also we will assume that advertising does not affect the end users' utility.[3] Suppose that content provider 2 provides more interesting or higher-quality content than content provider 1. The value of content from content provider 1 is equal to 1 and the value of content by content provider 2 is equal to $v > 1$.

Now, contrast one-sided and two-sided pricing. If the internet service provider is restricted to using a one-sided pricing strategy, it sets $p = v + 1$ and extracts the full expected surplus on the user side. The outcome is efficient. In contrast, with two-sided pricing, the monopoly ISP continues to set $p = v + 1$ for users and, in addition, charges a termination fee $t = R$ from content providers, where R is the rent that content providers would have received under one-sided pricing. End user choice is not affected, and, in this simple example, the profits of content providers and ISPs change by equal amounts in opposite directions. In other words, two-sided pricing leads to a redistribution of rents between content providers and the ISP.

The "Waterbed Effect"

The ability of internet service providers to charge termination fees to content providers for delivering their content raises an important question for policy: does a positive termination fee lead to an equal reduction in prices to users? The economic

[3] To simplify the interaction between advertisers and users who buy the products advertised on the internet, we will assume that advertisers extract the full surplus in the advertiser–user relationship and that users are neutral to advertising. Thus, we do not need to introduce consumer surplus from the advertised products in the welfare analysis.

intuition resembles common arguments about “pass-through” and in this context goes by the label “the waterbed effect.”

To illustrate, consider the situation in which the two content providers are identical, but users differ in their willingness to pay a subscription fee, which results in an elastic demand for access to the internet service provider. Thus, only some users may be willing to subscribe if the fee is high. When an ISP decides to charge a content provider for the (prioritized high-speed) traffic it generates, one may conjecture that the subscription fees paid by end users will decrease.

This result is obvious if the internet service provider is in a competitive setting, because its overall profits, from every source, are held to a normal level by the competitive process. However, the same result also holds for internet service providers with market power. If content providers are charged more, subscription fees will decrease because of the two-sided nature of the market. This will be to the advantage of end users, an aspect which is sometimes forgotten in the policy debate.

To develop some intuition for this result, consider the adjustment of prices on the user side with a monopoly internet service provider. Suppose that the two content providers are identical and can generate profits from advertising equal to R. Let the value of content for the first user be 1 for each content provider (and thus 2 overall). The corresponding value for the second user is denoted as $v > 1$ (note that 1 and v have now a different meaning than in the previous example). If the ISP is restricted to use a one-sided pricing strategy, it sets either $p = 2$ or $p = 2v$. In the former case, both users subscribe; in the latter case only the second user does so. It is profit-maximizing for the ISP to serve only the second user if $v > 2$, because the fee that it can charge to the second user more than compensates for the loss of the first user. The ISP uses its market power to raise the price such that some users (here user 1) prefer not to participate. In this case, the allocation is inefficient.

Now consider two-sided pricing. The internet service provider optimally charges the content provider a termination fee of $t = R$ to deliver content to each user. Thus, profits of the ISP on the content provider side are $4R$ if all content is delivered to all end users (since total transactions are 2 content providers × 2 users = 4 transactions) and $2R$ if all content is delivered to only one user. On the user side, the ISP again sets either $p = 2$ or $p = 2v$. In the former case, its overall profits are $4(1 + R)$ and in the latter $2(v + R)$. Thus, under two-sided pricing it is optimal for the ISP to set $p = 2$ and both users subscribe if $v \leq 2 + R$. Otherwise, $p = 2v$ and only user 2 subscribes.

Comparing one-sided to two-sided pricing, we see that users behave differently across the two regimes if $2 < v \leq 2 + R$. Here, a regulatory intervention to set $t = 0$ leads to a price *increase* on the user side from 2 to $2v$. This exemplifies a “waterbed effect,” which refers to a situation in which pressure on one side of the market leads to a corresponding change in prices on the other side of the market—as when pushing on one side of a waterbed causes a bulge to appear elsewhere. In this case, a higher termination fee results in a lower subscription price p for end users. In the present example, restricting internet service providers to one-sided pricing reduces total welfare and reduces the surplus received by the user, while content providers benefit from this regulatory intervention.

What economic mechanism is at work in this example? Given the opportunity to charge the content provider for termination, the internet service provider is more willing to decrease the subscription fee to end users, precisely because more end users can be attracted to join the platform, resulting in more transactions with content providers that are profitable for the ISP too. Net neutrality, instead, cuts all profits from content providers for the ISP, which is therefore not interested in generating additional traffic from them. This generates inefficiencies when it is more profitable to make more money from fewer subscribers.[4]

This illustration about waterbed effects can also help us understand why two-sided pricing is not always superior to one-sided pricing. Consider a similar situation, with two identical end users, but now allow content providers to generate different levels of profits from advertising, as occurs when content providers differ in their ability to engage users. In particular, content provider 1 generates profits from advertising, while content provider 2 generates profits R_2, with $R_1 > R_2$. The willingness to pay for content is 1 for each user and any content. How do one-sided and two-sided pricing compare?

Under one-sided pricing, the internet service provider charges $p = 2$ to each user and the allocation is efficient. Under two-sided pricing, the ISP sets $p = 1$ and $t = R_1$ if $R_1 > 2R_2 + 1$; otherwise it sets $p = 2$ and $t = R_2$. In the former case, the ISP is willing to sacrifice the delivery of content by content provider 2 despite also obtaining less from users, because of the incentive to extract rents from the content provider that generates the largest advertising profits. The resulting allocation under two-sided pricing is inefficient, as one of the content providers is excluded.[5] Due to the heterogeneity among content providers, the ISP might actually find it attractive to restrict access of some content providers, as a way to extract more money from giving termination to those content providers with a higher willingness to pay.[6]

[4] Note that, if the ISP could price discriminate and offer them different prices, user 1 would pay $p_1 = 2$ and user 2 would pay a higher price $p_2 = 2v$, both with and without net neutrality. In that case, there would be no real economic effects. Once again, with sufficiently "many prices," we would have a neutrality (of net neutrality) result.

[5] A very similar result emerges if the two content providers are identical but their advertising profits depend on the number of active content providers. If $R(m)$ denotes profits from advertising when m content providers are active, competition for advertisers among content providers implies that $R(1) > R(2)$. The analysis then is in line with the one in the previous example $t = R(1)$ and under two-sided pricing if $R(1) > 2R(2)$ +1. As a result, two-sided pricing here leads to less content in equilibrium and an efficiency loss compared to one-sided pricing. For a formal analysis of exclusive access as a means to reduce competition among content providers for advertisers, see Kourandi, Krämer, and Valletti (2015).

[6] Also here, with sufficiently "many prices," the earlier "neutrality (of net neutrality)" result would be at work. If the monopolist could perfectly target each advertiser, it would set $t_1 = R_1$ for content provider 1 and a lower $t_2 = R_2$ for content provider 2. Everybody would be on board, reaching the same level of welfare with and without net neutrality.

The trade-off between charging more to content providers or users has a clear economic intuition: one-sided pricing is welfare-superior if heterogeneity among content providers is particularly pronounced, whereas two-sided pricing is welfare-superior if heterogeneity among end users is particularly pronounced. In the former case, one-sided pricing tends to lead to more content providers being active; in the latter case, two-sided pricing tends to lead to more users enjoying content. How the trade-off plays out for end users crucially depends on the size and incidence of the waterbed effect in the presence of market power—that is, how much would allowing or eliminating a termination fee affect the price charged to subscribers. This is an open empirical question.

In many respects, it is a familiar question for economists, potentially lending itself to empirical analysis of the "pass-through" between termination fees and user rates.[7] Pass-through analysis is common in empirical studies of international trade and taxation. However, that alone would not be sufficient to settle arguments about the trade-offs, and the lack of experience with use of termination fees in practice makes this more of a forecast than an estimate on past behavior.

When Content Providers Charge End Users

The economic modeling becomes more challenging when content providers have direct relationships with end users and charge them a subscription fee. This market practice is becoming common among streaming firms and providers of "over-the-top" services, such as Amazon Prime, Netflix, YouTube Red, and subscription television services. The economic insights in this situation are sufficiently ambiguous that economists should be wary of advocates making bold policy prescriptions in favor of two-sided pricing or one-sided pricing when content firms offer subscriptions.

In general, with both one-sided and two-sided pricing by the internet service provider, we should expect inefficient outcomes, as both content providers and the ISP want to charge users for the complementary services they offer.[8] Moreover, the extent of the (in)efficiency of two-sided pricing (in comparison to one-sided pricing) depends crucially on whether an ISP can charge distinct termination fees to different content providers, and whether content providers have an ability to pass on the termination fees charged by ISPs. When ISPs can tailor termination fees to each content firm, then they have an instrument for "taxing" every content firm and extracting surplus, which will be passed on to users if content firms can do so. Thus tailored termination fees tend to lead to more efficiency.

More realistically, without the ability of an internet service provider to tailor fees in this way, then inefficiencies will arise, as any termination fee will induce exits

[7] For a formal treatment, see Armstrong (2006) and Weyl and Fabinger (2013). Empirical evidence of the "waterbed" effect is provided by Genakos and Valletti (2011, 2015) in the context of cellular phones.

[8] This feature was already present in our baseline example; for example, as summarized by Equation 3 in footnote 2. An efficient outcome would imply that the total price should be equal to marginal costs $c_{ISP} + c_{CP}$; instead Equation 3 says that there is a mark-up above such costs. The relevant question is therefore whether this inefficiency is more severe with or without net neutrality.

(from content firms) that would not have arisen in one-sided pricing.[9] In that case, one-sided pricing can outperform two-sided pricing when content providers make their revenues from charging end users and users differ in their willingness to pay for content. The intuition appeared in prior discussion: if an ISP cannot perfectly price discriminate across its users, it does not have incentive to account for the additional gain to users from access to additional content.

Contracting with Externalities

Stepping back from the details, we can observe a pattern in the analysis so far. The cases above are examples of situations where parties are "contracting with externalities." An internet service provider may increase bandwidth to subscribers without taking into account the advertising revenues that will accrue to content providers, who deliver content to such subscribers. Similarly, a content provider may introduce new applications desired by subscribers without taking into account the effect this has on the rents the ISP can extract. With suitably many payments (and symmetric information) between the parties, mild forms of regulation would be neutral in this setting: that is, relative prices might change, but not total prices paid/received by the parties involved—and regulation has little or no impact on final allocations.

Certain forms of neutrality regulation can lead to real effects, namely, when they impose sufficiently binding constraints on the contracts between internet service providers, users, and content providers. For instance, when prices are required to be uniform or zero between two types of parties, then numerous inefficiencies arise, as we demonstrated with the examples in the previous section (see also Gans 2015).

It is an unresolved question which type of economic effect dominates in practice. The analysis indicates the challenge for addressing the issue. Which insights are most empirically relevant in an environment where content firms use a mix of advertising and subscription models for generating revenue? Moreover, if regulators do impose strong constraints on contracts between the players in internet access markets, it becomes difficult to learn what pricing structures would arise in an alternative situation, for example, if parties had been permitted to negotiate.

[9] Let us revisit the example with ad-financed content providers, in which content providers generate different values for users. Imagine the content providers commit to their fees before the monopoly ISP sets its prices. (However, content providers do anticipate the pricing by the ISP.) As we will see, from a welfare perspective, one-sided pricing now tends to outperform two-sided pricing. Suppose that the user's willingness to pay for content by content provider 1 is equal to 1 and the willingness to pay for content by content provider 2 is equal to $v > 1$. Then, content provider 1 optimally sets $f_1 = 1$ and content provider 2 sets $f_2 = v$ irrespective of whether the ISP is required to use one-sided pricing or is allowed to engage in two-sided pricing. With one-sided pricing, the ISP cannot make a positive profit and sets $p = 0$; the allocation is efficient as both content providers would be active. With two-sided pricing, the ISP either decides to set $t = 1$ and $p = 0$ or $t = v$ and $p = 0$. With the former strategy, its profit is equal to 4 and with the latter, $2v$. If $v > 2$, the ISP optimally uses the latter strategy, content provider 2 prefers not to participate, and the allocation becomes inefficient. This model assumes that the ISP must set a nondiscriminatory termination rate—that is, the same for both content providers—which makes participation of the low-quality content provider unattractive.

Broadening the Model: Congestion, Investment, Competition

The discussion to this point has been heavily focused on choices about price-setting in different sets of circumstances. However, several important economic trade-offs are missing. Price setting interacts with the quality of service when networks regularly suffer congestion, for example. Congestion can be exacerbated by lack of investment, or by the (presence or absence of) rules governing prioritization of traffic. When incorporating investments, long-term trade-offs come into play.

These long-term trade-offs depend on the competitive setting, both horizontal competition (between internet service providers) and vertical competition (between ISPs integrated into content and other content providers). While these long-term issues are standard issues in industrial organization, setting them in a data network gives rise to novel trade-offs and concerns.

Congestion: Static Effects

The presence of potential congestion can result in some internet traffic being delivered with delay, making it potentially valueless. Standard economics suggests a strong analogy here with pricing automobile traffic congestion. However, the discussion above suggested the potential for a missing price. Thus, in a world of second best, interesting economic trade-offs should arise.

To begin, recognize that some types of traffic lose their value with delay (like Skype calls) while other types do not (downloading large files with BitTorrent). Delaying the latter would cause little social cost, as the material is not very time-sensitive, and the principal cost of delay is inconvenience. Some form of time-of-day pricing could induce delaying traffic until nonpeak hours, although such congestion pricing would be at odds with the strictest net neutrality requirements. Appropriate peak pricing could give incentives to users (in particular, content providers) to reduce congestion, and this should be welfare-enhancing if capacity is provided with priority to high-value traffic.

Time-of-day pricing was common during the era of dial-up internet service providers, but it is not common in the broadband era. One puzzle is why broadband carriers have not initiated experiments with such programs, especially in the era prior to political lobbying for net neutrality regulation.

Because congestion tends to arise only during peak load hours, a more controversial question concerns treating content providers unequally during such hours. Some carriers have proposed keeping a "slow lane" for free, while allowing for a paid-for "fast lane."

There are still few contributions that consider the impact of net-neutrality policies on high-volume and time-sensitive traffic (exceptions include Choi, Jeon, and Kim 2015a, 2015b; Peitz and Schuett 2015). The potential for efficiency gains arises because rationing traffic could lead to better performance for time-sensitive traffic in times of congestion. Whether users are better off depends on whether those gains outweigh the potential distortions that arise. As with the discussion above, the distortions depend crucially on the ability and incentives of the content provider to pass on to users the price for prioritized delivery, and the incentives of the ISP to adjust the subscription fee on the user side.

A major policy concern for prioritization is whether internet service providers manipulate congestion in self-interested ways that lead to (un)desirable outcomes. For example, Choi and Kim (2010), using a queuing model for traffic, illustrate how prioritized access could serve as a rent-extraction device.[10] In other words, the ISP engages in "menu pricing" (or second-degree price discrimination) by offering a choice between price plans, which lets content providers sort themselves by their choices. Prioritization then gives a "too large" market share to the content provider that opts for priority, while it would be socially preferable to have more equal shares and more content provision.

Economides and Hermalin (2012), considering a fixed pipe for time-sensitive traffic transmitted in times of congestion, find that charges for prioritized access can serve as a price-discrimination device. In their setting, the internet service provider extracts considerable surplus, which may lead to too little content provision. The analysis is reminiscent of the properties of third-degree price discrimination, insofar as welfare increases under the regime that allows a greater amount of content to be consumed.

Investment and the Dirt Road

Public debate has expressed a concern that, in the absence of net neutrality, an internet service provider might benefit from strategically degrading the quality of the nonpriority lane in order to drive traffic to a paid-for prioritized lane. In popular discussion, this possibility sometimes goes under the heading "fast lane versus the dirt road."

Economic analysis acknowledges distinct policy concerns. One set of concerns about the dirt road builds on a standard model of endogenous quality selection from a monopolist provider. When internet service providers offer multiple tiers of services and prices, a monopolist ISP could face incentives to shade the quality of lower-quality products in order to give incentives to users to upgrade to higher margins for the higher-quality products (Mussa and Rosen 1978). This incentive could manifest itself as lower investment in the capacity of lower-tier service, which users would experience as constrained capacity. Importantly, most practical net neutrality proposals permit tiered services to users, and, therefore, do not alter this incentive.

Additional policy concerns arise from selling prioritization to content providers for delivering data to users. Because monopoly providers of access to users may be the only channel through which content providers can reach users, internet service providers have incentives to invest in ways that raise the value of the prioritization

[10] Specifically, they employ the so-called M/M/1 queuing system, which is also used in other studies like Krämer and Wiewiorra (2012) and Bourreau, Kourandi, and Valletti (2015). This queuing system is considered a good proxy for actual congestion: 1) the total number of Internet users is large; 2) each user has a small impact on congestion; and 3) all Internet users can be assumed as independent. In reality, packets move through a complex network of routers, but economic models have abstracted from this.

sold to content providers. For example, Choi and Kim (2010) argue that the ISP may have *less* incentive to invest in network expansion in a regime with prioritization, because by doing so it can create scarcity that makes content providers more desperate to obtain priority. An ISP might benefit from strategically degrading (at least in relative terms) the quality of the nonpriority lane in order to extract higher profits from the priority lane. Bourreau, Kourandi, and Valletti (2015) argue that this risk of fast lane/dirt road is present even with competing ISPs.

Innovation introduces an additional dimension into this debate. A certain level of "quality of service" (performance of the network), may be needed to make innovative services by content providers feasible: for example, guaranteed delivery quality may be a key factor to make socially valuable major innovations in interactive e-learning, e-health care services, and e-mobility in the form of autonomous vehicles. In that case, a regime of fast and slow lanes allows internet service providers to extract additional revenues from content providers through priority fees. It is possible that innovation in content provider services will also increase: some highly congestion-sensitive applications, which were left out of the market under net neutrality, would enter when applications can make deals for high-priority lanes (Bourreau, Kourandi, Valletti 2015; Krämer and Wiewiorra 2012).[11]

One additional policy implication deserves to be mentioned: initiative by a regulatory authority to monitor traffic quality can help avoid the fast lane/dirt road problems by enforcing a minimally required floor. On a related note, if regulation of traffic quality is too complex or costly for the regulatory authority to monitor, a net neutrality regime might be a useful policy to avoid quality degradation of the traffic for nonpriority content providers.

Economic analyses on investment and network neutrality have mainly focused on the expansion of network capacity by internet service providers. However, an expansion of capacity by ISPs is not the only solution to resolving congestion problems. Major content providers such as Google, Netflix, and Amazon have developed other measures, such as advanced compression technologies, to ensure a sufficient quality-of-service. They have also deployed or rented content delivery networks. That raises the question of whether prioritized delivery and other investments in quality-of-service are substitutes or complements, which remains an important unresolved question. Choi, Jeon, and Kim (2014) take first steps in addressing it, by showing how the result depends on whether the ISP has a large or small installed network capacity.

One often hears the concern that strict net neutrality rules would help small innovative firms because large content providers are better able to pay for

[11] As noted by several writers, however, the opportunity cost of such services may be the underinvestment in public access networks, discouraging investment brought by entrepreneurial services that use the public network (Bourreau et al. 2015). The question then is the following: if the internet service provider can charge on the content side, and earns greater total profit by doing so, will it invest more in equilibrium (because it can appropriate a greater share of the surplus generated by the investment) or invest less (because the content side invests less)? According to Reggiani and Valletti (2016), there is a complementarity between investment by ISPs and total investments on the content provider side.

prioritization. However, as large content providers have other means to deal with the congestion issue, it may instead be the small innovative firms which need the possibility of prioritized access, because it does not require larger forms of up-front investments which they can ill afford.

Competition and Bottlenecks

It is often argued that spurring competition between internet service providers can remove the need for a regulatory approach to net neutrality. For example, the European Commission (2011) stated that "the significance of the types of problems arising in the net-neutrality debate is correlated to the degree of competition existing in the market." In the United States, the Federal Communications Commission in a 2010 ruling exempted mobile networks from most of the net neutrality rules, on the grounds that they face stronger capacity constraints than fixed networks, and that competition at least mitigates any negative effects of a departure from net neutrality.

Would introducing more competition eliminate the need for net neutrality regulation? This is an open question because few models address how bargaining between internet service providers and content providers changes when some of an ISP's users face competitive alternatives. For example, is the threat by some users to move between ISPs sufficient to alter an ISP's pricing and investment activity? What is the biggest competitive consequence from an ISP becoming larger through merger? Does it hamper the prospects of potential entrants or does it increase its strength when negotiating with content providers?

Recent theory contributions generally support the idea that lifting net neutrality regulation on competing platforms is welfare-increasing (Krämer and Wiewiorra 2012; Bourreau, Kourandi, and Valletti 2015). However, this outcome does not arise because competition reduces the incentives of ISPs to discriminate between content providers. In these models, each ISP has a unilateral incentive to introduce a priority lane, no matter what its rival does. Thus, price discrimination on the content provider side continues to be present. What then is the reason for the welfare gain? In a situation where end users subscribe to only one ISP (that is, they "single-home"), there is fierce competition for end users among ISPs when the ISP is allowed to charge content providers. The overall effect of more competition tends to be a better deal for users and lower overall price distortions. An increase in competition makes the ISP's firm-specific demand on the user side more elastic, but does not make much difference to the ISP's firm-specific demand on the content side (which is inelastic so long as end users single-home).

The exercise of monopoly power over content providers arises independent of competition for end users under two-sided pricing. This "termination bottleneck" problem is common in traditional telephony regulation (see also Economides and Tåg 2012). The problem is less pronounced if the content provider can reach some users on multiple platforms (that is, if some end-users "multi-home") or if the content provider has bargaining power so that it can negotiate its termination fees with the internet service provider (Armstrong 2002).

Competition and Vertical Issues

Internet service providers can decide to integrate into services other than delivery of data, like video on demand. There are two key questions for economic analysis of net neutrality. First, under what condition does vertical integration reflect some efficiency rationale, thus improving the experience of users? Second, in which circumstances does it lead to harm to the competitive process because users cannot access alternative content providers, who compete on an "uneven playing field?"[12] Once again, the economics of this topic depends mostly on speculative forecasts about carrier behavior, user elasticities, and content provider incentives, and not much in the way of regulatory case experience.

The economic arguments for efficiency in this setting are not unique to the net neutrality debate. At least in theory, when content providers and internet service providers offer complements, vertical integration may reduce prices, as absent integration neither party internalizes the profit loss inflicted on the other party by raising its price. Also, vertical integration may reduce the underinvestment that arises with independent parties, as the investment generates benefits for the firm producing the complementary product. The magnitude of these gains in practice is unclear, but the empirical literature from other industries identifies many examples of efficiency gains from vertical integration (Lafontaine and Slade 2007).

Anticompetitive concerns arise because an internet service provider may offer its own services and charge termination fees for competing content providers, potentially leading to partial or full exclusion. For example, Netflix's customers may use Comcast's network to download videos from Netflix, while Comcast also sells video services delivered through cable television. Similarly, both telecom and cable ISPs provide their own phone services that are also supplied by independent voice-over-IP providers such as Skype or Vonage. Without network neutrality rules (or interventions based on general competition law), ISPs may favor their own services and use price and possibly nonprice instruments to reduce competition.

A related concern about "uneven playing fields" arises in situations where internet service providers impose data caps on use. As stressed in the earlier discussion, caps arise as part of tiered pricing.[13] Data caps also shape competition between content providers. Data caps create an artificial scarcity, making users perceive different digital products from different content providers as substitutes (for scarce space within the cap). Thus, the cap "heightens" the degree of competition

[12] A notable example is the case regarding Madison River, a small telephone company accused of blocking ports used for voice-over-internet (VoIP) applications, thereby affecting customers' ability to use VoIP through one or more VoIP service providers. See "In the Matter of Madison River Communications, LLC and affiliated companies" (FCC 2005).

[13] For an economic estimate of the behavioral consequences of such caps, see Nevo, Turner, and Williams (forthcoming). They show that caps imposed on end users have important allocative effects. However data limitations do not allow them to consider unequal treatment of traffic from different sources, which remains an important area for further research (for a first step see Nurski, 2014, using UK data). See also Jullien and Sand-Zantman (2016) for a formal treatment.

experienced by content providers, who might otherwise have been able to differentiate from competitors satisfying heterogeneous user tastes (Economides and Hermalin 2015). If the ISP is allowed to charge the content provider directly for a priority lane, then an ISP may be able to raise its profits further with the "heightened" competitive setting, inducing content providers to bid more for the fast lane than they otherwise would have done in the absence of a cap.

One other concern with caps is their scope. Some internet service providers have adopted policies to exempt traffic for their own services in a policy known as "zero-rating." Simple general conclusions are hard to state because of the variety of situations. As one example, in 2012 Comcast exempted Xfinity app use from its data cap when watching through Xbox (Open Internet Advisory Committee 2013, Appendix 1, pp. 35–38). Other content providers raised questions about whether exemption of some traffic created an uneven playing field, while carriers claimed the practice generated efficiency gains. As a second example from a very different setting, some carriers have considered supplying a free bundled service to a wireless broadband subscription, such as a Spotify Premium for subscribers to T-Mobile. If a carrier cannot exempt its own service, is it permitted to do so with a business partner? A third example also raises issues about universal service. Zero pricing may seem to offer "free" services—often in the context of free access to a limited version of the internet for the poor. Does expansion of use provide a benefit that merits less concern about the competitive effects, or not?

Conclusion

Net neutrality has been on the agenda of policymakers and in the news in recent years. The Federal Communications Commission (2010) adopted is first order on this subject in November 2010. Its most recent order came in early 2015 (available at https://www.fcc.gov/document/protecting-and-promoting-open-internet-nprm). It covers three distinct areas of the behavior of a broadband internet service provider: rules to limit the right to block traffic; rules defining minimal transparency requirements for internet service providers; and rules for limiting discriminatory treatment of traffic. Both the precise meaning and the legal status of the 2015 order remain unsettled. Many regulators around the globe continue to debate how to implement these rules.

Economics has always had much to bring to the debate involving the provision of services that require high fixed costs and result in prices above variable expenses, so economic analysis on net neutrality can build on prior thinking. There are, however, a number of open research questions in this setting because the situation involves multiple participants in complementary economic relationships where they share the costs and benefits of actions, and users benefit from improvement and investment. It should come as no surprise, therefore, that the thrust of the conclusions from economic analysis tilt against simplistic declarations in favor or against net neutrality. This suggests that bold and sweeping recommendations and interventions, given the current state of empirical knowledge, have a substantial chance of being misguided.

■ *None of the authors has any financial stake in matters related to net neutrality policy. Greenstein served on the Open Internet Advisory Committee in 2013, and in 2014–1015 as special economic consultant to the Federal Communications Commission in the Comcast/TWC merger. Peitz has served as an economic expert for the German Federal Ministry of Economic Affairs on net neutrality from 2010 to 2012 and on a regulatory framework for the digital economy since 2015. We thank Jonathan Baker, Jay Pil Choi, Joshua Gans, Jan Krämer, Florian Schuett, and the editors for comments. We especially thank Timothy Taylor for very useful and detailed suggestions. The opinions are our own, and we are responsible for all remaining errors.*

References

Abbate, Janet. 1999. *Inventing the Internet.* MIT Press.

Armstrong, Mark. 2002. "The Theory of Access Pricing and Interconnection." Chap. 8 in *Handbook of Telecommunications Economics,* vol. 1, edited by Martin E. Cave, Sumit K. Majumdar, and Ingo Vogelsang. North Holland: Amsterdam.

Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *Rand Journal of Economics* 37(3): 668–91.

Bourreau, Marc, Frago Kourandi, and Tommaso Valletti. 2015. "Net Neutrality with Competing Internet Platforms." *Journal of Industrial Economics* 63(1): 30–73.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim. 2015a. "Net Neutrality, Network Capacity, and Innovation at the Edges." http://www.law.northwestern.edu/research-faculty/searlecenter/events/internet/documents/Kim_Choi_Jeon_NN-QoS-May%2022-2015-Revision.pdf.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim. 2015b. "Network Neutrality, Business Models, and Internet Interconnection." *American Economic Journal: Microeconomics* 7(3): 104–41.

Choi, Jay Pil, and Byung-Cheol Kim. 2010. "Net Neutrality and Investment Incentives." *Rand Journal of Economics* 41(3): 446–71.

Cisco. 2015. "Cisco Visual Networking Index: Forecast and Methodology, 2014–2019 White Paper." May 27. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html.

Cournot, Antoine Augustin. 1838 [1974]. *Recherches sur les Principes Mathématiques de la Théorie de la Richesse.* Paris: Calmann-Lévy (1974 edition).

Economides, Nicholas, and Benjamin E. Hermalin. 2012. "The Economics of Network Neutrality." *Rand Journal of Economics* 3(4): 602–29.

Economides, Nicholas, and Benjamin E. Hermalin. 2015. "The Strategic Use of Download Limits by a Monopoly Platform." *Rand Journal of Economics* 46(2): 297–327.

Economides, Nicholas, and Joacim Tåg. 2012. "Network Neutrality on the Internet: A Two-Sided Market Analysis." *Information Economics and Policy* 24(2): 91–104.

European Commission. 2011. "Communication from the Commission to the European Parliament, The Council, The Economic and Social Committee, and the Committee of the Regions: The Open Internet and Net Neutrality in Europe." April 19. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52011DC0222.

Federal Communications Commission (FCC). 2008. "Commission Orders Comcast to End Discriminatory Network Management Practices." August 1. https://apps.fcc.gov/edocs_public/attachmatch/DOC-284286A1.pdf.

Federal Communications Commission (FCC). 2005. "In the Matter of Madison River Communications, LLC and affiliated companies." Consent Decree DA-05-543. https://apps.fcc.gov/edocs_public/attachmatch/DA-05-543A2.pdf.

Federal Communications Commission (FCC). 2010. *Preserving the Open Internet Broadband Industry Practices.* FCC 10-201. https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-201A1.pdf.

Frischmann, Brett M. 2012. *Infrastructure: The Social Value of Shared Resources.* Oxford University Press.

Gans, Joshua. 2015. "Weak Versus Strong Net Neutrality." *Journal of Regulatory Economics* 47(2): 183–200.

Genakos, Christos, and Tommaso Valletti. 2011. "Testing the 'Waterbed' Effect in Mobile Telephony." *Journal of the European Economic Association* 9(6): 1114–42.

Genakos, Christos, and Tommaso Valletti. 2015. "Evaluating a Decade of Mobile Termination Rate Regulation." *Economic Journal* 125(586): 31–48.

Greenstein, Shane. 2015. *How the Internet Became Commercial: Innovation, Privatization, and the Birth of a New Network.* Princeton University Press.

Hastings, Reid. 2014. "Internet Tolls and the Case for Strong Net Neutrality." Netflix Media Center, March 20. https://media.netflix.com/en/company-blog/internet-tolls-and-the-case-for-strong-net-neutrality.

Jullien, Bruno, and Wilfried Sand-Zantman. 2016. "Internet Regulation, Two-Sided Pricing, and Sponsored Data." http://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2016/wp_12-327_2016.pdf.

Kourandi, Frago, Jan Krämer, and Tommaso Valletti. 2015. "Net Neutrality, Exclusivity Contracts, and Internet Fragmentation." *Information Systems Research* 26(2): 320–38.

Krämer, Jan, and Lukas Wiewiorra. 2012. "Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment, and Regulation." *Information Systems Research* 23(4): 1303–21.

Lafontaine, Francine, and Margaret Slade. 2007. "Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature* 45(3): 629–85.

Lee, Robin S., and Tim Wu. 2009. "Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality." *Journal of Economics Perspectives* 23(3): 61–76.

Leiner, Barry, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larrwrence G. Roberts, and Stephen Wolff. 2003. *A Brief History of the Internet.* Internet Society, http://www.isoc.org/internet/history/brief.shtml.

Marcus, J. Scott. 2014. *Network Neutrality Revisited: Challenges and Responses in the EU and in the US.* Requested by the European Parliament's Committee on the Internal Market and Consumer Protection. European Parliament. http://www.europarl.europa.eu/RegData/etudes/STUD/2014/518751/IPOL_STU%282014%29518751_EN.pdf.

Marcus, J. Scott, Pieter Nooren, Jonathan Cave, and Kenneth R. Carter. 2011. "Network Neutrality: Challenges and Responses in the EU and in the US." Requested by the European Parliment's Committee on Internal Market and Consumer Protection. European Parliament. http://www.europarl.europa.eu/RegData/etudes/etudes/join/2011/457369/IPOL-IMCO_ET%282011%29457369_EN.pdf.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." *Journal of Economic Theory* 18(2): 301–317.

Nevo, Aviv, John L. Turner, and Jonathan W. Williams. Forthcoming. "Usage-Based Pricing and Demand for Residential Broadband." *Econometrica.*

Nurski, Laura. 2014. "Net Neutrality and Online Innovation: An Empirical Study of the UK." Unpublished paper.

Open Internet Advisory Committee (OIAC), Working Group on Economic Impacts of Open Internet Frameworks. 2013. "Policy Issues in Data Caps and Usage Based Pricing." In *Open Internet Advisory Committee 2013 Annual Report*, by the Open Internet Advisory Committee Federal Communications Commission, pp. 14–38. http://transition.fcc.gov/cgb/oiac/oiac-2013-annual-report.pdf.

Ou, George. 2008. "Managing Broadband Networks: A Policymaker's Guide." The Information Technology and Innovation Foundation, December. Available at: http://www.itif.org/publications/2008/12/11/policymakers-guide-network-management.

Peitz, Martin, and Florian Schuett. 2015. "Net Neutrality and the Inflation of Traffic." TILEC Discussion Paper No. 2015-006, Tilberg Law and Economics Center, Tilburg University.

Rayburn, Dan. 2014. "Here's How the Netflix and Comcast Deal Is Structured, with Data and Numbers." StreamingMediaBlog.com, Feb. 27. http://blog.streamingmedia.com/2014/02/heres-comcast-netflix-deal-structured-numbers.html.

Reggiani, Carlo, and Tommaso Valletti. 2016. "Net Neutrality and Innovation at the Core and at the Edge." *International Journal of Industrial Organization* 45: 16–27.

Sidak, Gregory J. 2006. "A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet." *Journal of Competition Law & Economics* 2(3): 349–474.

Singleton, Micah. 2015. "The FCC Has Changed the Definition of Broadband." *The Verge,* January 29. http://www.theverge.com/2015/1/29/7932653/fcc-changed-definition-broadband-25mbps.

Svensson, Peter. 2007. "Comcast Blocks Some Internet Traffic." *Washington Post,* October 19. Associated Press. http://www.washingtonpost.com/wp-dyn/content/article/2007/10/19/AR2007101900842_pf.html.

van Schewick, Barbara. 2010. *Internet Architecture and Innovation.* MIT Press.

Waldrop, M. Mitchell. 2001. *The Dream Machine: J.C.R. Licklider and the Revolution That Made Computing Personal.* New York: Penguin.

Weyl, E. Glen, and Michal Fabinger. 2013. "Pass-Through as an Economic Tool: Principle of Incidence under Imperfect Competition." *Journal of Political Economy* 121(3): 528–83.

Wu, Tim. 2003. "Network Neutrality, Broadband Discrimination." *Journal of Telecommunications and High Technology Law* 2(1): 141–78.

Yoo, Christopher S. 2005. "Beyond Network Neutrality." *Harvard Journal of Law and Technology* 19(1): 1–77.

Zittrain, Jonathan. 2008. *The Future of the Internet—And How to Stop It.* Yale University Press; New Haven, CT.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 151–178

The Billion Prices Project: Using Online Prices for Measurement and Research

Alberto Cavallo and Roberto Rigobon

New data-gathering techniques, often referred to as "Big Data," have the potential to improve statistics and empirical research in economics. This paper presents one example of how this can be achieved by using the vast number of online prices displayed on the web. We describe our work with the Billion Prices Project at MIT, and emphasize key lessons that can be used for both inflation measurement and some fundamental research questions in macro and international economics. In particular, we show how online prices can be used to construct daily price indexes in multiple countries and to avoid measurement biases that distort evidence of price stickiness and international relative prices.

The basic procedure used in most countries to collect inflation data has remained roughly the same for decades. A large number of people working for national statistical offices visit hundreds of stores on a monthly or bimonthly basis to collect prices for a preselected basket of goods and services. The micro data are then processed and used to construct consumer price indexes and other related indicators. This process is expensive, complex, and often too slow for some users of the data. Infrequent sampling and slow updates to the baskets can complicate

■ *Alberto Cavallo is the Douglas Drane Career Development Professor of Information Technology and Management and an Associate Professor of Applied Economics, and Roberto Rigobon is the Society of Sloan Fellows Professor of Management and a Professor of Applied Economics, both at the Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts. Cavallo is a Faculty Research Fellow and Rigobon is a Research Associate at the National Bureau of Economic Research, Cambridge, Massachusetts. Cavallo is the corresponding author at acavallo@mit.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.151 doi=10.1257/jep.30.2.151

adjustments for quality changes and the introduction of new goods.[1] Groves (2011) further describes other challenges faced by traditional survey-based methods of data collection, including growing levels of nonresponse. Shrinking resources are straining the work of national statistical offices, while recent crises have prompted policymakers and other users of these statistics to demand faster and more accurate data.

Online prices have a natural appeal in this context. While the data are dispersed across hundreds of websites and thousands of webpages, advances in automated "scraping" software now allow anyone to design and implement large-scale data collections on the web. Detailed information can be collected for each good, and new and disappearing products can be quickly detected and accounted for. Online data collection is cheap, fast, and accurate, making it an ideal complement to traditional methods of collecting prices, particularly in categories of goods that are well-represented online.

The first use of online data to construct inflation indexes was motivated by the manipulation of inflation statistics in Argentina from 2007 to 2015. By 2007, it had become apparent that the official level of inflation reported by the national statistical office in Argentina did not reflect the actual changes in prices. Using online data collected every day from the websites of large retailers, Cavallo (2013) showed that while Argentina's government announced an average annual inflation rate of 8 percent from 2007–2011, the online data suggested it was actually over 20 percent, in line with the estimates of some provincial governments and local economists, and consistent with the results from surveys of household inflation expectations. The online price indexes used in that paper were automatically computed and published on a website every day from March 2008 onwards.[2] The ability to collect prices from outside the country proved particularly useful in 2011, when Argentina's government started to impose fines and to pressure local economists to stop collecting data independently. The manipulation of the official price index ended in December 2015 when a new government was elected.

Argentina's statistical debacle had a positive side effect: it showed us the potential that online prices had for inflation measurement applications. With this idea in mind, we created the Billion Prices Project at MIT in 2008 to extend our work to other countries, including the United States. The word "billion" was simply meant to express our desire to collect a massive amount of prices, though we in fact reached that number of observations in less than two years. By 2010, we were collecting 5 million prices every day from over 300 retailers in 50 countries. Half a million prices were collected every day in the United States alone (by comparison, the US Bureau

[1] For discussion of some of these measurement topics, see the "Symposium on Measuring the CPI" in the Winter 1998 issue or the "Symposium on the Consumer Price Index" in the Winter 2003 issue of this journal.

[2] See InflacionVerdadera (http://www.inflacionverdadera.com), which was created to provide alternative price indexes to the official ones in Argentina. The original website had two price indexes constructed with Argentina's official National Institute of Statistics and Censuses (INDEC is the abbreviation of the Spanish translation): a "Basic Food" index and a broader "Food and Beverages" index. The website also showed the time series of prices for every good used in the index.

of Labor Statistics collects approximately 80,000 prices on a monthly or bimonthly basis). Although gathering this massive amount of prices was cheaper online than with traditional methods, it required funding that could not be sustained through grants. Thus, in 2011 we started a company called PriceStats that now collects the data and produces high-frequency indexes for central banks and financial-sector customers. PriceStats greatly expanded both the quantity and quality of the data. The company currently uses about 15 million products from over 900 retailers to build daily inflation indexes in 20 countries. Its micro datasets contain information from an even larger number of retailers in over 60 countries, with varying degrees of coverage. The indexes and micro data from PriceStats are available to researchers working with the Billion Prices Project, as we explain later in this paper.

Many of the other attempts to use big data in economics rely on social media or search data to *forecast* the behavior of important economic indicators. Our approach is different because we focus on *measurement*, not on prediction. Our objective is to experiment with these new sources of information to improve the computation of traditional economic indicators, starting with the Consumer Price Index. We seek to understand whether online prices have distinct dynamics, their advantages and disadvantages, and whether they could be a reliable source of information in a "production" setting (not just for a one-time research application).

We start this paper with a description of the methodology used to collect online prices. A first-order aspect is to realize that although the amount of data online is massive, carefully selecting the categories and retailers to sample is still crucial. The goal is to obtain data that is representative of retail transactions, so we focus our data collection efforts on large multichannel retailers such as Walmart that sell both online and offline, instead of using online-only retailers that may have many products but a relatively small share of retail transactions. We also focus on categories of goods that are included in the official consumer price index baskets, for which consumer expenditure weights are available. After describing the sources of data, we discuss the advantages and disadvantages of online data relative to other large micro price databases (including scanner data and official price-index data), and highlight the results of a large-scale validation exercise to show how online-price levels and behaviors closely resemble those that can be obtained by physically visiting offline stores.

Next, we describe the methodology used to compute online price indexes and show how they co-move with consumer price indexes in most countries. We emphasize two characteristics in greater detail. First, online indexes have the ability to approximate hedonically adjusted price indexes in sectors with a large number of goods that come and go with overlapping life-cycles (as for instance, in electronics). Second, online indexes appear able to anticipate movements in the official consumer price index in many countries. This anticipation extends beyond the publication lags, which suggests that online prices often adjust sooner to aggregate shocks.

We then move on to research applications and discuss two areas in macro and international economics where online price data can have a major impact. First, we show that online price data, collected daily, can significantly alter some key results in the price-stickiness literature. In particular, we document that online prices exhibit a very different distribution of price changes compared to prices collected for official consumer price indexes and by scanner prices. The main reason for the difference is that online prices do not have time averages, common in scanner data, or imputed prices, common in official micro data, which create a large number of small spurious price changes. Second, using online data to test the "law of one price" (that there should not be large or persistent cross-country differences in the prices of identical goods when translated into a common currency) gives us a more nuanced picture of when and where this law works well. The existing consensus in the literature is that there are large and persistent deviations from the law of one price, with little pass-through from nominal exchange rates to relative prices, and vice-versa, causing persistent shocks to real-exchange rates that take years to dissipate. While deviations can also be large with online data, we find that the law of one price holds well across countries that use the same currency. We also show that, when goods are identically matched across countries, then relative prices and nominal exchange rates co-move more closely than previously thought. This implies higher pass-through rates and less persistent real-exchange rate dynamics.

Both research examples illustrate how using data collected by others, with different purposes in mind, can distort empirical findings. They also suggest we should not treat big data as simply a collection of large datasets created as a byproduct of something else. One of the greatest opportunities of big data is that anyone can now use new technologies such as web-scraping, mobile phones, satellite imaging, and all kinds of interconnected sensors to build customized datasets designed to fit specific measurement or research needs. We end the paper by describing how the Billion Prices Project data are publicly shared and by discussing why data collection is an important endeavor that macro- and international economists should pursue more often.

Collecting and Processing Online Price Data

A large and growing share of retail prices all over the world are posted online on the websites of retailers. This is a massive untapped source of retail price information. Collecting these prices is not trivial because they are posted on hundreds of different websites that lack a homogeneous structure and format. And retailers do not provide historical prices, so the data has to be collected continuously and consistently over time.

To collect and process online prices we follow a "data curation" approach. It involves carefully identifying the retailers that will serve as data sources; using web-scraping software to collect the data; then cleaning, homogenizing, categorizing, and finally extracting the information so it can be used in measurement and research applications.

The Selection of Retailers and Data Source

The starting point is to select the retailers and categories of goods to sample. These decisions are driven by our need to get prices that are representative of retail transactions. We therefore focus almost exclusively on large multichannel retailers (those retailers that sell both offline and online, such as Walmart) and tend to ignore online-only retailers (such as Amazon.com). The reason is that multichannel retailers still are involved in the majority of all retail sales in most countries. We are also careful when we choose what categories of goods to monitor within each retailer, concentrating on those categories that are part of traditional consumer price index baskets, and avoiding categories that are overrepresented online such as CDs, DVDs, cosmetics, and books.

We make an effort to collect the data directly from each retailer's website, rather than relying on third parties such as marketplaces, price aggregators, and price comparison websites. Data collection from individual retailers is far more challenging, but it maximizes our chances of obtaining prices linked to actual transactions and prevents third-parties from filtering or altering our samples. It also gives us full control of what we choose to collect and makes the whole process more robust, as it does not depend on a few sources of data.

Once the data are collected, we clean them, standardize them to fit a common database schema, classify individual products using consumer price index categories, and start computing simple indicators to evaluate their characteristics and performance over time.

We treat each retailer as a separate sampling unit or "stratum" with potentially unique characteristics and pricing behaviors. Before including a retailer in a price index, we usually monitor its behavior for over a year to identify any special characteristics in the data so that we can know whether it is a useful and reliable source of price information.

Most retailers that sell online have a single price for all shoppers in all locations within a country (though shipping costs and taxes may differ). Grocery retailers can sometimes show different prices for the same good depending on the zip code entered by the consumer. In such cases, we select a few zip codes corresponding to major cities and treat each case as an independent retailer.

The amount of data and the coverage of different categories that we can observe online vary across countries. For about 25 countries, our datasets have information on categories that cover at least 70 percent of the weights in consumer price index baskets.

Data Collection Using Web Scraping Software

After selecting the sources of data, the next step is to collect the information. The technology to collect online prices on a large scale—called "web scraping"—is quickly improving. Just a few years ago, it required researchers to write programs in languages such as Python and PHP (for an example, see the discussion in this journal by Edelman 2012). Today, there are many "point-and-click" software solutions that require almost no technical expertise. Users can simply use their mouse to teach the software what

Table 1
Alternative Micro-Price Data Sources

	Online data	*Scanner data*	*CPI data*
Cost per observation	Low	Medium	High
Data frequency	Daily	Weekly	Monthly
All products in retailer (Census)	Yes	No	No
Uncensored price spells	Yes	Yes	No
Countries with research data	~60	<10	~20
Comparable across countries	Yes	Limited	Limited
Real-time availability	Yes	No	No
Product categories covered	Few	Few	Many
Retailers covered	Few	Few	Many
Quantities or expenditure weights	No	Yes	Yes

Source: Table 1 from Cavallo (2015).
Notes: The Billion Prices Project (bpp.mit.edu) datasets contain information from over 60 countries with varying degrees of sector coverage. Nielsen US scanner datasets are available at the Kilts Center for Marketing at the University of Chicago. Klenow and Malin (2010) provide stickiness results with Consumer Price Index data sources from 27 papers in 23 countries. See Cavallo (2013) for more details.

pieces of information they want to collect from a webpage. The software then creates a "robot" that is able to extract information from any other webpage with a similar structure, storing the information in a database. It identifies relevant pieces of information on a page by finding special characters of HTML code (the language that is used to create webpages) that come before and after each relevant piece of information. These characters are relatively steady as long as the page does not change its look-and-feel. The challenge in web scraping is mostly to monitor the performance of the robots over time so any errors in the data can be quickly detected and fixed. The robots we construct always collect a product identification number, the name, description, brand, package size, category information, and the price. When available, we also collect other variables such as sale prices and stock indicators. We provide more details of the web-scraping process in the online Appendix available with this paper at http://e-jep.org.

Advantages and Disadvantages of Online Price Data

To understand the strengths and weaknesses of this scraped online data for measurement and research applications, Table 1 offers a comparison with two other sources of micro price data: traditional consumer price index data collected offline by national statistical organizations, and scanner data recorded from consumer purchases at point-of-sale terminals by companies such as Nielsen. Detailed descriptions of these other data sources can be found in ILO et al. (2004) and Feenstra and Shapiro (2003).

One of the most obvious advantages of online data is the low cost per observation. While the cost is not trivial, it is far cheaper to use web scraping than hire people to visit physical stores or buy information from commercial scanner data providers such as Nielsen.

A second major advantage is the daily frequency of data collection. It is easier to detect errors in the data when it is collected at such high frequency. This approach also avoids a need to use time averages, which can generate spurious price changes as we discuss later on.

Third, online data includes detailed information for all products being sold by the sampled retailers. The cross-section of prices available is therefore much larger within categories than in consumer price index data. Later, we discuss how this big data feature can be used to simplify quality adjustments and other traditional measurement problems.

Fourth, there are no censored price spells in online data. Prices are recorded from the first day a product is offered to consumers until the day it is discontinued from the store. Traditional data collection methods, in contrast, will typically start monitoring new goods only when the goods in the basket disappear from the stores. Knowing the full history of prices for individual goods can help to control for new-good biases, make both implicit and explicit quality adjustments, and study prices at time of product introductions.

Fifth, online data can be collected remotely. This is particularly useful in situations like the one experienced by Argentina in recent years, where the government was trying to prevent independent data collection for the computation of inflation. It also allows us to centralize the data collection and homogenize its characteristics.

Sixth, and related to the previous point, online datasets can be readily comparable across countries because prices can be collected with identical methods on matching categories of goods and time periods. This is useful in research applications that use cross-country comparisons.

Finally, online data are available in real time, without any delays to access and process the information. This is particularly useful for policymakers and anyone who needs up-to-date information.

One of the main disadvantages of online prices is that they currently cover a much smaller set of retailers and product categories than a government-run survey of consumer prices do. In particular, the prices of most services are still not available on the web, and the number and type of retailers is limited compared to official consumer price index data.

Another disadvantage is that online datasets lack information on quantities sold. Online prices must be combined with weights from official consumer expenditure surveys or other sources for expenditure-weighted applications. Scanner datasets, by contrast, have detailed information on quantities sold, and could potentially be a source of high-frequency expenditure weights in some categories of goods such as groceries.

Are Online Prices Different?

An important concern is whether online prices are different from offline prices; after all, most transactions still take place offline. The suspicion that online prices are different is fueled by reports that some online retailers use "dynamic pricing" strategies in which prices are varied for strategic purposes: for examples, see Mikians, Gyarmati, Erramilli, and Laoutaris (2012) and Valentino-DeVries, Singer-Vine, and Soltani (2012). In addition, many papers with "online prices" use data from online marketplaces such as Ebay or price-comparison websites such as Google Shopping. As Brynjolfsson and Smith (2000), Ellison and Ellison (2009), and Gorodnichenko, Sheremivov, and Talavera (2014) have shown, these prices seem to change more frequently and in smaller sizes than in Consumer Price Index data. However, the retailers in these datasets are mostly online-only stores participating in a fiercely competitive environment, not really the type of "online data" we use.

To better understand whether online and offline prices for multichannel retailers behave differently, Cavallo (2016) simultaneously collected prices on the websites and physical stores for over 24,000 products in 56 of the largest retailers in 10 countries. This large-scale comparison was possible thanks to the combination of a smartphone app, crowdsourced workers, and web-scraping techniques. More than 370 freelance workers used their phones to scan barcodes in physical stores, manually enter prices, take photos of the price tags, and upload the information to our Billion Prices Project servers. We then used the barcodes in the offline data to collect the prices for those exact same goods at the website of the same retailer within a seven-day time window.

This direct comparison between online and offline prices revealed a high degree of similarity in price levels, as well as in both the frequency and size of price changes. On average, about 70 percent of price levels were identical in the offline and online samples. The similarity was highest in retailers that sell electronics or apparel, and lowest in drugstores and office-supply retailers that also tend to price differently across offline stores. While price changes do not have the exact same timing online and offline, they tend to have similar frequency and average sizes. This suggests that the price spells for individual goods may not be synchronized online and offline, consistent with evidence to be discussed below that online prices may anticipate later price changes. Despite the general similarity between online and offline pricing, our results also revealed a great deal of heterogeneity among pricing behaviors, suggesting some validation is needed in papers with data from a limited number of retailers.

Inflation Measurement

Online prices are increasingly being used in inflation measurement applications. Besides the Billion Prices Project and PriceStats, many national statistical organizations are experimenting with the use of online data, including the US Bureau of Labor Statistics (Horrigan 2013a), the UK Office of National Statistics (Breton et al. 2015), Statistics Netherlands (Griffioen, de Haan, Willenborg 2014), Statistics New Zealand (Krsinich 2015), and Statistics Norway (Nygaard 2015).

In this section, we show that online price indexes can closely approximate the official consumer price index in a number of countries and settings. We then discuss how a large number of overlapping price series in the data can simplify quality adjustments in categories with frequent product turnover, such as electronics. Finally, we show that online price indexes can anticipate changes in the official inflation rate several months in advance.

Methodology for Comparison to Official Consumer Price Indexes

For multiple Latin American countries, Cavallo (2013) showed that online prices could be effectively used as an alternative source of price information to construct price indexes that mimic the behavior of official consumer price indexes. The methodology for these daily indexes was based on a combination of online data with standard techniques used in official price indexes, including expenditure weights for each sector where online data are available. This initial work included only data from food retailers and a handful of countries. In 2010, we founded PriceStats to expand the data collection and to compute inflation measures in real time in other sectors and countries. The company is currently publishing daily price indexes in 22 countries with only a three-day lag. In Figures 1 to 4, we plot these online indexes next to the all-item nonseasonally adjusted consumer price index in each country. We first highlight the cases of Argentina and the United States, and then show some selected cases in a larger set of countries.

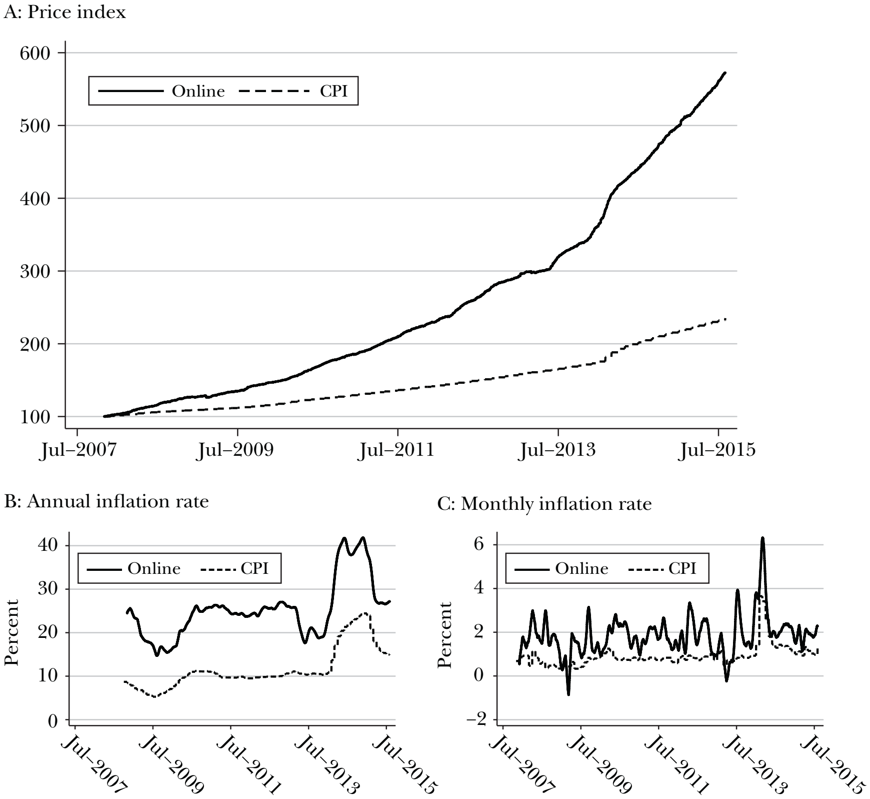
Figure 1 illustrates the case of Argentina from 2007 to 2015. Figure 1A compares a price index produced with online data to the official consumer price index.

The fact that the two measures of inflation in Figure 1A diverge so dramatically will not surprise anyone who knows the recent story of statistics in Argentina. In February 2007, the government intervened in the National Institute of Statistics and Census (INDEC) and fired the people responsible for computing the consumer price index. The index quickly stabilized, but many local economists claimed the government was manipulating the data. Household inflation expectations increased dramatically, closely tracking some alternative estimates of inflation produced by local economists and some independent provincial goverments, as shown in Cavallo, Cruces, and Perez-Truglia (2016). Suspicions were abundant, but before a measure of inflation based on online prices became available, there was no consistent way to confirm the magnitude of the discrepancy and track its evolution over time.

The manipulation in the official inflation data continued for almost nine years, ending in December 2015 when a new government was elected. During all this time, the monthly inflation rate shown in Figure 1C was consistently higher than the official data reported, with the exception of a few months in 2014 when, in response to a "motion of censure" issued by the IMF in 2013 (Rastello and Katz 2013), the Argentinian government decided to launch a new consumer price index. Unfortunately the change was temporary and the new official index quickly lost all credibility again.

Looking only at the discrepancy in the trend of the price index or the monthly inflation rates, however, misses an important point. The online index tracked the dynamic behavior of the annual inflation rate over time, as shown in Figure 1B. The

Figure 1
Argentina

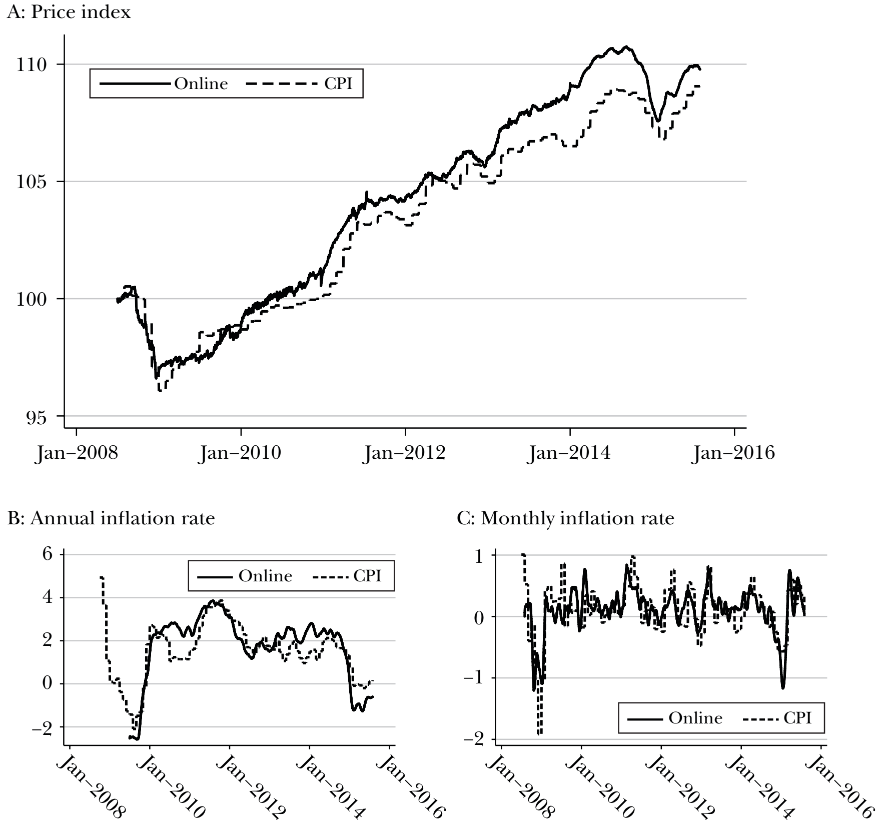


Source: Authors using online price index computed by PriceStats and the consumer price index from the national statistical office in Argentina (INDEC).
Notes: The figure compares a price index produced with online data to a comparable official consumer price index (CPI) for the case of Argentina from 2007 to 2015. It also looks at annual and monthly inflation rates using each source of data. Monthly inflation rates for the online index are computed as the percentage change in the average of the previous 30 days compared to the same average a month before. Annual inflation rates for the online index are computed as the percentage change in the average of the previous 30 days compared to the same average 365 days before. All price indexes are nonseasonally adjusted.

difference was mostly in the *level* of the annual inflation rate, not its movements over time. The online index also quickly reacted to aggregate shocks, such as the massive roadblocks by farmers who protested export tax hikes in 2008. This strongly suggested that online data was capable of capturing the fundamental dynamics of inflation and prompted us to collect data in other countries.[3]

[3] It also implied that the government was not using a particularly sophisticated algorithm to change the inflation rate. In Cavallo (2013), we showed that one could closely approximate the official index by simply dividing the online inflation rate by three.

Figure 2
United States



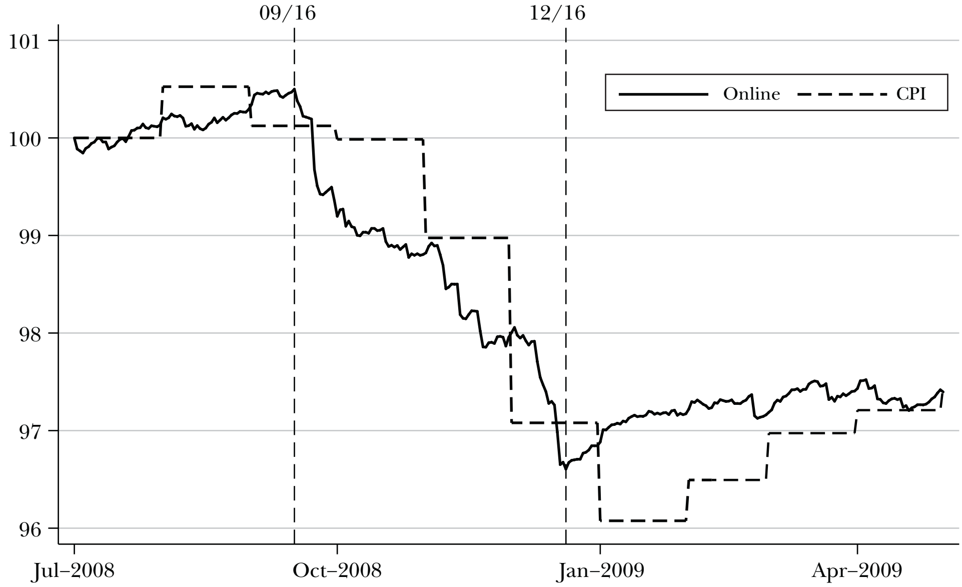
Source: Authors using online price index computed by PriceStats and a comparable consumer price index (CPI) from the US Bureau of Labor Statistics.
Notes: The figure compares a price index produced with online data to the US CPI for the United States January 2008 to January 2016. The monthly inflation rates for the online index are computed as the percentage change in the average of the previous 30 days compared to the same average a month before. Annual inflation rates for the online index are computed as the percentage change in the average of the previous 30 days compared to the same average 365 days before. All price indexes are non–seasonally adjusted.

The comparison of online and offline indices in other countries is completely different. The daily US index, shown in Figure 2, is a great example.

Despite the multiple reasons why we might expect inflation indexes based on online and offline prices to deviate, the US online index has co-moved closely with the official Consumer Price Index for over seven years. Although there are periods where the indexes diverge, the differences are relatively small and temporary. This can also be seen in the monthly and annual inflation rates in Figures 2B and 2C.

The US online index is particularly good at anticipating major changes in inflation trends. Predicting these changes is important for participants in financial

Figure 3
US Consumer Price Index around the Bankruptcy of Lehman Brothers



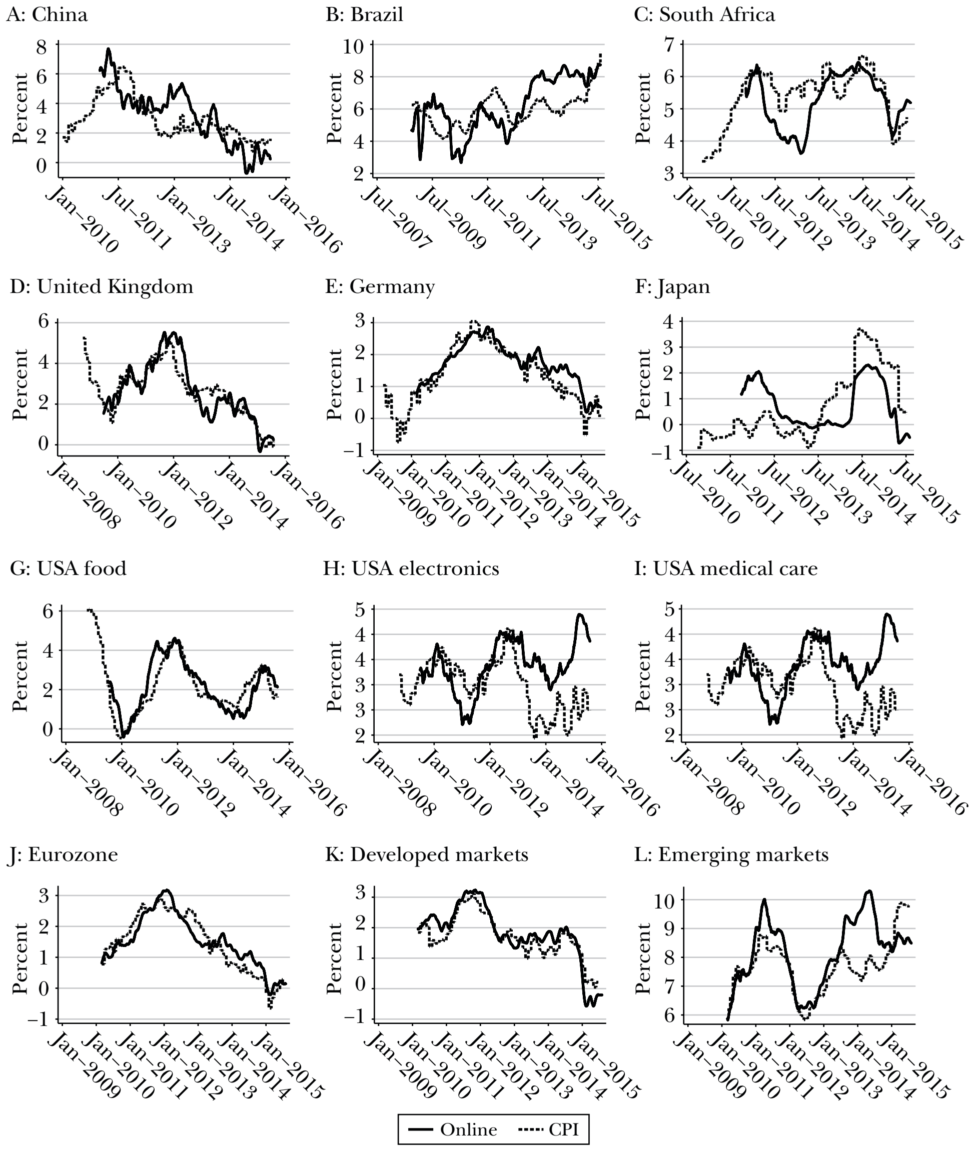
Source: Authors using online price index computed by PriceStats and the Consumer Price Index from the US Bureau of Labor Statistics.
Note: The figure highlights the events around the bankruptcy of Lehman Brothers, the fourth-largest investment bank in the United States, during September 2008.

markets, policymakers, and those economists who monitor the economy closely. One remarkable example of a turning point detected with online data months before it showed up in official US Consumer Price Index data was September 16, 2008, the Tuesday after Lehman Brothers filed for bankruptcy. As Figure 3 shows, the online price index peaked that day and started falling. By October 15th, it had lost almost 1.2 percent in a single month. On October 16th, the Consumer Price Index for September came out with only a 0.14 percent drop. When the official October Consumer Price Index numbers were published on November 19th, it had fallen another 1.01 percent. In other words, it took more than two months after Lehman's disaster for the official Consumer Price Index numbers to reflect the full impact on price levels. Two months later, on December 16, 2008, the online price index stopped falling and started to increase once again. The Consumer Price Index did not show this change in the trend until the estimates for January were published on February 20, 2009. We measure the degree of anticipation in online data more formally later on.

Figure 4 compares inflation as measured by online prices and by the offline prices in the official consumer price index for a selection of other countries and sectors. The main lesson of the figure is that the correspondence is reasonably close,

Figure 4

Online versus Consumer Price Index (CPI) Annual Inflation Rates



A: China

B: Brazil

C: South Africa

D: United Kingdom

E: Germany

F: Japan

G: USA food

H: USA electronics

I: USA medical care

J: Eurozone

K: Developed markets

L: Emerging markets

Source: Authors using online price indexes computed by PriceStats and consumer price indexes sourced from the national statistical office in each country.

Notes: Figure 4 compares inflation as measured by online prices and by the offline prices in the official consumer price index for a selection of countries, sectors, and regions. Annual inflation rates for daily online price indexes are computed as the percentage change in the average of the previous 30 days compared to the same average 365 days before. The series are nonseasonally adjusted. Indexes are "all-items" with the exception of China, where an online supermarket index is shown next to the official food index. Global aggregates in the last row are computed using 2010 consumption weights in each country and CPIs from official sources.

but some more specific insights are also possible. First, we do not find evidence that China has been systematically holding its official inflation rate below the rate based on online prices, though we can only compare some sectoral indexes because official expenditure weights are not publicly disclosed. Second, the difference between the online price index and the official consumer price index appears to be smaller in developed countries like the UK and Germany, and greater for countries like Brazil or South Africa, where the online sector seems to have more independent patterns. In Japan, we observed significantly more inflation after the March 2011 earthquake and an immediate impact of the sales tax changes in April 2014. While the online index in Japan does not follow its official consumer price index closely, is does seem to anticipate key changes in inflation trends.

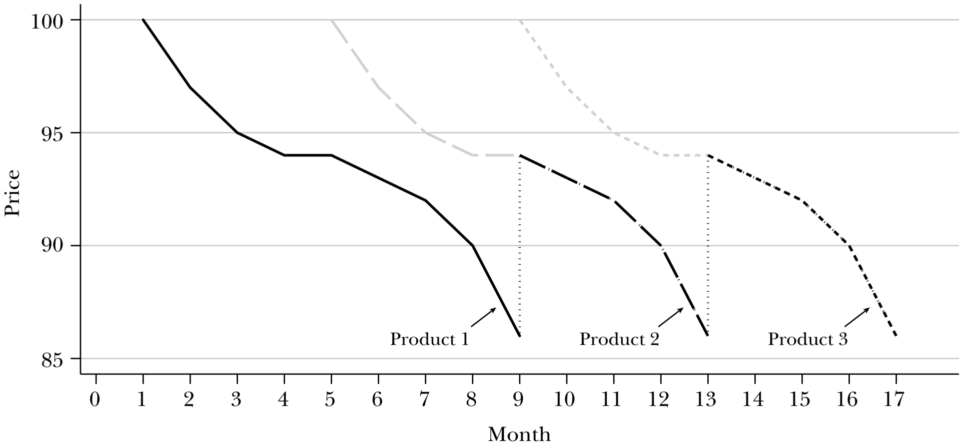
The third row of Figure 4 shows results for a few US sectors. As one might expect, the online data matches the US Consumer Price Index better in sectors such as food and electronics, for which online information is widely available. By contrast, some official inflation patterns seen in the medical care sector are not well-captured by online prices, mostly because many services cannot be monitored online. The fourth row shows that online data can be used to provide global aggregates using country consumption weights.

It may seem surprising to some readers that indexes based on online data have the ability to mimic official consumer price indexes in so many cases: for large and small countries, for developed and emerging markets, and for the aggregate and sectoral data. After all, the data differs significantly from traditional sources, and we do not apply many adjustments and methods used by national statistical organizations, such as hedonic quality adjustments. We believe there are two reasons for this closer-than-expected correspondence. First, as mentioned before, we carefully design and select the data that goes into these indexes to ensure that they are representative. Second, we learned that many sampling characteristics in our data made it *simpler* to deal with some traditional measurement problems. To illustrate this point, we next discuss how online data can simplify quality adjustments by providing a large number of uncensored and overlapping price spells.

Overlapping Quality Adjustments

Quality adjustment poses a problem for any measure of inflation: as is widely understood, if a good rises in both quality and price, then some of the price increase is presumably due to the quality changes and should not be attributed to inflation. National statistical organizations use different methods for quality adjustments, including seeking the closest comparable substitutes when a product disappears and often relying on adjustments with hedonic regressions (in which the price change is calculated while holding constant certain attributes of a good, like the memory or hard drive capabilities of a computer). Online datasets make it easier to deal with quality adjustments because they provide *uncensored* price spells for a large number of models and varieties of each good. With better underlying data, online price indexes can approximate the results of more sophisticated, and often impractical, hedonic-regression methods.

Figure 5
Traditional Consumer Price Index (CPI) Data with Censored Prices, and the Problem of New Products



Source: Authors.
Notes: This Figure shows hypothetical prices series for three goods. Each downward-sloping line reflects the prices of a different good. The shaded portions of these lines illustrate the data points that would be missed with traditional data collection methods.

To build some intuition for why this result holds true, consider a hypothetical example of a series of prices in Figure 5. It illustrates the data resulting from a traditional offline data collection process. Each line represents the price of a single good over time. Many models of electronic products, such as televisions, dishwashers, washing machines, and vacuum cleaners, tend to be introduced at relative high prices and then are discounted gradually over their life-cycle, with clearance sales occurring right before the product disappears from the stores (Silver and Heravi 2005).

With traditional data collection methods, it is too expensive to collect the prices for every good available for sale at each point in time within a sampled retailer. Instead, the data collector focuses on one (or a few) of the most popular models and records its price once per month until it disappears from the store. When a particular model is no longer available, the data collector starts to sample a different model, as shown by the vertical dashed line in the figure. But at the time of the shift, the previous prices for the new model are unknown (shown where the line is shaded more lightly on the figure). The problem is to decide how much of the price gap at that point in time is attributable to quality differences. This issue is exacerbated in goods that experience extreme price movements along their life-cycles and may have steep discounts right before disappearing from the shelves.

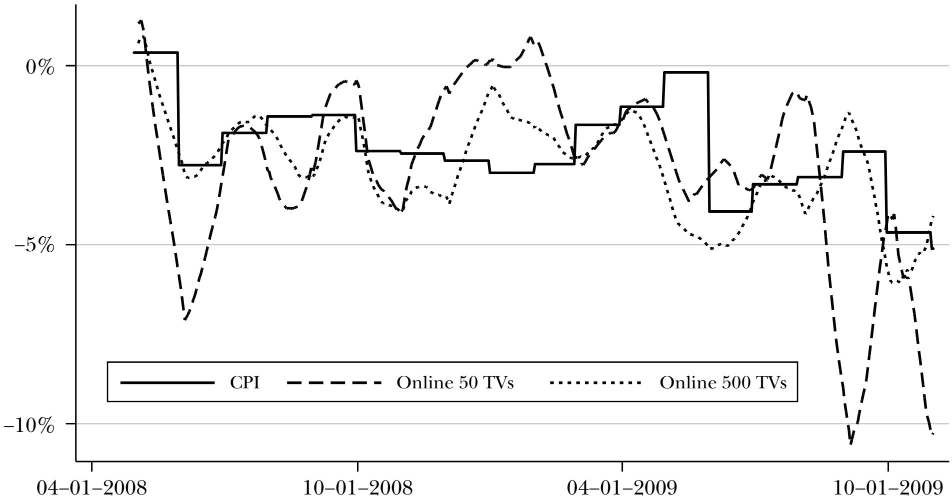
National statistical organizations have two main ways of dealing with this problem. One preferred method is to use hedonic techniques. Again, these involve setting up a regression with the price of a good on one side and actual attributes of the good on the other side, so that future changes in the price of the good can be calculated while holding constant the attributes. While hedonic techniques have become popular in recent decades, the question of what traits should be included in the regression, how they can be measured, and what specification should be used can make hedonic techniques too data intensive and complex to implement in practice.

A simpler alternative method is to use "overlapping qualities." As Armknecht and Weyback (1989) point out, if two goods coexist for some time, their overlapping prices can be used to obtain an estimate of quality change. In practice, this approach tends to assume that the price gap at the time of introduction of the new variety mostly reflects a quality difference. The problem with traditional data, however, is that the price of the new good is not observed at the time of introduction, but much later, when the old good disappears from the stores. This is noted in the *Consumer Price Index Manual* of the International Labor Organization (ILO et al. 2004, p. 27–28): "When there is overlap, simple linking … may provide an acceptable solution … In practice, however, this method is not used very extensively because the requisite data are seldom available. … [T]he information needed for this … will never be available if price collectors are instructed only to introduce a new quality when an old one is dropped."

Online prices offer a simple solution to this data problem by providing a large number of uncensored price spells for all models on sale at any point in time. With this type of data, a simple index using overlapping qualities can closely approximate official indexes that use complex hedonic quality-adjustment methods. Similar results were documented earlier in the price-index literature using scanner data. For example, Aizcorbe, Corrado, and Doms (2000, 2003) used scanner prices to demonstrate that, with high-frequency data, matched-model price indexes could yield results that are numerically close to those obtained using hedonic techniques in samples where product characteristics did not change much over time. More generally, the extent to which a simple matched-model price index can capture quality change will depend on several factors (Silver and Heravi 2005). First, both varieties of the product need to have a substantial degree of overlap in their prices. Second, there needs to be a large number of models so that continuing varieties can capture aggregate effects without being overly affected by idiosyncratic price movements of goods that enter and disappear from the sample.

As evidence of this effect, consider the data in Figure 6. It contains three price indexes for televisions in the United States from 2008 and 2009. The solid line shows the official Consumer Price Index for televisions as computed by the US Bureau of Labor Statistics using hedonic methods. The line with long dashes shows an online price index based on 50 distinct models of televisions from a large US retailer. The line with short dashes shows an online price index with 500 models from the same source. As we increase the number of models included in the index, we more closely approximate the results of the hedonic price index constructed by the US Bureau of Labor Statistics during this time period. Intuitively, the more

Figure 6
Hedonic Consumer Price Index (CPI) versus Online Index for US Televisions
(monthly inflation rate)



Source: Authors and the US Bureau of Labor Statistics.
Notes: The solid line shows the official Consumer Price Index for televisions as computed by the US Bureau of Labor Statistics using hedonic methods. The line with long dashes shows an online price index based on 50 distinct models of television from a large US retailer. The line with short dashes shows an online price index with 500 models from the same source.

overlapping price series being used, the less important the extreme price movements of goods being sold at clearance prices or newly introduced will be for our price index.

This example illustrates one of the *size* advantages of online datasets. We may not need or want to use every single data point available in these data, but being able to extract and use uncensored spells for a large number of models can greatly simplify measurements. Even if the goal is to run a hedonic regression, online data can supply the detailed information needed to make it practical. And with more data, simpler methods can be applied. For example, Krsinich (2015) showed that online data can be used to construct a time-product dummy index that is equivalent to a fully interacted time dummy hedonic index based on all product characteristics.

Anticipation of Future Changes in the Consumer Price Index

As mentioned before, online price indexes can sometimes anticipate changes in official inflation. In this section, we document this pattern more formally and conjecture about some possible explanations.

To document the degree of anticipation, we estimate a simple autoregression equation with the US Consumer Price Index as the dependent variable and our online price index as the exogenous variable, and compute an impulse response to see how shocks to the online index impact the official price index over time. The regression is expressed in monthly changes: specifically, we use monthly log changes in the Consumer Price Index and monthly log changes of the online index on the last day each month. We include six lags of each variable, plus the contemporaneous value of the online price index to account for the early availability of the online price information.[4]

Figure 7 shows the cumulative impulse response of the Consumer Price Index to a shock in the online index over time, together with the 95 percent confidence intervals. In the United States, it takes several months for the Consumer Price Index to fully incorporate the shock to the online price inflation. At the sector level, the impact is quickest in fuel (transportation) and slowest in food and electronics (see Appendix for details). The result is robust to the elimination of the contemporaneous effect of the official price index from the vector autoregression. In most cases, the anticipation significantly exceeds the typical publication delays in official statistics. Moreover, we find similar degrees of anticipation in other countries.

Possible reasons for why online prices can anticipate shocks in the consumer price index include delays embedded in the methodology used for the official data, differences in mixture of stores sampled, and faster adjustment of online prices in some sectors or retailers. Understanding what drives the anticipation is sill an open question for future research, but the patterns in Figure 7 suggest that online data can be a useful addition to inflation forecasting models. This is explored by Aparicio and Bertolotto (2016), who show that out-of-sample inflation forecasts using online data can outperform a large number of alternative forecasting models in the US and UK economies.

Lessons for Macro and International Research

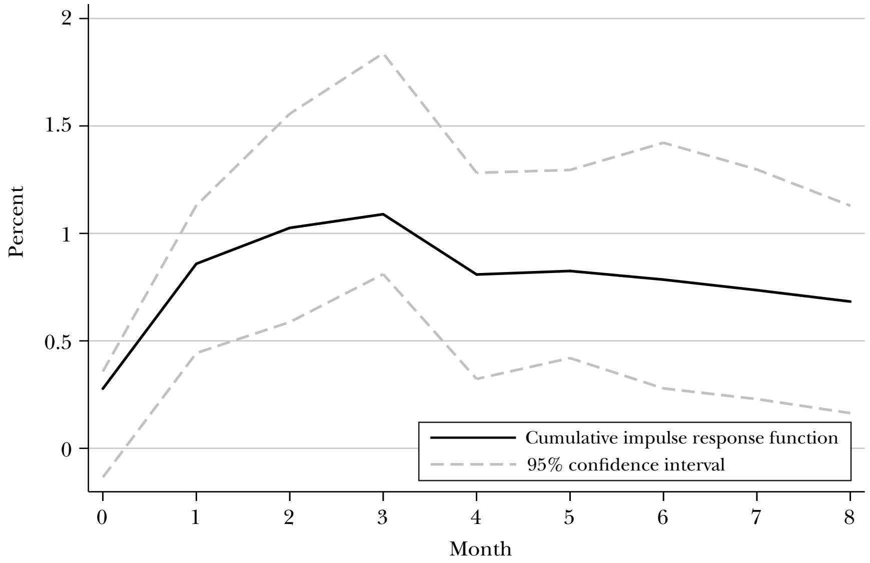
In this section, we use questions of price-stickiness and real exchange rate behaviors to illustrate how online data can change empirical results in macro and international research. Our main objective is to show how online datasets constructed to fit specific research needs can help mitigate biases and other empirical challenges that are so frequent in traditional datasets collected for other purposes.

[4] For each month t, the specification is as follows:

$$\Delta \ln(CPI_t) = \alpha + \beta \Delta \ln(Online_t) + \textstyle\sum_{i \in [1,6]} \alpha_i \, \Delta \ln(CPI_{t-i}) + \textstyle\sum_{i \in [1,6]} \beta_i \, \Delta \ln(Online_{t-\mathrm{i}})$$

The autoregressive distributed lag (ADL) specification is equivalent to a vector auto regression (VAR) with an exclusion restriction. The confidence bands are computed by bootstrapping in blocks. This specification gives the online price index the highest chance to explain the observed variation. There is, however, no unambiguous way of identifying the system given that under the null hypothesis both indexes are valid measures of the underlying inflation. We chose this specification because it matches the actual availability of data at the end of each month: the online index is immediately available, while the CPI has a publication lag of 15 days in most countries. The results are robust to the elimination of the contemporaneous effect of the online price index from the equation.

Figure 7
Cumulative Impulse Response of the US Consumer Price Index (CPI) to an Online Price Index Shock
(response to a 1% shock in the online index)



Source: Authors using online data computed by PriceStates and US Consumer Price Index.
Notes: The Consumer Price Index is a US city average, all items non–seasonally adjusted from the Bureau of Labor Statistics. Data from July 2008 to January 2015.

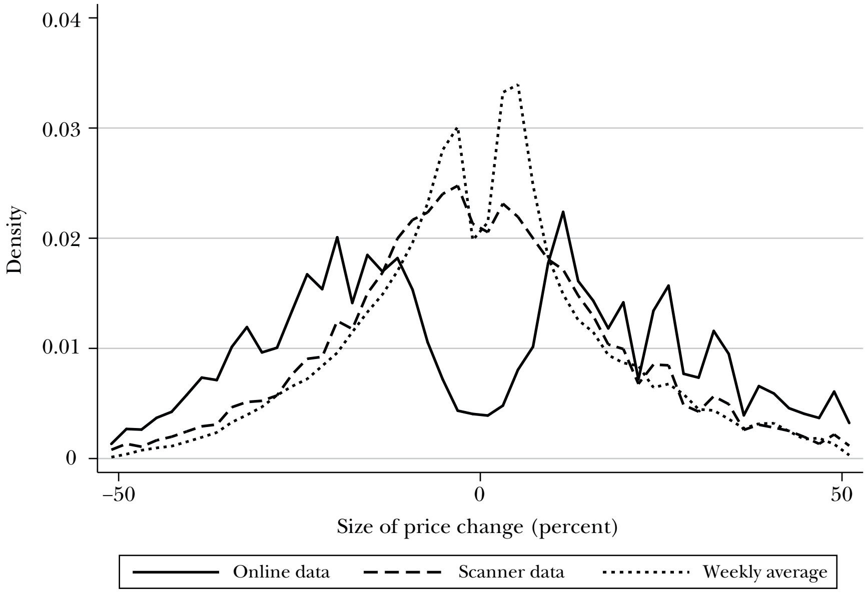
Price Stickiness and the Distribution of Price Changes

Sticky prices are a fundamental element of many macroeconomic models. In the past decade, a large empirical literature has tried to measure price stickiness and understand its microfoundations (for an example in this journal, see Dhyne et al. 2006; for a survey of the literature, see Nakamura and Steinsson 2013, and the references cited there). This research has been possible due to an unprecedented access to micro-level consumer price index data and scanner datasets in several countries. Over time, the literature has settled on a set of stylized facts, summarized by Klenow and Malin (2010). In Cavallo and Rigobon (2011) and Cavallo (2015), we use online data to argue that the sampling characteristics of official consumer price index and scanner data can introduce measurement biases that affect the stylized facts in the literature on patterns of price changes.

As one prominent example, a pattern that has received a lot of attention in the literature is the shape of the distribution of the size of price changes. Most papers using scanner or consumer price index data found bell-shaped (unimodal) distributions centered around zero percent, with a significant share of small price changes, which seemed inconsistent with standard menu-cost models that predict periods

Figure 8

The Distribution of the Size of Price Changes in the United States



Source: Cavallo (2015).
Notes: Online and scanner data was collected from the same retailer, zip code, and time period. Weekly averages were computed using the daily online data. Nielsen Scanner Data provided by the Kilts Marketing Center at Chicago Booth.

of unchanging prices followed by relatively large changes (a bi-modal distribution centered around zero). This finding motivated a surge in papers trying to adapt sticky-price models to account for this fact (for example, Woodford 2009; Midrigan 2011).

However, the shape of the distribution of price changes is greatly affected by the sampling characteristics of the data. This can be seen in Figure 8, where we show a distribution of prices changes for both online and scanner data obtained from exactly the same US retailer, zip code, and time period. While the raw prices that generate these distributions are in principle the same, the results are strikingly different. The online data distribution is strongly bimodal, with very few price changes close to zero percent. There is a simple explanation for the difference. Scanner data are reported as weekly averages. As noted by Campbell and Eden (2014), this can create a large number of spurious small changes. For example, in a three-week period with a single price change in the middle of the second week, taking weekly averages would yield two small price changes: one from the first week to the second, and another from the second week to the third. These spurious changes can be seen explicitly in Figure 8, where we approximate the shape of the scanner data distribution by simply taking weekly averages of the raw online data.

Something similar happens with Consumer Price Index data, although the source of measurement bias is different in nature, as discussed in Cavallo (2015).

In particular, micro data from the Consumer Price Index will often contain imputed prices for temporarily missing items, which is a sensible thing to do when measuring inflation. This imputation is often done with the average price change of related goods, resulting in an artificial pattern of many small changes. If these imputations are not identified or removed when generating the distributions, the result is a unimodal-shaped distribution similar to those found in the literature. Furthermore, other forms of measurement biases can have a similar impact. For example, Eichenbaum, Jaimovich, Rebelo, and Smith (2014) use Consumer Price Index and scanner data from multiple stores to show how "unit-value prices," which are reported as the ratio of sales revenue to the quantity sold, also create a large number of spurious small changes.

Controlling for measurement bias is important, but to better understand price stickiness and its determinants, the literature also needs data with similar characteristics from multiple countries and economic settings. This is very hard with traditional data sources. For example, to obtain frequency estimates in 24 countries, Klenow and Malin (2010) had to source them from 27 different papers, each with its own particular data and methodologies. Appearing in this journal, Dhyne et al. (2006) is one of the few papers with data from multiple countries, thanks to the coordination provided by the European Inflation Persistence Network. But even in this case, each European national statistical organization was unwilling to share its micro data with Eurostat, so the frequency analysis had to be conducted independently in each country by a different team, each facing a dataset with different characteristics.

Instead, online prices have the potential to provide datasets with identical sampling characteristics in a large number of countries. At the Billion Prices Project we are currently working to standardize stickiness statistics in all our data and to be able to produce them on a ongoing basis. The goal is not only to share with other researchers a range of indicators that can be used to study price stickiness, but also provide policymakers with more up-to-date information about its behavior over time.

International Prices and the Law of One Price

The global nature of online data also makes it appealing for research in international economics. In particular, the relation between relative prices and exchange rates is a classic question in international economics. A basic hypothesis is the "law of one price," which implies that there should not be large or persistent cross-country differences in the prices of identical goods when translated into a common currency. When considering a group of many traded goods, the law of one price implies that exchange rates and relative prices will adjust to maintain stable purchasing power parities ("PPP"). Modest deviations from PPP are not surprising in a world with transport costs and other barriers to arbitrage. However, a huge literature documents failure of the law of one price for many traded goods at retail prices, resulting in significant volatility in the relative cost of consumption across countries. This failure occurs not only in price levels ("absolute PPP"), but also in changes over time ("relative PPP"). Furthermore, nominal exchange rate shocks tend to have persistent effects on the real exchange rate, leading to what Kenneth Rogoff called the "PPP puzzle." At the core of this puzzle is the fact that relative prices do not seem to adjust quickly to nominal exchange rate shocks. Many papers have documented the slow response of prices by measuring very low exchange "pass through" rates.[5]

The literature concerning the law of one price and PPP is hampered by the formidable difficulties in obtaining prices for a large number of identical goods sold simultaneously in a large number of countries, as discussed by Taylor (2001). In practice, researchers are forced to settle on having prices for identical goods from two countries (typically the US and Canada) or using price indexes from a large number of countries (constructed with different methods and baskets, and precluding any price level comparisons). Some micro sources of data, such as an index published by *The Economist* magazine based on the prices of McDonald's Big Mac sandwiches, provide information on many countries but are limited to a single good. The World Bank's International Comparison Program makes a worldwide effort to collect price data and to estimate PPP-adjusted GDPs in dozens of countries. But carrying out this task with traditional methods of collecting prices and adjusting for quality is so daunting that it can only be done every five years or so, severely limiting its use for research on real-exchange rate levels and dynamics.

In principle, online prices can be obtained in high frequency, for a large number of goods, in dozens of countries. The main challenge is not in the raw data collection, but rather in matching identical products across countries, as product identification codes in the data tend to be specific to the good, country, and retailer where the product is sold.

In Cavallo, Neiman, and Rigobon (2014) we addressed the matching problem by using prices collected from global retailers such as Apple, IKEA, Zara, and H&M, who sell identical goods with the same identifying information in several dozen countries. This allowed us to directly study conditions under which the law of one price holds. Much to our surprise, we found that the law of one price only holds well in countries that share the same currency: for example, countries within the euro area, or countries that use the US dollar such as El Salvador and Ecuador. What really seems to matter for these global retailers is simply whether prices have to be shown to customers in the same currency, not whether countries are physically close, in a trade union, or even strongly pegging their currencies. In Cavallo, Neiman, and Rigobon (2015), we used the introduction of the euro in Latvia in January 2014 to show that the adjustment towards the law of one price can take place within a matter of days after a country joins a currency union. This type of price convergence was, after all, one of the objectives of the euro.

The main implication of this line of work is that choice of currency units is far more important for defining the boundaries between markets for goods than has previously been suspected. Conversely, factors that were traditionally thought to be

[5] See Rogoff (1996) for a description of the "PPP puzzle" and Taylor and Taylor (2004) for a review of the PPP literature. Burstein and Gopinath (2013) provide a review of the empirical literature on relative prices and exchange rates, and a discussion of some theoretical advances, including accounting for nontradeables or tradeables that are only locally consumed, variable markups, and pricing-to-market.

important—such as physical distance, political and tax territories, language, and culture—do not seem to matter as much. Furthermore, these patterns also point to the importance of customer psychology, organizational structure, and the internet for price-setting behavior. For example, firms may fear antagonizing customers who see prices posted on the web in the same currency across borders. Such considerations do not yet feature prominently in most macroeconomic models.

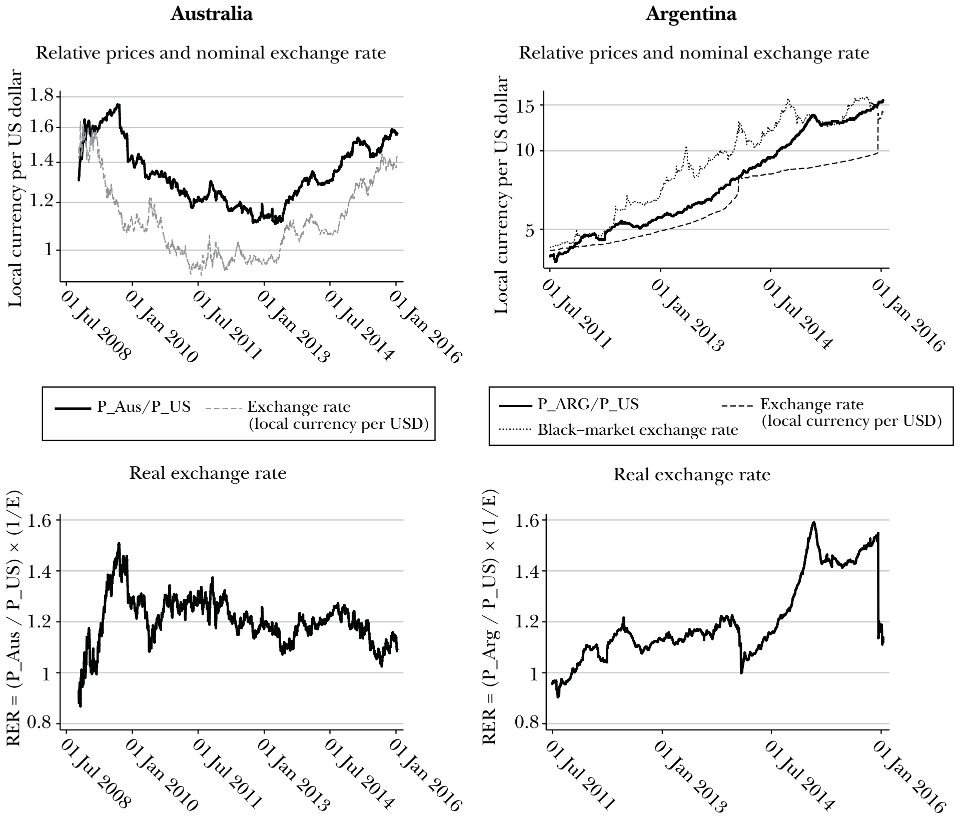
Ideally though, for some applications we need to have both global and local retailers. So since 2014, PriceStats has been expanding the product matching to include local retailers as well, classifying over 30,000 individual goods into 300 product categories. The challenge is to classify a large set of heterogeneous individual products (with varying package sizes, flavors, and retailers) into narrowly defined product categories such as "Basmati White Rice, 1kg" or "LG Basic Blu-Ray Player, 1 unit." This is achieved by using supervised machine learning (specifically a "Naive Bayes" classifier) that trains on language-specific, hand-categorized items. The process is described in detail in Bertolotto (2016). The output resembles a collection of hundreds of "Big Mac"-type indexes for different kinds of goods.

These matched indexes can be used to study real-exchange-rate levels and dynamics, as in Cavallo and Neiman (2016). To illustrate this, Figure 9 shows PPP metrics constructed by PriceStats for an average of more than 250 goods in food, electronics, and fuel in Argentina and Australia relative to the United States (examples for other countries are provided in the online Appendix available with this paper at http://e-jep.org). The top panel shows the relative prices (in local currencies) and the nominal exchange rate (defined as local currency per US dollar). For the case of Argentina, we also plot the black-market exchange rate. The bottom panel shows the real exchange rate constructed from the other two variables (as the ratio between the relative prices and the nominal exchange rate). This is simply the relative cost of the basket when expressed in the same currency.

A common finding, present also in other countries, is that relative prices co-move closely with the nominal exchange rate movements. For example, as the Australian dollar appreciated from 2008 to 2011, relative prices in Australia fell to compensate, and when the Australian dollar started to depreciate again in 2013, relative prices rose. In Argentina, the steady increase in relative prices was matched by the overall trend of depreciation in the currency, which is gradual in the black-market and lumpy on the official exchange rate. There are long periods where prices kept rising and the official exchange rate was held fixed by the government, causing "deviations" in the real-exchange rate, but there were sudden adjustments in the two occasions when the country devalued its currency, in January 2014 and December 2015.

This co-movement between relative prices and exchange rates implies high rates of pass-through, which can go in both directions. In Australia, there is evidence that nominal exchange rate shocks affect retail prices (as the literature tries to capture in traditional "pass-through" estimates). In Argentina, retail price movements tend to precede nominal exchange rate adjustments.

Figure 9
Relative Prices and Exchange Rates



Source: Authors.
Notes: The top panel shows the ratio of relative prices (in local currencies, P/P_{US}) and the nominal exchange rate (E, defined as local currency per US dollar). The bottom panel shows the real exchange rate computed as $(P/P_{US}) \times (1/E)$. It is the relative cost of the basket in each country relative to the US, when expressed in the same currency. Real exchange rates and relative price series are computed by PriceStats at the product level and aggregated using a Fisher index with official expenditure weights for food, fuel, and electronics.

Another unique feature of online data is that they provide information on relative price *levels,* which are not available when using consumer price indexes. For example, the real exchange rates in Figure 9 show that the basket tends to be 20 percent more expensive in Australia relative to the United States. In Argentina, the cost is about 10 percent higher than in the United States when the currency is allowed to float. Recognizing these patterns is useful for estimating the degree of currency misalignment at different points in time, particularly in countries with managed exchange rates.

For example, in December 2015 the new government of Argentina wanted to remove all foreign-exchange market restrictions. It was unclear what the free-market exchange rate would be, and what effect it would have on tradable prices. The nominal exchange rate implied by purchasing power parity was 14.3 pesos per dollar, suggesting that the official rate of 9.6 pesos per dollar was greatly overvalued while the black market rate of 15 pesos per dollar was slightly undervalued. When the market was freed, the new exchange rate quickly settled around 14 pesos per dollar, closely matching the implied PPP exchange rate (the ratio of relative prices). This can be seen in the jump of the official exchange rate in the top right panel of Figure 9.

While we do not expect these metrics to help predict exchange rates so closely in every country and situation, they can provide better measures of the amount of deviation of real-exchange rates from "normal" levels at a given point in time.

So far, our micro data has only been matched for seven countries and the time series is still too short to make strong inferences, but it is clear that some key puzzles in international economics and macroeconomics that emerged from studies using official price indexes appear quite different when viewed through the perspective of online data.

Access to the Billion Prices Project Data

As an academic project, we share as much data and results as possible on our webpage (bpp.mit.edu). Most of the micro data and indexes used in our papers are currently available to download on that page, together with detailed scripts that allow others to replicate and extend our results. The micro data are posted with little pretreatment, so other researchers can apply their own methods. We will upgrade the shared data periodically, both increasing the number of databases and retailers and also expanding the time series.

The US and Argentina inflation indexes used in this paper are published with a 30-day lag on the Billion Prices Project website, while the PPP exchange rate information discussed in the previous section are currently published with a one-year lag on the PriceStats website. The raw micro data collected by PriceStats are not publicly available but can be shared with academic researchers who collaborate with the Billion Prices Project and sign a data-access agreement.

Final Remarks

The need for economists to get involved in data collection was eloquently pointed out many years ago (1985) by Zvi Griliches (also see his Presidential Address at the American Economic Association in 1994). In his words,

> … [W]e have shown little interest in improving [the data], in getting involved in the grubby task of designing and collecting original data sets of our own.

> Most of our work is on "found" data, data that have been collected by somebody else, often for quite different purposes. … "They" collect the data and are responsible for all their imperfections. "We" try to do the best with what we get, to find the grain of relevant information in all the chaff.

Big data technologies are finally providing macro and international economists with opportunities to stop treating the data as "given" and get personally involved with data collection. We can now build datasets customized to fit specific measurement and research needs. This will help mitigate issues in empirical research such as sample selection, endogeneity, omitted variables, and error-in-variables, which are so frequent in traditional datasets.

The Billion Prices Project is just one example of the use of big data in economics.[6] Although online price data are the focus of this paper, we hope to have convinced other economists and perhaps a few policymakers of the benefits of experimenting with alternative data sources. Other examples include various types of "scraped" data, such as labor and real estate information available on the web, along with data from mobile phones, satellite images, GPS signals, and many other sensors that are increasingly part of our daily lives.

While many governments have been active in searching for alternative data sources, hoping to increase the quality of statistics and to reduce cost, their use will require not only the will of policymakers and statisticians working on the field, but also the involvement of more economists and academics who can help identify the best ways to collect, treat, and use these new sources of information.

■ *We thank Manuel Bertolotto, Maria Fazzolari, Diego Aparicio, Belen Bazano, and Augusto Ospital for outstanding research assistance at various stages of our Billion Prices Project research. We thank Pilar Iglesias and all the team at PriceStats for sharing their indexes and micro data. We also thank the large number of students who helped us over the years through MIT's Undergraduate Research Opportunities Program. We thanks Chris Sims for detailed comments on an earlier version of this paper, as well as seminar participants at the MIT, IMF, IADB, Census Bureau, BEA, ECB, CEMLA, NYU, Federal Reserve Bank of San Francisco, UNECE/ILO CPI Experts Conference, NBER CIRW Conference, Bureau of Labor Statistics, and the State Street Annual Research Conference. All remaining errors are ours. Financial support for this research was provided by MIT Sloan. Publicly available Billion Prices Project datasets can be downloaded at http://bpp.mit.edu/datasets/.*

[6] Einav and Levin (2014) provide a more general discussion of this topic, including new granular data sources, computational techniques such as machine learning, and the role of theory in analyzing large, unstructured datasets. In this journal, Varian (2014) describes in detail some new big data techniques that are useful to analyze large datasets.

References

Aizcorbe, Ana M., Carol A. Corrado, and Mark E. Doms. 2000. "Constructing Price and Quantity Indexes for High Technology Goods." Paper prepared for the CRIW Workshop on Price Measurement at the NBER Summer Institute, July 31–August 1, 2000.

Aizcorbe, Ana M., Carol A. Corrado, and Mark E. Doms. 2003. "When Do Matched-Model and Hedonic Techniques Yield Similar Measures?" Federal Reserve Bank of San Francisco Working Paper, no. 2003-14.

Aparicio, Diego, and Manuel Bertolotto. 2016. "Forecasting Inflation with Online Prices." Unpublished paper, MIT.

Armknecht, Paul A., and Donald Weyback. 1989. "Adjustments for Quality Change in the US Consumer Price Index." *Journal of Official Statistics* 5(2): 107–23.

Bertolotto, Manuel. 2016. "Real Exchange Rates Using Online Data." Working Paper, Universidad de San Andrés.

Breton, Robert, Gareth Clews, Liz Metcalfe, Natasha Milliken, Christopher Payne, Joe Winton, and Ainslie Woods. 2015. "Research Indexes Using Web Scraped Data." Office for National Statistics, UK.

Brynjolfsson, Erik, and Michael D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46(4): 563–85.

Burstein, Ariel, and Gita Gopinath. 2013. "International Prices and Exchange Rates." NBER Working Paper 18829.

Campbell, Jeffrey R., and Benjamin Eden. 2014. "Rigid Prices: Evidence from U.S. Scanner Data." *International Economic Review* 55(2): 423–42.

Cavallo, Alberto. 2013. "Online and Official Price Indexes: Measuring Argentina's Inflation." *Journal of Monetary Economics* 60(2): 152–65.

Cavallo, Alberto. 2015. "Scraped Data and Sticky Prices." NBER Working Paper 21490.

Cavallo, Alberto. 2016. "Are Online and Offline Prices Similar?" 2016 NBER Working Paper 22142.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia. 2016. "Learning from Potentially-Biased Statistics: Household Inflation Perceptions and Expectations in Argentina." NBER Working Paper 22103.

Cavallo, Alberto, and Brent Neiman. 2016. "Real Exchange Rate Behavior: Evidence from Online Retailers in Nine Countries." Unpublished Paper.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2014. "Currency Unions, Product Introductions, and the Real Exchange Rate." *Quarterly Journal of Economics* 129(2): 529–95.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2015. "The Price Impact of Joining a Currency Union: Evidence from Latvia." *IMF Economic Review* 63(2): 281–97.

Cavallo, Alberto, and Roberto Rigobon. 2011. "The Distribution of the Size of Price Changes." NBER Working Paper 16760.

Dhyne, Emmanuel, Luis J Álvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lünnemann, Fabio Rumler, and Jouko Vilmunen. 2006. "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data." *Journal of Economic Perspectives* 20(2): 171–92.

Edelman, Benjamin. 2012. "Using Internet Data for Economic Research." *Journal of Economic Perspectives* 26(2): 189–206.

Eichenbaum, Martin, Nir Jaimovich, Sergio Rebelo, and Josephine Smith. 2014. "How Frequent Are Small Price Changes?" *American Economic Journal: Macroeconomics* 6(2): 137–55.

Einav, Liran, and Jonathan Levin. 2014. "Economics in the Age of Big Data." *Science* 346 (6210).

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica* 77(2): 427–52.

Feenstra, Robert C., and Matthew D. Shapiro, eds. 2003. *Scanner Data and Price Indexes.* NBER.

Gorodnichenko, Yuriy, Viacheslav Sheremirov, and Oleksandr Talavera. 2014. "Price Setting in Online Markets: Does IT Click?" NBER Working Paper 20819.

Griffioen, Robert, Jan de Haan, and Leon Willenborg. 2014. "Collecting Clothing Data from the Internet." Proceedings of *Meeting of the Group of Experts on Consumer Price Indexes*, May 26–28. UNECE. Available at: http://www.unece.org/stats/documents/2014.05.cpi.html#/.

Griliches, Zvi. 1985. "Data and Econometricians—The Uneasy Alliance." *American Economic Review* 75(2): 196–200.

Groves, Robert M. 2011. "Three Eras of Survey Research." Public Opinion Quarterly 75(5): 861–71.

Horrigan, Michael W. 2013. "Big Data: A Perspective from the BLS." *Amstat News*, January 1. http://magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/.

ILO, IMF, OECD, Eurostat, UNECE, and the World Bank. 2004. *Consumer Price Index Manual: Theory and Practice.*

Klenow, Peter J., and Benjamin A. Malin. 2010. "Microeconomic Evidence on Price-Setting." Chap. 6 in *Handbook of Monetary Economics*, vol. 3, edited by Benjamin M. Friedman and Michael Woodford. Elsevier.

Krsinich, Frances. 2015. "Price Indexes from Online Data Using the Fixed-Effects Window-Splice (FEWS) Index." Paper presented at the Ottawa Group, Tokyo, Japan, May 20–22, 2015. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s2p7_pap.pdf.

Midrigan, Virgiliu. 2011. "Menu Costs, Multi-product Firms, and Aggregate Fluctuations." *Econometrica* 79(4): 1139–80.

Mikians, Jakub, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. "Detecting Price and Search Discrimination on the Internet." In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, 79–84. HotNets-XI. ACM. http://doi.acm.org/10.1145/2390231.2390245.

Nakamura, Emi, and Jón Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics* 5: 133–63.

Nygaard, Ragnhild. 2015. "The Use of Online Prices in the Norwegian Consumer Price Index." Paper prepared for the meeting of the Ottowa Group, Tokyo, Japan, May 20–22, 2015. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s2p5_pap.pdf.

Rastello, Sandrine, and Ian Katz. 2013. "Argentina Is First Nation Censured by IMF for Economic Data." Bloomberg, February 2. http://www.bloomberg.com/news/articles/2013-02-01/argentina-becomes-first-nation-censured-by-imf-on-inflation-data.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34(2): 647–68

Silver, Mick, and Saeed Heravi. 2005. "A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment Using Scanner Data." *Journal of Business & Economic Statistics* 23(3): 269–81.

Taylor, Alan M. 2001. "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price." *Econometrica* 69(2): 473–98.

Taylor, Alan M., and Mark P Taylor. 2004. "The Purchasing Power Parity Debate." *Journal of Economic Perspectives* 18(4): 135–58.

Valentino-DeVries, Jennifer, Jeremy Singer-Vine, and Ashkan Soltani. 2012. "Websites Vary Prices, Deals Based on Users' Information." *Wall Street Journal*, December 24. http://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Woodford, Michael. 2009. "Information-Constrained State-Dependent Pricing." *Journal of Monetary Economics* 56(Supplement): S100–S124.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 179–200

The Masking of the Decline in Manufacturing Employment by the Housing Bubble

Kerwin Kofi Charles, Erik Hurst, and Matthew J. Notowidigdo

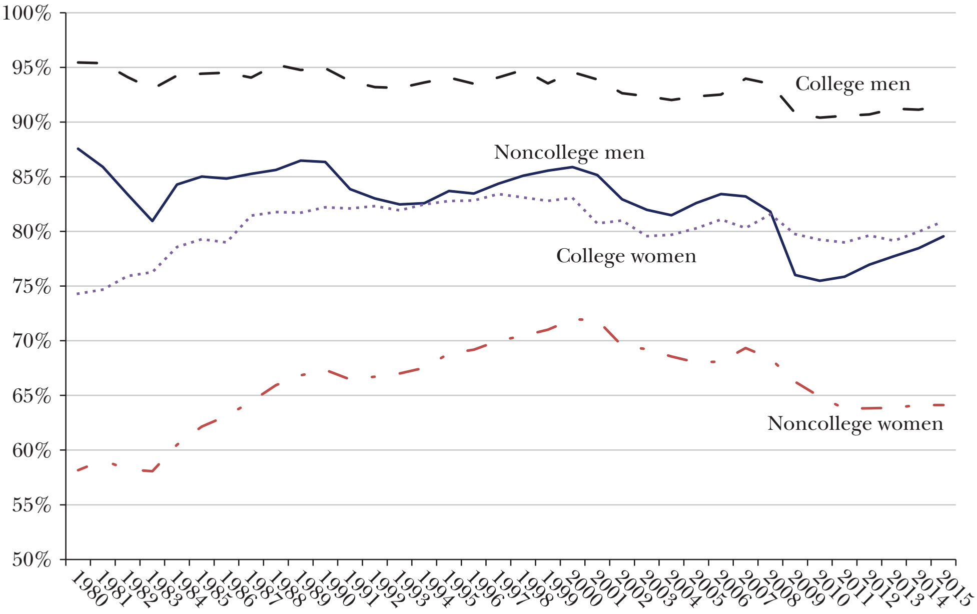
The employment-to-population ratio among prime-aged adults aged 25–54 has fallen substantially since 2000. Many authors have commented on this decline (including Hall 2011; Moffitt 2012; Davis and Haltiwanger 2014). Hall (2014) describes it as the defining feature of the labor market since 2000. Similarly, Acemoglu et al. (2016) call the employment decline of the 2000s the "Great US Employment Sag."

The magnitude of the fall in employment and its distribution across the population can be seen in Figure 1. This figure shows the employment rate from 1980 to 2015, separately by gender and education level, using annual data from the March Current Population Survey (CPS). Most of the reduction in the employment rate since about 2000 has come from those without a four-year college degree, who we refer to as "noncollege" throughout this paper. The employment rate for prime-aged, noncollege men hovered around 85 percent from 1980 to 2000, but in 2014 was only 79 percent, fully 6 percentage points below the 2000 level. The employment rates for prime-aged, noncollege women fell from roughly 70 percent in the late 1990s to 64 percent in 2015, also a decline of about 6 percentage points relative

■ *Kerwin Kofi Charles is Edwin and Betty L. Bergman Professor, University of Chicago Harris School of Public Policy, Chicago, Illinois. Erik Hurst is the V. Duane Rath Professor of Economics and the John E. Jeuck Faculty Fellow, University of Chicago Booth School of Business, Chicago, Illinois. Matthew J. Notowidigdo is Associate Professor of Economics, Northwestern University, Evanston, Illinois. All three authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are kerwin.charles@gmail.com, erik.hurst@chicagobooth.edu, and noto@northwestern.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.179 doi=10.1257/jep.30.2.179

Figure 1
Employment Rate, Prime-Age Individuals, 1980–2015 Current Population Survey



Source: This figure uses data from 1980–2015 March Current Population Survey.
Note: : The "College" men and women are all prime-age adults (aged 25–54) who have at least a four-year college degree. Adults with lower levels of education are "Noncollege." The figures calculate employment rates as share of total population of each group using individual-level survey weights.

to the 2000 level. These large and seemingly persistent reductions in employment among the less-educated over the course of the 2000s were much larger than those for both prime-aged men and women with four-year college degrees, whose employment rates fell by only about 2 percentage points between 2000 and 2014.

The explanations proposed for the decline in the employment-to-population ratio have been of two broad types. Focusing on the fact that the employment rate decline was especially sharp from 2007 to 2010—that is, during and immediately after the Great Recession—one set of explanations emphasizes cyclical factors associated with the recession, including temporary declines in labor demand, economic and policy uncertainty, "mismatch" between unemployed workers and jobs, and the availability of unemployment insurance for extended periods. The second set of explanations focuses on the role of longer-run structural factors, the potential importance of which is suggested by the reduction in the employment rate even before the start of the Great Recession began, and the persistently low employment-to-population rates for low-skilled workers years after the official end of the recession. Structural explanations focus on long-term secular trends, such as the falling demand for routine tasks performed by workers in many manufacturing jobs. However, if structural factors indeed explain much of the decline in employment since 2000, it is not immediately clear why the effect of these long-term factors should have been so modest from 2000 to 2007, only to appear with sudden and pronounced effect during the Great Recession.

In this paper, we argue that while the decline in manufacturing and the consequent reduction in demand for less-educated workers put downward pressure on their employment rates in the pre-recession 2000–2006 period, the increased demand for less-educated workers because of the housing boom was simultaneously pushing their employment rates upwards (Charles, Hurst, and Notowidigdo 2016, 2015). For a few years, the housing boom served to "mask" the labor market effects of manufacturing decline for less-educated workers. When the housing market collapsed in 2007, there was a large, immediate decline in employment among these workers, who faced not only the sudden disappearance of jobs related to the housing boom, but also the fact that manufacturing's steady decline during the early 2000s left them with many fewer opportunities in that sector than had existed at the start of the decade.

We begin with a short overview of various cyclical and structural arguments about the decline in the employment-to-population ratio. We then present several different pieces of evidence which support our hypothesis that the masking and unmasking of manufacturing decline by the housing boom and bust play an important role in changes in the employment rate since 2000. The evidence we present in support of this argument includes aggregate time-series results; local labor markets evidence that exploits the large variation in the size of manufacturing decline and in the size of the housing boom and bust across different metropolitan areas in the United States; and individual-level evidence using data about the re-employment rates of displaced manufacturing workers in the Displaced Workers Survey. Our focus throughout is on prime-aged, noncollege men. However, in the conclusion we briefly discuss employment masking for prime-aged, noncollege women, as well. Our conclusion also discusses why the presence of masking and the distinction between cyclical and structural forces is important for policymaking.

Reviewing Cyclical and Structural Explanations of Employment Rate Changes Since 2000

Most of the decline in the employment-to-population ratio for lower-skilled workers since 2000 occurred during the Great Recession, as shown in Figure 1. Perhaps because of this pattern, several recent papers seeking to understand employment changes have focused on "cyclical" explanations.

One strand of this work studies the role of the negative shocks to household and bank balance sheets that arose from the recession. Using cross-region data, Mian and Sufi (2014) find that local areas that experienced larger declines in household net worth had larger reductions in employment in nontradable sectors during the 2007–2009 period. Chodorow-Reich (2014) links the decline in employment to disruptions in the banking sector, by showing that firms that had pre-recession relationships with distressed banks were much less likely to secure credit during the recession, and were much more likely to shed employment during 2007–2009.[1] Mondragon (2015) estimates the effect of local credit supply shocks on employment and finds a large effect.[2] Also broadly related to this area of research is the work of Giroud and Mueller (2015), who find that high firm leverage before the start of the Great Recession also contributed to large employment losses during the recession.

Several other papers in this literature assess other cyclical factors that were likely changed because of the recession, including increased economic and policy uncertainty, increases in sectoral and spatial mismatch, and changes in the duration and generosity of unemployment benefits and other transfer programs. Baker, Bloom, and Davis (2015) show that measures of aggregate uncertainty were high during the Great Recession relative to historical levels, and argue that this increased uncertainty can account for some of the decline in the employment rate. Similar explanations are also found in Fernández-Villaverde et al. (2015). Sahin, Song, Topa, and Violante (2014) examine the extent to which search frictions that affect the ease with which workers can move between occupations and locations may have increased the unemployment rate and reduced the employment rate. Their results suggest that these mismatch forces may explain as much as one-third of the rise in the unemployment rate between 2007 and 2010, with the effect diminishing by 2012.

The growing literature relating the decline in aggregate employment to the expansion of unemployment benefits during the Great Recession has come to mixed conclusions. Rothstein (2011) and Farber and Valletta (2013) find that although unemployment benefit extension may have propped up the unemployment rate by delaying exits from the labor force, benefit extension did not have much of an effect on the employment rate. However, Johnston and Mas (2015) and Hagedorn, Karahan, Manovski, and Mitman (2013) find larger effects of unemployment benefit extensions. Additionally, Mulligan (2012) discusses how broader policy changes that occurred during the recession—such as the expansion of the Supplemental Nutritional Assistance Program (often known as Food Stamps)—could have discouraged individual labor supply and thus reduced the employment rate.

Finally, another program that could have had an important effect on aggregate employment during the recession and thereafter is Social Security Disability Insurance (SSDI). Even before the Great Recession, there were staggering increases in enrollment, due both to reduced screening stringency and higher demand for the partial wage insurance provided by the program (Autor and Duggan 2006). Although there were no significant changes to the program during the Great Recession, research has documented a strong link between labor market conditions and SSDI application rates (Autor and Duggan 2003; Sloane 2015), and strong effects of SSDI on

[1] A large theoretical literature examines the role of tightening borrowing constraints on households and firms and how that translates into declining aggregate employment: for example, see Eggertsson and Krugman (2012) and Guerrieri and Lorenzoni (2011).

[2] Greenstone, Mas, and Nguyen (2015) also examine the relationship between local credit supply shocks to local banks and local employment outcomes. They show that while credit supply shocks do reduce credit to small firms, the employment losses of small firms have little effect on local employment rates.

both employment and earnings (Maestas, Mullen, and Strand 2013). Sloane (2015) documents large increases in disability rates between 2007 and 2011 in local markets with large increases in unemployment rates during the Great Recession. Given that disability tends to be persistent, this could explain some of the low employment rates after the recession. Her estimates, however, suggest such effects are modest. Additionally, unlike unemployment insurance and many other social insurance programs, there are often large waiting times to get onto SSDI, which means that denied SSDI applicants typically search for work after long periods of detachment from the labor market. The delay in processing applications appears to generate duration dependence in nonemployment (it is more difficult to become employed the longer one has already not been employed). So it is harder for rejected applicants to return to the labor market (Autor, Maestas, Mullen, and Strand 2015).[3] As a result, the large number of rejected SSDI applicants during the Great Recession may experience lower employment rates in the longer-run (Maestas, Mullen, and Strand 2015).

While the preceding papers seem to explain a meaningful share of the employment changes observed over the course of the Great Recession, a problem for the notion that cyclical factors are the main explanation for the full pattern of observed employment changes since 2000 is that cyclical factors cannot readily explain the *persistence* of reduced employment among prime-age lower-skilled individuals: that is, the fact that rates have remained low long after the impact of cyclical factors from the recession should have ended. Despite growing evidence of market normalization in the years since the end of the Great Recession—stabilization of housing prices, favorable lending conditions, declines in aggregate uncertainty, return of labor market mismatch to pre-recessionary levels, and the cessation of extended unemployment benefits—the employment rate remains significantly below pre-recessionary levels.

Alongside the literature studying cyclical factors, another literature has emerged studying the role of structural factors in explaining recent changes in employment. How long-term changes in underlying demographics, such as the ageing of the population, have contributed to the decline in labor force participation has been the focus of one strand of research (for example, Aaronson, Hu, Seifoddini, and Sullivan 2015). On the whole, results from this work indicate that demographic factors explain a portion of the overall decline in employment and labor force participation. Our work focuses throughout on the population of prime-aged, noncollege men, so the declines we document and attempt to explain cannot be accounted for by changes in demographics and must be due to other factors. The focus on these other factors complements existing work studying demographics.

[3] For simplicity, this section focuses on explanations of changes in employment during the 2000s that can be readily categorized into either cyclical or structural explanations. However, factors such as duration dependence belong to a class of explanations that suggest important interactions between the two. For example, Kroft, Lange, Notowidigdo, and Katz (2016) calibrate a search and matching model to show that the low job finding rate in the aftermath of the Great Recession may partly be due to duration dependence in unemployment. As a result, the sharp decline in vacancies generated an increase in long-term unemployment and—through duration dependence in unemployment—reduced both the overall job-finding rate and aggregate employment.

Another strand of the work studying longer-term structural factors has linked declining demand for routine tasks (Autor, Levy, and Murnane 2003) and job polarization (Autor and Dorn 2013) to declining employment rates for less-educated workers during the 2000s. Autor, Dorn, and Hanson (2013), Charles, Hurst, and Notowidigdo (2016), and Acemoglu et al. (2016) all discuss how the decline in manufacturing during the 2000s depressed employment rates for less-educated workers. International trade appears to account for some of the decline in manufacturing employment. Autor, Dorn, and Hanson (2013) provide local labor markets evidence that increased import competition from increased trade with China reduced manufacturing employment during the 1990s and 2000s. Similarly, Pierce and Schott (forthcoming) provide evidence that the "surprisingly swift" decline of manufacturing during the 2000s is linked to changes in trade policy that eliminated potential tariff increases for Chinese imports. Consistent with their interpretation of the role of changes in trade policy, there is a clear divergence in manufacturing employment trends between the United States and the European Union following this policy change.

These papers suggest an important role for structural factors in understanding the pattern of employment changes since 2000. However, it is not clear how these relatively slow-moving structural shifts could explain the sudden reduction in employment rates after 2008. Furthermore, since any structural forces affecting employment likely operated steadily throughout the 2000s, one would have expected their influence to reduce employment substantially before the recession. Yet employment rates in the early 2000s declined only modestly.

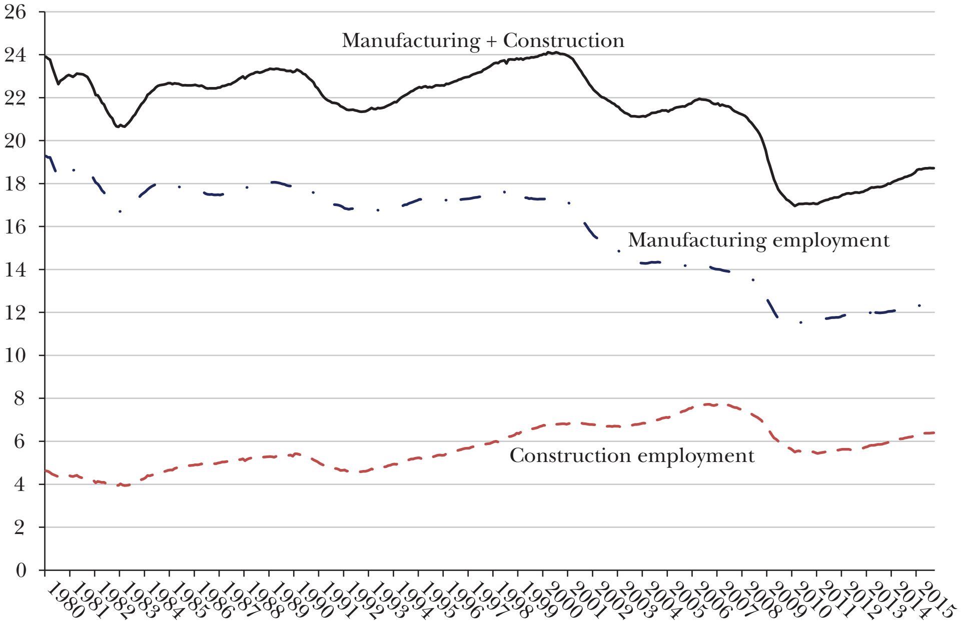
The following sections present an array of evidence that employment losses arising from the structural decline in manufacturing were "masked" by positive employment effects for lower-skilled labor associated with the national housing boom during the 2000–2006 period, and then "unmasked" when the housing market reverted to be closer to its normal state after 2007. This explanation reconciles the key facts about the full pattern of changes in employment since 2000 for prime-aged noncollege adults, including the sudden large decline in 2008 after a period of relatively little change, and the persistently low levels of employment several years after the end of the recession.

Masking: Evidence from Aggregate Time-Series Data

The counterbalancing patterns of long-term job decline in manufacturing and the surge in construction jobs during the housing boom is apparent in time-series data. Figure 2 shows the patterns of total jobs in manufacturing and construction from 1980 to 2015. Manufacturing jobs were in slow decline through the 1980s and 1990s, then entered a period of rapid decline from around 1999 through 2010, and have since leveled out.[4] During the 15-year period between 2000 and 2015,

[4] For analyses of the decline in US manufacturing during the 1980s and 1990s, useful starting points are Bound and Holzer (2000) and Berman, Bound, and Griliches (1994).

Figure 2
Total Monthly US Manufacturing Employment 1980M1–2015M9
(in millions)



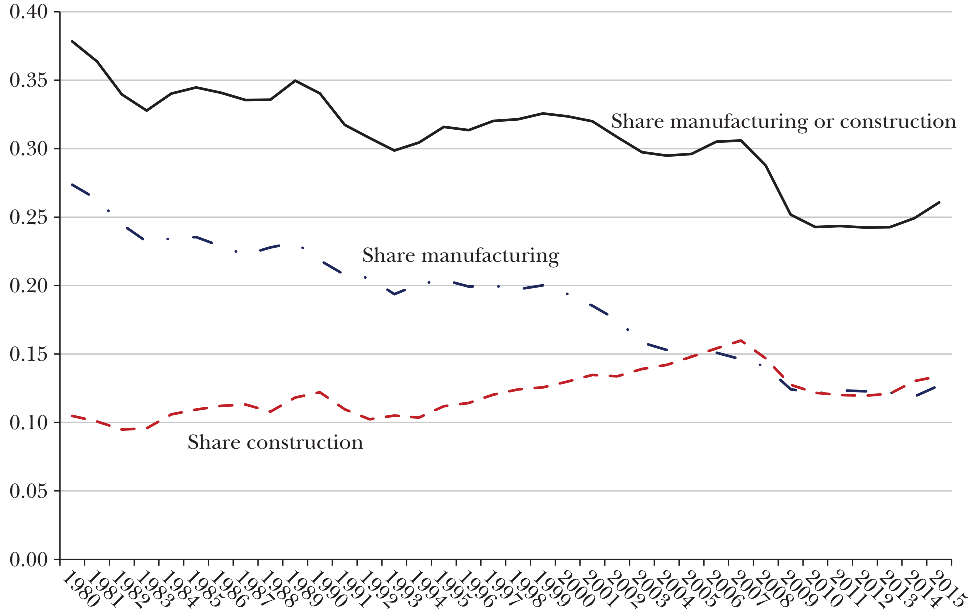
Source: Authors using aggregate data from the Bureau of Labor Statistics on monthly employment in manufacturing and construction sectors.

the US economy lost roughly one-third of the manufacturing jobs that had existed in 2000.

The national housing boom—marked by massive increases in housing prices, new construction and renovations, and real estate transactions—began in the late 1990s and completely collapsed over a short time period beginning in 2007. The boom changed employment opportunities in many sectors, but in this section we focus only on the number of jobs in the construction sector, which expanded and contracted significantly over the course of the housing boom and bust. This can be seen in the lower line in Figure 2, which plots total monthly construction jobs between 1980 and 2015, using data from the Bureau of Labor Statistics (BLS). From 1980 to the mid-1990s, the total number of construction jobs fluctuated between four and five million. However, between the mid-1990s and the mid-2000s total construction jobs surged by three million, peaking at nearly eight million jobs in 2006. When the boom ended in 2007, construction employment collapsed with it. By 2010, the number of construction jobs in the economy had returned to their 1996 levels and have remained close to those levels ever since.

The top line in Figure 2 is the total number of jobs in the economy in either manufacturing or construction from 1980 to 2015. The figure shows that between

Figure 3
Construction and Manufacturing Employment Shares, Noncollege Men Aged 25–54; 1980–2015



Source: This figure uses data from 1980–2015 from the March Current Population Survey, restricted to prime-age men with education below a four-year college degree.
Note: The figures calculate employment shares as a share of total population using individual-level survey weights.

2000 and 2006, the surge in the number of construction jobs substantially offset job losses in manufacturing, leaving the total number of jobs accounted for by the two sectors essentially constant during this period. After 2007, the total number of jobs in construction and manufacturing declined sharply, as construction collapsed to long-term historical levels following the housing bust and as the number of manufacturing jobs continued to decline. The job gains from the housing boom meant that the decline in the number of jobs because of long-term, sectoral decline that otherwise would have been apparent in aggregate data on the total number of jobs was not evident until 2008, although the decline started years earlier.

Working in either manufacturing or construction has long been very important in the life experience of prime-aged noncollege men, who we have shown had particularly pronounced changes in employment in the last 20 years. Using individual-level data from the Current Population Survey (CPS), Figure 3 shows that among all noncollege prime-aged men, including those not working at all, roughly 30 percent were working in either manufacturing or construction at any time between 1980 and 2007.[5] In effect, during the 2000–2007 period, as the share of all prime-aged, noncollege men working in manufacturing fell substantially, surging opportunities in construction from the housing boom almost exactly made up for the lost manufacturing employment for these men. Since 2007, with the bust in construction and continued decline in manufacturing, the share of all prime-aged noncollege men employed in construction or manufacturing has fallen sharply, going from roughly 30 percent in 2008 to 23 percent in 2014.[6]

The aggregate evidence suggests that when the housing boom ended, and the construction jobs associated with it disappeared, many prime-aged noncollege men who had been working in construction simply left the labor force. Figure 4 shows the fraction of all prime-aged, noncollege men who are working in construction, working in manufacturing, or are not employed (that is, either unemployed or not in the labor force) has been incredibly stable over time. Historically, when the manufacturing plus construction employment share for prime-aged, noncollege men has gone up, the incidence of nonemployment among such men has gone down; when the manufacturing/construction share has been flat, as it was from the mid-1990s to mid-2000s, nonemployment has been flat; and when the manufacturing plus construction share has gone down, as it did sharply after 2007, nonemployment has surged. The negative association between the two series can be clearly seen in the top line in the figure, which shows how remarkably constant the fraction of all prime-aged, noncollege men engaged in these three activities has been over time. This time-series evidence also suggests that many of the men not working in the construction and manufacturing sectors since 2007 have ceased being employed altogether.

Masking: Local Labor Markets Evidence

An obvious concern about time-series evidence is that a temporal association between the different series need not reflect a causal relationship. In particular, it could be that other unmeasured, national-level shocks account for both the upward pattern in nonemployment of prime-aged, noncollege males after 2007 and its sustained low level though 2015.

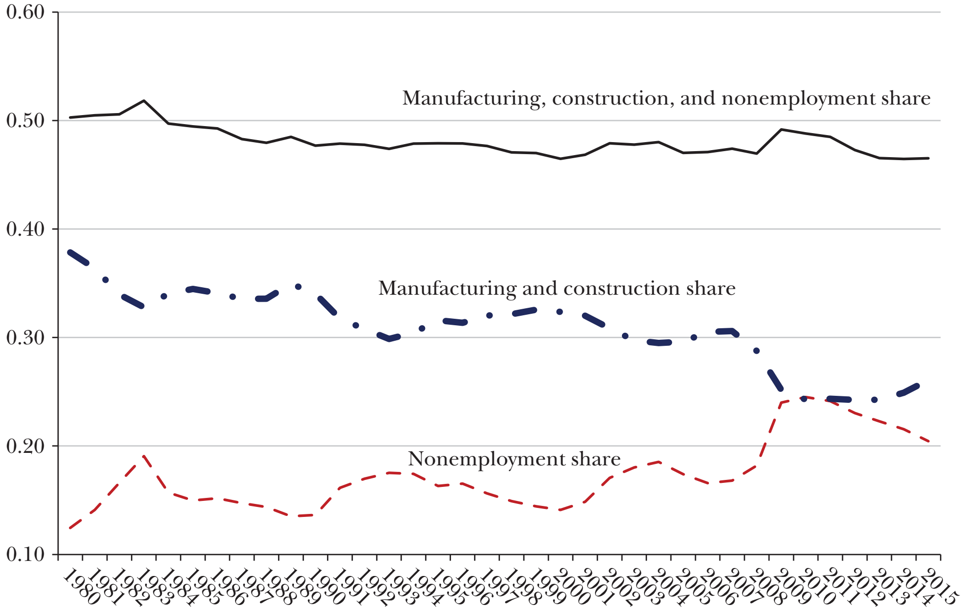
To investigate these concerns, we exploit variation across urban areas in the size of manufacturing decline and the size of the local housing boom they experienced. We create a panel of metropolitan statistical areas (MSAs) using data from the 2000 Census and from various years of the American Community Survey (ACS) individual-level and household-level extracts from the Integrated Public

[5] For the results in Figures 3 and 4, we use data from the March CPS, which are downloaded from the IPUMS-CPS (Ruggles et al. 2015) at https://cps.ipums.org/cps/.

[6] See Charles, Hurst, and Notowidigdo (2016) for discussion that although some of decline in manufacturing employment share from 1982 to 1999 was the result of increase in the size of prime-aged noncollege population, the population was constant in the 2000–2006 interval, so the decline in manufacturing employment share was exclusively the result of sectoral decline. The fact that manufacturing jobs are lost during the 1990 and 2000 recession is highlighted in Jaimovich and Siu (2012).

Figure 4

Construction, Manufacturing, and Nonemployment Shares, Noncollege Men Aged 25–54; 1980–2015



Source: This figure uses data from 1980–2015 from the March Current Population Survey, restricted to prime-age men with education below a four-year college degree.
Note: The figures calculate nonemployment rates and employment shares as a share of total population using individual-level survey weights.

Use Microsamples database (Ruggles et al. 2015). The analysis extends from 2000 (the first year during the boom with reliable information in the Census at the level of metropolitan statistical areas) to 2012 (the midpoint of 2011–2013 ACS data). These years span the 2000–2006 housing boom, the 2007–2009 housing bust, and several years after the end of the housing bubble and the Great Recession. We compute employment rates and employment shares in various occupations in each metropolitan statistical area.[7] The primary analysis sample consists of noninstitutionalized, prime-aged men aged 25–54 without a four-year college degree.

Our measure of the decline in manufacturing in any given metropolitan statistical area, ΔM_k, is the change in the fraction of the prime-aged, noncollege male

[7] For the 2000 numbers, these means are from the 2000 Census. For the 2006 numbers, we pool the American Community Survey data from 2005 to 2007 to increase the precision of the metropolitan statistical area estimates. Similarly, we pool the 2011–2013 American Community Survey for the 2012 numbers.

population in the Census/ACS employed in manufacturing industries over the relevant time period. In a simple model of housing demand and supply, the effect of a shock that shifts housing demand will be a weighted sum of the change in the price of housing and the change in housing supply, which can be proxied by the amount of housing built. Our measure of the housing demand change, ΔH_k, is therefore the (log) change in the average price of houses sold in the metropolitan statistical area (MSA) plus the (log) change in the number of building permits approved in the MSA. We use house price data from the Federal Housing Finance Agency (FHFA), mapping FHFA metro areas to the Census/ACS metro areas by hand. We use data on housing permits from the Census Building Permits Survey, and match the MSA codes in the permits data to the Census/ACS metro area codes by hand.[8]

Changes in both house prices and in the housing stock can affect employment. House prices affect household wealth or liquidity and thus households' demand for goods and services produced in the local market (Mian and Sufi 2011). Changes in the amount (or quality) of housing necessarily involves construction activity such as demolition, renovation, home improvements, or new construction. Our housing demand measure captures all of these effects.

Table 1 reports summary statistics for our sample of 275 metropolitan statistical areas with nonmissing labor market and housing market data. Our specific approach is to consider the effects of two shocks to local markets during the years from 2000 to 2006: the change in the share of population employed in manufacturing and our proxy for housing demand (based on changes in housing prices and construction permits). We look at the effects of these changes on the employment rate for noncollege, prime-age men and on the share of construction employment for this group in two different time periods: the immediate effect of the shocks from 2000–2006 and the long-run effect from 2000–2012. This set of variables will let us look at masking during the 2000–2006 period, and the extent to which such effects might persist through 2012.

Panel A of Table 1 presents the means and standard deviation of the two local labor market shocks we study. The top row shows that the average decline over the 2000–2006 period in the manufacturing employment share across urban areas was –2.6 percentage points. The standard deviation of 1.9 indicates that there was substantial variation across urban areas in this mean decline; indeed, our analysis exploits this variation. The mean change in the housing demand proxy across urban areas between 2000 and 2006 was 0.66 log points with a standard deviation of 0.58. Because the housing proxy is the sum of the log changes in housing prices and building permits, this can be interpreted as meaning that the sum of housing prices and building permits rose by more than 50 percent across metropolitan statistical areas, on average, with substantial variation across metro areas.

Panel B of Table 1 presents summary statistics for the change in the employment rate and in the construction employment share for prime-aged noncollege

[8] See Charles, Hurst, and Notowidigdo (2015) for more details on the matching of the house price and housing permit data to the Census/ACS data.

Table 1
Descriptive Statistics of Manufacturing Decline and Housing Booms across Cities

	N	*Mean*	*Standard deviation*
Panel A: Manufacturing decline and changes in housing demand (two shocks)			
2000–2006 Change			
In share of population employed in manufacturing, ΔM_k	275	–0.026	0.019
In housing demand, ΔH_k	275	0.657	0.585
Panel B: Changes in total employment and construction employment			
2000–2006 Change			
In employment rate for noncollege men	275	0.019	0.039
In construction employment share for noncollege men	275	0.027	0.019
2000–2012 Change			
In employment rate for noncollege men	275	–0.021	0.048
In construction employment share for noncollege men	275	–0.002	0.022

Notes: This table reports the summary statistics for the baseline sample of 275 metropolitan areas (MSAs) across the time periods studied in the regressions that use the Census/American Community Survey data. The housing demand variable is constructed by adding the change in housing prices (from FHFA house price index) to the change in housing permits (from Census Building Permits Survey). This procedure creates a proxy for the change in housing demand in an MSA between 2000 and 2006; see Charles, Hurst, and Notowidigdo (2015) for more details. All of the reported sample statistics are computed using the 2000 population of prime-aged, noncollege men in the MSA (from Census/ACS) as weights, since these weights are used in all of the regressions.

men for different periods between 2000 and 2012. These means are consistent with the aggregate patterns discussed before. Across metropolitan statistical areas, the employment rate and overall construction share rose during the 2000–2006 boom, then fell sharply after 2007. By 2012, the share of noncollege men working in construction had returned to levels seen in 2000 in the average metro area, but their employment rate remained substantially below 2000 levels, long after the end of the housing cycle.

In Table 2, we will be investigating the relationship between employment changes in a metropolitan statistical area and manufacturing decline and housing demand shocks by estimating:

$$\Delta E_k = \beta_0 + \beta_1 \Delta M_k + \beta_2 \Delta H_k + X_k \Gamma + \eta_k,$$

where ΔE_k is either the change in employment in metropolitan statistical area or the change in the construction share for one of the two time periods 2000–2006 or 2000–2012; the ΔM_k and ΔH_k variables represent the local labor market shocks in manufacturing and housing, which occurred over 2000–2006; X_k is a vector of control variables; and η_k is a mean-zero regression error. The parameters β_1 and β_2 measure, respectively, the effect of a change in local manufacturing employment and a change in local housing demand on the change in employment. For simplicity, the results we present here are estimated with these two coefficients in a single ordinary least squares regression model.[9] The analysis is conducted in first differences and thus accounts for time-invariant differences across metropolitan statistical areas. In each specification, the *X* vector follows our paper Charles, Hurst, and Notowidigdo (2016) and includes controls for the share of employed workers with a college degree, the share of women in the labor force, and the population of the metropolitan statistical area. All standard errors are clustered by state and are weighted by the prime-age, noncollege male population in 2000.

Table 2 presents the estimates of regression equations for the 2000–2006 and 2000–2012 time periods. Each column in each panel reports the estimates from a separate equation. The point estimates in the first column of the top panel imply that a one standard deviation decrease in manufacturing employment, given as 0.019 in Table 1, would, multiplied by the coefficient 0.471 from Table 2, decrease the employment rate among non-prime-aged college men by about 0.9 percentage point during 2000–2006. Likewise, over the same period, a one standard deviation increase in housing demand of 0.58 multiplied by the relevant coefficient from Table 2 would increase the employment rate of prime-aged noncollege men by about 1.3 percentage points during the 2000–2006 period.

In column 2, we assess how the two local shocks affect the share of noncollege men working in construction in the metropolitan statistical area. In the top panel, we find no statistically significant relationship between the manufacturing shock and construction employment. By contrast, construction employment of noncollege men increased the larger the housing demand shock in the metro area. The portion of the employment increases experienced by noncollege men as a result of the housing boom that was attributable to the employment in the construction sector is the estimated effect of the housing demand change in column 2 divided by the effect in column 1, or approximately 78 percent (that is, 0.018 divided by 0.023).

The regressions in the bottom panel of Table 2 examine how local manufacturing decline and local housing market changes between 2000 and 2006 affect the longer-term change during the 2000–2012 period in outcomes for noncollege men. The results indicate that the effect of manufacturing decline during the 2000–2006

[9] As we discuss in Charles, Hurst, and Notowidigdo (2016), our results are similar in a more complicated two-equation model that allows for both the direct effect of manufacturing decline on labor market outcomes as well as an indirect effect of manufacturing decline on labor market outcomes coming through the effect of manufacturing on housing demand. In the more complicated two-equation model, we can identify both the direct and indirect effect under the assumption that changes in local housing demand do not affect local manufacturing activity directly, which we show appears to be a reasonable assumption in our setting, since the housing boom has no significant effect on manufacturing employment. We also show that the main results are similar using an instrumental variable for changes in housing demand that is formed by using sharp structural breaks in local housing prices that are interpreted as proxies for speculative activity. This instrument is described in detail in Charles, Hurst, and Notowidigdo (2015). The similarity of the two-stage least squares estimates to the main ordinary least squares estimates reported in this paper is consistent with limited endogeneity bias during the 2000–2006 period, plausibly because this is a time period when a very large share of changes in housing demand is due to speculative activity rather than due to other changes in local labor demand.

Table 2

Manufacturing Decline, Housing Booms, and Cross-City Masking: Regression Results

Sample: Prime-age, noncollege men

	Dependent variable:	
	Employment Rate (1)	*Construction Employment Share* (2)
Panel A: The dependent variable is Change in Employment Rate, or Change in Construction Employment Share, over *2000–2006*		
The independent variables (shocks) are change over 2000–2006:		
In share of population employed in manufacturing, ΔM_K	0.471	0.009
	(0.090)	(0.63)
In housing demand, ΔH_K	0.023	0.018
	(0.006)	(0.004)
R^2	0.76	0.45
N	275	275
Include baseline controls	Yes	Yes
Panel B: The dependent variable is Change in Employment Rate, or Change in Construction Employment Share, over *2000–2012*		
The independent variables (shocks) are change over *2000–2006*:		
In share of population employed in manufacturing, ΔM_K	0.653	–0.057
	(0.156)	(0.089)
In housing demand, ΔH_K	0.004	–0.001
	(0.011)	(0.005)
R^2	0.60	0.29
N	275	275
Include baseline controls	Yes	Yes

Notes: This table reports the coefficients from estimating $\Delta E_k = \beta_0 + \beta_1 \Delta M_k + \beta_2 \Delta H_k + X_k\Gamma + \eta_k$, by ordinary least squares for various samples, where ΔE_k is either the change in employment in metropolitan statistical area or the change in the construction share for one of the two time periods 2000–2006 or 2000–2012; the ΔM_k and ΔH_k variables represent the local labor market shocks in manufacturing and housing, which occurred over 2000–2006; X_k is a vector of control variables; and η_k is a mean-zero regression error. A 0.01 unit decrease in the Share of Population Employed in Manufacturing corresponds to a 1 percentage point decline in share of prime-age (25–54) non-college-educated male population employed in manufacturing. A 1-unit change in housing demand measure corresponds to 1 log point increase in housing demand proxy. The baseline controls include the initial (year 2000) values of the share of employed workers with a college degree, the share of women in the labor force, and the log population in the metropolitan statistical area. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state, are in parentheses.

period on the long-term employment of noncollege men was, in fact, quite durable. Indeed, the effects of the manufacturing decline during 2000–2006 on employment growth between 2000 and 2012 were fairly similar to the effects shown for 2000–2006. The results for the employment effects of housing demand changes during 2000–2006, however, differed sharply over 2000–2006 and the longer 2000–2012 period. In particular, we find that changes in estimated housing demand during the 2000–2006 housing boom period had no significant long-term effect on employment of noncollege men over the 2000–2012 period, either for the overall employment rate or for the share of employment in construction.

This evidence across metro areas suggests that the 2000–2006 housing boom had a masking effect on the loss of manufacturing jobs during those years, but this masking was undone during the housing bust. The negative employment effects of the housing bust were similar in magnitude to the positive employment effects of the preceding housing boom. Over the entire period from 2000 to 2012, the strong relationship between the local decline in manufacturing and the employment rate of noncollege men in a metropolitan statistical area was not affected by changes in housing demand in the metro area during the 2000–2006 boom period.

Individual-Level Masking: Evidence from Displaced Manufacturing Workers

Our local labor markets analysis suggests that masking occurred both within and between metropolitan areas.[10] What is not clear is the extent to which this masking within metro areas was because different types of workers were affected by manufacturing and housing market shocks, and how much, if any, was because some of the specific workers who lost jobs in manufacturing found employment in housing during the boom, only to lose them when housing collapsed.

To determine the extent to which the specific workers displaced from manufacturing because of the decline in that sector were re-employed in housing-related sectors, we use individual-level data from the Displaced Worker Survey, which is conducted every two years as part of the Current Population Survey.[11] This survey focuses on individuals who have been displaced from a job at some point during the preceding three years. In addition to the standard battery of questions about current employment and demographics, respondents are asked detailed questions

[10] In Charles, Hurst, and Notowidigdo (2016), we present results from more in-depth analysis to assess how much of masking is between-city and how much within-city. We find evidence of both types of masking; many cities experienced either a large housing boom or manufacturing decline between 2000–2006 but not both. Within cities, we find manufacturing affected older adults relatively more than younger adults, while our estimates suggest that the housing boom affected employment rates of older and younger adults similarly.

[11] See Farber (2015) for more information on Displaced Worker Survey data and a detailed investigation of labor market outcomes of workers displaced during the Great Recession (compared to earlier recessions).

about their previous job. We construct a sample consisting of all men aged 25–54 without a college degree in the 1994–2006 waves of the Displaced Worker Survey who were displaced from jobs in the manufacturing sector. Displacements in this sample occurred between 1992 and 2005.

The resulting sample of 2,161 persons is relatively small, but it contains geographic identifiers that allow us to sort displaced workers by the size of the housing boom that their local metropolitan statistical area experienced. We create an indicator variable to denote displacement between 1997 and 2005, which are years in the midst of the national housing boom. Persons for whom this indicator was zero were therefore displaced between 1991 and 1996 (in the years before the housing boom). For each displaced worker in our sample, we also know whether they lived in a "housing boom metropolitan statistical area," which we define to be those areas whose especially large housing booms placed them in the top one-third in the distribution of the housing demand change measure, ΔH_k.[12] And we also create an indicator variable for these areas. Intuitively, this captures the metro areas that had especially large increases in housing demand. We estimate a model of the form:

$$y_{ikt} = \beta_1 1\{Housing\ Boom\ MSA_k\} \times 1\{Boom\ Period\} + \alpha_k + \delta_t + X_{ikt}\Gamma + e_{ikt}$$

where y is either (in different specifications) re-employment or re-employment in construction of a displaced worker i in market k at time t. The terms α_k and δ_t are metropolitan statistical area and time period fixed effects, respectively, and the vector X contains individual-level controls like years of education, union status in the last job, and a fifth-order polynomial in age.

The coefficient β_1 from this regression is a difference-in-difference estimate of the effect of being in a metropolitan statistical area with a large housing boom on the probability of becoming re-employed for a worker displaced from manufacturing during the years of the housing boom. We study two outcomes: whether the person reported employment as of the survey year, and whether the person was employed in construction as of the survey year.

Table 3 presents the estimated effects, with associated standard errors clustered by state. For each outcome in Table 3, we present two difference-in-difference specifications. The first specification (in columns 1 and 3) includes fixed effects for metropolitan statistical areas and adds fixed effects for each year of displacement. The second specification (columns 2 and 4) adds the individual-level controls to the specifications in columns 1 and 3. The results for employment suggest a substantial amount of "individual-level masking." We find that manufacturing workers displaced in markets with especially large housing demand increases during the 2000–2006

[12] The evidence is fairly similar using other thresholds such as the top quartile or top 10 percent. It is also robust to residualizing housing demand change to manufacturing decline proxy and other controls, so the definition of a housing boom metro area is based on change in housing demand that is above and beyond what one would predict from manufacturing decline and other variables. See Charles, Hurst, and Notowidigdo (2016) for more details.

Table 3

Displaced Manufacturing Workers, Housing Booms, and Individual-Level Masking: Regression Results

Sample: Noncollege men, age 25–54, manufacturing workers displaced 1992–2005

	Dependent variable:			
	Employed		*Employed in Construction*	
	(1)	(2)	(3)	(4)
Difference-in-difference estimate of effect of housing boom:				
Independent variable:				
(Displaced between 1997 and 2005) ×	0.094	0.093	0.045	0.045
(Housing Boom MSA)	(0.045)	(0.043)	(0.019)	(0.019)
Mean of dependent variable	0.693	0.693	0.056	0.056
N	2,161	2,161	2,161	2,161
R^2	0.144	0.151	0.119	0.125
Include MSA fixed effects	y	y	y	y
Include displacement year fixed effects	y	y	y	y
Include individual-level controls		y		y

Source: Current Population Survey (CPS) Displaced Worker Surveys, 1994–2006.
Notes: This table reports the coefficients from an ordinary least squares regression of the equation $y_{ikt} = \beta_1 1\{Housing\ Boom\ \ MSA_k\} \times 1\{Boom\ \ Period\} + \alpha_k + \delta_t + X_{ikt}\Gamma + e_{ikt}$. The first row reports the difference-in-difference estimate of the effect of being displaced during housing boom time period within a metropolitan statistical area that was experiencing a housing boom. If a displaced worker is not in one of the MSAs with housing market data or is in a non-metro region, then this indicator is set to 0. The additional individual-level controls in columns (2) and (4) are the following: education, union status in last job, and 5th-degree polynomial in age. Standard errors are clustered by state and are in parentheses.

period were around 9 percentage points more likely to be re-employed. This result holds across various specifications, and is large relative to the mean of the outcome variable of 69 percent. These estimates imply that, compared to displaced workers in other markets, individuals displaced from manufacturing in a metropolitan statistical area with a large housing boom were roughly 13 percent (9/69) more likely to be re-employed relatively quickly after being displaced.

The results for construction re-employment are also striking. In the results in columns 3 and 4, the point estimates suggest that displaced manufacturing workers were much more likely to be employed in construction if they became displaced in markets with big housing demand increases. The point estimate of 0.045 suggests displaced manufacturing workers in markets during the years of the housing boom in markets with big booms were likely to find re-employment in construction at a rate that was roughly 50 percent of the overall employment effect. These results suggest that a meaningful share of the employment "masking" for noncollege men at the individual level came through construction employment.

Collectively, these results provide evidence of individual-level masking. Had there been no temporary housing boom from the late 1990s through the mid-2000s, many workers displaced from manufacturing because of the ongoing decline in that sector would have been significantly more likely to end up in nonemployment during this time period.

Conclusion

This paper argues that employment gains from the recent national housing boom "masked" the adverse employment effects of declining manufacturing in the years before the Great Recession. What has been called the national "employment sag" that began in 2000 would therefore have been even larger in the absence of this masking in the years before the Great Recession. We show that aggregate masking occurred overall in the national time series, both between and within cities, and at the individual level. The sharp decline in employment that occurred during the Great Recession was due not only to cyclical forces, but also to the fact that the massive housing bust, which coincided with the start of the recession, "unmasked" the adverse employment effects of more than a decade of systematic manufacturing decline. Persistently low employment in the several years after the end of the recession points to the ongoing importance of these structural factors.

Our analysis focuses on noncollege men in the United States, but the mechanism we have highlighted is more broadly relevant. For example, many prime-aged noncollege women also lost jobs from declining manufacturing. When we do an analysis across metropolitan areas similar to that presented for their male counterparts, we find that the local decline in manufacturing had an effect on the employment rate of noncollege women that is roughly two-thirds the size of the effect for noncollege men. While housing booms in local labor markets increased employment for noncollege women, as well, for them virtually all of the increased employment came in services and related sectors (such as finance, insurance, and real estate) rather than in construction. Outside the United States, Hoffman and Lemieux (2016) have emphasized the perhaps surprising explanatory power of construction employment in accounting for cross-country patterns in employment growth in the aftermath of the Great Recession. We therefore speculate that housing booms may have "masked" the adverse effects of manufacturing decline in other countries, as well.

Our results shed light on the question of how much of the recent decline in employment rates are the result of cyclical factors, and how much from structural factors like the long-term decline in manufacturing and associated losses in routine jobs. The answer to this question is crucially important because of its implications for policy response to the falling employment. Traditional monetary and fiscal policy tools, such as temporary interest rate cuts, tax rebates or increases in government spending, are designed to provide a temporary boost to labor demand. These tools can thus offset temporary declines in hiring arising from cyclical factors like short-lived tightening of credit markets or transitory increases in uncertainty, temporarily boosting employment until the economy returns to its normal level. By contrast, there is little reason to suppose that these traditional monetary and fiscal tools can satisfactorily address employment decline arising from structural factors.

What policies might be effective in the future at raising employment rates, particularly among the relatively less-skilled? One set of options might be policies that encourage skill investment among the noncollege educated, thereby directly addressing their skill deficits and hopefully raising their long-term employment prospects. It has historically proved difficult to get people to alter their human capital choice using traditional policies like targeted taxes or educational subsidies. It is therefore likely that in order to encourage workers who have traditionally worked in routine occupations to obtain the skills demanded in the current economic environment, there will have to be experimentation with new policies ideas to spur schooling investment.

Some have suggested that the portion of employment decline attributable to structural factors might be addressed by the undertaking of large-scale, publicly financed infrastructure projects (for example, Summers 2014). A potential benefit of the public undertaking of such projects would be the boost to employment opportunities they would provide for less-educated workers, particularly if the projects raised the overall demand for such workers rather than simply reallocated them from the private to public sectors. Such investments could yield longer-term gains for lower-skilled workers if these investments in infrastructure led to broader productivity gains within the economy.

However, publicly financed infrastructure investments are not without important costs. One set of these are the various efficiency costs associated with raising public funds, even in this period of historically low interest rates. Another potential cost is that the necessarily temporary nature of infrastructure projects might have the unintended effect of adversely affecting skill-upgrading among noncollege persons. Unless infrastructure construction translates into permanent increases in labor demand for lower-skilled workers, the temporary gain in employment such projects would provide would be similar to the gain in employment from the hot housing market in the early 2000s, in the sense that underlying weakness in the labor market would be masked for a time before being revealed when the projects ended.

In other work, we have found evidence suggesting that the abundant but temporary employment opportunities provided by the housing boom during the 2000s caused some young noncollege persons to delay college-going, presumably because they erroneously believed that housing-related job opportunities would exist in the longer-term (Charles, Hurst, and Notowidigdo 2015). Many of these individuals did not return to college when those labor market opportunities vanished during the housing bust, thereby delaying the chance to obtain skills demanded in sectors like high-tech manufacturing and tradeable services. There is thus a persistently lower level of college attendance among the specific birth-year cohorts who were of college-going age during the housing boom period. Employment masking from temporary public construction projects in the future could have a similar effect.

The structural interpretation of the declining employment-to-population ratio highlights another layer to rising inequality in the United States in income and consumption between more- skilled and less-skilled workers during the last three decades. Our work highlights changes in employment inequality between higher- and lower-skilled workers since the early 2000s. While this fact has been highlighted by others (Aguiar and Hurst 2009; Attanasio, Hurst, and Pistaferri 2015), our work shows that the decline in manufacturing employment has contributed to the increased inequality in employment propensities. To the extent that employment propensities translate to earnings, our results account for some portion of the increased earnings inequality between higher- and lower-skilled workers that has occurred since the early 2000s.

Our paper is silent on the welfare implications of the masking phenomenon that we document. As noted above, in companion work, we have documented how the housing boom altered skill acquisition for lower-skilled men and women during the early 2000 boom years. However, there may have been benefits of the masking phenomenon. For example, we show evidence that masking from the housing boom postponed and thereby softened the economic costs of structural transition. The fact that the boom appears to have ameliorated the employment losses that would have otherwise occurred because of manufacturing decline may have given regions time to engage in difficult reallocation of workers across sectors, thereby easing some of the costs of adjustment. One possible benefit of masking may have been added time to develop new tradeable industries better fitted to the changing landscape of import competition. How this benefit compares to the costs of altering human capital decisions has not been studied. In future work, it would be useful to quantify the importance of these different factors.

We close with the observation that the phenomenon of employment masking studied in this paper may be important for understanding the economic consequences of sectoral shifts more broadly. A growing literature finds that large structural shifts such as the shift from agriculture to manufacturing work, and from routine jobs to nonroutine jobs, have important macroeconomic effects and may significantly affect economic growth. In some cases, these structural shifts proceeded with what appears to have been minimal effects on aggregate employment. Results from our work and that of others suggest, by contrast, that the decline of manufacturing may be associated with significant adverse effect on aggregate employment. Whether adverse macroeconomic effects arise from a structural shift may depend on the ability of workers to shift between sectors and occupations, either immediately or perhaps with some modest delay after retraining or some other form of human capital accumulation. Understanding the process of how workers switch sectors in response to large structural shifts is an important area for future research.

■ *This paper contains some results that appeared in a previously distributed NBER working paper entitled "Manufacturing Decline, Housing Booms, and Non-Employment" (NBER Working Paper 18949). We thank Hank Farber, Matt Gentzkow, Gordon Hanson, Enrico Moretti, Ann Norman, and Timothy Taylor for their comments, feedback, and suggestions.*

References

Aaronson, Daniel, Luojia Hu, Arian Seifoddini, and Daniel G. Sullivan. 2015. "Changing Labor Force Composition and the Natural Rate of Unemployment." Chicago Fed Letter 338, Federal Reserve Bank of Chicago.

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2016. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics.* 34(S1): S141–S198.

Aguiar, Mark, and Erik Hurst. 2009. *The Increase in Leisure Inequality: 1965–2005.* Washington, DC: AEI Press.

Attanasio, Orazio, Erik Hurst, and Luigi Pistaferri. 2015. "The Evolution of Income, Consumption and Leisure Inequality in the United States, 1980–2010." Chap. 4 in *Improving the Measurement of Consumer Expenditures*, edited by Christopher Carroll, Thomas Crossley, and John Sabelhaus. National Bureau of Economic Research.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103(5): 1553–97.

Autor, David H., David Dorn, and Gordon Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.

Autor, David H., and Mark Duggan. 2006. "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." *Journal of Economic Perspectives* 20(3): 71–96.

Autor, David H., and Mark Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118(1): 157–206.

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.

Autor, David H., Nicole Maestas, Kathleen Mullen, and Alexander Strand. 2015. "Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants." NBER Working Paper 20840.

Baker, Scott R., Nickholas Bloom, and Steven J. Davis. 2015. "Measuring Economic Policy Uncertainty." NBER Working Paper 21633.

Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers." *Quarterly Journal of Economics* 109(2): 367–97.

Bound, John, and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics* 18(1): 20–54.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "Housing Booms, Manufacturing Decline, and Labor Market Outcomes." March. http://faculty.wcas.northwestern.edu/noto/research/CHN_manuf_decline_housing_booms_mar2016.pdf.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2015. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." NBER Working Paper 21587.

Chodorow-Reich, Gabriel. 2014. "The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008–09 Financial Crisis." *Quarterly Journal of Economics* 129(1): 1–59.

Davis, Steven J., and John Haltiwanger. 2014. "Labor Market Fluidity and Economic Performance." NBER Working Paper 20479.

Eggertsson, Gauti B., and Paul Krugman. 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics* 127(3): 1469–1513.

Farber, Henry S. 2015. "Job Loss in the Great Recession and its Aftermath: U.S. Evidence from the Displaced Worker Survey." Working Paper 589, Industrial Relations Section, Princeton University.

Farber, Henry S., and Robert G. Valetta. 2013. "Do Extended Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the U.S. Labor Market." NBER Working Paper 19048.

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez. 2015. *American Economic Review* 105(11): 3352–84.

Giroud, Xavier, and Holger M. Mueller. 2015. "Firm Leverage and Unemployment during the Great Recession." NBER Working paper 21076.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Q. Nguyen. 2015. "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2187521.

Guerrieri, Veronica, and Guido Lorenzoni. 2011. "Credit Crises, Precautionary Savings and the Liquidity Trap." NBER Working Paper 17583.

Hagedorn, Marcus, Fatih Karahan, Iourii Manovski, and Kurt Mitman. 2013. "Unemployment Benefits and Unemployment in the Great

Recession: The Role of Macro Effects." NBER Working Paper 19499.

Hall, Robert E. 2011. "The Long Slump." *American Economic Review* 101(2): 431–69.

Hall, Robert E. 2014. "Secular Stagnation." http://web.stanford.edu/~rehall/SecStag.

Hoffman, Florian, and Thomas Lemieux. 2016. "Unemployment in the Great Recession: A Comparison of Germany, Canada and the United States." 34(S1): S95–139.

Jaimovich, Nir, and Henry E. Siu. 2012. "The Trend is the Cycle: Job Polarization and Jobless Recoveries." NBER Working Paper 18334.

Johnston, Andrew, and Alexandre Mas. 2015. "Potential Unemployment Insurance Duration and Labor Supply: The Individual and Market-Level Response to a Benefit Cut." Working Paper 590, Princeton University, Industrial Relations Section.

Kroft, Kory, Fabian Lange, Matthew Notowidigdo, and Lawrence F. Katz. 2016. "Long-Term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Nonparticipation." *Journal of Labor Economics* 34(S1, Part 2): S7–S54.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103(5): 1797–1829.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2015. "Disability Insurance and the Great Recession." *American Economic Review* 105(5): 177–82.

Mian, Atif, and Amir Sufi. 2014. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82(6): 2197–2223.

Mian, Atif and Amir Sufi. 2011. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis." *American Economic Review* 101(5): 2132–56.

Moffitt, Robert A. 2012. "The Reversal of the U.S. Employment–Population Ratio in the 2000s: Facts and Explanations." *Brookings Papers for Economic Activity*, Fall, 201–264.

Mondragon, John. 2015. "Household Credit and Employment in the Great Recession." http://www.voxeu.org/article/household-credit-and-employment-great-recession.

Mulligan, Casey B. 2012. *The Redistribution Recession: How Labor Market Distortions Contracted the Economy.* New York: Oxford University Press.

Pierce, Justin R., and Peter K. Schott. Forthcoming. "The Surprisingly Swift Decline of U.S. Manufacturing Employment." *American Economic Review.*

Rothstein, Jesse. 2011. "Unemployment Insurance and Job Search in the Great Recession." *Brookings Papers on Economic Activity*, Fall, 143–96.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota. https://usa.ipums.org/usa/.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni Violante. 2014. "Mismatch Unemployment." *American Economic Review* 104(11): 3529–64.

Sloane, Carolyn M. 2015. "Where Are the Workers? Technological Change, Rising Disability and the Employment Puzzle of the 2000s: A Regional Approach." https://sites.google.com/site/sloanecarolyn/research/working-papers/where%20are%20the%20workers_text_main_tables.pdf?attredirects=0&d=1.

Summers, Lawrence H. 2014. "U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." Keynote Address at the NABE Policy Conference, February 24, 2014. *Business Economics* 49(2): 65–72.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 201–218

Going for the Gold: The Economics of the Olympics

Robert A. Baade and Victor A. Matheson

In summer 2016, the eyes of the world will turn to Rio de Janeiro as it hosts the Games of the XXXI Olympiad, better known as the Summer Olympics. Unfortunately, the price tag of well over $10 billion for the event is adding to the already considerable strain on government budgets in Brazil. Faced with a nasty recession, cuts in public services, and rising unemployment, throngs of Brazilians have turned out to protest what is seen as wasteful spending and a misallocation of resources on the Olympics. Throw in the growing threat of the Zika virus and Brazil may end up with larger crowds of agitators protesting the government than of sports fans cheering on the athletes. But are these complaints about Olympic spending justified? The quadrennial Summer Olympic Games is one of the world's premier sporting events, with over 10,000 athletes representing 204 countries, 300 individual events in 28 different sports, over 10 million tickets sold to spectators, and a worldwide television audience in the billions. On a somewhat smaller scale, the most recent Winter Olympic Games held in 2014 in Sochi, Russia, welcomed nearly 3,000 athletes from 88 countries to compete in 98 events in 15 disciplines while generating large revenues and massive television ratings.

While most viewers tune in to watch the competition among the athletes, the battle among cities to be selected to host these events can be just as fierce. Although bidding cities have numerous reasons for wanting to host, none seems more prevalent than the desire for an economic windfall. In this paper, we explore the costs and

■ *Robert A. Baade is A. B. Dick Professor of Economics, Lake Forest College, Lake Forest, Illinois. Victor A. Matheson is Professor of Economics, College of the Holy Cross, Worcester, Massachusetts. Their email addresses are baade@mx.lakeforest.edu and vmatheso@holycross.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.201 doi=10.1257/jep.30.2.201

benefits of hosting the Olympic Games. On the cost side, there are three major categories: general infrastructure such as transportation and housing to accommodate athletes and fans; specific sports infrastructure required for competition venues; and operational costs, including general administration as well as the opening and closing ceremony and security. Three major categories of benefits also exist: the short-run benefits of tourist spending during the Games; the long-run benefits or the "Olympic legacy" which might include improvements in infrastructure and increased trade, foreign investment, or tourism after the Games; and intangible benefits such as the "feel-good effect" or civic pride.

Each of these costs and benefits will be addressed in turn, but the overwhelming conclusion is that in most cases the Olympics are a money-losing proposition for host cities; they result in positive net benefits only under very specific and unusual circumstances. Furthermore, the cost–benefit proposition is worse for cities in developing countries than for those in the industrialized world. In closing, we discuss why what looks like an increasingly poor investment decision on the part of cities still receives significant bidding interest and whether changes in the bidding process of the International Olympic Committee (IOC) will improve outcomes for potential hosts.

The Costs of Hosting the Olympics

The modern Summer Olympic Games date back to 1896, and the Winter Games commenced in 1924. The host cities are selected roughly seven years before the event through an open-bidding process. The host cities are responsible for the entire bill for organizing the event, although the International Olympic Committee typically provides some funds to help defray the costs. Historically, host cities have come almost exclusively from rich, industrialized nations. Between 1896 and 1998, over 90 percent of all host cities came from Western Europe, the United States, or Canada, Australia, and Japan. Only Mexico City, Moscow, and Seoul—hosts of the 1968, 1980, and 1988 Summer Games, respectively—and Sarajevo, host of the 1984 Winter Games, bucked this trend.

More recently, the International Olympic Committee has encouraged bids from developing countries and has awarded the games on multiple occasions to cities outside the regions that had traditionally served as hosts. The 2008 Summer Games were hosted by Beijing, China, which will in turn host the Winter Olympics in 2022. The 2016 Summer Olympics will be held in Rio de Janeiro, Brazil, the first time the event has taken place in South America. The 2014 Winter Olympics were in Sochi, Russia, with PyeongChang, South Korea, to follow in 2018.

As seen in Table 1, the composition of countries submitting formal bids has also changed dramatically in recent decades. Only 18 percent of the bids submitted for the Summer Games prior to 2000 came from the developing world or the former Soviet sphere of influence. Since that time, however, over half of all bids have come from this group, including applications by Istanbul, Bangkok, Havana, Doha, and Cape Town, as well as the successful bids from Beijing and Rio de Janiero. For

Table 1
Number of Bids for Summer and Winter Olympic Games

	Bidders			*Hosts*		
Event	*Industrialized countries*	*Developing countries*	*Eastern European/ Former Soviet states*	*Industrialized countries*	*Developing countries*	*Eastern European/ Former Soviet states*
Summer Olympics:						
1896–1996	71 (82%)	9 (10%)	7 (8%)	20 (87%)	2 (9%)	1 (4%)
2000–2020	23 (49%)	21 (44%)	4 (7%)	4 (67%)	2 (33%)	0 (0%)
Winter Olympics:						
1924–1998	51 (93%)	1 (2%)	3 (5%)	17 (94%)	0 (0%)	1 (6%)
2002–2022	21 (56%)	4 (9%)	12 (34%)	4 (67%)	1 (17%)	1 (17%)

the Winter Olympics, the past decade has witnessed for the first time bids from Kazakhstan, Georgia, China, Slovakia, and Poland.

Bidding for the Olympics is no small undertaking. A key to the bidding process involves a visit by the Evaluation Commission of the International Olympic Committee which assesses the condition of the applicant city. A significant portion of the bidding expense relates to the preparations the applicant city undertakes to impress the Evaluation Commission, and these plans, including detailed architectural renderings, financial estimates, and pre-event marketing, are likely to be extensive since it cannot be known what the preparations of the other applicant cities will be. Chicago, for example, spent at least $70 million and perhaps over $100 million on its unsuccessful application to host the 2016 Games (Pletz 2010; Zimbalist 2015). But the costs of the formal bidding process pale in comparison to the expenses a region will incur should it actually be selected by the International Olympic Committee.

A first set of major expenses involves general infrastructure to accommodate the anticipated wave of tourists and athletes that descend upon the chosen city. The International Olympic Committee requires that the host city for the Summer Games have a minimum of 40,000 hotel rooms available for spectators and an Olympic Village capable of housing 15,000 athletes and officials. In addition, the city needs to have both internal and external transportation facilities that can get tourists to the city itself and then to the individual sports venues within the region. Hotel capacity alone can be a major challenge. Rio de Janeiro, already one of the most popular tourist destinations in South America, still required the construction of over 15,000 new hotel rooms for the 2016 Summer Games. While investment in the hospitality industry can in theory pay long-term dividends once the Games are over, heavy expenditures to meet a two-week period of peak demand may result in severe overcapacity once the event is over. For example, following the 1994 Winter Olympics in Lillehammer, Norway, 40 percent of the town's full-service hotels went bankrupt (Teiglund 1999).

The Olympics also require spending on specialized sports infrastructure. Because of the somewhat obscure nature of many of the events, most cities do not have the facilities in place to host all of the competitions, especially if large spectator viewing areas are desired. Even modern cities in high-income countries may need to build or expand an existing velodrome, natatorium, ski-jumping complex, or speed skating oval. Furthermore, modern football and soccer stadiums are generally incompatible with a full-size Olympic track, because including space for such a track would cause an undesirably large separation between the fans and the playing field. For this reason, Boston's failed bid to host the 2024 Summer Games had proposed $400 million to build an entirely new stadium for the track and field events, despite the presence of four large existing outdoor sports stadiums in the area.

Once the facilities are in place, the Games require spending for operations including event management, the opening and closing ceremonies, and security. The Olympics have long been a target for terrorists and have suffered deadly attacks in both 1972 in Munich and 1996 in Atlanta. In the era of post-September 11, 2001, security costs have escalated rapidly. Security costs for Sydney's Games in 2000 totaled $250 million, while four years later in Athens, security expenditures topped $1.6 billion, four times the initial budget, and have stayed near this figure for the past decade (Matheson 2013).

An accurate financial accounting of Olympic expenditures in various cities is hard to find for multiple reasons. It can be difficult to disentangle spending on Olympic building projects from planned infrastructure improvements that might not be attributable directly to the games. Moreover, concerns about cost overruns or corruption may prompt officials to limit the release of accurate data. The true final cost of the 1998 Nagano Winter Olympics will never be known, because the host committee ordered a portion of the event's financial records to be burned (Jordan and Sullivan 1999). While we keep these concerns in mind, Table 2 shows some cost estimates for recent Olympic Games as provided by the International Olympic Committee, host committees, and various academic or public media sources with spending on sports infrastructure and general infrastructure broken out where possible.

Finally, it is important to note the Olympics have consistently produced final costs that exceeded their original budgets. From 1968 to 2012, every single Olympic Games ended up costing more than originally estimated. The median Games were 150 percent over the original budget, with the worst offenders—Montreal 1976 and Sarajevo 1984—exceeding initial estimates by more than ten-fold (Flyvbjerg and Stewart 2012). The 2012 London organizers originally won the bid in 2005 with a cost estimate of £2.4 billion, which was revised upwards within two years to £9.3 billion. Then, when the final costs came in at a mere £8.77 billion, the organizers laughably claimed the event had come in under budget (BBC 2013).

The Short-Run Benefits of Hosting the Olympics

Although the costs of hosting can be daunting, the local Organizing Committees for the Olympic Games point to both a short-run boost from the construction phase preceding the event as well as tourism bumps during the Games and the long-run

Table 2
Costs of Hosting Recent Olympic Games

	Type of spending	*Spending (billions, 2015$)*	*Source*
Summer Olympics			
Seoul, 1988	Sports infrastructure	$2.067	Preuss (2004, Table 7.8 and Figure 9.1)
	General infrastructure	$3.523	
	Total cost	**$ 6.503**	
Barcelona, 1992	Sports infrastructure	$1.485	Preuss (2004)
	General infrastructure	$12.457	
	Total cost	**$16.409**	
Atlanta, 1996	Sports infrastructure	$.765	Preuss (2004)
	General infrastructure	$.959	
	Total cost	**$3.576**	
Sydney, 2000	Sports infrastructure	$1.761	Preuss (2004)
	General infrastructure	$1.817	
	Total cost	**$6.926**	
Athens, 2004	**Total cost**	**$13.800** (est.)	Tagaris (2014)
Beijing, 2008	Sports infrastructure	$2.315	Preuss (2004)
	Total cost (est.)	**$45.000** (est.)	Fowler and Meichtry (2008)
London, 2012	**Total cost**	**$11.401**	BBC (2012b)
Rio 2016	**Total cost**	**$11.100** (est.)	Leme (2015)
Winter Olympics			
Nagano, 1998	**Total cost**	**$15.250**	Longman (1998); *The Economist* (1998)
Salt Lake City, 2002	**Total cost**	**$2.500** (approx.)	US GAO (2001)
Torino, 2006	**Total cost**	**$4.350** (approx.)	Payne (2008); Flyvbjerg and Stewart (2012)
Vancouver, 2010	Sports infrastructure	$.715	VanWynsberghe (2011)
	General infrastructure	$3.497	
	Total cost	**$7.556**	
Sochi, 2014	Sports infrastructure	$6.700 (est.)	Farhi (2014)
	Total cost	**$51.000** (est.)	

legacy effect of the Games as an economic justification for hosting these events. In addition, the Olympics do generate significant sponsor, ticketing, licensing, and media revenues that can be used to offset the costs of staging the event.

Table 3 shows data on revenues generated by the International Olympic Committee and the organizing committees for the Vancouver and London Games from the most recent IOC four-year budget cycle. In theory, the revenues generated from the Games can be divvied up any way the organizers see fit, but ultimately the IOC exercises complete control over the event and can share as much or as little of largesse as they deem fit subject to the constraint of finding

Table 3
Direct Revenues and Hosting Costs from Olympic Games
($ millions)

	IOC 2009–12	*Vancouver 2010 organizing committee*	*London 2012 organizing committee*
Revenue source			
Broadcast rights	$2,723	$414	$713
International sponsors	$475	$175 (est.)	$300 (est.)
Domestic sponsors	$0	$688	$1,150
Ticketing	$0	$250	$988
Licensing	$0	$51	$119
Total	**$3,198**	**$1,578**	**$3,270**
Hosting costs	-	$7,556	$11,401

Source: IOC (2014b).
Notes: Table 3 shows data on revenues generated by the International Olympic Committee and the organizing committees for the Vancouver and London Games over the 2009–2012, the most recent IOC budget cycle. It also shows hosting costs for the Vancouver and London Games.

a city willing to host the event. Most recently, television rights have represented nearly half of total revenues with the IOC sharing less than 30 percent of the total with the local Organizing Committees. Revenues from international sponsors are split between the International Olympic Committee and the Organizing Committees, while ticket revenue, domestic sponsorships, and licensing fees are kept by the host city. Obviously, the IOC could provide more generous subsidies to cities in order to defray the costs of hosting their tournaments, and international sports governing bodies, including the IOC, are often known for their lavish expenses. However, in the case of the London and Vancouver Games, the direct revenues generated by the games represented only a fraction of the total costs of hosting the event and would not have come close to covering the total costs even if the IOC had committed all revenue streams to the host committees, so one must rely on other sources of benefits to provide an economic justification for the events.

Any large public works project such as the Olympics can lead to a short-run increase in economic activity in the run-up to the opening, depending on the level of slack in a region's labor and capital markets, and act as a form of an expansionary fiscal policy. It is perhaps telling to note that at the same time David Cameron's government in the United Kingdom was promoting the supposed expansionary effects of fiscal austerity in the wake of the Great Recession, the same government was touting the stimulative effects of increased government spending on London's Olympic preparations (Mullholland 2012). However, unless policymakers can predict recessions years ahead of time—given that the International Olympic Committee awards the Games seven years in advance—using the Olympics to pull a country out of recession would rest more on dumb luck rather than prudent planning. Otherwise, the spending involved with the Games is as likely to redistribute spending in an economy near full employment as it is to lift an economy out of recession. Indeed, unless unemployment is high, employment gains in construction are not an important economic benefit since they come at the cost of employment losses in other industries.

That being said, various economic impact studies done in advance of the Olympic Games have often produced large estimates of economic gains. An InterVISTAS Consulting (2002) report on the 2010 Vancouver Winter Olympics predicted $10.7 billion (Canadian) in new economic output and 244,000 jobs compared to $4.8 billion (in 2002 dollars) and 35,000 job-years predicted in Salt Lake City eight years earlier by the state government of Utah (IOC 2010). The 1996 Atlanta Games were predicted to generate 77,026 jobs and $5.142 billion (in 1996 dollars) in economic activity, while the London Olympics promised £1.936 billion in economic activity and an additional 8,164 full-time equivalent jobs created (Humphreys and Plummer 1995; Blake 2005).

The variation alone in these estimates suggests some reason for concern about their accuracy; indeed, these before-the-Games predictions are rarely matched by reality when economists look back at the data. Table 4 shows academic studies of various Olympic Games. Overwhelmingly, the studies show actual economic impacts that are either near-zero or a fraction of that predicted prior to the event. Nearly all of the analyses follow the same pattern. Researchers collect any type of regional economic data that is readily available such as employment, personal income, GDP, tax collections, or tourism figures, and then analyze the data before, during, and after the Olympics in search of any changes that occur either during the event or in the preparation stages. The observed changes in economic variables are then compared to the predictions made by the Olympic organizers prior to the event.

For example, as noted previously, the Utah state government predicted the 2002 Winter Olympics would generate 35,000 job-years, concentrated primarily in the year of the event itself. Baumann, Engelhardt, and Matheson (2012) examine monthly employment overall as well as in a variety of specific industries such as retail trade and leisure between 1990 and 2009 in Utah using employment in several adjacent states to control for regional employment trends around the time of the Olympics. They find no identifiable increase in employment either before or after the Olympics, and while they find a statistically significant bump in employment during the actual Games, the increase was 4,000 to 7,000 jobs, or roughly one-quarter to one-tenth the number claimed by Utah officials. Considering that the federal government spent $342 million directly on the 2002 Olympics and at least another $1.1 billion on infrastructure improvements leading up the Games, this amounts to about $300,000 in federal government spending per job created. Other studies listed in Table 4 find similar outcomes. Indeed, these results lend credence to a common rule-of-thumb often used by economists who study mega-events: If one wishes to know the true economic impact of an event, take whatever numbers the promoters are touting and move the decimal point one place to the left.

Table 4
Academic Studies of the Economic Impact of the Olympic Games

Study	*Event*	*Results*
Baade and Matheson (2002)	1984 Summer Games (Los Angeles) and 1996 Summer Games (Atlanta)	5,043 new jobs in Los Angeles. Between 3,467 and 42,448 new jobs in Atlanta.
Jasmand and Maennig (2008)	1972 Summer Games (Munich)	No impact on employment in host regions. Positive impact on income.
Porter and Fletcher (2008)	1996 Summer Games (Atlanta) and 2002 Winter Games (Salt Lake City)	No impact on taxable sales, hotel occupancy, or airport usage. Significant increase in hotel prices.
Baade, Baumann, and Matheson (2010)	2002 Winter Games (Salt Lake City)	Taxable sales in restaurants and hotels up by $70.6 million but taxable sales at general merchandisers down by $167.4 million.
Giesecke and Madden (2011)	2000 Summer Games (Sydney)	Household consumption in Australia reduced by $2.1 billion.
Baumann, Engelhardt, and Matheson (2012)	2002 Winter Games (Salt Lake City)	Increase in employment of 4,000–7,000 jobs for one year compared to predictions of 35,000 full-time equivalent job-years.
Hotchkiss, Moore, and Zobay (2003)	1996 Summer Games (Atlanta)	Increase in employment of 293,000 jobs. Increase in employment growth rate by 0.2%.
Feddersen and Maennig (2013)	1996 Summer Games (Atlanta)	29,000 jobs added during month of Olympics only.

These results beg the question: Why do before-the-Games economic impact studies rarely stand up to after-the-Games scrutiny? One obvious answer is that economic impact studies are often commissioned by groups who have a vested interest in their outcome, and these groups choose firms that are likely to produce a favorable result. Estimates can be easily manipulated by making unrealistic assumptions about costs and benefits. The resulting claim of a large economic windfall may be used to curry public favor or to justify a large taxpayer subsidy.

Even when a highly positive estimate of Olympic benefits is not the explicit goal of an economic impact study, the methodology used in most studies is flawed in a way that biases the economic impact upwards. First, economic impact studies often ignore the "substitution effect" that occurs when local residents shift their spending from other goods in the local economy to the Olympics. If the study counts the purchase of a ticket by a local resident to an Olympic event without accounting for what would have been purchased in the absence of the Games, the impact of the Olympics will be overstated. For this reason, economists studying the effect of sporting events on local economies often advocate eliminating expenditures by local residents entirely.

Second, the "crowding out effect" occurs when the crowds and congestion associated with a mega-event dissuades other regular tourists or business travelers from visiting the host region. Even when the number of out-of-town Olympics spectators is large, hotel rooms in the host city may normally be nearly full so that the net increase in visitor arrivals to the region is likely to be much smaller and perhaps even negative. For example, the UK Office for National Statistics (2015) reported that the number of international visitors to the country fell to 6,174,000 visitors in July and August 2012, the months of the Olympics, from 6,568,000 the year before, and some popular shows in London's theater district actually shut down during the Games. Similarly, Beijing reported a 30 percent drop in international visitors and a 39 percent drop in hotel occupancy during the month of the 2008 Games compared to the previous year. Utah ski resorts noted a 9.9 percent fall in skier days in the 2001–02 season during which the Salt Lake City Winter Games occurred, compared to the previous year along with a drop in taxable sales collections at these locations (Zimbalist 2015; Baade, Baumann, and Matheson 2010). Taxable sales and skier visits rebounded the following season, after the departure of the Olympic fans and athletes. Other host cities that have experienced an increase in visitors during the Olympics still routinely report net increases in tourism that are significantly below expectations—and typically lower than the number of identified ticket buyers. American baseball player Yogi Berra's famous quip, "Nobody goes there anymore. It's too crowded," may apply here.

The third main failing of standard before-the-fact economic impact analysis is the problem of choosing an appropriate multiplier for expenditures. Clearly some level of tourist spending will recirculate in the economy as local businesses and workers re-spend a portion of any Olympic windfall that comes their way. At a very basic level, standard macroeconomic analysis suggests that the expenditure multiplier will be

$$\Delta Y/\Delta \text{spending} = \Delta Y/\text{Olympic spending} = 1/(1 - \text{marginal propensity to consume}),$$

so that a \$1 increase in spending due to the Olympics will result in 1/(1 − MPC) extra dollars in total output for the host city. While every city and industry is different, it is common to see multipliers of roughly 2 applied to visitor spending, so that an initial increase in direct spending leads to a similar level of indirect spending and a doubling of the total economic impact.

Several tools can be used to potentially produce more precise economic impact estimates including the Regional Input-Output Multiplier System (RIMS II) provided by the US Bureau of Economic Analysis and IMpact for PLANning (IMPLAN), a commercially available software package. Both models use input-output tables for specific industries grounded in interindustry relationships within regions based upon an economic area's normal production patterns. But as Matheson (2009) notes: "During an event like the Olympics, however, the economy within a region may be anything but normal, and therefore, these same inter-industry relationships may not hold. Since there is no reason to believe that the usual economic multipliers are the same during mega-events, any economic analyses based upon these multipliers may, therefore, be highly inaccurate."

The hotel industry offers case in point. Even critics of the Olympics like Porter and Fletcher (2008) concede that the Olympics typically cause a substantial increase in room rates. The wages paid to a hotel's desk clerks and room cleaners, however, are likely to remain roughly unchanged. As a hotel's revenue increases without a corresponding increase in labor costs, the return to capital rises while the return to labor falls as a percent of revenues. To the extent that hotels (as well as chain restaurants, car rental agencies, airlines, and similar firms) are nationally or internationally owned, this increase in corporate profits doesn't stick in the host city but instead leaves the area in which the profits were earned. In effect, due to these increased leakages, the MPC in the host city falls, thus reducing the multiplier effect during mega-events.

Replacing input-output models with computable general equilibrium (CGE) models that account for capacity constraints, displacement, expenditure shifting, price changes, and changing economic conditions can lead to improved estimates for the economic impact of the Olympics, although the use of these models is a much more difficult undertaking. As one example, Giesecke and Madden (2011) carried out a retrospective examination of the 2000 Sydney Olympics using a CGE model. They found a *reduction* in total consumption for Australia of $2.1 billion; in contrast, before-the-Games estimates that didn't account for the degree of slack in labor markets and assumed no displacement of international tourism predicted increases in consumption of $2.5 billion over the same period.

While spending directly associated with the Olympics is typically insufficient to cover the costs of staging the Games, short-run intangible benefits must also be considered. Host cities frequently experience a "feel-good effect" both in the run-up to and in the wake of mega-events. For example, 80 percent of respondents surveyed by the BBC (2012) immediately after the 2012 Olympics reported that the event "made them more proud to be British." Several studies have attempted to quantify the intangible benefits of the Olympics through the use of contingent valuation methodology, which constructs a set of survey questions that are designed to elicit the monetary value people place on whether certain events occur or do not occur. Using this approach, both Atkinson, Mourato, Szymanski, and Ozdemiroglu (2008) and Walton, Longo, and Dawson (2008) undertook sophisticated contingent valuation surveys using best practices for the 2012 London Olympics and found that persons both within London and throughout the United Kingdom expressed a willingness to pay to host the Games over and above any costs associated with actually attending any of the events. The total intangible value identified to UK residents in the studies was approximately £2 billion (or roughly $3.4 billion at the exchange rates at the time of the study). This amount is clearly substantial, but it is well below the cost of hosting the Games.

Given the expenses associated with specialized venues and event operations, especially security, it is difficult for the revenues directly generated by the

Olympics or the surrounding tourism to cover the cost of the event. Allowing for a "feel-good effect" doesn't close the gap, either. Thus, an economic justification for the Olympics must rest on including additional benefits from the long-run legacy of the Games.

The Long-Run Benefits of Hosting the Olympics

The arguments that the Olympics bring long-term benefits fall into several categories. First, the Games might leave a legacy of sporting facilities that can be used by future generations. Second, investments in general infrastructure can provide long-run returns and improve the livability of host cities. Third, the media attention surrounding the Games can serve as an advertising campaign that serves to promote the area as a destination for future tourism. Finally, the Olympics can promote foreign direct investment and increased international trade, as the Olympics causes investors and companies worldwide to become familiar with the area.

A positive legacy of sporting facilities is the least promising of these claims. Academic studies of sports facilities on host communities are nearly unanimous in finding little or no economic benefits associated with stadiums and arenas (Coates and Humphreys 2008). Furthermore, due to the nature of the sporting events sponsored by the Olympics, host cities are often left with specialized sports infrastructure that has little use beyond the Games, so that in addition to the initial construction costs, cities may be faced with heavy long-term expenses for the maintenance of "white elephants." Many of the venues from the Athens Games in 2004 have fallen into disrepair. Beijing's iconic "Bird's Nest" Stadium has rarely been used since 2008 and has been partially converted into apartments, while the swimming facility next door dubbed the "Water Cube" was repurposed as an indoor water park at a cost exceeding $50 million (Farrar 2010). The Stadium at Queen Elizabeth Olympic Park in London, the site for most of the track and field events as well as the opening and closing ceremonies in 2012, was designed to be converted into a soccer stadium for local club West Ham United in order to avoid the "white elephant" problem. Before the Games, the stadium had an original price tag of £280 million. Cost overruns led to a final construction cost of £429 million, and then the conversion cost to remove the track and prepare the facility to accommodate soccer matches topped £272 million, of which the local club is paying only £15 million (Sky Sports 2015).

General infrastructure improvements clearly have the potential for better returns. The athletes' villages in both Atlanta and Los Angeles were converted into new dormitories for local universities in their respective cities, and Utah wound up with expanded highways between its major population center in Salt Lake City and the popular ski resorts in the mountains to its east. But here, too, a caveat is in order. It is often argued that the Olympics can serve as a catalyst for urban redevelopment and to generate the political will required to undertake needed infrastructure investments. However, there is no reason to believe that the investments required to host the Olympics will provide higher returns than alternative infrastructure projects that could have been carried out instead. Also, while the firm deadlines provided by the Olympics may constrain cities to follow projects through to timely completion, the same deadlines may raise costs due to time pressures and labor constraints.

The Olympics can serve to "put a city on the map" as a tourist destination. In 1990, Barcelona was the 13th most popular tourist destination in Europe with fewer than half the number of bed nights as its neighboring rival, Madrid. Following the 1992 Summer Olympics that also highlighted many nonsports venues in the region, the city experienced the fastest growth in tourism among large European cities, so that by 2010 the city was the fifth most popular destination on the continent and had eclipsed Madrid in bed nights (Zimbalist 2015). Similarly, ski resorts in Utah experienced a 20.4 percent increase in skier visits between the year before the Salt Lake City Games in 2000–01 and 2014–15, outpacing Colorado's 8.0 percent growth over the same period.

However, the results in Salt Lake City and Barcelona have not been replicated in other host cities. The explanation for their success may be that both of these locations can be seen as "hidden gems," locations that are highly attractive to tourists but that had been previously passed over for their better-known neighbors in Colorado and Madrid. This strategy won't necessarily work for many other potential host cities. Lillehammer, Norway, the venue for the Winter Games in 1994, offered few attractions to tourists outside of the Olympic events and was therefore unattractive to tourists after the Games left town. By 1997, the increase in international guest-nights in Lillehammer was only 8 percent higher than the increase in foreign tourism in Norway overall (Tiegland 1999). Similarly, the 1988 Calgary Winter Olympics significantly raised international awareness of the city, but without a lasting ability to attract tourists, the enhanced image of the city rapidly faded (Richie and Smith 1991). Conversely, London, with over 18 million international visitors per year, was already the most popular tourist destination in the world prior to the 2012 Olympics, and it was never likely that the event would raise its already impressive profile. The success of the Olympics in developing a city as a tourist destination should not be rejected out of hand, but neither is it a surefire way to ensure a steady stream of visitors after the closing ceremonies.

A final economic justification for hosting the Olympics is that the Games can serve as positive signal to businesses and consumers about the future state of the economy. Using regression analysis of time-series panel data, Rose and Spiegel (2011) examine exports from 196 countries and territories between 1950 and 2006 and find that countries that host the Olympics experience an increase in exports of over 20 percent. Using a similar methodology, Brückner and Pappa (2015) examine consumption, investment, and output data over a similar time frame and range of countries and discover that all three measures of economic activity rise significantly around the time that the host country makes its initial bid as well as two to five years before the event actually takes place. On the surface, these results appear to vindicate the massive expenditures that are routinely incurred when hosting the Games. However, the same studies also show that unsuccessfully bidding for the Olympics appears to have similar effects on these economic variables.

There are several possible explanations for these surprising results. Rose and Spiegel (2011) suggest that it is not the event itself or the resulting tourism or advertising that increases exports, but rather that the very act of bidding serves as a credible signal that a country is committing itself to trade liberalization that will permanently increase trade flows. Brückner and Pappa (2015) theorize that the announcement of a bid for the Olympics represents a news shock predicting increases in future government investment.

While signaling and news shocks may be important drivers of modern economies, it is a bit hard to swallow the claim that the mere act of a single city within a country bidding for the right to throw a three-week party seven years in the future can result in enormous nationwide increases in trade, investment, and income. A more plausible answer is that countries are not randomly chosen to bid for the Games, but rather that bidding nations are almost exclusively drawn from a set of countries with sound economies and bright prospects for the future—a clear case of selection bias. To test for spurious correlation, Maennig and Richter (2012) and Langer, Maennig, and Richter (2015) note that when bidding countries are appropriately compared with countries that are otherwise similar but did not bid for the Games using propensity matching techniques, the significant Olympic effects on trade, consumption, investment, and income all disappear. Again, the long-run benefits of hosting the Games prove to be elusive.

Why Do Countries Continue To Host?

If the Olympic Games tend to offer only a low chance of providing host cities with positive net benefits, why do cities keep lining up to host these events? At least three possibilities arise. First, even if the overall effect of holding the Games is typically negative, large projects will still create winners and losers. Boston's ultimately unsuccessful bid to host the 2024 Summer Games was spearheaded by leaders in the heavy construction and hospitality industries, the two sectors of the economy that stood the most to gain from the city hosting the Olympics.

Second, economic concerns may only play a small role in a country's decision whether or not to stage the Olympics. The desire to host the Games may be driven by the egos of a country's leaders or as a demonstration of a country's political and economic power. It is difficult to explain Russia's $51 billion expenditure on the 2014 Sochi Games or China's $45 billion investment in the 2008 Beijing Summer Olympics otherwise. In countries where the government is not accountable to voters or taxpayers, it is quite possible for the government to engage in wasteful spending that enriches a small group of private industrialists or government leaders without repercussions. In the bidding for the 2022 Winter Olympics, four of the cities in liberal western democracies that initially indicated interest in staging the Games—Oslo, Stockholm, Krakow, and Munich—withdrew from the bidding after local voters expressed opposition to the bids, leaving the International Olympic Committee to choose which autocratic regime would hold the event: Beijing, China, or Almaty, Kazakhstan. In the bidding for the 2024 Summer Olympics, both Boston and Hamburg withdrew their bids in the face of public opposition.

Finally, it is possible to ascribe a portion of the economic failings of the Olympics to the "winner's curse," the result in auction theory that when parties are bidding on an asset of uncertain value (like rights to offshore oil leasing tracts), the winner will tend to be the bidder who is most prone to overestimating the value of the asset—which means that the winner is likely to be systematically disappointed (for an overview of the "winner's curse," see Thaler 1988). The 1970s witnessed a decline in enthusiasm among cities willing to host the Games. In 1972, voters in Denver, after having been initially awarded the 1976 Winter Olympics, rejected a $5 million bond referendum that would have been used to finance the Games, requiring the International Olympic Committee to rescind its offer. Following the financial debacle of the 1976 Montreal Olympics, by the time it came to award the 1984 Summer Games, Los Angeles was the only bidder. Given the resulting bargaining position, the Los Angeles Organizing Committee was able to dictate the terms of bid to the International Olympic Committee. For example, it insisted on utilizing the area's existing sports infrastructure, including the 60-year old Los Angeles Coliseum for the premier track and field events as well as the opening and closing ceremonies, and the heavy use of corporate sponsors to finance the Games. The focus on restraining costs resulted in total expenditures for the Games of a "mere" $546 million ($1,244 million in 2015 dollars), less than one-quarter of that spent by Montreal eight years earlier. The 1984 Los Angeles event managed to become one of the only profitable Games in Olympic history, with a final profit of $232.5 million (Walker 2014).

When Los Angeles had shown the possibility of profits from the Games, it led multiple cities to enter the bidding process, each hoping to cash in on the potential Olympic windfall. However, this crop of new entrants meant that bargaining power shifted back to the International Olympic Committee. No longer could cities design bids based solely on expected revenues and the expenses necessary to stage the event. Instead, applicant cities needed to consider how to beat competing bids from other potential hosts. Not only did the competition among cities to host create a bidding environment prone to corruption, but it became commonplace for bidders to attempt to impress the International Olympic Committee with spectacular new architectural monuments like Beijing's Bird's Nest or the £269 million London Aquatics Centre. The estimated cost of the new, ultra-modern National Olympic Stadium in Tokyo, planned as the centerpiece of the 2020 Games, eventually rose to $2.02 billion—which for perspective was nearly twice the cost, even after accounting for inflation, of the entire 1984 Los Angeles Games—before public outcry led to a massive redesign (Ripley and Hume 2015).

Solutions to the Economic Viability Problem

The Olympic Games as currently conducted are not economically viable for most cities. The most important reasons include infrastructure costs relating to the venues hosting the events; the monopoly rents that flow to the International Olympic Committee; poor management; corruption; and the specter of unreasonable and unrealizable economic expectations for the host city and nation. Concerns about costs are nothing new. Even Salt Lake City's $1.9 billion in expenditures in 2002 ($2.5 billion in 2015 dollars), which seem almost quaint by today's standards, raised concerns among organizers. Then-President of the International Olympic Committee, Jacques Rogge, expressed the "need to streamline costs and scale down the Games so the host cities are not limited to wealthy metropolises.... The scale of the Games is a threat to their quality," he said. "In a way, they risk becoming a victim of their own success" (as quoted in Roberts 2002).

Costs of staging the Games have skyrocketed in the years since those comments were made. The Olympics have reached a tipping point where the majority of potential host nations and cities in the industrialized, democratic West have come to the realization that hosting is more likely to drain rather than to enhance financial resources. Even before Boston and Hamburg's withdrawals as applicant cities for the 2024 Summer Olympic Games, and even before only two applicant cities emerged as contenders for the 2022 Winter Olympic Games, the International Olympic Committee had been considering major changes to its strategic vision. Its *Olympic Agenda 2020*, which was unanimously passed at the IOC's 127th Session in Monaco in December 2014, included 40 recommendations for reform, many of which promoted increased economic sustainability for host cities.

The recommendations provide at least some semblance of solutions to the problems relating to the economic viability of the Olympic Games. Specifically, they propose to: 1) shape the bidding process as an invitation; 2) evaluate bid cities by assessing key opportunities and risks; 3) reduce the cost of bidding; 4) include sustainability in all aspects of the Olympic Games; 5) include sustainability within the Olympic Movement's daily operations; and 6) reduce the cost and reinforce the flexibility of Olympic Games management (IOC 2014a). In addition, *Olympic Agenda 2020* seeks to reduce corruption by increasing transparency.

Recommendations, of course, must be translated into action. The International Olympic Committee has yet to complete a full bidding cycle under their new guidelines, but some cities are taking its recommendation seriously. Los Angeles, which emerged as the US bid city for 2024 following Boston's exit, has proposed using existing college dormitories at UCLA and the University of Southern California for athlete housing during the Games, thus eliminating over $1 billion in costs for an athletes' village from their original plans. Of course, if the IOC again finds itself lured into selecting the city with the fanciest accommodations for athletes (and, of course, for IOC executives), the most glamorous new stadiums, and the most elaborate ceremonies over simpler but more economically rational bids like what may be emerging in Los Angeles, then the clear signal will be that it is business as usual for the Olympics. Furthermore, blame for such an outcome should not be directed solely at the International Olympic Committee. Managing expectations is critical. Promising that hosting the Olympics will provide a significant boost to a host city and nation's economy is very likely to result in disappointment. Host cities and nations have to be more proactive, rather than permitting economic interests who stand to benefit from the Games to serve as the primary spokespersons for economic impact. Officials from national Organizing Committees should do more hands-on-management to ensure that the promises of vested interests are reasonable and achievable.

The problem posed by the extraordinary sports facilities costs can be solved through one or a few permanent locations for the Olympic Games. The original home of the Olympics in Greece is sometimes proposed. Alternatively, the IOC could designate, perhaps, four Summer Olympic and three Winter Olympic venues throughout the world that would rotate the staging duties. As yet another alternative, the IOC might award two successive Games to the same host, so that facilities could at least be used twice. Any of these proposals would serve to ensure that Olympic sports venues have a useful life of more than just one three-week event.

The fact that Los Angeles profited from the Olympics in 1984 and Barcelona experienced an economic revival of sorts as a consequence of hosting the Games in 1992 has added currency to claims that the Games can be economically transformative. But hosting the Games has become an increasingly expensive gambit; indeed, as the rules for bidding currently stand, the entire structure of the Olympic Games shouts "potential host beware." Issues start with the excesses of the bidding process, and are then followed by the construction of expensive and ostentatious sports infrastructure and the expensive opening and closing spectacles. If the commercial dimension of the Games has become too embedded to eliminate, then the costs must be managed better; infrastructure has to be made less expensive and reused; host nations and cities have to play the lead role in defining and achieving reasonable economic outcomes; and corruption has to be targeted through increased transparency and broader involvement. The goal should be that the costs of hosting are matched by benefits that are shared in a way to include ordinary citizens who fund the event through their tax dollars. In the current arrangement, it is often far easier for the athletes to achieve gold than it is for the hosts.

■ *The authors are thankful to Robert Baumann and Andrew Zimbalist for useful discussions in previous projects that have informed much of the analysis here and to Gordon Hanson, Enrico Moretti, Ann Norman, and Timothy Taylor for useful comments on earlier drafts. Responsibility for interpretation of the data, as well as for any errors, is the authors' alone.*

References

Atkinson, Giles, Susana Mourato, Stefan Szymanski, and Ece Ozdemiroglu. 2008. "Are We Willing to Pay Enough to 'Back the Bid'?: Valuing the Intangible Impacts of London's Bid to Host the 2012 Summer Olympic Games." *Urban Studies* 45(2): 419–44.

Baade, Robert, Robert Baumann, and Victor Matheson. 2010. "Slippery Slope? Assessing the Economic Impact of the 2002 Winter Olympic Games in Salt Lake City, Utah." *Région et Développement* no. 31, pp. 81–91.

Baade, Robert, and Victor Matheson. 2002. "Bidding for the Olympics: Fool's Gold?" In *Transatlantic Sport: The Comparative Economics of North American and European Sports*, edited by Carlos Pestana Barros, Muradali Ibrahímo, and Stefan Szymanski, 127–51. London: Edward Elgar.

Baumann, Robert, Bryan Engelhardt, and Victor Matheson. 2012. "Employment Effects of the 2002 Winter Olympics in Salt Lake City, Utah." *Journal of Economics and Statistics* 232(3): 308–17.

Billings, Stephen B., and J. Scott Holladay. 2012. "Should Cities Go for the Gold? The Long-Term Impacts of Hosting the Olympics." *Economic Inquiry* 50(3): 754–72.

Blake, Adam. 2005. "The Economic Impact of the London 2012 Olympics." Unpublished paper, Christel DeHaan Tourism and Travel Research Institute, Nottingham University Business School, Jubilee Campus.

British Broadcasting Corporation (BBC). 2012. "Post Olympic Spirits High But May Fizzle Out—Survey." August 14. http://www.bbc.com/news/uk-19246044.

British Broadcasting Corporation (BBC). 2013. "London 2012: Olympics and Paralympics £528m Under Budget." July 19. http://www.bbc.com/sport/0/olympics/20041426.

Brückner, Markus, and Evi Pappa. 2015. "News Shocks in the Data: Olympic Games and Their Macroeconomic Effects." *Journal of Money, Credit and Banking* 47(7): 1339–67.

Coates, Dennis, and Brad R. Humphreys. 2008. "Do Economists Reach a Conclusion on Subsidies for Sports Franchises, Stadiums, and Mega-Events?" *Econ Journal Watch* 5(3): 294–315.

Economist, The. 1998. "Downhill All the Way." February 5. http://www.economist.com/node/112549.

Farhi, Paul. 2014. "Did the Winter Olympics in Sochi Really Cost $50 billion? A Closer Look at that Figure." *Washington Post,* February 10.

Farrar, Lara. 2010. "Beijing's Water Cube Now Has Slides, Rides, A Wave Pool and Spa." CNN.com, August 11. http://travel.cnn.com/explorations/play/beijings-watercube-water-park-now-open-040746.

Feddersen, Arne, and Wolfgang Maennig. 2013. "Mega-Events and Sectoral Employment: The Case of the 1996 Olympic Games." *Contemporary Economic Policy* 31(3): 580–603.

Flyvbjerg, Bent, and Allison Stewart. 2012. "Olympic Proportions: Cost and Cost Overrun at the Olympics 1960–2012." Saïd Business School Working Paper, University of Oxford.

Fowler, Geoffrey A., and Stacy Meichtry. 2008. "China Counts the Cost of Hosting the Olympics." *Wall Street Journal*, July 16.

Giesecke, James, and John Madden. 2011. "Modelling the Economic Impacts of the Sydney Olympics in Retrospect—Game Over for the Bonanza Story?" *Economic Papers* 30(2): 218–32.

Hotchkiss, Julie, Robert Moore, and Stephanie M. Zobay. 2003. "Impact of the 1996 Summer Olympic Games on Employment and Wages in Georgia." *Southern Economic Journal* 69(3): 691–704.

Humphreys, Jeffrey, and Michael Plummer. 1995. *The Economic Impact on the State of Georgia of Hosting the 1996 Summer Olympic Games.* Selig Center for Economic Growth, University of Georgia.

International Olympic Committee (IOC). 2010. "Factsheet: Legacies of the Games, Update—January 2010." http://www.olympic.org/Documents/Reference_documents_Factsheets/Legacy.pdf.

International Olympic Committee (IOC). 2014a. "Olympic Agenda 2020: 20 + 20 Recommendations." http://www.olympic.org/Documents/Olympic_Agenda_2020/Olympic_Agenda_2020-20-20_Recommendations-ENG.pdf.

International Olympic Committee (IOC). 2014b. "Olympic Marketing Fact File." 2014 Edition. http://www.olympic.org/Documents/IOC_Marketing/OLYMPIC_MARKETING_FACT_%20FILE_2014.pdf.

InterVISTAS Consulting. 2002. "The Economic Impact of the 2010 Winter Olympics and Paralympic Games: An Update." British Columbia Ministry of Competition, Science and Enterprise. November, 20. http://www.intervistas.com/downloads/Economic_Impact_of_Hosting_2010_Winter_Games.pdf.

Jasmand, Stephanie, and Wolfgang Maennig. 2008. "Regional Income and Employment Effects of the 1972 Munich Olympic Summer Games." *Regional Studies* 42(7): 991–1002.

Jordan, Mary, and Kevin Sullivan. 1999. "Nagano Burned Documents Tracing '98 Olympics Bid." *Washington Post*, January 21, p. A1. http://www.washingtonpost.com/wp-srv/digest/daily/jan99/nagano21.htm.

Langer, Viktoria C. E., Wolfgang Maennig, and Felix Richter. 2015. "News Shocks in the Data: Olympic Games and their Macroeconomic Effects—Reply." Working Paper 52, Hamburg Contemporary Economic Discussions, University of Hamburg.

Leme, Luisa. 2015. "Weekly Chart: How Much Will Rio's 2016 Summer Olympics Cost?" *Americas Society/Council of the Americas*, August 5. http://www.as-coa.org/articles/weekly-chart-how-much-will-rios-2016-summer-olympics-cost.

Longman, Jere. 1998. "Nagano 1998: Seven Days to Go; High Costs and High Expectations." *New York Times*, January 30.

Maennig, Wolfgang, and Felix Richter. 2012. "Exports and Olympic Games: Is There a Signal Effect?" *Journal of Sports Economics* 13(6): 635–41.

Matheson, Victor. 2009. "Economic Multipliers and Mega-Event Analysis." *International Journal of Sport Finance* 4(1): 63–70.

Matheson, Victor. 2013. "Assessing the Infrastructure Impact of Mega-Events in Emerging Economies." In *Infrastructure and Land Policies*, Gregory K. Ingram and Karin L. Brandt, eds., 215–32. Cambridge, MA: Lincoln Land Institute.

Mullholland, Hélène. 2012. "David Cameron Claims London 2012 Will Bring £13bn 'Gold for Britain.'" *The Guardian*, July 5.

Payne, Bob. 2008. "The Olympics Effect: When the Games Are Over, Which Cities Win Big—And Which Stumbled?" *MSNBC*, August 3. http://today.msnbc.msn.com//id/26042517#.UBwk104gcsd.

Pletz, John. 2010. "Chicago 2016's Final Tally: $70.6M Spent on Olympics Effort." *Crain's Chicago Business*, May 17.

Porter, Philip K., and Deborah Fletcher. 2008. "The Economic Impact of the Olympic Games: Ex Ante Predictions and Ex Poste Reality." *Journal of Sport Management* 22(4): 470–86.

Preuss, Holger. 2004. *The Economics of Staging the Olympic Games.* London: Edward Elgar.

Ripley, Will, and Tim Hume. 2015. "Japan Scraps Plans for Controversial 'Bike Helmet' Olympic Stadium." CNN, July 17. http://www.cnn.com/2015/07/17/asia/japan-tokyo-olympic-stadium-scrapped/.

Ritchie, J. R. Brent, and Brian H. Smith. 1991. "The Impact of a Mega-Event on Host Region Awareness: A Longitudinal Study." *Journal of Travel Research* 30(1): 3–10.

Roberts, Selena. 2002. "Olympics: Notebook; I.O.C.'s Rogge Steps Into The Cold." *New York Times,* February 4.

Rose, Andrew K., and Mark M. Spiegel. 2011. "The Olympic Effect." *Economic Journal* 121(553): 652–77.

Sky Sports. 2015. "Olympic Stadium Costs Soar Ahead of West Ham Move." June 19. http://www.skysports.com/football/news/11685/9890173/olympic-stadium-costs-soar-ahead-of-west-ham-move.

Tagaris, Karolina. 2014. "Ten Years On, Athens 2004 Gives Greece Little to Cheer." Reuters, August 7. http://uk.reuters.com/article/2014/08/07/uk-olympics-greece-idUKKBN0G70Y220140807.

Teigland, Jon. 1999. "Mega-Events and Impacts on Tourism; The Predictions and Realities of the Lillehammer Olympics." *Impact Assessment and Project Appraisal* 17(4): 305–317.

Thaler, Richard. 1988. "Anomalies: The Winner's Curse." *Journal of Economic Perspectives* 2(1): 191–202.

UK Office for National Statistics (ONS). 2015. *Overseas Travel and Tourism.* Various months. http://www.ons.gov.uk/ons/rel/ott/overseas-travel-and-tourism—quarterly-release/index.html.

US General Accounting Office (GAO). 2001. "Olympic Games: Costs to Plan and Stage the Games in the United States." Report to the Ranking Minority Member Subcommittee on the Legislative Branch Committee on Appropriations U.S. Senate, November.

VanWynsberghe, Rob. 2011. "Olympic Games Impact (OGI) Study for the 2010 Olympic and Paralympic Winter Games: Games-time Report." http://cfss.sites.olt.ubc.ca/files/2011/10/The-Olympic-Games-Impact-Study-Games-time-Report-2011-11-21.pdf.

Walker, Alissa. 2014. "How L.A.'s 1984 Summer Olympics Became the Most Successful Games Ever." *Gizmodo.com*, February 6. http://gizmodo.com/how-l-a-s-1984-summer-olympics-became-the-most-success-1516228102.

Walton, Harry, Alberto Longo, and Peter Dawson. 2008. "A Contingent Valuation of the 2012 London Olympic Games: A Regional Perspective." *Journal of Sports Economics* 9(3): 304–17.

Zimbalist, Andrew. 2015. *Circus Maximus: The Economic Gamble Behind Hosting the Olympics and the World Cup.* Brookings Institution Press.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 219–236

Retrospectives
How Economists Came to Accept Expected Utility Theory: The Case of Samuelson and Savage

Ivan Moscati

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

In the expected utility approach to decision-making under risk, the utility of a risky prospect is given by the sum of the utilities u of the alternative possible outcomes of the prospect, each weighted by the probability that the outcome will occur. Thus, for example, the utility of a lottery yielding either a trip to London with probability p or a payment of \$2,000 with probability $(1 - p)$ is given by $(u(\text{trip to London}) \times p) + (u(\$2{,}000) \times (1 - p))$. According to expected utility theory, the decision maker will choose among these risky prospects, or lotteries, and will choose the one with the highest expected utility. This theory dominated the economic analysis of individual decision-making under risk from the early 1950s to the 1990s. Beginning in the late 1970s, the accumulation of robust experimental evidence against the expected utility hypothesis prompted decision theorists

■ *Ivan Moscati is Associate Professor of Economics, University of Insubria, Varese, Italy. He also teaches History of Economics at Bocconi University, Milan, Italy. His email address is ivan.moscati@uninsubria.it.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.2.219 doi=10.1257/jep.30.2.219

to advance a number of alternative theories, such as prospect theory (Kahneman and Tversky 1979), Choquet expected utility (Schmeidler 1989), maxmin expected utility (Gilboa and Schmeidler 1989), rank-dependent utility (Quigging 1993), or the smooth model of ambiguity aversion (Klibanoff, Marinacci, and Mukerji 2005). However, none of these alternative theories has yet reached the dominant status that expected utility theory once enjoyed. As Gilboa and Marinacci (2013, p. 232) have argued in a recent survey of the decision-theoretic literature, it is not clear that a single theory of decision-making under uncertainty will replace expected utility theory, and "even if a single paradigm will eventually emerge, it is probably too soon to tell which one it will be." Because of the absence of a clear single alternative, and thanks also to its simplicity and adaptability, expected utility theory remains the primary model in numerous areas of economics dealing with risky decisions, such as finance, the theory of asymmetric information, and game theory.

Expected utility theory was originally advanced by Daniel Bernoulli in the eighteenth century and was adopted by some nineteenth-century economists, such as Alfred Marshall, but it came under sustained criticism from the 1930s to the early 1950s. During this period, some economists argued that individuals evaluate risky alternatives by looking at the mean, the variance, and possibly other elements of the distribution of uncertain payoffs, rather than using expected utility (Hicks 1931). Others noted that an individual who places probabilities and utilities on a range of outcomes, and then calculates the weighted average of utilities, is engaging in cardinal measurability of utility, which contrasts with the ordinal conception of utility that dominated utility analysis in the 1930s (Tintner 1942). Along with approaches based on the distribution of payoffs were models based on the "minimax" criterion (Wald 1950) or on the idea that individuals focus only on the best-possible and the worst-possible outcomes of risky alternatives (Shackle 1949).

The fortunes of expected utility theory began to recover when John von Neumann and Oskar Morgenstern introduced a set of axioms of rational individual decision-making that implied expected utility theory in their *Theory of Games and Economic Behavior* (1944; second edition with an explicit proof of the expected utility theory theorem, 1947). Among the early supporters of the expected utility hypothesis in the von Neumann–Morgenstern version were Milton Friedman and Leonard Jimmie Savage (1948), both based at the University of Chicago, and Jacob Marschak (1948, 1950), a leading member of the Cowles Commission for Research in Economics.

Paul Samuelson of MIT was initially a severe critic of expected utility theory. Between mid-April and early May 1950, Samuelson (1950a, b, c) composed three papers in which he attacked von Neumann and Morgenstern's axiomatic system for its lack of transparency, contested the capacity of expected utility theory to explain empirical phenomena, identified and named what he said was a hidden axiom behind expected utility theory—the "Independence Axiom"—and claimed this axiom was untenable. This axiom states that if an individual prefers a trip to London over a trip to Venice then, for any probability p and any amount of money $\$K$, he should also prefer the lottery yielding a trip to London with probability p and $\$K$

with probability (1 – *p*), to the lottery yielding a trip to Venice with probability *p* and $*K* with probability (1 – *p*). Samuelson argued that, rather than satisfy the Independence Axiom, the individual's ordinal preferences over risky alternatives should satisfy only one property besides completeness, transitivity, and continuity—namely, what today we call monotonicity with respect to first-order stochastic dominance. In the case of choices over lotteries with monetary payoffs, monotonicity means that raising a payoff without changing the other payoffs, or increasing the probability of a larger payoff at the expense of the probability of a smaller payoff, raises preferability. For Samuelson (1950a, p. 169) any assumption beyond monotonicity would impose an arbitrary "straight-jacket" on individual preferences over risky alternatives.

By 1952, however, Samuelson had somewhat unexpectedly become a resolute supporter of the expected utility hypothesis. In a prominent conference on decision theory held in Paris in May 1952, he joined Friedman, Savage, and Marschak in advocating expected utility theory against the attacks of Maurice Allais and other opponents of the theory. In 1952, Samuelson also organized a symposium on expected utility theory that was published in the October 1952 issue of *Econometrica* and was instrumental in stabilizing expected utility theory as the dominant economic model of choice under risk.

Why did Samuelson change his mind? Accounts of Samuelson's conversion to expected utility theory based on published materials and personal recollections have been provided by Samuelson himself (for example, Samuelson 1947 [1983]), and by Fishburn and Wakker (1995) in their essay on the origin of the Independence Axiom.[1] The present article fills out these accounts by employing, for the first time, letters and other unpublished materials collected in the Samuelson Papers held at Duke University, the Savage Papers at Yale University, and the Friedman Papers at the Hoover Institution. These archives reveal that Samuelson's change of mind occurred mainly through an exchange of letters with Savage and, to a lesser extent, with Marschak and Friedman, between May and September 1950. This correspondence shows that Samuelson accepted expected utility theory only when Savage persuaded him of the normative force of the Independence Axiom. Samuelson frankly admitted his capitulation in a letter to Friedman dated August 25, 1950 (Samuelson Papers, Box 31):

> Dear Milton: … [L]et me make an important surrender. Savage's patient letters and the induced cogitation have convinced me that he is right on the only important difference between us. … I called the [Independence] assumption gratuitous, arbitrary, etc. … etc. (You know how I can lay it on when I get

[1] Other cases of prominent economists explicitly admitting to having changed their mind on some substantive issue are rare but not nonexistent. For instance, David Ricardo (1821) famously changed his mind about the effects of new machinery on the demand for labor, and John Maynard Keynes (1936) repudiated many of the views about employment, interest, and money he had held before writing the *General Theory*. On at least one other prominent issue, Samuelson also admitted to having been wrong, namely in the debate on the re-switching of production techniques (Samuelson 1966).

> going.) But now I must eat my words. As you know I hate to change my mind, but I hate worse to hold wrong views, and so I have no choice.

The correspondence also shows that, for Savage, his exchange with Samuelson modified his thinking about expected utility theory. Samuelson's arguments prompted Savage to streamline the normative defense of expected utility theory and formulate the Sure-Thing Principle, which is the central assumption of the subjective version of expected utility theory that Savage later advanced in *The Foundations of Statistics* (1954).

Based on the correspondence between Samuelson, Savage, Marschak, and Friedman, this article reconstructs the joint intellectual journey that led Samuelson to accept expected utility theory and Savage to revise his motivations for supporting it. The article is organized around the main issues those four economists discussed in their correspondence: 1) identifying the importance of the Independence Axiom as a key point under dispute; 2) the nature of the cardinal function featuring in expected utility theory, and the relationship between the Independence Axiom and the idea that the utilities of different commodities are independent; 3) the descriptive validity of expected utility theory; and 4) the normative appeal of the Independence Axiom. More detail on the history of expected utility theory between 1930 and 1950 is available in Moscati (forthcoming).

Identifying the Importance of the Independence Axiom

In *Theory of Games* (1944 [1947]), von Neumann and Morgenstern state a series of axioms about the individual's preferences over indifference classes of lotteries and offer a proof that an individual obeying these axioms will then follow expected utility theory. Their axiomatization of the expected utility hypothesis theory includes the completeness, transitivity, and continuity of preferences but does not feature an assumption corresponding to what today we call the Independence Axiom. In the first of Samuelson's three 1950 papers, completed in April 1950 and called the "Japanese paper" because it was later published in a Japanese journal, Samuelson declared that he found von Neumann and Morgenstern's axioms opaque. On the one hand, he observed, they concern preference relations and seem therefore to be ordinal in nature; on the other hand, the axioms imply that the preference relations can be represented by the expected-utility formula, which features a cardinal utility function.[2] Samuelson was puzzled by this apparent contradiction and could not understand how von Neumann and Morgenstern's ordinal axioms imply expected utility theory: "I am simply confused," he admitted (1950a, p. 172).

Samuelson guessed that von Neumann and Morgenstern had "implicitly added a hidden and unacceptable premise to their axioms" (p. 172), but in April 1950 he

[2] A utility function $U(x)$ is cardinal if it is unique up to positive linear transformations of the form $aU(x) + b$, where $a > 0$.

was unable to identify this hidden assumption. Nevertheless, Samuelson christened the hidden premise the "independence assumption" (p. 170, footnote 7) because he associated it with the assumption that the utilities of different commodities are independent or additively-separable.[3] Since the early twentieth century, this assumption had been discredited in utility analysis because it rules out the substitutability and complementarity of goods. In his *Foundations of Economic Analysis* (1947 [1983]), Samuelson had extensively criticized additive separability as farfetched, and shown that it implies the cardinal measurability of utility and further implausible features of the demand functions for commodities.

In Samuelson's second 1950 paper, completed by May 5 and published as a RAND Corporation memorandum on May 24, Samuelson made the hidden premise of von Neumann–Morgenstern axiomatics explicit, naming it the "Special Independence Assumption."[4] The adjective "special" was intended in a pejorative sense, emphasizing the dubiously restrictive character of the assumption. Samuelson (1950b, pp. 6–7) stated the assumption in terms of indifference:

> *Special Independence Assumption*: If two situations A and B are indifferent, so that V(A) = V(B) [the utility V(A) of A is equal to the utility V(B) of B], then… V(A, C) = V(B, C) [the utility of the probability mixture of situation A and situation C is equal to the utility of the probability mixture of situation B and situation C] *for all* C's.

In late April–early May 1950, Samuelson was unaware that he was not the first to have stated the Independence Axiom. Marschak (1948, 1950), RAND researcher Norman Dalkey (1949), and John Nash (1950), then still a PhD student in mathematics at Princeton, had all put forward axiomatizations of expected utility theory including versions of the Independence Axiom (Fishburn and Wakker 1995; Bleichrodt, Li, Moscati, and Wakker forthcoming). Notably, in an article published in the April 1950 issue of *Econometrica*, Marschak (1950) called it Postulate IV$_2$.[5]

In Samuelson's third paper of 1950, completed in early May and published as a RAND memorandum on June 13, Samuelson referred to an oral communication in which Marschak had called his attention to Postulate IV$_2$. Accordingly, in a footnote Samuelson (1950c, p. 2) noted that what he had called the Special Independence Axiom "seems to be his [Marschak's] Postulate IV$_2$."

[3] This means that the utility U of commodity bundle $(x_1, \ldots, x_n)$ can be expressed as $U(x, \ldots, x_n) = \sum_{i=1}^{n} U_i(x_i)$, where U_i is the utility function relative to commodity i.

[4] The RAND Corporation is a private think tank originally funded by the US Army Air Force in 1946 through the Douglas Aircraft Company with the goal of bringing together civil scientists from different backgrounds to work on interdisciplinary research projects with possible military applications. RAND became an independent nonprofit corporation in 1948, and Samuelson began collaborating with RAND in 1949.

[5] Formally, Marschak's Postulate IV$_2$ reads as follows: let A, B and C, be three different risky prospects or lotteries; if A $\sim$ B, then pA + $(1 - p)$C $\sim$ pB+$(1 - p)$C, where $0 < p < 1$ (1950, p. 120).

Samuelson circulated the Japanese paper to colleagues, requesting comments and criticism. On May 1, 1950, he sent a copy to Milton Friedman and asked him to forward a second copy to Leonard (Jimmie) Savage. Friedman did so in early May 1950, accompanying Samuelson's paper with a perplexed comment to Savage: "Dear Jimmie: … Can you figure out what it is about? I must confess I cannot" (Friedman Papers, Box 99). On May 19, 1950, Savage sent a long letter to Samuelson containing extensive comments on the Japanese paper; the letter was carbon-copied to Friedman.

In the Japanese paper, Samuelson had also questioned the axiomatization of expected utility theory that Friedman and Savage (1948) had advanced. This axiomatization consisted of three assumptions that Friedman and Savage (p. 288) claimed implied expected utility theory and are logically equivalent to von Neumann and Morgenstern's axioms. Notably, Friedman and Savage's assumptions do not include the Independence Axiom. In the Japanese paper, Samuelson (1950a, pp. 121–23) argued that the von Neumann–Morgenstern and the Friedman–Savage axiomatic systems were not equivalent, and further doubted that the three Friedman–Savage assumptions actually implied expected utility theory. In his letter of May 19, 1950, to Samuelson, Savage acknowledged that Samuelson was right (Samuelson Papers, Box 67):

> Dear Professor Samuelson: … On reexamination I find that β [the Friedman-Savage axiomatic system] does not imply α [expected utility theory]. This is because Milton and I slipped in leaving out of it something very like what … you call the basic hypothesis.

In the second part of his letter, Savage provided a new set of axioms, collectively labeled as β', and proved that they actually imply expected utility theory. In particular, Axiom 2″ of β' is a version of the Independence Axiom, albeit expressed in terms of preference rather than indifference.[6]

After some discussion, which occupied their correspondence between May and July 1950 and which was hampered by terminological misunderstandings, by late July 1950 Samuelson and Savage came to agree that Samuelson's Special Independence Assumption, Savage's Axiom 2, and Marschak's Postulate IV_2 are fundamentally equivalent. They also agreed that, if the preference relation over lotteries is complete, transitive, and continuous, the Independence Axiom is a necessary and sufficient condition for expected utility theory. At that point, the only important difference between them concerned the plausibility of the Independence

[6] Savage's Axiom 2″ reads as follows: let A, B and C be three different gambles; if $A \prec B$, then $pA + (1-p)C \preccurlyeq pB + (1-p)C$, where $0 < p < 1$.

Axiom. Thus, in his letter to Savage of July 20, 1950, Samuelson refocused the discussion on this latter point (Savage Papers, Box 29):

> Dear Dr. Savage: … I shall be interested in knowing what you think of the "plausibility" of the postulate V(A) = V(B) implies V(A,C) = V(B,C) for arbitrary C and arbitrary mixtures.

Samuelson himself maintained that the postulate was a "gratuitously-arbitrary-special-implausible hypothesis."

Independence Axiom, Independent Utilities, and Cardinal Utility

On April 20, 1950, Samuelson sent the Japanese paper to Jacob Marschak, who replied in a letter dated May 11, 1950 (Samuelson Papers, Box 66). Marschak accepted that his Postulate IV$_2$ and Samuelson's Special Independence Assumption are equivalent. However, he rejected Samuelson's association of the Independence Assumption, and thus Postulate IV$_2$, with the discredited hypothesis that the utilities of different commodities are additively separable. Marschak stressed that while the latter has to do with the joint consumption of different goods—for example, beer *and* pretzels—Postulate IV$_2$ relates to the consumption of different goods in mutually-exclusive situations where a choice is being made between outcomes—that is, *either* beer *or* pretzels. Thus a man who is indifferent between beer and tea might well prefer the commodity bundle beer-and-pretzels to the commodity bundle tea-and-pretzels and, at the same time, be indifferent between a lottery consisting of "either beer or pretzels" and another lottery consisting of "either tea or pretzels." Marschak wrote:

> I should not expect… [the] man to tell me that the mere co-presence in the same lottery bag of tickets inscribed "pretzels" with tickets inscribed "tea" will contaminate (or enhance) the enjoyment of either the liquid or the solid that will be the subject's lot.

In his rejoinder of May 15, 1950 (Samuelson Papers, Box 66), Samuelson accepted that different lottery prizes are mutually exclusive in a probability sense. But he then returned to his earlier concern about the additive separability of utilities, and wrote that he could not understand why this "*ex ante* preference pattern" toward lotteries could generate, as is the case in expected utility theory, a utility indicator that "impute[s] an *independent* numerical score to each possible prize." The expected utility of a "beer or pretzels" lottery is in fact expressed by $(u(\text{beer}) \times p) + (u(\text{pretzels}) \times (1 - p))$. But the expressions u(beer) and u(pretzels) seem to suggest that the utilities of beer and pretzels are independent of each other. For Samuelson, this was "a pun on words."

The solution came from Milton Friedman. In May 1950, William Baumol, then a young assistant professor at Princeton University, submitted to the *Journal of*

Political Economy a paper arguing that expected utility theory involved a return to a cardinal conception of utility, which was eventually published as Baumol (1951). In a letter to Baumol dated June 3, 1950, and carbon-copied to Savage and Samuelson, Friedman disclosed that he would serve as a referee for this paper, and commented on it (Samuelson Papers, Box 15). Friedman wrote the report on Baumol's paper between June and August 1950, and forwarded copies to Savage and Samuelson.

In opposition to Baumol's claims, Friedman contended that expected utility theory "does not commit you in any way to 'cardinal utility' whatever that may mean" (Savage Papers, Box 29). Friedman labelled as g(A) the expected utility of lottery A, and noticed that any monotonically increasing transformation G of g(A) continues to represent the preference order between lotteries:

> It is obvious that we can take as a utility function any member of the set G[g(A)], where the set is subject only to the restriction that G′ be greater than zero.[7]

For Friedman, therefore, expected utility theory was not in contradiction with the ordinal approach to utility. Friedman argued that the cardinal function u featuring in the expected-utility formula and the utility function U expressing the individual's preferences over riskless outcomes are two different functions. According to Friedman, Baumol failed to see this difference. In order to avoid further confusion, Friedman suggested giving another name to the function u and proposed calling it "the choice generating function."

Friedman's interpretation of the nature of u, which was articulated by Friedman and Savage (1952) and quickly became the official view among utility theorists, has a number of important consequences. First, although expressions such as u(beer) and u(pretzels) in the expected-utility formula may suggest that the riskless utilities of beer and pretzels are independent of each other, this inference is unwarranted: the form of the function u carries no implications over the complementarity or substitutability of beer and pretzels in riskless situations. Second, even if the utilities of beer and pretzels were independent and could be represented by an additively-separable utility function U, this function would still differ from the choice-generating function u, in the specific sense that u need not be a positive linear transformation of U.[8]

Between May and August 1950, Samuelson apparently came to see expected utility theory as innocent of the charge that it involved a return to a cardinal conception of utility, and the Independence Axiom as innocent of any necessary relationship to the hypothesis that utilities are additively separable. This supposition

[7] More explicitly, if A is a lottery that yields payoff x_i with probability p_i, with $i = 1, \ldots n$, then g(A) $= \sum_{i=1}^{n} p_i\, u(x_i)$. I have slightly modified Friedman's notation to make it more consistent with that used in Samuelson's correspondence.

[8] A further consequence of Friedman's interpretation is that the possible concavity of the function u, which is associated with risk aversion, cannot be conceived of as an expression of the decreasing marginal utility of riskless outcomes—for example, of money. This question, however, was not discussed in the Samuelson–Savage–Marschak–Friedman correspondence reconstructed in the present paper.

is backed by the fact that in the two papers on expected utility theory that Samuelson completed in 1952 (1952 [1966]; 1952), he insisted on the ordinal nature of the Independence Axiom and the other assumptions underlying expected utility theory, stressed the fact that the expected utility hypothesis is about mutually exclusive outcomes rather than joint consumption of different goods, and pointed out that "independence in probability situations puts *no* restriction whatsoever upon the dependence or independence that holds in the nonstochastic situation" (1952, p. 673). Friedman's proposed distinction seems to have eliminated one significant obstacle in Samuelson's accepting of the Independence Axiom and expected utility theory, but did not offer him any positive reasons to endorse them.

The Descriptive Validity of Expected Utility Theory

Based on data from the US Bureau of Labor Statistics, the National Bureau of Economic Research, books about the history of lotteries in different countries, and casual observation, Friedman and Savage (1948) identified three basic facts that a satisfactory theory of choice under risk should be able to explain: 1) individuals of all income levels buy insurance; 2) individuals of all income levels purchase lottery tickets or engage in similar forms of gambling; and 3) most individuals both purchase insurance and gamble. Friedman and Savage (p. 297) claimed that, by assuming that the utility curve of money is first concave, then convex, and then concave again, expected utility theory can rationalize these three facts. In the Japanese paper, Samuelson (1950a, p. 168) contested this claim, noticing that, for instance, expected utility theory cannot explain "the perfectly possible case of a man who refuses fair small bets at all income levels and yet buys lottery tickets." More generally, Samuelson contended that the phenomena associated with gambling are "infinitely richer" than the expected utility hypothesis permits, and that there is as much to be learned about gambling "from Dostoyevsky as from Pascal."[9]

In his letter to Samuelson of May 19, 1950, Savage replied that the fact that expected utility theory is not consistent with every conceivable sort of behavior shows that the theory "is not simply tautological" (Samuelson Papers, Box 67). More forcefully, Savage advanced a simplicity defense, namely that expected utility theory should be accepted as a simple and acceptable approximation to reality: "It is … the simplest theory of gambling behavior which has come to my attention and which seems at all consistent with the facts in any reasonably extensive range of contexts."

Savage also argued that the understanding of expected utility theory as a handy approximation of reality was also shared by Friedman, and also by von Neumann,

[9] Fyodor Dostoyevsky authored the autobiographical novel "The Gambler" (1867), whose protagonist was addicted to roulette. In his *Pensées* (1670), Blaise Pascal made a famous argument for believing in God based on the wager that if you do not believe in a God that exists, you risk eternal punishment, but if you do believe in a God that does not exist, the costs are much less dire.

whom he had known in the academic year 1941–1942 when studying in Princeton as a post-doctoral fellow: "I have repeatedly heard von Neumann express the idea in the most emphatic language. His interest, like Milton's and mine, in the theory stems from the belief that it is a skillfully chosen zero approximation to reality."

In his response to Savage of July 20, 1950, carbon-copied to Friedman and Marschak, Samuelson ironically rejoined that he found no particular merit in the fact that expected utility theory is nontautological (Savage Papers, Box 29): "Both you and Milton express in separate letters pride that you have labored like lions and produced a *non-tautology*. I am sure there is a category of people who must be told that this is not necessarily a crime.... I do not see that I qualify for this category."

To refute Savage's simplicity argument, Samuelson contrasted the expected utility hypothesis with the theory of decision-making under risk that he had advocated in the Japanese paper, which theory is based on the hypothesis that preferences over risky alternatives are monotonic with respect to first-order stochastic dominance. Both theories make only one further assumption besides those concerning the completeness, transitivity, and continuity of preferences, namely expected utility theory uses the Independence Axiom while Samuelson's theory uses stochastic-dominance monotonicity. Therefore, at the formal level, both theories are equally simple.

Concerning the empirical implications of the two theories, Samuelson argued that where these implications differ "there is *no* factual evidence in favor of the special theory [that is, expected utility theory] and some against it." Samuelson also commented cursorily on the pioneering experiment to test the validity of the expected utility hypothesis conducted between 1948 and 1949 by Frederick Mosteller, a Harvard statistician associated with Friedman and Savage, and Philip Nogee, then a Harvard PhD student in psychology.[10] Samuelson disparaged the design of the experiment—"[Nobody] could expect anything from this pitiful set-up"—and thus implicitly dismissed Mosteller and Nogee's claim that their experimental findings supported the empirical validity of the expected utility hypothesis. Thus, argued Samuelson, Savage's simplicity criterion backfires, supporting his own theory of decision-making under risk rather than expected utility theory. He concluded: "On the matter of simplicity, I know how I would wield Occam's Razor."

Arguably, Samuelson's argument had some effect on Savage who, in his subsequent letters to Samuelson, no longer insisted on the simplicity and descriptive power of expected utility theory. Friedman, by contrast, remained convinced that expected utility theory was empirically valid. In a letter to Samuelson dated September 13, 1950 (Samuelson Papers, Box 31), Friedman stressed that he accepted expected utility theory, not because he judged the axioms underlying it particularly plausible,

[10] The findings of the experiment were published in Mosteller and Nogee (1951). On the Mosteller–Nogee experiment and the role Friedman and Savage played in its design, see Moscati (2016).

but because he considered it a simple theory whose implications are not only far from obvious, but also consistent with much common experience:

> Dear Paul: …It has never seemed to me obviously true or necessary that individual's reactions to complicated gambles should be completely predictable from their reactions to two-side ones—which has always seemed to me the fundamental empirical content of the B[ernoulli]–M[arshall] hypothesis—and it still does not. At the same time, it has seemed… the simplest and most direct way to extend the usual utility analysis to choices involving risk, and not inconsistent with much common experience.

Notably, Friedman explicitly admitted that certain phenomena related to gambling cannot really be explained by expected utility theory, and predicted that "to handle some experience" the theory would "need complication."

The Normative Plausibility of the Independence Axiom

In his letter to Samuelson of May 11, 1950 (Samuelson Papers, Box 66), Marschak identified behavior satisfying the assumptions underlying expected utility theory, including the Independence Axiom (or, equivalently, Postulate IV_2), with rational behavior. He compared the argument for expected utility theory to Euclidean geometry, and behavior violating expected utility theory axioms with non-Euclidean geometry.[11] Marschak admitted that among actual people the observation of "non-Euclidean habits" is likely, but argued that such conduct is not advisable: "It may be *usual* for village carpenters… to deviate from the advice of Euclidian geometers…All the same, they would be better advised to behave rationally by following Euclid." As an example of non-advisable behavior, Marschak took what he called "love for danger"—that is, a violation of preference monotonicity with respect to stochastic dominance. Marschak rhetorically asked Samuelson whether, as a factory owner, he would hire a statistician "whose formula for quality control would be based on 'love for danger,' i.e., on rather liking the prospect of the factory being blown up, with 5% probability?"

Samuelson's rejoinder soon arrived. In his letter to Marschak of May 15, 1950 (Samuelson Papers, Box 66), Samuelson declared that, like Marschak, he was interested in "locating the 'natural' discontinuities which fence-out 'irrational' from 'rational' behavior." However, he contested the claim that the Independence Axiom or Postulate IV_2 should be included among the axioms defining rational,

[11] In economic discourse in the first half of the twentieth century, the analogy with Euclidean and non-Euclidean geometries was far from infrequent. It was used by, among others, Pigou (1920), J. M. Clark (1921), and, most famously, Keynes (1936, p. 16) in a passage of the *General Theory* in which he compared the "the classical theorists" to "Euclidean geometers in a non-Euclidean world." Echoing Keynes, Samuelson (1942, p. 593) opposed the Euclidean to the non-Euclidean world in a paper on fiscal policies.

or Euclidean, behavior. For Samuelson, "IV_2 stands out like a sore-thumb as arbitrary and alien." Accordingly, he dismissed "any identification of 'non-Euclideanism' with non-IV_2-ism." With respect to the love-for-danger example, Samuelson correctly rejected Marschak's identification of love for danger with a violation of the Independence Axiom:

> I would not hire Dostoyevsky to be my quality-control statistician; but until it can be shown that there is an iota of connection between "love of danger" and "*ex ante* lack of additive independence" [the Independence Axiom], the analogy is more confusing than clarifying.

Savage also advanced some normative arguments in defense of the Independence Axiom in letters from May to July 1950. However these arguments failed to impress Samuelson.[12] As already mentioned, in his letter to Savage of July, 20, 1950, Samuelson still dismissed the Independence Axiom as a "gratuitously-arbitrary-special-implausible hypothesis." It seemed that discussion had reached a deadlock, but then Savage came up with a new argument.

Savage's letter to Samuelson, dated August 12, 1950 (Samuelson Papers, Box 67), began by alluding to "a simple but important idea" not yet "set down in our correspondence to date." Savage considered three incomes A, B, C, two mutually exclusive events E and E′, and two contracts I and II reading as follows:

> I. In the event E Jimmie's income shall be C, and in the event E′ it will be A.
>
> II. In the event E Jimmie's income shall be C, and in the event E′ it will be B.[13]

Savage argued that if he (as Jimmie) prefers income B to income A, he would certainly prefer contract II to I. The reason is that, by choosing contract II, "I guarantee that whichever of the [two] events occurs I will have nothing to reproach myself for." If one accepts the argument as compelling, continued Savage, one should also accept the Independence Axiom, because "if E and E′ are disjoint random events of probabilities $(1 - p)$ and p" the Independence Axiom "is a special case."

The reader familiar with Savage's work will recognize that his "simple but important idea" is none other than the Sure-Thing Principle—that is, the central

[12] The main argument concerned a hypothetical individual whom Samuelson named Ysidro, and whose preferences satisfy monotonicity with respect to stochastic dominance but not the Independence Axiom. Savage argued that Ysidro's preferences are less rational than Samuelson believed, because they could lead Ysidro to accept a "Dutch-book," which is an expression for a bet yielding a sure loss. However, the argument did not convince Samuelson. In a letter of August 16, 1950 (Savage Papers, Box 29), Samuelson wrote Savage that "as yet, I cannot tell that Ysidro has anything 'to reproach himself for.'" Only after accepting the Independence Axiom did Samuelson accept Savage's Dutch-book argument against Ysidro's preferences.

[13] I have slightly modified the symbols Savage used to make his notation more consistent with that used in the rest of this article.

assumption of the subjective version of expected utility theory he later advanced in *The Foundations of Statistics* (1954). In this book (pp. 21–24), contracts I and II became acts *f* and *g*, and the Sure-Thing Principle was formalized as Postulate 2. The Postulate states that the preference between acts *f* and *g* should not depend on the situations in which the acts have the same consequence, such as event E in the above example, but only on the situations, like event E′, in which two acts have different consequences. The letter to Samuelson of August 12, 1950, appears to be Savage's first statement of the Sure-Thing Principle.

Savage's argument provoked Samuelson, who, in a letter dated August 16, 1950 (Savage Papers, Box 29), replied: "Dear Jimmie, I have read your letter hastily and translated its contents into the terminology defined in my earlier letters." However, Samuelson's translation was misguided and, on August 18, 1950 (Samuelson Papers, Box 67), Savage wrote back to correct Samuelson's misunderstanding:

> Dear Paul, your points mistake the meaning of my last letter.… It therefore seems in order to give the argument… once more. If in every event which can possibly occur the consequence of action I is not preferred to that of action II, and if in some possible event the consequence of II is preferred to that of I, then any sane preferer would prefer II to I. Your [Special Independence Assumption] is a very special case of this.

Samuelson did not reply to Savage's letter, probably because he was going to meet Savage two weeks later at a meeting of the Econometric Society held at Harvard University. But we know that Savage's repetition of his argument hit home from the letter that Samuelson sent to Friedman on August 25, 1950 (Samuelson Papers, Box 31), in which, as quoted earlier, Samuelson admitted that Savage was right "on the only important difference between us." As Samuelson explained to Friedman, by this point the only important difference between him and Savage concerned the normative issue of whether the Independence Axiom should be included among the assumptions "defining 'rational behavior.'" Initially Samuelson believed that this was not the case, but Savage's idea had finally persuaded him:

> I have had to review my notion of what it is "reasonable" to postulate … Is it reasonable to postulate that V(A) < V(B) implies V(A,C) < V(B,C) for all C? I must answer with a reluctant but firm Yes.

Samuelson as an Advocate of Expected Utility Theory

Samuelson and Savage both attended the Econometric Society meeting at Harvard in early September 1950. In the paper Samuelson presented at the meeting (a draft of which can be found in Samuelson Papers, Box 152), and for which Savage served as a discussant, he expanded on what he had already written to Friedman. In particular, he declared that thanks to Savage he had come to see the Independence Axiom as a compelling requisite of rational behavior:

> What should be our definition of behavior by a "rational" man? … Dr. Savage has helped persuade me that the Independence Axiom *is not* so much a sore thumb as compared to weaker axioms as I had believed.

However, while Samuelson had come to accept expected utility theory because of the normative force of the Independence Axiom, he remained skeptical about the descriptive power of the theory. Thus, he maintained that expected utility theory does not provide "a very illuminating explanation" of the facts concerning gambling or investment behavior, not even "as a first approximation."

Between the Harvard meeting and the Paris conference of May 1952, Samuelson did not publish on expected utility theory. In Paris, he continued to downplay the descriptive power of the theory, arguing that "from the standpoint of explaining *actual behavior* of men on this planet, the Bernoulli utility hypothesis appears to me of rather trifling importance" (1952 [1966], p. 128). Samuelson also repeated that the decisive reason for his eventual endorsement of expected utility theory was that, thanks to Savage, he had come to see the Independence Axiom as "a natural if not inevitable concept" in the realm of choices between lotteries (p. 130). Samuelson even changed the label for the postulate from "Special Independence Assumption" to "Strong Independence Axiom" (p. 133). The term "special," which Samuelson had used to stress the restrictive and arbitrary character of the assumption, disappeared. It was replaced with the term "strong," by which Samuelson meant to point out that the Independence Axiom was now expressed in terms of preference rather than indifference.[14]

In 1952, Samuelson also organized a symposium on expected utility theory, which was published in the October issue of *Econometrica*. In this symposium, Malinvaud (1952) clarified how the Independence Axiom is hidden in von Neumann and Morgenstern's axiomatization of expected utility theory. Specifically, Malinvaud showed that the Independence Axiom is implied by the fact that von Neumann and Morgenstern's assumptions concern preferences over indifference classes of lotteries rather than preferences over single lotteries. Savage explained why an argument against the Independence Axiom put forward by Wold was "a non sequitur" (Wold, Shackle, and Savage 1952). Samuelson (1952, p. 672) restated the paper he had given at the Paris conference, and explicitly presented the Independence Axiom as "a version of what Dr. Savage calls the 'sure-thing principle.'"

[14] Samuelson's Strong Independence Axiom reads as follows: for all lotteries A, B, and C, $A \succcurlyeq B$ if and only if $pA + (1 - p)C \succcurlyeq pB + (1 - p)C$, where $0 < p < 1$. For more on the Paris conference, see Mongin (2014).

The Evolution of Savage's Views

The discussions of summer 1950 between Samuelson and Savage changed not only Samuelson's thinking about expected utility theory but also Savage's. Samuelson's arguments induced Savage to stop advocating the theory by invoking its simplicity and descriptive power. More importantly, Samuelson's skepticism about the Independence Axiom pushed Savage to formulate the normative justification of it expressed by the Sure-Thing Principle.

Savage explicitly acknowledged the importance that the controversy with Samuelson had on the development of his own ideas. On July 3, 1951, he sent Samuelson the first draft of *The Foundations of Statistics* (Savage Papers, Box 29). In the letter accompanying the manuscript, Savage wrote: "Dear Samuelson: Attached is some dittoed material which I hope soon to complete and redraft as a book. … This work owes much to the written and oral discussions you and I had last summer."

The evidence from the Samuelson–Savage correspondence contradicts an often-told "normative retreat story" about Savage (for example, in Jallais and Pradier 2005). According to this story, Savage retreated to the normative defense of expected utility theory presented in *The Foundations of Statistics* only after violating the theory himself at the Paris conference; this happened when, during a conference break, Allais presented Savage with the choice situations later associated with the expression "Allais paradox" (Allais and Hagen 1979). Savage affirmed that he preferred Lottery 1, which pays 100 million Francs with probability 1, to Lottery 2, which yields 500 million Francs with probability 0.10, 100 million Francs with probability 0.89, and 0 Francs with probability 0.01. He also stated that, between Lottery 3 yielding 100 million Francs with probability 0.11 and 0 Francs with probability 0.89, and Lottery 4 yielding 500 million Francs with probability 0.10 and 0 Francs with probability 0.90, he preferred Lottery 4. However, this pair of preferences—Lottery 1 preferred to Lottery 2, and Lottery 4 preferred to Lottery 3—violates expected utility theory.[15] In *The Foundations of Statistics*, Savage (1954, p. 103) argued that the preferences he had expressed in Paris were in conflict with the Sure-Thing Principle and were therefore erroneous. Accordingly, he corrected himself and argued that, upon reflection, he preferred Lottery 3 to Lottery 4.

Savage's letters to Samuelson show that for him the normative force of the Sure-Thing Principle, and therefore of the Independence Axiom, was a crucial motivation for endorsing expected utility theory well before the Paris conference, and independently of the Allais paradox. If there was one person responsible for Savage's normative turn, it was Samuelson, not Allais.

[15] To see why, notice that, according to expected utility theory, preferring Lottery 1 to Lottery 2 implies that $u(100) > 0.10u(500) + 0.89u(100) + 0.01u(0)$. On the other hand, preferring Lottery 4 to Lottery 3 implies that $0.10u(500) + 0.90u(0) > 0.11u(100) + 0.89u(0)$. It is easy to see that there exists no utility function u satisfying both inequalities, which implies that that pair of preferences cannot be rationalized by expected utility theory. More on Allais and his paradox in the essay in this journal by Munier (1991).

Summary and Conclusion

The Paris conference and the *Econometrica* symposium of 1952 marked the acceptance of expected utility theory as the mainstream model for risky choices in economics, and were instrumental in establishing the "Independence Axiom" as the standard name for the key underlying postulate. Indeed, most of the arguments in favor and against expected utility theory discussed in Paris and in the *Econometrica* symposium had been already addressed by Samuelson, Savage, Marschak, and Friedman in their intense correspondence between May and September 1950.

But while these four major economists all came to accept expected utility theory, their reasons were not the same. Among them, only Friedman accepted expected utility theory because he judged it empirically valid. Marschak, in contrast, accepted the theory because he found the axioms underlying it normatively appealing (see also Marschak 1951). Samuelson remained skeptical about the descriptive power of expected utility theory, and only came to accept the theory when, through the lens of Savage's Sure-Thing Principle, he came to view the Independence Axiom as a requisite for rational behavior in conditions of risk, and thus as normatively compelling. Savage initially advocated expected utility theory by appealing to its simplicity, empirical validity, and normative plausibility, but his controversy with Samuelson induced him to focus on the normative defense of the theory, which he perfected by formulating the Sure-Thing Principle.

The correspondence between Samuelson and Savage of May–August 1950 enhanced the fortunes of expected utility theory in at least two important ways. First, it won over to the cause of expected utility theory a prominent economist, namely Samuelson, who after 1950 contributed to stabilizing the theory as the dominant economic model of choice under risk. Second, it induced Savage to articulate the Sure-Thing Principle, which later became the central normative argument in favor of the theory.

■ *I would like to acknowledge the following for helpful comments on earlier drafts of this paper: Roger Backhouse, Simon Cook, Marco Dardi, Martin Dufwenberg, Itzhak Gilboa, Nicola Giocoli, Catherine Herfeld, Jean-Sébastien Lenfant, Michael Mandler, Massimo Marinacci, Steve Medema, Philippe Mongin, Aldo Montesano, David Teira, Peter Wakker, and Carlo Zappia. I also thank Enrico Moretti, Gordon Hanson, and particularly Timothy Taylor for their very insightful indications as I reshaped the manuscript for this journal. I would also like to acknowledge the kind and helpful assistance of the staff at Yale University Library, the Hoover Institution Archives, and especially the David M. Rubenstein Rare Book & Manuscript Library at Duke University. Finally, I gratefully acknowledge financial support by Fondazione Cariplo and the European Society for the History of Economic Thought.*

References

Allais, Maurice, and Ole Hagen, eds. 1979. *Expected Utility Hypotheses and the Allais Paradox.* Dordrecht: Reidel.

Baumol, William J. 1951. "The Neumann–Morgenstern Utility Index—An Ordinalist View." *Journal of Political Economy* 59(1): 61–66.

Bleichrodt, Han, Chen Li, Ivan Moscati, and Peter P. Wakker. Forthcoming. "Nash Was a First to Axiomatize Expected Utility." *Theory and Decision.*

Clark, John Maurice. 1921. "Soundings in Non-Euclidean Economics." *American Economic Review* 11(1): 132–43.

Dalkey, Norman Crolee. 1949. "A Numerical Scale for Partially Ordered Utilities." RAND Corporation, Research Memorandum 296.

Fishburn, Peter, and Peter Wakker. 1995. "The Invention of the Independence Condition for Preferences." *Management Science* 41(7): 1130–44.

Friedman, Milton. No date. Papers. Hoover Institution Archives, Hoover Institution.

Friedman, Milton, and Leonard J. Savage. 1948. "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy* 56(4): 279–304.

Friedman, Milton, and Leonard J. Savage. 1952. "The Expected-Utility Hypothesis and the Measurability of Utility." *Journal of Political Economy* 60(6): 463–74.

Gilboa, Itzhak, and Massimo Marinacci. 2013. "Ambiguity and the Bayesian Paradigm." In *Advances in Economics and Econometrics: Theory and Applications*, Vol. 1, edited by D. Acemoglu, M. Arellano, and E. Dekel, pp. 179–242. Cambridge University Press.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-Unique Prior." *Journal of Mathematical Economics* 18(2): 141–53.

Hicks, John R. 1931. "The Theory of Uncertainty and Profit." *Economica* 32(2): 170–89.

Jallais, Sophie, and Pierre-Charles Pradier. 2005. "The Allais Paradox and Its Immediate Consequences for Expected Utility Theory." Chap. 2 in *The Experiment in the History of Economics*, edited by Philippe Fontaine and Robert Leonard. New York: Routledge.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–92.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money.* London: Macmillan.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica* 73(6): 1849–92.

Malinvaud, Edmond. 1952. "Note on von Neumann-Morgenstern's Strong Independence Axiom." *Econometrica* 20(4): 679.

Marschak, Jacob. 1948. "Measurable Utility and the Theory of Assets." Cowles Commission for Research in Economics, Economics Discussion Paper 226.

Marschak, Jacob. 1950. "Rational Behavior, Uncertain Prospects, and Measurable Utility." *Econometrica* 18(2): 111–41.

Marschak, Jacob. 1951. "Why 'Should' Statisticians and Businessmen Maximize 'Moral Expectation'?" In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, pp. 493–506. Berkeley: University of California Press.

Mongin, Philippe. 2014. "Le paradoxe d'Allais: comment lui rendre sa signification perdue." *Revue économique* 65(5): 743–79.

Moscati, Ivan. 2016. "Measuring the Economizing Mind in the 1940s and 1950s: The Mosteller–Nogee and Davidson–Suppes–Siegel Experiments to Measure the Utility of Money." *History of Political Economy* 48(annual supplement).

Moscati, Ivan. Forthcoming. *Measuring Utility: From the Marginal Revolution to Neuroeconomics.* Oxford University Press.

Mosteller, Frederick, and Philip Nogee. 1951. "An Experimental Measurement of Utility." *Journal of Political Economy* 59(5): 371–404.

Munier, Bertrand R. 1991. "Nobel Laureate: The Many Other Allais Paradoxes." *Journal of Economic Perspectives* 5(2): 179–99.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18(2): 155–62.

Pigou, Arthur Cecil. 1920. *The Economics of Welfare.* London: Macmillan.

Quiggin, John. 1993. *Generalized Expected Utility Theory—The Rank-Dependent Model.* Amsterdam: Kluwer.

Ricardo, David. 1821. *On the Principles of Political Economy and Taxation*, 3rd edition. London: John Murray.

Samuelson, Paul A. No date. Papers. David M. Rubenstein Rare Book & Manuscript Library, Duke University.

Samuelson, Paul A. 1942. "Fiscal Policy and Income Determination." *Quarterly Journal of Economics* 56(4): 575–605.

Samuelson, Paul A. 1947 [1983]. *Foundations of Economic Analysis.* Harvard University Press.

Samuelson, Paul A. 1950a. "Probability and the Attempts to Measure Utility." *Economic Review* 1(3): 167–73.

Samuelson, Paul A. 1950b. "Measurement of Utility Reformulated." RAND Corporation, Research Memorandum D-0765.

Samuelson, Paul A. 1950c. "Two Queries about Utility and Game Theory." RAND Corporation, Research Memorandum D-0774.

Samuelson, Paul A. 1952 [1966]. "Utility, Preference, and Probability." In *The Collected Scientific Papers of Paul Samuelson*, Vol. 1, edited by Joseph E. Stiglitz, pp. 127–36. MIT Press.

Samuelson, Paul A. 1952. "Probability, Utility, and the Independence Axiom." *Econometrica* 20(4): 670–78.

Samuelson, Paul A. 1966. "A Summing up." *Quarterly Journal of Economics* 80(4): 568–83.

Savage, Leonard J. No date. Papers. Manuscripts and Archives Collection, Yale University Library.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: Dover.

Schmeidler, David. 1989. "Subjective Probability and Expected Utility without Additivity." *Econometrica* 57(3): 571–87.

Shackle, George L. S. 1949. *Expectation in Economics.* Cambridge University Press.

Tintner, Gerhard. 1942. "A Contribution to the Non-Static Theory of Choice." *Quarterly Journal of Economics* 56(2): 274–306.

von Neumann, John, and Oskar Morgenstern. 1944 [1947]. *Theory of Games and Economic Behavior.* Princeton University Press.

Wald, Abraham. 1950. *Statistical Decision Functions.* New York: Wiley.

Wold, Herman, George L. S. Shackle, and Leonard J. Savage. 1952. "Ordinal Preferences or Cardinal Utility?" *Econometrica* 20(4): 661–64.

Journal of Economic Perspectives—Volume 30, Number 1—Winter 2016—Pages 237–244

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate 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., Saint Paul, Minnesota, 55105.

Annual Reports

The 2016 *World Development Report* from the World Bank is focused on the theme of "Digital Dividends." "Digital technologies have spread rapidly in much of the world. Digital dividends—the broader development benefits from using these technologies—have lagged behind." "Perhaps the greatest contribution to growth comes from the internet's lowering of costs and thus from raising efficiency and labor productivity in practically all economic sectors. Better information helps companies make better use of existing capacity, optimizes inventory and supply chain management, cuts downtime of capital equipment, and reduces risk. . . . Vietnamese firms using e-commerce had on average 3.6 percentage point higher TFP

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives, *based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

http://dx.doi.org/10.1257/jep.30.2.237 doi=10.1257/jep.30.2.237

[total factor productivity] growth than firms that did not use it. Chinese car companies that are more sophisticated users of the internet turn over their inventory stocks five times faster than their less savvy competitors. And Botswana and Uruguay maintain unique ID and trace-back systems for livestock that fulfill requirements for beef exports to the EU, while making the production process more efficient." "The biggest gains from digital technologies for the poor are likely to come from lower information and search costs. Technology can inform workers about prices, inputs, or new technologies more quickly and cheaply, reducing friction and uncertainty. That can eliminate costly journeys, allowing more time for work and reducing risks of crime or traffic accidents. Using technology for information on prices, soil quality, weather, new technologies, and coordination with traders has been extensively documented in agriculture…" Available at http://www.worldbank.org/en/publication/wdr/wdr-archive.

The Council of Economic Advisers discusses its own history in Chapter 7 of the 2016 *Economic Report of the President,* titled "The 70th Anniversary of the Council of Economic Advisers." The chapter includes a dose of detailed history of the CEA with quotations and some mini-essays from previous CEA chairs. "CEA has consistently advanced a perspective that emphasizes the importance of decentralized decisions to the effective functioning of our market economy, but which also recognizes that the Federal Government has an important role in macroeconomic stabilization, in correcting market failures, and in ensuring that everyone participates sufficiently in the economy's benefits. Indeed, the Council's very first Report rejected both complete laissez-faire and overreliance on fiscal and monetary remedies as approaches to macroeconomic policy, denoting these two positions, respectively, as the 'Spartan Doctrine of Laissez Faire' and the 'Roman Doctrine of an External Remedy.'" "Joseph Stiglitz claims that 'the money saved from just one of the many bad projects the CEA had helped stop…would have been enough to provide us with a permanent endowment…' Former CEA Chairman Ben Bernanke also emphasized this function when describing economists' role in policymaking more generally, while at the same time emphasizing the limitations of economics: 'Economics is a highly sophisticated field of thought that is superb at explaining to policymakers precisely why the choices they made in the past were wrong. About the future, not so much. However, careful economic analysis does have one important benefit, which is that it can help kill ideas that are completely logically inconsistent or wildly at variance with the data. This insight covers at least 90 percent of proposed economic policies…'" At https://www.whitehouse.gov/sites/default/files/docs/ERP_2016_Book_Complete%20JA.pdf.

The Office of Financial Research, an agency established by the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010, has published its *Financial Stability Report.* The report emphasizes three main risks facing the US economy. "First and most important, credit risks are elevated and rising for U.S. nonfinancial businesses and many emerging markets.…In 2015, U.S. nonfinancial business debt continued to grow rapidly, fueled by highly accommodative credit and underwriting standards. The ratio of that debt to gross domestic product has moved above pre-crisis highs, and corporate leverage continues to rise. . . . The combination of higher corporate leverage, slower global growth and inflation, a stronger dollar, and the plunge in commodity prices is pressuring corporate earnings and weakening the debt-service capacity of many U.S. and emerging market borrowers. A shock that significantly further impairs U.S. corporate or emerging market credit quality could potentially threaten U.S. financial stability." "Second, the low interest rate environment may persist for some time, with associated excesses that could pose financial stability risks. . . . The persistence of low rates contributes to excesses that could pose financial stability risks, including investor reach-for-yield behavior, tight risk premiums in U.S. bond markets, and, as noted, the high level and rapid growth of U.S. nonfinancial business debt." "Third, although the resilience of the financial system has improved significantly in the past five years, it is uneven. . . . Financial activity and risks continue to migrate, challenging existing regulations and reporting requirements. Market liquidity appears to be episodically fragile in major U.S. financial markets, diminishing sharply under stress. Run and fire-sale risks persist in securities financing markets." December 2015, https://financialresearch.gov/financial-stability-reports/files/OFR_2015-Financial-Stability-Report_12-15-2015.pdf.

Symposia

Bernard Hoekman has edited a collection of 19 short and readable essays on *The Global Trade Slowdown: A New Normal?* From his overview for the volume: "One of the 'stylised facts' of the last six decades is that international trade has grown faster than global production and income, in contrast to previous time periods, when the elasticity of trade with respect to output was much lower . . . The ratio between trade and income or output is not a constant. . . . The period between the mid-1980s and the mid-2000s was a major outlier on the upside. It spanned two major geopolitical developments and one economic one: (i) the fall of the Berlin Wall and the re-integration of central and eastern European nations with western Europe; (ii) the re-integration of China into the global economy, following the adoption of an export-oriented growth strategy that culminated with the accession of China to the WTO; and (iii) a great expansion in the use of so-called global value chains (GVCs) by large manufacturers and retailers, involving the outsourcing of parts of the production process to firms located in different countries. Starting in the early 2000s, the rate of global trade growth slowed relative to income growth . . . Indeed, the most recent data suggest that trade is not even keeping up with global output growth and has started to decline . . . Trade is falling across the board, in contrast to the period immediately following the 2008 financial crisis in the US and Europe, when trade by the BRIICS was relatively dynamic. . . . [N]ot only has China's import demand for commodities been falling, but it is also importing fewer manufactured goods, with knock-on effects for major OECD countries and other Asian economies." The 19 chapters that follow explore potential causes of the slowdown—including cyclical, structural, protectionist, and global value chain factors—and look at implications and consequences. 2015. A VoxEU.org eBook, at http://www.voxeu.org/sites/default/files/file/Global%20Trade%20Slowdown_nocover.pdf.

In the *Journal of Economic Education*, five economists put the concept of opportunity cost under a definitional microscope. From the introduction to the symposium by David Colander: "It begins with one by Michael Parkin ['Opportunity Cost: A Reexamination'] who looks at the historical use of opportunity costs and distinguishes between a value specification and a quantity specification. In his article, he traces the history of both specifications and argues in favor of using a quantity specification. That article is followed by three others by economists who have been active in the opportunity cost debate." The economists are Daniel G. Arce, Rod O'Donnell, and Daniel F. Stone. Parkin then closes the symposium with a reply. Winter 2016, http://www.tandfonline.com/toc/vece20/47/1.

The Winter 2016 issue of the *Journal of Policy Analysis and Management* contains a "Point/Counterpoint" exchange on the effects of the 1996 welfare reform legislation. In his overview "Welfare Reform: A 20-Year Retrospective," Richard V. Burkhauser offers some reminders of the intensity of the rhetoric back in 1996 when the welfare reform bill was on the verge of being signed into law. For example, here's Democratic New York Senator Daniel Patrick Moynihan: "The welfare bill terminates the basic federal commitment to support dependent children. It endangers children with absolutely no evidence that this radical idea has even the slightest chance of success… The current batch in the White House have only the flimsiest grasp of social reality, thinking anything doable and equally undoable. As, for example, the horror of this legislation." Ron Haskins takes the glass-half-full position in "TANF at Age 20: Work Still Works." Sandra K. Danziger, Sheldon Danziger, Kristin S. Seefeldt, and H. Luke Shaefer takes the glass-half-empty position in "From Welfare to a Work-Based Safety Net: An Incomplete Transition." The authors then respond to each other. At http://onlinelibrary.wiley.com/doi/10.1002/pam.2016.35.issue-1/issuetoc.

The March 2016 issue of *Finance & Development*, published by the International Monetary Fund, contains five articles on demographic shifts. David E. Bloom contributed the lead article, called "Demographic Upheaval: The world will struggle with population growth, aging, migration, and urbanization." Bloom writes: "The world continues to experience the most significant demographic transformation in human history. Changes in longevity and fertility, together with urbanization and migration, are powerful shapers of our demographic future, and they presage significant social, political, economic, and environmental consequences. … It is unlikely that the worst fears associated with rapid population growth and graying populations will be realized. But a great deal of analysis, debate, behavioral adaptation, and policy reform—in both the public and private spheres—must occur before we can be sure." The titles and subtitles of the other four papers are "Older and Smaller: The fiscal consequences of shrinking and aging populations threaten advanced and emerging market economies alike," by Benedict Clements, Kamil Dybczak, and Mauricio Soto; "She Is the Answer: Women can help offset the problems of an aging population and a shrinking workforce," by Yuko Kinoshita and Kalpana Kochhar;

"Age and Inflation: Baby boomers drove down inflation when they joined the workforce and will drive it up as they retire," by Mikael Juselius and Elöd Takáts; and "Surf the Demographic Wave: Sub-Saharan Africa could reap significant benefits from its growing population—if the transition is well managed," by Vimal Thakoor and John Wakeman-Linn. http://www.imf.org/external/pubs/ft/fandd/2016/03/index.htm.

Potpourri

Charles Bean, Christian Broda, Takatoshi Ito, and Randall Kroszner have written a short book titled *Low for Long? Causes and Consequences of Persistently Low Interest Rates*. "[W]e are led to conclude that there is no single driver of the decline in long-term risk-free real interest rates over the past two decades. Instead, different factors seem to have been more important at different times. In particular: • Demographic pressure associated with increased longevity and lower fertility is likely to have been important, especially during the first half of the period. The surge in Chinese savings is likely to be a particular reflection of these demographic forces. But these pressures are likely to wane in coming years, as the population share of the high-saving middle-aged relative to that of dissaving retirees is presently around its peak. • The gradual integration of China into global financial markets may have also placed downward pressure on the global real interest rate.... • While a decline in the propensity to invest seems less convincing as an explanation of the pre-crisis downward trend in real interest rates, it does seem likely to have played a role in explaining developments since 2008. • Shifts in the supply of, and demand for, safe assets may also have placed downward pressure on the risk-free real rate, particularly since the financial crisis." Geneva Reports on the World Economy 17, October 2015, published by the International Center for Monetary and Banking Studies and the Centre for Economic Policy Research, at http://www.voxeu.org/sites/default/files/file/Geneva17_27oct.pdf.

The Commission on a Global Health Risk Framework for the Future, an international group chaired by Peter Sands, spells out *The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises*. "The World Bank has estimated the economic impact of a severe pandemic (that is, one on the scale of the influenza pandemic of 1918–1919) at nearly 5 percent of global gross domestic product (GDP), or roughly $3 trillion. Some might see this as an exaggeration, but it could also be an underestimate. Aggregate cumulative GDP losses for Guinea, Sierra Leone, and Liberia in 2014 and 2015 are estimated to amount to more than 10 percent of GDP. This huge cost is the result of an epidemic that, for all its horror, infected only about 0.2 percent of the population of Liberia, roughly 0.25 percent of the population of Sierra Leone, and less than 0.05 percent of the population of Guinea, with 11,287 total deaths. The Commission's own scenario modeling, based on the World Bank parameters, suggests that during the 21st century global pandemics could cost in excess of $6 trillion, with an expected loss of more than $60 billion per year.... Against this, we propose incremental spending of about $4.5 billion per year—a fraction of what we spend on other risks to humankind." 2016. National Academies Press, at http://www.nap.edu/catalog/21891/the-neglected-dimension-of-global-security-a-framework-to-counter.

B. Zorina Khan discusses "Inventing Prizes: A Historical Perspective on Innovation Awards and Technology Policy." "The use of prizes and bounties was common in the colonial period, and the Continental Congress in 1783 'recommended to the Legislatures of the several states to … encourage the establishment of useful manufactures either by premiums or by such other means as they may find most effectual.' … The framers of U.S. policies were aware of the options that had prevailed in the colonial period and in Europe, but rejected the use of 'premiums' in favour of property rights in patents." "Whereas, the majority of organizations that had specialized in granting prizes for industrial innovations ultimately became disillusioned with this policy, and the practice of bestowing technology awards declined among both private and public institutions … Judges had to combine technical and industry-specific knowledge with impartiality, but even the most competent personnel could not ensure consistency; decision-making among panels was complicated by differences in standards, interpretation, capture, and risk-aversion." "In England, by the 1820s the Royal Society realized the inefficiencies associated with prizes, and instead switched to lobbying in favour of patents. … The system of inducement prizes in France and England was typically replaced by research grants to underwrite the costs of R&D inputs into the technology production process. Both institutions also switched their mandate towards the provision of information and technical education. The RSA even refused to accept further funding from benefactors who wished to designate prizes, because such endowments hampered their desire to reform their policies away from such targeted awards and towards more productive endeavours for 'the advancement of Natural Knowledge.'" "In any event, history indicates that the evolution of the institution of innovation prizes over the past three centuries serves as a cautionary tale rather than as a success story." *Business History Review*, Winter 2015, vol. 89, no. 4, pp. 631–60.

Interviews

Renee Haltom has an "Interview" with Emi Nakamura. "[I]f one abstracts from the huge number of sales in retail price data, then prices look a lot less flexible than they first appear. … It turns out sales have quite special characteristics that suggest that they do not contribute much to aggregate price flexibility—for example, they are very transient; they often return to the original price after a sale. … To me, the key consequence of sticky prices is that demand shocks matter. Demand shocks can come from many places: house prices, fiscal stimulus, animal spirits, and so on. But the key prediction is that prices don't adjust rapidly enough to eliminate the impact of demand shocks." "I think the Great Recession has actually increased the emphasis in macroeconomics on traditional Keynesian frictions … The models that have been successful in explaining the Great Recession have typically been the ones that have combined nominal frictions with a financial shock of some kind to households or firms. One can also see the effects of traditional Keynesian factors in other countries. Jón [Steinsson] is from Iceland, which experienced a massive exchange rate devaluation during its crisis. Other countries that were part of the euro, such as Spain, did not. I think this probably mattered a lot; if prices and wages were flexible, the distinction between a fixed and flexible exchange rate wouldn't matter. Another example is Detroit. If Detroit had had a flexible exchange rate with the rest of the United States, a devaluation would have been possible to lower the relative wages of autoworkers, which might have been very helpful. Much of what happened during the Great Recession felt like a textbook example of the consequences of Keynesian frictions." *Econ Focus*, published by the Federal Reserve Bank of Richmond (Third Quarter 2015, pp. 26–30). https://www.richmondfed.org/publications/research/econ_focus/2015/q3/interview.

James Guszcza conducts a lively interview in "The Importance of Misbehaving: A Conversation with Richard Thaler." For example, Thaler says: "Economists assume that the people they study, so called homo economicus, or what I call Econs, are really smart. They know as much economics as the best economist. They make perfect forecasts, have no self-control problems and are complete jerks. They'll steal your money if they can and get away with it. Most of the people I meet don't have any of those qualities. They have trouble balancing their checkbook without a spreadsheet. They eat too much and save too little. But nevertheless they'll leave a tip at a restaurant even if they don't plan to go back. So for the last four decades I've been pleading with economists that we should be studying Humans, not these mythical Econ creatures." "Keep in mind that I am still an economist at heart. I would like markets to be more efficient. … I'm a believer in rational behavior as a goal. I just don't think people are very good at it on their own, so we should help if we can." *Deloitte Review*, Issue 18, published January 26, 2016. http://dupress.com/articles/behavioral-economics-richard-thaler-interview.

Conversation Starters

Holly Fretwell proposes "The NPS Franchise: A Better Way to Protect Our Heritage." "Decades of neglect have left the national parks crumbling in disrepair. Rundown infrastructure; encroaching non-native invasive species; unarchived artifacts; poor air quality; dilapidated roads, trails, and public transportation; and overcrowding plague units in the system. While the agency struggles to make ends meet, the size of the agency, the acreage under its control, and number of units it manages continue to grow. Instead of continually adding more acreage for the agency to steward, what if NPS [National Park Service] offered a franchise for entrepreneurs to run new park sites that were deemed of national significance? The land and structures would remain in private hands but be given 'national park' stature. … Don't misunderstand. This is not the April Fool's joke depicting McDonald's Golden Arches National Park or the Nike swoosh on Yosemite's Half Dome. It is quite the opposite. This is a serious strategy to add value to the NPS brand and protect new areas without spreading the NPS budget any thinner. Franchising opportunities would allow individuals advocating for a new park area to drive the management of that park. Rather than hand newly protected areas to a struggling federal agency, conservationists could take responsibility to ensure its protection." *George Wright Forum*, 2015, vol. 32, no. 2, 114–22, http://www.georgewright.org/322fretwell.pdf.

Carol Boyd Leon describes "The Life of American Workers in 1915" along many dimensions of experience including health, income, work, and everyday consumption. As one example, here are some facts about housing. "If you were alive in 1915, chances are you rented your house or apartment; the ratio of renters to homeowners was about 4 to 1 in 1920. In contrast, by 2004, 69 percent of American families owned rather than rented their residence. … The cost of a home in 1915 was about $3,200 ($75,600 in 2015 dollars), compared with today's median home value of $183,500. … Mortgages were typically for just 5 to 7 years and required downpayments ranging from 40 to 50 percent of the home purchase price. In contrast, the median downpayment on a new mortgage in 2015 was 10 percent of the purchase price. Ethnic groups formed their own loan associations because banks could raise the mortgage rate, reduce the loan term to 3 years, and foreclose after two late payments." "Whether or not your abode was a single-family home or a crowded tenement, it probably was heated by a potbelly stove or by a coal furnace in the basement. It wasn't until the coal shortage during World War I that oil or gas-powered central heating became a popular replacement for the hand-fired coal furnaces and stoves. Your home probably wasn't yet wired for electricity; less than a third of homes had electric lights rather than gas or kerosene lamps. However, electricity was the byword of new middle-class homes, which sported electric toasters and coffee pots … Telephones could be found in at least a few million homes. However, direct dialing did not exist until the 1920s. If your home had an indoor toilet, the toilet likely was located in a closet or a storage area. It would be a few more years until it was common for toilets, sinks, and bathtubs to share a room … Although some households had running water in 1915, many rural families and city dwellers did not. Less affluent residents still heated a boiler full of water on a coal or wood range, rubbed clothes on a washboard, used a hand ringer, and hung clothes to dry. Homes without gas or electric heat were harder to clean because of soot from the fireplace or wood stove." *Monthly Labor Review*, published by the US Bureau of Labor Statistics, February 2016, at http://www.bls.gov/opub/mlr/2016/article/pdf/the-life-of-american-workers-in-1915.pdf.

Irwin Collier is building up a collection of original materials focused on the history of graduate education in economics at his website "Economics in the Rear-View Mirror: Archival Artifacts from the History of Economics." Interested in a contemporary article and photo about the AEA Twenty-fifth Anniversary Celebration, held in New York City in 1909? Want to see the PhD exam questions that Jacob Viner wrote at Chicago in 1928? How about Paul Samuelson's reading list for a 1943 course on business cycles? The website http://www.irwincollier.com offers a collection of several hundred of these kinds of items.

Journal of Economic Perspectives—Volume 30, Number 2—Spring 2016—Pages 245–248

Correspondence

To be considered for publication in the Correspondence section, letters should be relatively short—generally less than 1,000 words—and should be sent to the journal offices at jep@jepjournal.org. The editors will choose which letters will be published. All published letters will be subject to editing for style and length.

Scoring Social Security Proposals

One of the responsibilities of Social Security's Office of the Chief Actuary is to project the effects of policy proposals on the program's finances, known as "scoring" a proposal. That these projections are not useful is claimed by Konstantin Kashin, Gary King, and Samir Soneji in "Systematic Bias and Nontransparency in US Social Security Administration Forecasts" (Spring 2015, 29(2): 239–58). But their claim is wrong and the argument apparently behind it seems to rest on a basic error in statistical reasoning.

The authors assert that, when the actuaries assess the impact of potential policy reforms, "the lower bound on the magnitude of [their] forecasting errors exceeds the estimated effect of the reforms" and draw the conclusion that policy discussions using these scores are not well-grounded. They reach this conclusion by using the realized forecast errors for the entire system as a lower bound on the magnitude of forecast errors for policy changes.

Before turning to the apparent logical error, I note the implausibility of their statement as a general claim from the following example (provided by Jeff Brown). Assume the policy change is for the Treasury to transfer $1 million dollars to the Social Security Trust Fund next year. This policy would increase the Trust Fund's balance next year by $1 million with certainty, while the impact on the remainder of the 75-year projection horizon would vary with future interest rates and the date when the Trust Funds are exhausted. The projection of the entire system depends significantly on many other variables as well (which are documented in the annual Social Security Trustees' reports). Indeed, uncertainty about the baseline projection of the Trust Fund's balance next year, by itself, exceeds the $1 million transfer of this policy.

Interpreting the logic of their claim as being that the variance of the forecast error of a policy score is at least as large as the variance of the forecast error of the Social Security baseline provides a possible source of error. The projected impact of a policy change is the difference between the baselines with and without the policy change. Only one of these baselines is observed. The authors treat the forecast error of the baseline without the policy as a lower bound on the forecast error of the difference between the two baselines. To see a problem with this claim, assume that the two baseline forecasts of the cost rate are equal to the expected values of the baseline cost rates, given the stochastic process generating Social Security financial outcomes. Then, the variances in forecast errors equal the variances in outcomes. As a difference between two random variables, the variance of the forecast error of a policy score equals the sum of the variances of the forecast errors of the two baselines less twice the covariance. By ignoring the covariance one would conclude that the variance of the forecast error of the policy score is bounded below by the variance of the forecast error in the baseline. But the covariance should not be ignored. Indeed, for a policy change that is small, the covariance is close to each of the variances, and the variance in the error in scoring the policy change is small relative to the variances in the forecast errors of the baselines. Thus, observations on realized baseline errors do not inform the size of the error distribution of policy scores.

By apparently ignoring the covariance, Kashin, King, and Soneji compare the marginal effect of a specific policy change to the overall uncertainty associated in forecasting Trust Fund finances. As policy scoring by the Office of the Chief Actuary plays a critical role in the discussion of Social Security policy, it is important to correct this erroneous attack on its value, based on an elementary error.

Peter Diamond
Massachusetts Institute of Technology
Cambridge, MA

Response from Kashin, King, and Soneji

We are grateful to Peter Diamond for his interest in our article, which offered the first systematic evaluation, by anyone in or out of government, of the Social Security Office of the Chief Actuary's demographic and financial forecasts and policy scores. We demonstrated that these forecasts depend on nontransparent, unreplicable, and antiquated methods and, as a result, are systematically biased and overconfident.

To clarify what is at issue here, the Office of the Chief Actuary makes baseline forecasts for the future of Social Security, and also estimates the effects of proposed policy changes. It does not offer any uncertainty estimates. Our paper makes claims about severe bias in the baseline estimates, and further claims that similar or greater bias exists in estimates concerning proposed policy changes. We also offer estimates for the extent of uncertainty implied by these biases. Diamond's letter offers no objection to our claims about bias in the baseline estimates or policy proposal, or about our arguments concerning uncertainty surrounding the baseline estimates. Diamond's criticism focuses on the two paragraphs in our article that seek to provide the first uncertainty estimates ever for the gap between the policy counterfactual C and the baseline estimate B.

Diamond offers a thought experiment about estimating the uncertainty in the 75-year forecast around a policy change involving a $1 million payment in the present. His analysis is a special case built on three underlying assumptions, two of which are incorrect in the present setting and a third which depends on an arbitrary choice of a theory of inference. We appreciate the opportunity to clarify how uncertainty depends on the magnitude of the policy shock and covariance, which of course we did not ignore.

First, we switch from Diamond's hypothetical small policy to the more realistic actual massive proposals evaluated by the Office of the Chief Actuary. These include (at the median over the last 15 years) five major provisions, 28 complicated interactions, and an estimated change in the actuarial balance of 100 percent. The uncertainty over time in the costs of these counterfactual proposals C equals i) the uncertainty in the baseline estimates B, plus ii) the uncertainty due to assumptions about each proposal's provisions, interactions, and never-before-observed effects. As a result, the standard deviation of C is much larger than B, that is, $a \equiv \sqrt{V(C)}/\sqrt{V(B)} \gg 1$. An estimate for this ratio, using all proposals evaluated by the Office since 2000, is $a = 3.5$, or $a = 4.3$ after adjusting for characteristics of policies and proposers. Yet, even if a is as small as 2, $V(B)$ is a lower bound of $V(C - B)$, as we claimed. The bound is obtained by rewriting $V(C - B) = V(C) + V(B) - 2\,\mathrm{Cov}(B,C) = [1 + a(a - 2r)]\,V(B)$.

Second, as Diamond writes, his analysis "assume[s] that the two baseline forecasts of the cost rate are equal to the expected values of the baseline cost rates." This unbiasedness claim has been false for 15 years, as documented in our article. Biases in estimates of baseline forecasts B are large and increasing (even though the Office of the Chief Actuary had the luxury of basing its forecasts on a long observed historical record). To claim that forecasts of counterfactual proposals C (based on no observed history) are somehow less biased than B, or to claim that we know that these biases in some way cancel each other out, requires believing in implausible and unobservable coincidences.

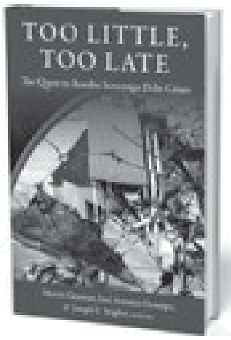
Third, consider the correlation r between errors in B and C across reruns of policy changes across a range of plausible worlds with implementation at year 0 and measurement 75 years later. How one thinks about this correlation actually depends on one's chosen theory of inference. Under frequentist theory, the true potential outcomes are fixed (and so cannot contribute to the variance) and the forecasts are random but almost identical. In this setting, the kind of hypothetical small policy described by Diamond may have r close to 1. In effect, this theory results in recognizing error in observable quantities, while implausibly assuming perfect foresight and no uncertainty for unobserved quantities—that is, in effect assuming that $V\ (C - B) \approx 0$. In contrast, under our preferred Bayesian theory, the true outcome has a random distribution and the forecasts are fixed thus, if a correlation between the errors of B and C is induced at year 0, that correlation can degrade over time. If the correlation degrades after 75 years to $r = 0.5$, our claim holds even in the unlikely case that $a = 1$, or more generally if $a/r \geq 2$. The forecasts of the Office of the Chief Actuary implicitly take the Bayesian view: across all proposals evaluated since 2000, the empirical correlation between their forecasts B and C is only $r = 0.51$, or $r = 0.36$ after we adjust for characteristics of policies and proposers. At one point, Diamond's letter also seems to express support for this Bayesian view, when he writes that "the variances in forecast errors equal the variances in *outcomes*" (emphasis added).

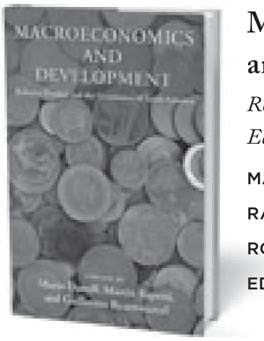
We hope future researchers will improve our uncertainty estimates. Ignoring uncertainty, or assuming it away, does a disservice to science, public policy, and millions of current and future retirees. If the Social Security Administration would follow scientific standards, the replication movement in academia, and recent Executive Orders requiring openness and transparency, more proposals and more science could become part of the political debate.

Konstantin Kashin, Harvard University,
Cambridge, Massachusetts

Gary King, Harvard University, Cambridge, Massachusetts

Samir Soneji, Dartmouth College, Hanover, New Hampshire





Attention
All
Students!

The *Journal of Economic Perspectives*: Proposal Guidelines

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*.

JEL
economics
sustainability
institutions
globalization
growth
game
markets
microfoundations
migration
oligopoly
common knowledge
asymmetric information
wealth
identification
sunspots
welfare
asset pricing
finance
pigouvian tax
arbitrage
property rights
existence
optimization
optimal taxation
time series models
value at risk
externalities
speculation
mechanism design
collusion
forecasting
taxes
welfare
behavorial economics
globalization
real business cycles
rebalancing
strategic trade
inequality
learning
game theory
panel data
finance
agglomeration
export led growth
markets
auctions

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary *JEP* for your e-reader, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $20.00, $30.00, or $40.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

FSC™
FSC
www.fsc.org
MIX
Paper from
responsible sources
FSC™ C101537

The Journal of

Economic Perspectives

Spring 2016, Volume 30, Number 2

AMERICAN ECONOMIC ASSOCIATION