Structural Transformation and the Rural-Urban Divide*

Viktoria Hnatkovska† and Amartya Lahiri‡

March 2013

Abstract

Development of an economy typically goes hand-in-hand with a declining importance of agriculture in output and employment. Given the primarily rural population in developing countries and their concentration in agrarian activities, this has potentially large implications for inequality along the development path. We examine the Indian experience between 1983 and 2010, a period when India has been undergoing such a transformation. We find a significant *decline* in the wage differences between individuals in rural and urban India during this period. However, individual characteristics such as education, occupation choices and migration account for *at most* 40 percent of the wage convergence. We use a two-sector model of structural transformation to rationalize the rest of the rural-urban convergence in India as the consequence of two factors: (i) differential sectoral income elasticities of demand along with productivity growth; and (ii) higher labor supply growth in urban areas. Quantitative results suggest that the model can account for 70 percent of the unexplained wage convergence between rural and urban areas.

JEL Classification: J6, R2

Keywords: Rural urban disparity, education gaps, wage gaps

---

*We would like to thank IGC for a grant funding this research. Thanks also to Paul Beaudry, Robin Burgess, Berthold Herrendorf, and seminar participants at UBC, Wharton, Philadelphia Fed and the IGC-India 2012 conference in Delhi for helpful comments. An online Appendix to this paper is available from the authors' websites.

†Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada and Wharton School, University of Pennsylvania. E-mail address: hnatkovs@mail.ubc.ca.

‡Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail address: amartyalahiri@gmail.com.

1 Introduction

The process of economic development typically involves large scale structural transformation of economies. As documented by Kuznets (1966), structural transformations typically involve a contraction in the agricultural sector accompanied by an expansion of the non-agricultural sectors – manufacturing and services. In as much as the contracting agricultural sector is primarily rural while the expanding sectors mostly urban, the structural transformation process has potentially important implications for the evolution of economic inequality within such developing economies. The process clearly induces large reallocation of workers across sectors as well as requires, possibly, re-training of workers to enable them to make the switch. Not surprisingly, in a recent cross-country study on a sample of 65 countries, Young (2012) finds that around 40 percent of the average inequality in consumption is due to urban-rural gaps.

In this paper we examine the consequences of structural transformation for rural-urban inequality by focusing on the experience of India between 1983 and 2010. Several features of India during this period make it particularly appropriate and informative for understanding the consequences of economic development. First, during this period India has had a very well publicized take-off in macroeconomic growth. As we shall show below, this growth take-off has also been accompanied by a structural transformation of the Indian economy along the lines of the stylized facts documented in Kuznets (1966). Second, the size of the rural sector in India is huge with upwards of 800 million people still residing in the primarily agrarian rural India in 2011. Hence, the scale of the potential disruption and reallocation unleashed by this process is massive.

Our study has two parts. In the first part, we document that there has been a significant decrease in the wage gaps between rural and urban India between 1983 and 2010 with the median wage premium of urban workers declining from 59 percent to 13 percent. However, we also find that conventional covariates of wages including demographics, education, occupations and migration explain *at most* 40 percent of the observed wage convergence. In the second part, we develop a model that can jointly account for the structural transformation of the economy as well as explain the urban-rural wage convergence. Under non-homothetic preferences stemming from a minimum consumption requirement of the agricultural good, our model explains these facts by incorporating two observed features in the Indian data: agricultural productivity growth and faster urban labor force growth relative to rural labor force growth. We show that the model can account for 70 percent of the wage convergence that is left unexplained by the standard covariates of wages.

The empirical analysis in the paper uses six rounds of the National Sample Survey (NSS) of households in India between 1983 and 2010. We start by showing that there has been a significant decline in labor income differences between rural and urban India during this period. Using a simple decomposition exercise we show that almost all of the measured convergence is due to shrinking wage gaps, both between and within occupations, rather than due to labor reallocation across occupations.

The mean wage premium of the urban worker over the rural worker fell significantly from 51 percent to 27 percent while the corresponding median wage premium declined from 59 percent to 13 percent between 1983 and 2010.

What accounts for the wage convergence between rural and urban India? The natural candidates are individual characteristics of workers such as their education levels and occupation choices. We find evidence of significant convergent trends in both education attainment rates as well as the occupation choices of rural workers toward those of urban workers. However, using the decomposition methods of DiNardo, Fortin, and Lemieux (1996) and Firpo, Fortin, and Lemieux (2009) for the entire wage distribution, we show that converging individual characteristics including education and occupation choices can explain *at most* 40 percent of the observed wage convergence between rural and urban areas. Hence, most of the convergence remains unexplained.[1]

A related narrative in the structural transformation literature suggests an important role for migration of workers from rural to urban areas in the process of moving from agriculture to industrial activities. Using the NSS surveys, we find that 5-year net flow of workers from rural to urban areas is small and has remained relatively stable at around 1 percent of all full-time employed workforce. We also find that migrants from rural to urban areas do not earn significantly lower wages than their urban non-migrant counterparts. Moreover, the wage differential between rural and urban non-migrant workers has been narrowing at the same rate as the overall wage gap between rural and urban workers. These results indicate to us that migration did not play an important role in inducing convergent dynamics between urban and rural areas.

Given the large residual wage convergence left unaccounted for by conventional covariates of wages, the second part of the paper focuses on providing a structural explanation for it. In view of the well documented aggregate growth and productivity take-off that occurred in the Indian economy since the 1980s, the model we develop examines the explanatory power of aggregate shocks in accounting for the unexplained wage convergence. Our choices of model building blocks are dictated by the joint requirements of accounting for both the ongoing structural transformation of the economy as well as the rural-urban wage convergence.

Our examination of the aggregate data suggests two key features that may have been important in understanding the dynamic behavior of the urban-rural wage gap and the simultaneous process of structural transformation in India during this period. First, the period between 1983 and 2010 was marked by agricultural productivity growth. Second, the urban labor force grew faster than the rural labor force during this period. While rural to urban migration accounted for some of this relatively

[1] We also examine the effect of an important rural employment program introduced in 2005 called National Rural Employment Guarantee Act (NREGA) on the rural-urban wage gaps. We use a state level analysis and find that the state-level wage and consumption gaps between rural and urban areas did not change disproportionately in the 2009-10 survey round, relative to their trend during the entire period 1983-2010. We also find that states that were more rural, and hence more exposed to the policy, did not exhibit differential responses of the percentile gaps in wages in 2009-10, relative to trend. We conclude that the effect of this program on the gaps was muted. These results are available in an online appendix.

faster increase in urban labor, the majority of it was due to a process of urban agglomeration which led to a number of rural areas getting reclassified as urban due to growth or assimilation into contiguous urban areas due to urban sprawl. Between these two factors, we find that urban sprawl was possibly a bigger contributor to urban growth. This caused previously rural workers to become urban workers in subsequent periods but without having changed their physical location. Importantly, this change in the rural-urban labor force distribution was the outcome of aggregate developments that induced urban agglomeration and hence, is exogenous to the individual worker.[2]

We embed these two exogenous shocks into a model with two sectors (agriculture and non-agriculture) and two factors of production (rural labor and urban labor). Given our finding of low and stable net migration flows and their limited effects on the wage gaps, we shut down all migration possibilities in the model. Individuals are exogenously determined as being either rural or urban and cannot endogenously change that state. To allow for structural change we introduce a minimum consumption need of the agricultural good which makes the income elasticity of demand for the agricultural good lower than the income elasticity of demand for the non-agricultural good.

In our environment, a rise in agricultural productivity releases labor from agriculture which induces the structural transformation of the economy. While this mechanism is well known, it is somewhat less noted that this effect also tends to raise the urban wage while lowering the rural wage. Hence, the rise in agricultural productivity *widens* the wage gap, which is counterfactual.[3] The increase in the relative supply of urban to rural labor, on the other hand, tends to lower the relative wage of urban labor and hence narrows the wage gap. Using a calibrated version of the model we show that these two factors can jointly account for 70 percent of the unexplained wage convergence between rural and urban areas. As the discussion makes clear, neither shock alone can generate both the structural transformation and the wage convergence simultaneously. Hence, one needs both shocks to jointly account for the two data features.

In our model the exogenous increase in the relative supply of urban to rural labor due to urban agglomeration is key to understanding the dynamics of the urban-rural wage gap. More specifically, allowing for endogenous migration and thereby endogenous changes in the relative supply of urban labor is insufficient to generate a *narrowing* of the urban-rural wage gap in response to an increase in agricultural productivity. An increase in agricultural productivity effectively raises the relative demand for urban labor which is used intensively in the non-agricultural sector. Allowing for migra-

[2] This process is important to incorporate into the model both due to the invariant definitions of "rural" and "urban" settlements in the dataset and to endogeneize the changing nature of these formerly "rural" areas. To be precise, in accordance with the Census, NSS Organization of India defines an "urban" area as all places with a Municipality, Corporation or Cantonment and places notified as town area; or all other places which satisfied the following criteria: (i) a minimum population of 5000; (ii) at least 75 percent of the male working population are non-agriculturists; (iii) a density of population of at least 1000 per sq. mile (390 per sq. km.).

[3] A notable and influential exception to this is the work of Caselli and Coleman (2001) who were the first to augment the demand side effect of non-homotheticity with a supply side channel for agricultural labor in order to match prices as well as quantitities in the context of the US structural transformation. Our work is complementary to their's since we too match both quantities and prices simultaneously.

tion makes the relative supply of urban labor an upward sloping function of the relative urban wage with the slope of the schedule depending on the migration cost, amongst other factors. With zero migration cost, the schedule is infinitely wage elastic while with infinite migration cost it is vertical, i.e., has zero wage elasticity. The upward shift of the relative demand for urban labor results in a higher relative urban wage as the economy moves up the relative labor supply schedule. Hence, the best outcome that one can generate by allowing migration in the model is under a perfectly elastic relative urban labor supply schedule which would imply no change in the urban-rural wage gap. Clearly, to generate a *decline* in the wage gap one needs the relative labor supply schedule to shift as well. That is precisely what urban agglomeration does.[4]

Our mechanism for generating structural change relies on lower income elasticity of demand for agricultural goods due to the non-homotheticity in preferences introduced by the minimum consumption need for the agricultural good. This is a *demand-side* effect generated by changing incomes. There is a *supply-side* mechanism that has also been proposed in the literature (dating back to Baumol (1967)) which relies on differential sectoral productivity growth. In particular, Ngai and Pissarides (2007) use a multi-sector model to show that as long as the elasticity of substitution between final goods is less than unity, over time factors would move to the sector with the lowest productivity growth. In the Indian case, this mechanism leads to a counterfactual implication. As we show, productivity growth in non-agriculture was faster than in agriculture. Hence, the Ngai and Pissarides (2007) mechanism would imply that factors should have migrated to the agricultural sector over time while the data shows the opposite. One could get around this by assuming that the elasticity of substitution between final goods is greater than unity. However, given the lack of precise estimates on this elasticity, it seems heroic to put the entire onus of the explanation on the configuration of a poorly measured parameter. Consequently, we shut down this channel by assuming that the elasticity of substitution between final goods is unity. This also implies that the sole reason for structural transformation in the model is the non-homotheticity in preferences introduced by the minimum consumption of agriculture. While we do introduce faster productivity growth in non-agriculture relative to the agricultural sector, its main role in our set-up is to generate an increase in the relative price of agriculture, a feature that characterizes the data during this period.[5]

Instead, the supply-side channel we formalize is complementary to the skill acquisition cost mechanism proposed by Caselli and Coleman (2001) in their study of regional convergence between the North and South of the USA. Like our urban agglomeration shock, in their model a fall in the cost of acquiring skills to work in the non-agricultural sector induces a fall in farm labor supply and leads

[4]In related work, Michaels, Rauch, and Redding (2012) propose a model of urbanization and structural transformation and test it using US and Brazilian data. Crucially, their process of urbanization works through labor migration. As we argued above, migration, by itself, is insufficient for generating wage convergence, which is one of our key facts of interest. For that we need an additional margin that shifts labor supply.

[5]See Laitner (2000), Kongsamut, Rebelo, and Xie (2001) and Gollin, Parente, and Rogerson (2002) for a formalization of the non-homothetic preference mechanism. The assumption of unitary substitution elasticity between final goods also eliminates the factor deepening channel for structural transformation formalized in Acemoglu and Guerrieri (2008). An overview of this literature can be found in Herrendorf, Rogerson, and Valentinyi (2013a).

to an increase in farm wages and relative prices.

Our focus on rural-urban gaps probably is closest in spirit to the work of Young (2012) who has examined the rural-urban consumption expenditure gaps in 65 countries. Like us, he finds that only a small fraction of the rural-urban inequality can be accounted for by individual characteristics, such as education differences. He attributes the remaining gaps to competitive sorting of workers to rural and urban areas based on their unobserved skills.[6] This process, however, relies on rural-urban migration of workers, which, as we showed, underwent little change in India. Our work is also related to an empirical literature studying rural-urban gaps in different countries (see, for instance, Nguyen, Albrecht, Vroman, and Westbrook (2007) for Vietnam, Wu and Perloff (2005) and Qu and Zhao (2008) for China and others). These papers generally employ household survey data and relate changes in urban-rural inequality to individual and household characteristics. Our study is the first to conduct a similar analysis for India and for multiple years, as well as extend the analysis to consider aggregate factors.

Overall, our paper makes three key contributions. First, we believe this is the first paper that provides a comprehensive empirical documentation of the trends in rural and urban disparities in India since 1983 in wages, education and occupation distributions as well as an econometric attribution of the changes in the rural-urban wage gaps to measured and unmeasured factors. Second, we provide a structural explanation for the observed wage convergence which is largely unexplained by the standard covariates of wages. Third, our results suggest a common driving process behind both structural transformation and rural-urban inequality. This latter connection has been largely omitted in the literature.

The rest of the paper is organized as follows: the next section presents the data and some motivating statistics. Section 3 presents the main results on changes in the rural-urban gaps as well as the analysis of the extent to which these changes were due to changes in individual characteristics of workers and their migration decisions. Section 4 presents our model and examines the role of aggregate shocks in explaining the patterns. The last section contains concluding thoughts.

2 Empirical motivation

We start by focusing on differences in labor income between urban and rural areas and trends therein since 1983.[7] Our data comes from successive rounds of the Employment & Unemployment surveys of the National Sample Survey (NSS) of households in India. The survey rounds that we include

[6]Young's explanation based on selection is complementary to Lagakos and Waugh (2012). Our finding of unexplained changes in rural-urban wage gaps over time also finds an echo in the work of Gollin, Lagakos, and Waugh (2012) who find large and unexplained differences in value-added per worker in agriculture relative to non-agriculture in developing countries.

[7]Since a large fraction of rural workers in India may be self-employed and thus do not report wage income, we also consider per capita consumption expenditures, and find that our findings are generally robust, especially for the lower percentiles of the consumption distribution. These results are presented in the online appendix.

in the study are 1983 (round 38), 1987-88 (round 43), 1993-94 (round 50), 1999-2000 (round 55), 2004-05 (round 61), and 2009-10 (round 66). Since our interest is in determining the trends in wages and determinants of wages such as education and occupation, we choose to restrict the sample to individuals in the working age group 16-65, who are working full time (defined as those who worked at least 2.5 days in the week prior to being sampled), who are not enrolled in any educational institution, and for whom we have both education and occupation information. We further restrict the sample to individuals who belong to male-led households.[8] These restrictions leave us with, on average, 140,000 to 180,000 individuals per survey round. Details on our data are provided in Appendix A.1.

The key sample statistics are given in Table 1. The table breaks down the overall patterns by individuals and households and by rural and urban locations. Clearly, the sample is overwhelmingly rural with about 77 percent of individuals on average being resident in rural areas. Rural residents are sightly less likely to be male, more likely to be married, and belong to larger households than their urban counterparts. Lastly, rural areas have more members of backward castes as measured by the proportion of scheduled castes and tribes (SC/STs).

The panel labeled "difference" reports the differences in individual and household characteristics between urban and rural areas for all our survey rounds. Clearly, the share of the rural labor force has declined over time. There were also significant differences in age and family size in the two areas. The average age of individuals in both urban and rural areas increased over time, although the increase was faster in rural areas. The families have also become smaller in both sectors, but the decline was more rapid in urban areas leading to a large differential in this characteristic between the two areas. The shares of male workers, probability of being married and the share of SC/STs have remained relatively stable in both rural and urban areas over time.

Our focus on full time workers may potentially lead to mistaken inference if there have been significant differential changes in the patterns of part-time work and/or labor force participation patterns in rural and urban areas. To check this, Figure 1 plots the urban to rural ratios in labor force participation rates, overall employment rates, as well as full-time and part-time employment rates. As can be see from the Figure, there was some increase in the relative rural part-time work incidence between 1987 and 2010. Apart from that, all other trends were basically flat.

To obtain a measure of labor income we need wages and the occupation distribution of the labor force. Our measure of wages is the daily wage/salaried income received for the work done by respondents during the previous week (relative to the survey week), if the reported occupation during that week is the same as worker's usual occupation (one year reference).[9] Wages can be paid in cash or kind, where the latter are evaluated at current retail prices. We convert wages into real terms using state-level poverty lines that differ for rural and urban sectors.[10] We express all wages in

[8] This avoids households with special conditions since male-led households are the norm in India.

[9] This allows us to reduce the effects of seasonal changes in employment and occupations on wages.

[10] Using poverty lines that differ between urban and rural areas may generate real wage convergence if urban prices

Table 1: Sample summary statistics

	(a) Individuals				(b) Households	
Urban	age	male	married	proportion	SC/ST	hh size
1983	35.03	0.87	0.78	0.22	0.16	5.01
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1987-88	35.45	0.87	0.79	0.21	0.15	4.89
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1993-94	35.83	0.87	0.79	0.23	0.16	4.64
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1999-00	36.06	0.86	0.79	0.23	0.18	4.65
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	36.18	0.86	0.77	0.25	0.18	4.47
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2009-10	36.96	0.86	0.79	0.27	0.17	4.27
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Rural						
1983	35.20	0.77	0.81	0.78	0.30	5.42
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1987-88	35.36	0.77	0.82	0.79	0.31	5.30
	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1993-94	35.78	0.77	0.81	0.77	0.32	5.08
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1999-00	36.01	0.73	0.82	0.77	0.34	5.17
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
2004-05	36.56	0.76	0.82	0.75	0.33	5.05
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
2009-10	37.66	0.77	0.83	0.73	0.34	4.77
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Difference						
1983	-0.17***	0.11***	-0.04***	-0.55***	-0.15***	-0.41***
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
1987-88	0.09	0.10***	-0.03***	-0.58***	-0.16***	-0.40***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1993-94	0.04	0.10***	-0.02***	-0.54***	-0.16***	-0.44***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1999-00	0.05	0.13***	-0.04***	-0.53***	-0.16***	-0.52***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	-0.39***	0.10***	-0.05***	-0.51***	-0.15***	-0.58***
	(0.10)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
2009-10	-0.70***	0.09***	-0.04***	-0.47***	-0.17***	-0.50***
	(0.12)	(0.00)	(0.00)	(0.00)	(0.01)	(0.03)

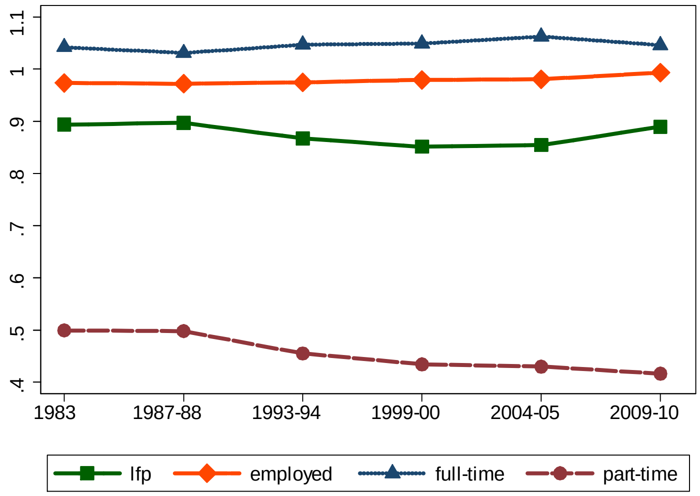
Notes: This table reports summary statistics for our sample. Panel (a) gives the statistics at the individual level, while panel (b) gives the statistics at the level of a household. Panel labeled "Difference" reports the difference in characteristics between rural and urban. Standard errors are reported in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

1983 rural Maharashtra poverty lines.[11] To assess the role played by labor reallocation across jobs, we aggregate the reported 3-digit occupation categories in the survey into two broad occupation

are growing faster than rural prices. This is indeed the case in India during our study period. However, only a small fraction of the observed real wage convergence is driven by the price dynamics. In the online appendix we show that nominal wages are converging slightly faster than real wages (except at the mean) during 1983-2010 period.

[11] In 2004-05 the Planning Commission of India changed the methodology for estimation of poverty lines. Among other changes, they switched from anchoring the poverty lines to a calorie intake norm towards consumer expenditures more generally. This led to a change in the consumption basket underlying poverty lines calculations. To retain comparability across rounds we convert the 2009-10 poverty lines obtained from the Planning Commission under the new methodology to the old basket using a 2004-05 adjustment factor. That factor was obtained from the poverty lines under the old and new methodologies available for the 2004-05 survey year. As a test, we used the same adjustment factor to obtain the implied "old" poverty lines for the 1993-94 survey round for which the two sets of poverty lines are also available from the Planning Commission. We find that the actual old poverty lines and the implied "old" poverty lines are very similar, giving us confidence that our adjustment is valid.

Figure 1: Labor force participation and employment gaps



Note: "lfp" refers to the ratio of labor force participation rate of urban to rural workers; "employed" refers to the ratio of employment rates for the two groups; while "full-time" and "part-time" are, respectively, the ratios of full-time employment rates and part-time employment rates of the two groups.

categories: *non-agricultural* occupations which include white-collar occupations like administrators, executives, managers, professionals, technical and clerical workers and blue-collar occupations such as sales workers, service workers and production workers; and *agrarian* occupations collecting farmers, fishermen, loggers, hunters etc..

We define labor income per worker in Rural (R) or Urban (U) location as the sum of labor income in the two occupations in each location – non-agricultural jobs (occ 1), and agrarian jobs (occ 2):

$$w_t^j = w_{1t}^j L_{1t}^j + w_{2t}^j L_{2t}^j, \tag{2.1}$$

where L_{it}^j is employment share of occupation i in location j, and w_{it}^j is average daily wage in occupation i in location j, with $i = 1, 2$ and $j = U, R$. Also $L_{1t}^j + L_{2t}^j = 1$. The labor income gap between urban and rural areas can then be expressed as

$$\begin{aligned}\frac{w_t^U - w_t^R}{w_t^R} \quad &= \quad \frac{\left(w_{1t}^U - w_{1t}\right) L_{1t}^U + \left(w_{2t}^U - w_{2t}\right) L_{2t}^U}{w_t^R} - \frac{\left(w_{1t}^R - w_{1t}\right) L_{1t}^R + \left(w_{2t}^R - w_{2t}\right) L_{2t}^R}{w_t^R} \\ &+ \frac{\left(w_{1t} - w_{2t}\right)\left(L_{1t}^U - L_{1t}^R\right)}{w_t^R},\end{aligned}$$

where w_{it} is the economy-wide average daily wage in occupation $i = 1, 2$. The decomposition above shows that the urban-rural labor income gap can arise due to two channels. First, the gap may occur if wages and employment within each occupation are different across urban and rural areas (first row on the right in the expression above). We refer to this channel as the *within-occupation* channel. Second, the gap may arise if there is cross-occupation inequality in wages and employment shares (second row in the expression above). This is the *between-occupation* channel.[12]

The last expression above allows us to decompose the change in the labor income gap between period t and $t-1$ as

$$\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \Delta\mu_{1t}^U \bar{L}_{1t}^U + \Delta\mu_{2t}^U \bar{L}_{2t}^U - \Delta\mu_{1t}^R \bar{L}_{1t}^R - \Delta\mu_{2t}^R \bar{L}_{2t}^R$$
$$+ \left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{2t}\right]$$
$$+\Delta L_{1t}^U \left(\bar{\mu}_{1t}^U - \bar{\mu}_{2t}^U\right) - \Delta L_{1t}^R \left(\bar{\mu}_{1t}^R - \bar{\mu}_{2t}^R\right) + (\overline{\eta_{1t} - \eta_{2t}})\Delta\left(L_{1t}^U - L_{1t}^R\right) \tag{2.2}$$

Appendix A.2 presents detailed derivations of this decomposition. Here $\mu_{it}^j \equiv \left(w_{it}^j - w_{it}\right)/w_t^R$, $\eta_{it} \equiv w_{it}/w_t^R$, $\bar{x}_t = \left(x_t + x_{t-1}\right)/2$, and $\Delta x_t = x_t - x_{t-1}$. This decomposition breaks up the change in the labor income gap over time into two components: changes in wages and changes in employment. In addition, the wage component is further split up into a within-occupation component and a between-occupation component. These are, respectively, the first and second rows of equation (2.2). The first row of equation (2.2) summarizes the change in the labor income gap attributable to changes in rural and urban wages in each occupation for a given level of employment. Thus, if rural wages are converging to urban wages in each occupation, so will the overall labor income gap. This is the within-occupation wage convergence component. The second row in equation (2.2) implies that convergence in labor incomes may occur if wages in different occupations converge, i.e., there is between-occupation wage convergence. Lastly, row three gives the part of labor income convergence attributable to changes in urban and rural employment in various occupations for a given average wage. This is the labor reallocation component.

Table 2: Decomposition of labor income gap, 1983-2010

	wage component		labor reallocation	total
	within	between	component	
non-agri	-0.139	-0.177	0.080	-0.235
agrarian	0.010			0.010
total	-0.130	-0.177	0.080	-0.226
% contribution	57.4	78.2	-35.6	100.0

Note: This table presents the decomposition of the change in the urban-rural labor income gap between 1983 and 2010. The decomposition is based on equation (2.2).

Table 2 presents the results of the decomposition by occupations and components. During the 1983-2010 period, the average labor income gap between urban and rural areas declined by 0.226. All of this decline is due to a convergence of wages, with a larger contribution of the between-occupation component relative to the within-occupation component. More precisely, convergence of rural and

[12] This decomposition is similar in spirit to that used by Caselli and Coleman (2001).

urban wages within each occupation has led to a 0.13 (or 57 percent) decline in the labor income gap between the two sectors. The between-occupation wage convergence in urban and rural areas produced an additional 0.18 (or 78 percent) decline in labor income gap. This convergence driven by wages was somewhat offset by reallocation of workers across occupations. The latter has led to an increase of the labor income gap by 0.08.

Clearly, convergence between urban and rural *wages* (both between and within agricultural and non-agricultural jobs) is key to understanding the narrowing labor income gap between the two areas. Motivated by this observation we next investigate wage convergence in rural and urban areas in greater detail by focusing on convergence patterns across the entire wage distribution as well as the factors behind this convergence.

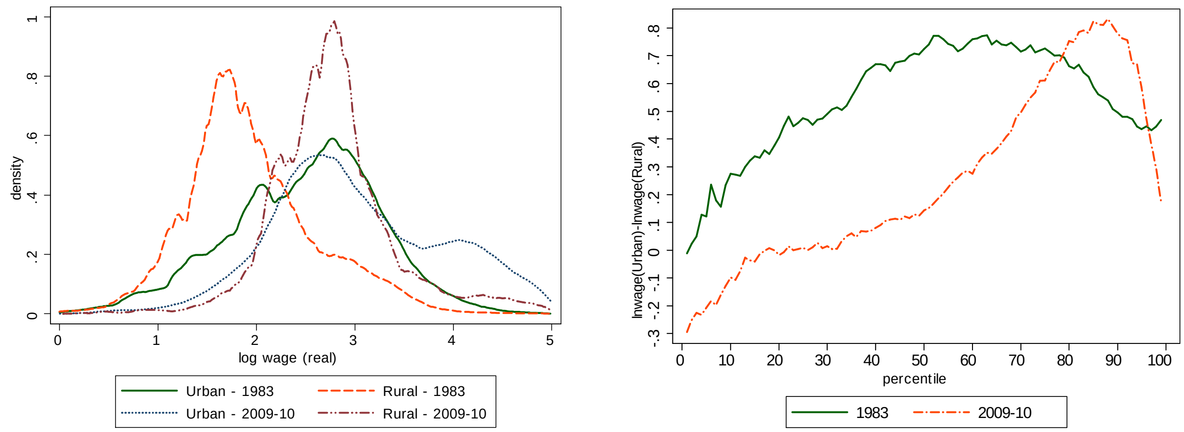
3 Rural-Urban Wage Gaps

We first examine the distribution of log wages for rural and urban workers in our sample. Panel (a) of Figure 2 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2009-10 survey rounds.[13] The plot shows a very clear rightward shift of the wage density function for rural workers during this period. The shift in the wage distribution for urban workers is much more muted. In fact, the mode almost did not change, and most of the changes in the distribution took place at the two ends. Specifically, a fat left tail in the urban wage distribution in 1983, indicating a large mass of urban labor having low real wages, disappeared. Instead a fat right tail has emerged.

Panel (b) of Figure 2 presents the percentile (log) wage gaps between urban and rural workers for 1983 and 2009-10. The plots give a sense of the distance between the urban and rural wage densities functions in those two survey rounds. An upward sloping schedule indicates that wage gaps are rising for richer wage groups. A rightward shift in the schedule over time implies that the wage gap has shrunk. The plot for 2009-10 lies to the right of that for 1983 till the 75th percentile indicating that for most of the wage distribution, the gap between urban and rural wages has declined over this period. Panel (b) shows that the median log wage gap between urban and rural wages fell dramatically. Between the 75th and 90th percentiles however, the wage gaps are larger in 2009-10 as compared to 1983. This is driven by the emergence of a large mass of people in the right tail of the urban wage distribution in 2009-10 period, as we discussed above. A last noteworthy feature is that in 2009-10, for the bottom 20 percentiles of the wage distribution, rural wages were actually higher than urban wages. This was in stark contrast to 1983 when urban wages were higher than

[13] The Mahatma Gandhi National Rural Employment Guarantee Act (NREGA) was enacted in 2005. NREGA provides a government guarantee of a hundred days of wage employment in a financial year to all rural household whose adult members volunteer to do unskilled manual work. This Act could clearly have affected rural and urban wages. To control for the effects of this policy on real wages, we also perform all evaluations on two subsamples: the pre-NREGA and post-NREGA periods. We find that the introduction of NREGA did not change the trends in real wages. Therefore, we proceed by presenting the results for the entire 1983-2010 period. The results for the pre- and post-NREGA subsamples are provided in an online Appendix.

Figure 2: The log wage distributions for urban and rural workers in 1983 and 2009-10



(a) densities of log-wages (b) difference in percentiles of log-wages

Notes: Panel (a) shows the estimated kernel densities of log real wages for urban and rural workers, while panel (b) shows the difference in log-wages between urban and rural workers by percentile. The plots are for the 1983 and 2009-10 NSS rounds.

rural wages for all percentiles.

Figure 2 suggests wage convergence between rural and urban areas. To test whether this is statistically significant, we estimate regressions of the log real wages of individuals in our sample on a constant, controls for age (we include age and age squared of each individual) and a rural dummy for each survey round. The controls for age are intended to account for potential life-cycle differences between urban and rural workers. We perform the analysis for different unconditional quantiles as well as the mean of the wage distribution.[14]

Table 3: Wage gaps and changes

	Panel (a): Rural dummy coefficient					**Panel (b): Changes**		
	1983	1993-94	1999-2000	2004-05	2009-10	83 to 94	94 to 10	83 to 10
10th quantile	-0.208***	-0.031***	-0.013	0.017	0.087***	0.177***	0.118***	0.295***
	(0.010)	(0.009)	(0.008)	(0.012)	(0.014)	(0.013)	(0.017)	(0.017)
50th quantile	-0.586***	-0.405***	-0.371***	-0.235***	-0.126***	0.181***	0.279***	0.460***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.012)	(0.012)	(0.013)
90th quantile	-0.504***	-0.548***	-0.700***	-0.725***	-1.135***	-0.044***	-0.587***	-0.631***
	(0.014)	(0.017)	(0.024)	(0.028)	(0.038)	(0.022)	(0.042)	(0.040)
mean	-0.509***	-0.394***	-0.414***	-0.303***	-0.270***	0.115***	0.124***	0.239***
	(0.008)	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)	(0.014)	(0.014)
N	63981	63366	67322	64359	57440			

Note: Panel (a) of this table reports the estimates of coefficients on the rural dummy from RIF regressions of log wages on rural dummy, age, age squared, and a constant. Results are reported for the 10th, 50th and 90th quantiles. Row labeled "mean" reports the rural coefficient from the corresponding OLS regression. Panel (b) reports the changes in the estimated coefficients over successive decades and the entire sample period. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Panel (a) of Table 3 reports the estimated coefficient on the rural dummy for the 10th, 50th and

[14] We use the Recentered Influence Function (RIF) regressions developed by Firpo, Fortin, and Lemieux (2009) to estimate the effect of the rural dummy for different points of the wage distribution.

90th percentiles as well as the mean for different survey rounds.[15] Clearly, rural status significantly reduced wages for almost all percentiles of the distribution across the rounds. However, the size of the negative rural effect has become significantly smaller over time for the 10th and 50th percentiles as well as the mean (see Panel (b)).[16] The largest convergence occurred for the median. Furthermore, the coefficient on the rural dummy for the 10th percentile has turned positive, indicating a gap in favor of the rural poor. At the same time, the wage gap actually increased over time for the 90th percentile. These results corroborate the visual impression from Figure 2: the wage gap between rural and urban areas fell between 1983 and 2010 for all but the richest wage groups.

3.1 The role of education and occupation

What explains the falling urban-rural wage gaps? We consider two explanations. First, wage convergence may have arisen due to convergence of wage covariates like education and occupation choices. Second, the wage levels of urban and rural workers may have been brought closer together through worker migration between urban and rural areas.

3.1.1 Education trends

Education in the NSS data is presented as a category variable indicating the highest education attainment level for each individual. In order to ease the presentation we proceed in two ways. First, we construct a variable for the years of education. We do so by assigning years of education to each category based on a simple mapping: not-literate = 0 years; literate but below primary = 2 years; primary = 5 years; middle = 8 years; secondary and higher secondary = 10 years; graduate = 15 years; post-graduate = 17 years. Diplomas are treated similarly depending on the specifics of the attainment level.[17] Second, we use the reported education categories but aggregate them into five broad groups: 1 for illiterates, 2 for some but below primary school, 3 for primary school, 4 for middle, and 5 for secondary and above. The results from the two approaches are similar.

Table 4 shows the average years of education of the urban and rural workforce across the six rounds in our sample. The two features that emerge from the table are: (a) education attainment rates as measured by years of education were rising in both urban and rural sectors during this period; and (b) the rural-urban education gap shrank monotonically over this period. The average number of years of education of the urban worker was 164 percent higher than for the typical rural worker in 1983 (5.83 years to 2.20 years). This advantage declined to 78 percent by 2009-10 (8.42 years to 4.72 years). To put these numbers in perspective, in 1983 the average urban worker had

[15] Due to widespread missing rural wage data for 1987-88, we chose to drop that round from the study of wages.

[16] The decline in the mean wage gap reported in Table 3 is slightly higher than the decline in Table 2. This is because we report conditional wage gaps (with controls for age and age squared) in Table 3 and unconditional wage gaps in Table 2.

[17] We are forced to combine secondary and higher secondary into a combined group of 10 years because the higher secondary classification is missing in the 38th and 43rd rounds. The only way to retain comparability across rounds then is to combine the two categories.

slightly more than primary education while the typical rural worker was literate but below primary. By 2009-10, the average urban worker had about a middle school education while the typical rural worker had almost reached primary education. While the overall numbers indicate the still dire state of literacy of the workforce in the country, the movements underneath do indicate improvements over time with rural workers improving faster.[18]

Table 4: Education Gap: Years of Schooling

	Average years of education			Relative education gap
	Overall	**Urban**	**Rural**	**Urban/Rural**
1983	3.02	5.83	2.20	2.64***
	(0.01)	(0.03)	(0.01)	(0.02)
1987-88	3.21	6.12	2.43	2.51***
	(0.01)	(0.03)	(0.01)	(0.02)
1993-94	3.86	6.85	2.98	2.30***
	(0.01)	(0.03)	(0.02)	(0.02)
1999-2000	4.36	7.40	3.43	2.16***
	(0.02)	(0.04)	(0.02)	(0.02)
2004-05	4.87	7.66	3.96	1.93***
	(0.02)	(0.04)	(0.02)	(0.01)
2009-10	5.70	8.42	4.72	1.78***
	(0.03)	(0.04)	(0.03)	(0.01)

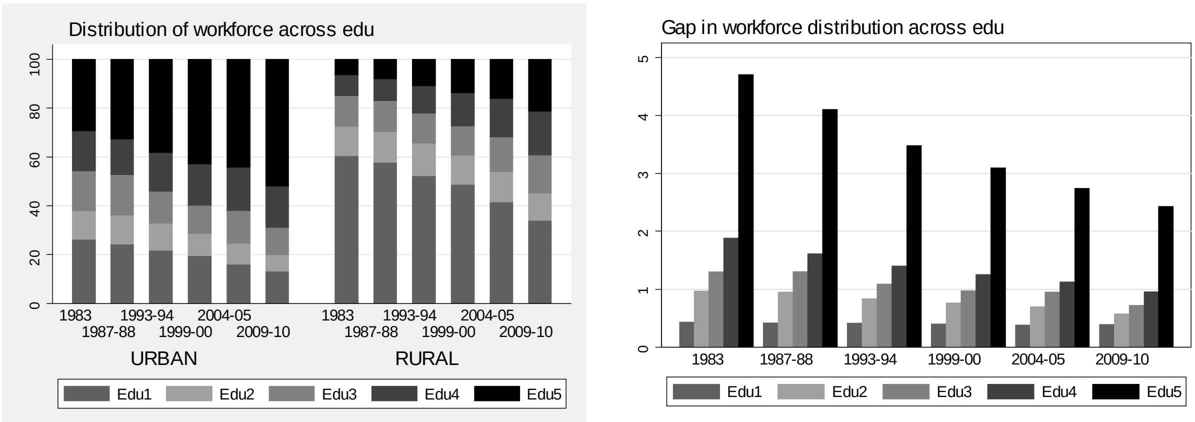
Notes: This table presents the average years of education for the overall sample and separately for the urban and rural workforce; as well as the relative education gap obtained as the ratio of urban to rural education years. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

The time trends in years of education potentially mask the changes in the quality of education. In particular, they fail to reveal what kind of education is causing the rise in years: is it people moving from middle school to secondary or is it movement from illiteracy to some education? While both movements would add a similar number of years to the total, the impact on the quality of the workforce may be quite different. Further, we are also interested in determining whether the movements in urban and rural areas are being driven by very different categories of education.

Panel (a) of Figure 3 shows the distribution of the urban and rural workforce by education category. Recall that education categories 1, 2 and 3 are "illiterate", "literate but below primary education" and "primary", respectively. Hence in 1983, 55 percent of the urban labor force and over 85 percent of the rural labor force had primary or below education, reflecting the abysmal delivery of public services in education in the first 35 years of post-independence India. By 2010, the primary and below category had come down to 30 percent for urban workers and 60 percent for rural workers. The other notable trend during this period is the perceptible increase in the secondary and above category for workers in both sectors. For the urban sector, this category expanded from about 30 percent in 1983 to over 50 percent in 2010. Correspondingly, the share of the secondary and higher educated rural worker rose from just around 5 percent of the rural workforce in 1983 to above 20 percent in 2010. This, along with the decline in the proportion of rural illiterate workers from 60

[18] We have also examined rural-urban gaps in years of education by age and birth cohorts. While we don't report those results here, our principal findings are (i) the gaps have been narrowing over time for all cohorts; and (ii) the gaps are smaller for younger and newer cohorts.

Figure 3: Education distribution



(a) (b)

Notes: Panel (a) of this figure presents the distribution of the workforce across five education categories for different NSS rounds. The left set of bars refers to urban workers, while the right set is for rural workers. Panel (b) presents relative gaps in the distribution of urban relative to rural workers across five education categories. See the text for the description of how education categories are defined (category 1 is the lowest education level - illiterate).

percent to around 30 percent, represent the sharpest changes in the past 27 years.

Panel (b) of Figure 3 shows the changes in the relative education distributions of the urban and rural workforce. For each survey year, the Figure shows the fraction of urban workers in each education category relative to the fraction of rural workers in that category. Thus, in 1983 urban workers were over-represented in the secondary and above category by a factor of 5. Similarly, rural workers were over-represented in the education category 1 (illiterates) by a factor of 2. Clearly, the closer the height of the bars are to one the more symmetric is the distribution of the two groups in that category. As the Figure indicates, the biggest convergence between 1983 and 2010 was in categories 4 and 5 (middle and secondary and above) where the bars shrank rapidly. The trends in the other three categories were more muted as compared to the convergence in categories 4 and 5.

While the visual impressions suggest convergence in education, are these trends statistically significant? We turn to this issue next by estimating ordered multinomial probit regressions of education categories 1 to 5 on a constant and the rural dummy. The aim is to ascertain the significance of the difference between rural and urban areas in the probability of a worker belonging to each category as well as the changes over time in these differences. Table 5 shows the results.

Panel (a) of the Table shows that the marginal effect of the rural dummy was significant for all rounds and all categories. The rural dummy significantly raised the probability of belonging to education categories 1 and 2 while it significantly reduced the probability of belonging to categories 4-5. In category 3 the sign on the rural dummy had switched from negative to positive in 2004-05 and stayed that way in 2009-10.

Panel (b) of Table 5 shows that the changes over time in these marginal effects were also significant

Table 5: Marginal Effect of rural dummy in ordered probit regression for education categories

	Panel (a): Marginal effects, unconditional						Panel (b): Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	2009-10	83 to 94	94 to 10	83 to 10
Edu 1	0.352***	0.340***	0.317***	0.303***	0.263***	0.229***	-0.035***	-0.088***	-0.123***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Edu 2	0.003***	0.010***	0.021***	0.028***	0.037***	0.044***	0.018***	0.023***	0.041***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Edu 3	-0.047***	-0.038***	-0.016***	-0.001*	0.012***	0.031***	0.031***	0.047***	0.078***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Edu 4	-0.092***	-0.078***	-0.065***	-0.054***	-0.044***	-0.020***	0.027***	0.045***	0.072***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Edu 5	-0.216***	-0.234***	-0.257***	-0.276***	-0.268***	-0.284***	-0.041***	-0.027***	-0.068***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
N	164979	182384	163132	173309	176968	136826			

Notes: Panel (a) reports the marginal effects of the rural dummy in an ordered probit regression of education categories 1 to 5 on a constant and a rural dummy for each survey round. Panel (b) of the table reports the change in the marginal effects over successive decades and over the entire sample period. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

for all rounds and all categories. The trends though are interesting. There are clearly significant convergent trends for education categories 1, 3 and 4. Category 1, where rural workers were over-represented in 1983 saw a declining marginal effect of the rural dummy. Categories 3 and 4 (primary and middle school, respectively), where rural workers were under-represented in 1983 saw a significant increase in the marginal effect of the rural status. Hence, the rural under-representation in these categories declined significantly. Categories 2 and 5 however were marked by a divergence in the distribution. Category 2, where rural workers were over-represented saw an increase in the marginal effect of the rural dummy while in category 5, where they were under-represented, the marginal effect of the rural dummy became even more negative. This divergence though is not inconsistent with Figure 3. The figure shows trends in the relative gaps while the probit regressions show trends in the absolute gaps.

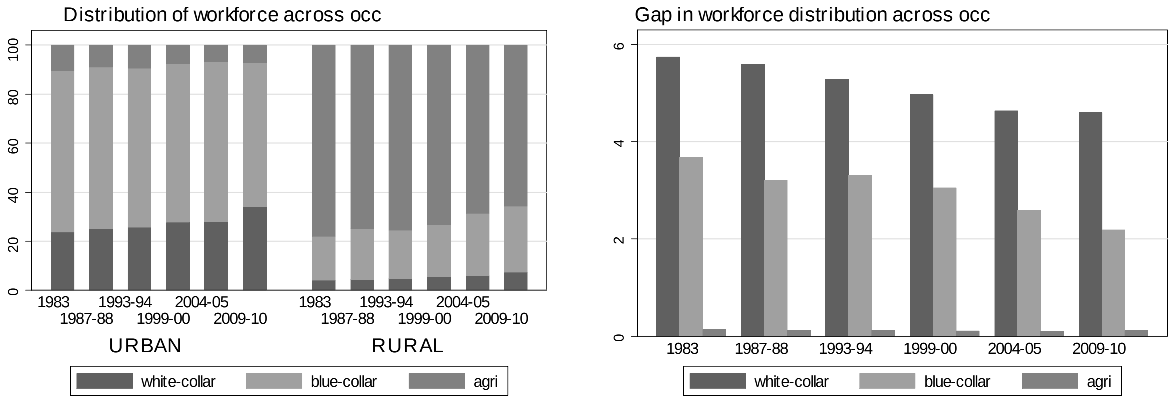
In summary, the overwhelming feature of the data on education attainment gaps suggests a strong and significant trend toward education convergence between the urban and rural workforce. This is evident when comparing average years of education, the relative gaps by education category as well as the absolute gaps between the groups in most categories.

3.1.2 Occupation Choices

We now turn to the occupation choices being made by the workforce in urban and rural areas. To examine this issue, we consider three occupation categories: white-collar occupations, blue-collar occupations, and agricultural occupations, as defined in Section 2. Panel (a) of Figure 4 shows the distribution of these occupations in urban and rural India across the survey rounds while panel (b) depicts the urban-rural gap in these occupation distributions.

The urban and rural occupation distributions have the obvious feature that urban areas have a much smaller fraction of the workforce in agrarian occupations while rural areas have a minuscule share of people working in white-collar jobs. Moreover, the urban sector clearly has a dominance

Figure 4: Occupation distribution



(a) (b)

Notes: Panel (a) of this figure presents the distribution of workforce across three occupation categories for different NSS rounds. The left set of bars refers to urban workers, while the right set is for rural workers. Panel (b) presents relative gaps in the distribution of urban relative to rural workers across the three occupation categories.

in the share of the workforce in blue-collar jobs that pertain to both services and manufacturing. Importantly though, the share of blue-collar jobs has been rising in rural areas. In fact, as Panel (b) of Figure 4 shows, the shares of both white-collar and blue-collar jobs in rural areas are rising faster than their corresponding shares in urban areas. Overall, these results suggest that the expansion of the rural non-farm sector has led to rural-urban occupation convergence.[19]

Is this visual image of convergent trends in occupations statistically significant? We examine this by estimating a multinomial probit regression of occupation choices on a rural dummy and a constant for each survey round. The results for the marginal effects of the rural dummy are shown in Table 6. The rural dummy has a significant negative marginal effect on the probability of being in white-collar and blue-collar jobs, while having significant positive effects on the probability of being in agrarian jobs. However, as Panel (b) of the Table indicates, between 1983 and 2010 the negative effect of the rural dummy in blue-collar occupations has declined (the marginal effect has become less negative) while the positive effect on being in agrarian occupations has become smaller, with both changes being highly significant. Since there was an initial under-representation of blue-collar occupations and over-representation of agrarian occupations in rural areas, this indicate an ongoing process of convergence across rural and urban areas in these two occupation. At the same time, the urban-rural gap in the share of the workforce in white-collar jobs has widened.

Overall, these results show that the employment distribution between urban and rural areas was becoming more uneven in white-collar jobs, in terms of *absolute differences*. In terms of the *relative*

[19] Most of the relative increase in rural blue-collar jobs is accounted for by a two-fold expansion in the share of rural production and transportation jobs. While sales and service jobs in the rural areas expanded as well, the increase was much less dramatic. The relative expansion of rural white collar jobs was spread across most categories of white-collar jobs though the sharpest change was in administrative jobs.

Table 6: Marginal effect of rural dummy in multinomial probit regressions for occupations

	Panel (a): Marginal effects, unconditional						Panel (b): Changes		
	1983	1987-88	1993-94	1999-2000	2004-05	2009-10	83 to 94	94 to 10	83 to 10
white-collar	-0.196***	-0.206***	-0.208***	-0.222***	-0.218***	-0.267***	-0.012***	-0.059***	-0.071***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
blue-collar	-0.479***	-0.453***	-0.453***	-0.434***	-0.400***	-0.318***	0.026***	0.135***	0.161***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)	(0.006)
agri	0.675***	0.659***	0.661***	0.655***	0.619***	0.585***	-0.014***	-0.076***	-0.090***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
N	164979	182384	163132	173309	176968	133926			

Note: Panel (a) of the table presents the marginal effects of the rural dummy from a multinomial probit regression of occupation choices on a constant and a rural dummy for each survey round. Panel (b) reports the change in the marginal effects of the rural dummy over successive decades and over the entire sample period. Agrarian jobs is the reference group in the regressions. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

differences, however, the occupation distribution was converging in white-collar jobs, as Figure 4 shows. Blue-collar and agrarian jobs have shown convergence over time in both absolute and relative terms.

3.1.3 Decomposition of wage gaps

How much of the wage convergence documented above is driven by a convergence of measured covariates? We examine this using two approaches.

DFL decompositions Our first approach is to use the procedure developed by DiNardo, Fortin, and Lemieux (1996) (DFL from hereon) to decompose the overall difference in the observed wage distributions of urban and rural labor within a sample round into two components – the part that is explained by differences in attributes and the part that is explained by differences in the wage structure of the two groups. To obtain the explained part, for each set of attributes we construct a counterfactual density for urban workers by assigning them the rural distribution of the attributes.[20]

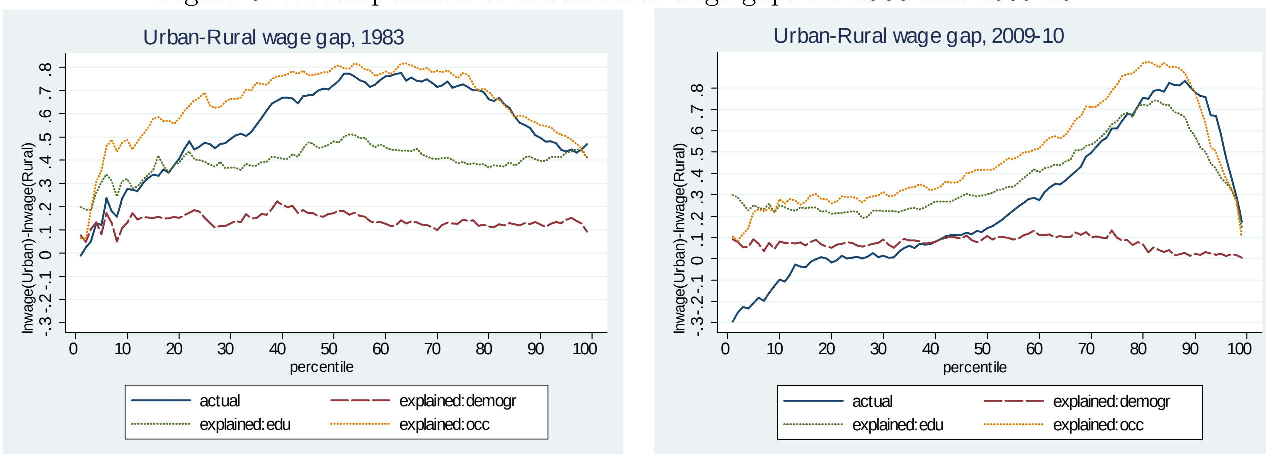
We consider several sets of attributes. First, we evaluate the role of individual demographic characteristics such as age, age squared, a dummy for the caste group (SC/ST or not) and a geographic zone of residence. The latter are constructed by grouping all Indian states into six regions – North, South, East, West, Central and North-East. We control for caste by including a dummy for whether or not the individual is an SC/ST in order to account for the fact that SC/STs tend to be disproportionately rural. Given that they are also disproportionately poor and have little education, controlling for SC/ST status is important in order to determine the independent effect of rural status on wages. Second, we add education to the set of attributes and obtain the incremental contribution

[20] The DFL method involves first constructing a counterfactual wage density function for urban individuals by giving them the attributes of rural households. This is done by a suitable reweighting of the estimated wage density function of urban households. We choose to do the reweighting this way to avoid a common support problem, i.e., there may not be enough rural workers at the top end of the distribution to mimic the urban distribution. The counterfactual urban wage density is then compared with the actual urban wage density to assess the contribution of the measured attributes to the observed wage gap.

of education to the observed wage convergence. Lastly, we evaluate the role played by differences in the occupation distribution for the urban-rural wage gaps.[21]

Figure 5 presents our findings for 1983 (panel (a)) and 2009-10 (panel (b)). The solid line shows the actual urban-rural (log) wage gaps for the entire wage distribution, while the broken lines show the gaps explained by differences in attributes of the two groups, where we introduced the attributes sequentially.

Figure 5: Decomposition of urban-rural wage gaps for 1983 and 2009-10



(a) (b)

Notes: Each panel shows the actual log wage gap between urban and rural workers for each percentile, and the counterfactual percentile log wage gaps when urban workers are sequentially given rural attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus occupations ("occ"). The left panel shows the decomposition for 1983 while the right panel is for 2009-10.

Figure 5 shows that demographic characteristics explain a small fraction of the urban-rural wage gap. Moreover, this fraction remains stable at around 0.1 along the entire distribution in both 1983 and 2009-10. In 1983 differences in education account for almost the entire wage gap at the bottom of the distribution, while differences in occupation explain the wage gap for the upper 50 percent of the distribution. Put differently, education and occupation choices can jointly account for almost the entire wage gap distribution in 1983. Turning to 2009-10 however, the picture is different. Here

[21] Our occupation controls include 7 disaggregated occupation categories. Within the blue-collar jobs we distinguish *sales workers*, which include manufacturer's agents, retail and wholesales merchants and shopkeepers, salesmen working in trade, insurance, real estate, and securities; as well as various money lenders; *service workers*, including hotel and restaurant staff, maintenance workers, barbers, policemen, firefighters; and *production and transportation workers and laborers*, which include among others miners, quarrymen, and various manufacturing workers. The white-collar group is disaggregated into three categories of workers as well. First group consists of *professional, technical and related workers* who include, for instance, chemists, engineers, agronomists, doctors and veterinarians, accountants, lawyers and teachers. The second is *administrative, executive and managerial workers*, which include, for example, officials at various levels of the government, as well as proprietors, directors and managers in various business and financial institutions. The third type of occupations consists of *clerical and related workers*. These are, for instance, village officials, book keepers, cashiers, various clerks, transport conductors and supervisors, mail distributors and communications operators. The seventh group is *agricultural workers*.

differences in education attainments between urban and rural workers explain a large fraction of the gap at the top end of the distribution (70th percentile and above). However, for those below the 70th percentile, covariates such as demographic characteristics, education and occupation choices systematically *over-predict* the actual wage gaps. This is particularly stark for the bottom 15 percent where the actual wage gap is negative while the demographic characteristics, education endowments and differences in occupations predict that the urban-rural gap should be positive 30 percent.

These results suggest that a large part of the observed convergence in wage differences cannot be explained by standard covariates of wages. Hence, the wage structure of urban and rural workers and changes therein during the sample period play an important role in our data. The unexplained component remains large when we consider the wage gaps for each occupation separately. The unexplained component is particularly pronounced in blue-collar and agrarian jobs. Similarly, we find the unexplained component of the between-occupation wage gaps to be large as well.[22] Therefore, both between- and within-occupation components of urban-rural wage gaps contribute to our finding of large wage structure effects.

RIF regressions Our second approach aims to understand the time-series evolution of wage gaps between urban and rural workers. We proceed with an adaptation of the Oaxaca-Blinder decomposition technique to decompose the observed changes in the mean and quantile wage gaps into explained and unexplained components as well as to quantify the contribution of the key individual covariates. We employ Ordinary Least Squares (OLS) regressions for the decomposition at the mean, and Recentered Influence Function (RIF) regressions for decompositions at the 10th, 50th, and 90th quantiles.[23]

Our set of explanatory factors, as before, includes demographic characteristics such as individual's age, age squared, caste, and geographic region of residence. Additionally, we control for the education level of the individual by including dummies for education categories.[24]

Table 7 shows the results of the decomposition exercise. Bootstrapped standard errors are in parenthesis.[25] The Table shows the decomposition of the change in measured gap (column (i)) into the explained and unexplained components (columns (ii) and (iii)), as well as the part of the gap that is explained by education alone (column (iv)). The results indicate that the part of the wage

[22] These results are not presented, but are available in the online appendix.

[23] The inter-temporal decomposition at the mean is in the spirit of Smith and Welch (1989). All decompositions are performed using a pooled model across rural and urban sectors as the reference model. Following Fortin (2006) we allow for a group membership indicator in the pooled regressions. We also used 1983 round as the benchmark sample. Details of the decomposition method can be found in the Appendix A.3.

[24] We do not include occupation amongst the explanatory variables since it is likely to be endogenous to wages. This is a problem for the RIF and OLS regressions, since they impute occupations in the decomposition based on the estimated coefficients, but less so for the DFL decomposition which uses reported individual occupations. In doing so we followed the original application of the DFL method in DiNardo, Fortin, and Lemieux (1996) who include occupation dummies in their estimation of the effects of unionization on the wage distribution in the USA.

[25] In the computations we accounted for the complex survey design of the NSS data. We also use adjusted sampling weights that account for the pooled sampling (over rounds) in our decompositions. The variance is estimated using the resulting replicated point estimates (see Rao and Wu (1988) and Rao, Wu, and Yue (1992)).

Table 7: Decomposing changes in rural-urban wage gaps, 1983 to 2009-10

	(i) measured gap	(ii) explained	(iii) unexplained	(iv) explained by education
10th quantile	-0.371***	-0.096***	-0.275***	-0.059***
	(0.036)	(0.016)	(0.040)	(0.013)
50th quantile	-0.568***	-0.202***	-0.366***	-0.166***
	(0.022)	(0.014)	(0.019)	(0.012)
90th quantile	0.332***	0.229***	0.103***	0.284***
	(0.041)	(0.046)	(0.045)	(0.044)
mean	-0.263***	-0.115***	-0.148***	-0.078***
	(0.019)	(0.014)	(0.017)	(0.012)

Note: This table presents the change in the rural-urban wage gap between 1983 and 2009-10 and its decomposition into explained and unexplained components using the RIF regression approach of Firpo, Fortin, and Lemieux (2009) for the 10th, 50th and 90th quantiles and using OLS for the mean. The table also reports the contribution of education to the explained gap (column (iv)). Bootstrapped standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

gap that is explained by the included covariates varies from 25 percent for the bottom 10 percent to about 90 percent for the top 10 percent. Based on the explained component of the mean and median urban-rural wage gaps, *at most* 40 percent of the decrease in the gap is explained by the included covariates. Importantly, education alone accounts for the majority of the explained component along every point of the distribution.

Overall, our conclusion from the wage data is that wages have converged significantly between rural and urban India during since 1983 for all except the very top of the income distribution. Education has been an important contributor to these convergent patterns. However, on average over 60 percent of the convergence is due to unmeasured factors.

3.2 The Role of Migration

A natural explanation for the narrowing of the wage gaps that we have documented above is migration from rural to urban areas. Rural migration to urban areas would tend to raise rural wages as long as the marginal product of labor in agriculture is positive while simultaneously putting downward pressure on urban wages. This would induce a narrowing of the rural-urban wage gaps.

In order to assess the contribution of migration to wage gaps, we examined the migration data contained in the NSS surveys. Unfortunately, migration particulars are not available in all the survey rounds that we study as questions on migration were not asked at all in most of them. Specifically, we have information on whether a surveyed individual migrated during the previous five years leading up to the survey date for the 38th round (1983) and 55th round (1999-00). We also have this information for the smaller 64th survey round conducted by the NSS in 2007-08. We use information from these three rounds to form an assessment of the role of migration.[26]

Table 8 shows the main patterns of migration for these three rounds. The first feature to note is that the number of recent migrants (those who migrated during the preceding five years) as a share

[26] We identify migrants as individuals who reported that their place of enumeration is different from the last usual residence and who left their last usual place of residence within the previous five years. These variables are available on a consistent basis across the three survey rounds. For these individuals we also know the reason for leaving the last usual residence and its location.

of all full-time employed workers has declined from 7.2 percent in 1983 to 6.2 percent in 2007-08. Of these migrants, the largest single group were those who moved between rural areas, although the share of rural-to-rural migration in overall migration flows has declined from about 50 percent in 1983 to just below 38 percent in 2007-08. The share of urban migrants to rural areas has stayed relatively unchanged around 9-10 percent during this period. In contrast, urban areas have experienced an increase in migration inflows from both rural and urban areas. Thus, the share of rural-to-urban migration in total migration flows has increased from 22 percent in 1983 to about 30 percent in 2007-08. Urban-to-urban migration, which stood at 19 percent in 1983, rose to 23 percent in 2007-08. Interestingly, the majority of the increase in migration to urban areas took place in the latter half of our sample – since 1999-00.

To put these flows in perspective, the rural-to-urban migrants account for around 7 percent of the urban full-time workforce. This share has remained stable over the period. Note that the net flow of workers from rural to urban areas is lower as there is some reverse flow as well.[27] In particular, the net inflow of migrants from rural to urban areas in the five years preceding 1983 was about 4.5 percent of all urban full-time employed workers, while in 2007-08 the corresponding number was 5 percent. As a share of all full-time employed workers, net migration flows from rural to urban areas were about 1 percent in 1983 and 1.3 percent in 2010. While not insignificant, the share of migrant workers from rural areas in the urban workforce is relatively small. Overall, between 1978 and 1983, about 2.1 million people moved from rural to urban locations, on net. During 1995-2000, the corresponding number was 3 million people. Between 2003 and 2008, the net inflow of migrants into urban areas from rural locations was about 6 million people.[28]

Table 8: Migration trends: 1983-2008

	migrant	migrants				*net* rural-to-urban	for job
	total ft	rural-to-urban	urban-to-urban	rural-to-rural	urban-to-rural	urban ft	rural-to-urban
1983	0.072	0.224	0.185	0.496	0.087	0.045	0.778
	(0.001)	(0.005)	(0.005)	(0.006)	(0.003)	(0.002)	(0.010)
1999-00	0.068	0.230	0.182	0.468	0.106	0.037	0.740
	(0.001)	(0.006)	(0.005)	(0.007)	(0.004)	(0.002)	(0.012)
2007-08	0.062	0.301	0.227	0.379	0.084	0.050	0.810
	(0.001)	(0.007)	(0.007)	(0.008)	(0.004)	(0.002)	(0.011)

The last column of Table 8 also shows that the majority of the rural-to-urban migration is job related. The rest is mostly for marriage reasons. The same is true for urban-to-urban migration flows. Interestingly, job related migration from rural to urban areas appears to have increased in 2007-08 relative to 1999-2000 despite the introduction of the rural employment program NREGA in

[27] These bidirectional migration flows were emphasized also in Young (2012).

[28] These numbers were obtained by multiplying the net flow as a share of full-time employed workers by the share of full time employment in the population in that year equal to 0.31. The shares were computed using 1983 NSS survey. Lastly, the resulting number was multiplied by the population in India which was equal to 683.3 million people according to 1981 Census. The corresponding numbers for 1999-00 were: the share of full time employment in the population – 0.35; population in 2001 Census – 1028.7 million; while in 2007-08: the share of full time employment in the population – 0.37; population in 2011 Census – 1210.2 million.

2005. Migration to rural areas is in equal proportion for job, marriage and other reasons.[29]

What do the wage profiles of these recently migrated workers look like? We perform a simple evaluation of migrant workers wages and their effect on urban-rural wage convergence by amending our wage regression specifications in Section 3.1.3 to include four additional dummy variables, each identifying a migration flow between rural and urban areas. We also re-define the rural dummy to identify rural non-migrant workers only. If migration flows contribute significantly to the urban-rural gaps, we should see the coefficient on the rural dummy change in value and/or significance after migration flow dummies are introduced.

Table 9 reports our results for (log) wages. We find that dummies for migration flows from urban areas have coefficients that are positive and significant, suggesting that urban migrants earn more than the benchmark group – urban non-migrants. Migrants from rural areas, in contrast, earn less than urban non-migrants, but the difference is significant mainly for rural-to-rural migration flows. Note also that the negative effects on wages for this group is declining over time, in line with the aggregate wage convergence. Wages of migrants who moved from rural to urban areas are no different than the wages of urban non-migrants.[30] These results apply to both mean and median wages. Do these migration flows contribute to the urban-rural wages gap convergence? A comparison of regression coefficients on the rural dummy in Table 9 and in the benchmark specification without migration flows dummies in Table 3 reveals that they are practically the same. We find that this result also holds for individuals at the two ends of the wage distribution (see Table A1 in Appendix A.4). This suggests that the wage gap between urban and rural non-migrants has been narrowing at the same rate as the overall urban-rural gap.

Table 9: Wage gaps: Accounting for migration

	mean			median		
	1983	1999-00	2007-08	1983	1999-00	2007-08
rural	-0.507***	-0.398***	-0.279***	-0.586***	-0.360***	-0.213***
	(0.008)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
rural-to-urban	-0.021	-0.027	-0.046**	0.035	0.062**	0.020
	(0.021)	(0.021)	(0.023)	(0.024)	(0.025)	(0.024)
urban-to-urban	0.367***	0.529***	0.506***	0.257***	0.261***	0.319***
	(0.024)	(0.041)	(0.033)	(0.025)	(0.019)	(0.022)
rural-to-rural	-0.279***	-0.205***	-0.069***	-0.361***	-0.231***	-0.032
	(0.020)	(0.023)	(0.025)	(0.025)	(0.024)	(0.025)
urban-to-rural	0.258***	0.213***	0.340***	0.113***	0.125***	0.269***
	(0.045)	(0.050)	(0.053)	(0.037)	(0.044)	(0.040)
N	63981	67322	69862	63981	67322	69862

Note: This table reports the estimates of coefficients on the rural dummy and dummies for rural-urban migration flows from the OLS and median RIF regressions of log wages on a set of aforementioned dummies, age, age squared, and a constant. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Overall, we do not find significant evidence that migration may have contributed to the shrinking

[29] Other reasons include natural disaster, social problems, displacement, housing based movement, health care, etc..

[30] The only exception is 2007-08 round where wages of rural-to-urban migrant workers are significantly lower than wages of urban non-migrants, but the difference is small.

wage gaps between rural and urban areas. Of course this conclusion is subject to an obvious caveat that the migration decision itself is endogenous to wage gaps between rural and urban areas. Such an analysis is left for future research.

4 The Role of Aggregate Shocks

The previous results suggest that a majority of the convergence between rural and urban India cannot be accounted for by convergence in the individual characteristics of the two groups. What then explains the convergent trends? One possibility is that aggregate developments during this period may have played a role. Specifically, the period between 1983 and 2010 was marked by deep economic reforms in trade and industrial policy in India, a sharp increase in the aggregate growth rate as well as a structural transformation of the economy. Could these aggregate changes have contributed to the changing rural-urban gaps? In this section we examine this possibility by exploring their effects through the lens of a structural model.

A natural starting point for examining the role of aggregate changes is the traditional theories of structural transformation. They rely on aggregate productivity growth and non-homothetic preferences. These theories imply that as an economy grows the demand for agricultural goods and, therefore, farm labor declines. Thus, they emphasize *demand-side* reasons for structural change. While these models can potentially match the sectoral changes in employment and output, they also imply a decline in the price of agricultural goods and farm relative wages, both of which are inconsistent with the factual movements in sectoral relative prices and wages in India, as we showed above for wages and will discuss below for prices.

To account for the data we augment the standard structural transformation model with a key *supply-side* effect. Specifically, we allow for differential labor force growth in urban and rural areas. This is an important feature of Indian data for the period 1983-2010. It induces an increase in the relative supply of labor to non-farm activities and can, therefore, potentially overturn the counterfactual movements in the agricultural terms of trade and wages implied by the standard demand-side channels of structural transformation.

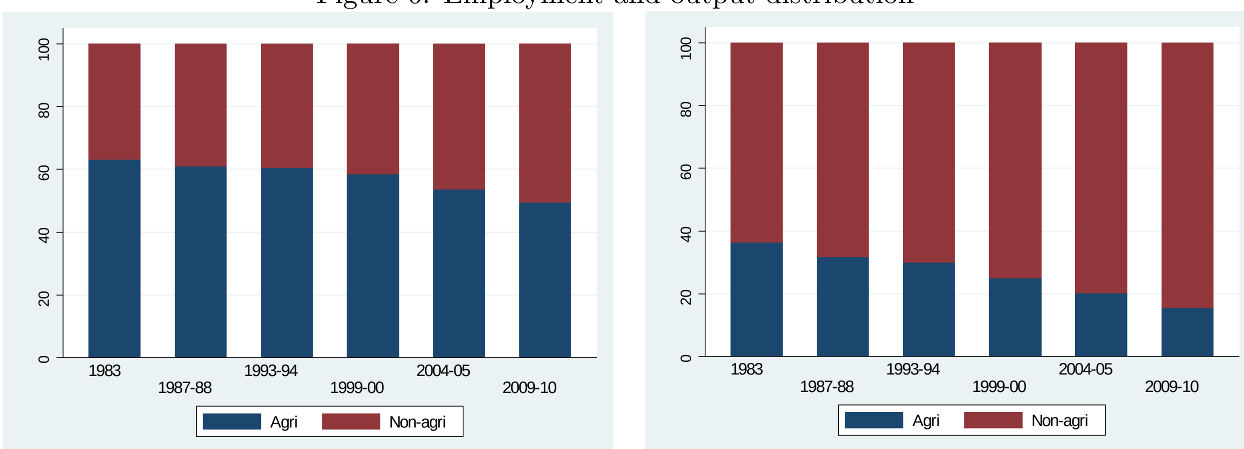
4.1 Key aggregate facts

Before presenting the model it is useful to summarize some key aggregate developments in India during the 1983-2010 period in terms of the structural composition of employment and output, sectoral productivities and relative prices. We want the model to be consistent with these facts.

Note that below we present aggregate facts for industries rather than occupations. This is innocuous since we will only distinguish between agriculture and non-agriculture based activities, and because the vast majority of agricultural jobs are in the agriculture industry. This guarantees a tight mapping between occupations and industries.

The ongoing process of structural transformation of the Indian economy during the 1983-2010 period can be seen through Figure 6 which shows employment shares (panel (a)) and output shares (panel (b)) in agriculture, and non-agriculture. As is easy to see, agriculture has been releasing workers, and its share of output has been declining over time. The non-agricultural sector, on the other hand, has expanded both as a share of employment and as a share of output. These are the textbook features of structural transformation. More precisely, the share of agriculture in total employment has declined from 63 percent in 1983 to 49 percent in 2010. The decline of agriculture in total output was even more pronounced with its share falling from 36 percent in 1983 to 16 percent in 2010.

Figure 6: Employment and output distribution



(a) employment shares (b) output shares

Notes: Panel (a) of this Figure presents the distribution of workforce across agricultural and non-agricultural sectors for different NSS rounds. Panel (b) presents distribution of output across the two sectors.

Underlying this process of structural transformation were changing patterns of sectoral productivity. Figure 7 presents labor productivity and total factor productivity (TFP) in agriculture and non-agriculture during the 1983-2010 period.[31] It is easy to see that productivity in both agriculture and non-agriculture was increasing during this period, with non-agricultural productivity expanding at a much faster pace. More precisely, labor productivity in non-agriculture grew by 163 percent during 1983-2010 period, while it increased by only 50 percent in agriculture. The patterns for TFP are very similar, with agricultural TFP growing by 24 percent between 1983 and 2010, and non-agricultural TFP expanding by a remarkable 119 percent during the same period.

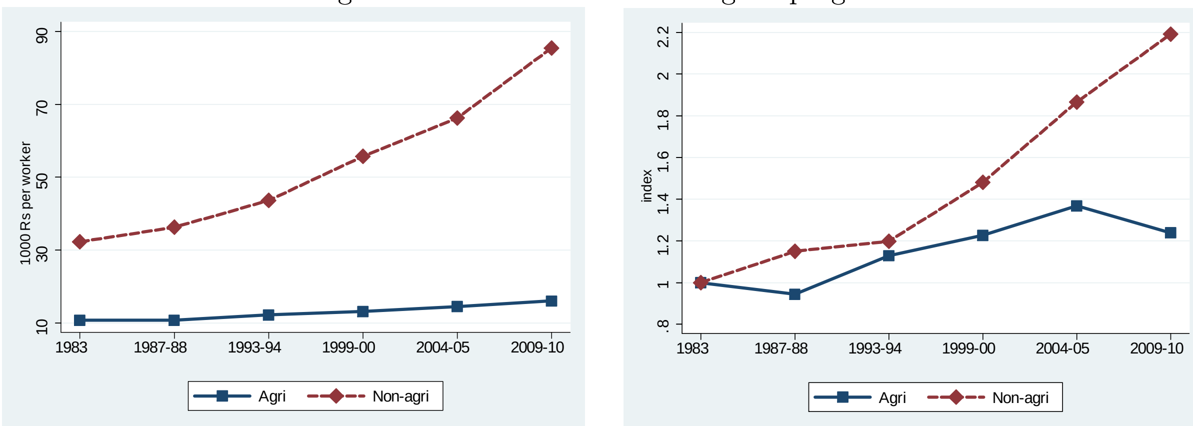
Lastly, Figure 8 presents the evolution of sectoral relative prices during the period 1983-2010. This period was characterized by a 25 percent decline in the relative price of non-agricultural output.[32]

Another fact that we already discussed but highlight here again was the gradual increase in the

[31] Data description and details on how TFP was computed can be found in Appendix A.5.

[32] These numbers were obtained using nominal and real output series from the National Accounts Statistics provided by the Ministry of Statistics and Programme Implementation (MOSPI) of Government of India.

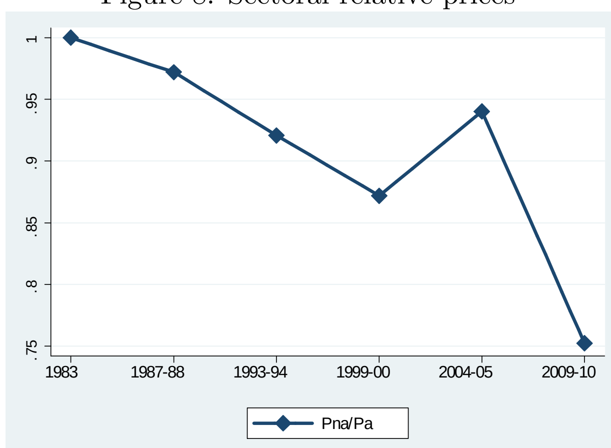
Figure 7: Sector-biased technological progress



(a) Labor productivity per worker (b) TFP

Notes: Panel (a) shows sectoral labor productivity during 1983-2010 period, while panel (b) shows sectoral total factor productivity (TFP) during the same time period.

Figure 8: Sectoral relative prices



Notes: This figure shows the price of non-agricultural output relative to agricultural output.

share of urban labor in the overall Indian labor force. As we showed in Table 1 using the NSS data, the proportion of the urban full time employed labor grew from 22 percent of all full time employed workers to 27 percent between 1983 and 2010. This increase in urban labor finds an echo in the Census figures for India for the overall population where the urban population share rose from 23 to 31 percent between 1981 and 2011. In general, urban labor force growth can occur due to three factors: natural growth due to fertility and death rate differentials; migration; and agglomeration of rural areas into urban areas. In India, natural growth was, and still is, higher in rural areas. Rural-to-urban net migration, as we showed above, was rather small. As we discuss later, even after subtracting cumulated migration flows, the urban labor force share in India rose from 22 to 29 percent between 1983 and 2010.

Instead, the faster rate of urban labor force growth in India occurred through a process of urban agglomeration that led to a number of rural settlements getting reclassified as urban due to growth or assimilation into adjoining urban areas. Evidence of this trend can be found in three interconnected facts. First, according to the decennial censuses, the number of towns and cities in India grew from 3245 in 1981 to 7935 in 2011 marking a startling 145 percent growth in the last 30 years. This is in sharp contrast to a tepid expansion in the number of cities in the seventy preceding years when the number of cities grew from 1811 in 1901 to 2476 in 1971. Second, urban population growth was concentrated in large cities with populations exceeding one million. Both the number of such cities and their share of the urban population have expanded over the last 30 years. In 1981 there were just 12 cities in India with million plus population and they accounted for 26 percent of the urban population. By 2011, the number of million plus cities rose to 53 and they collectively accounted for 43 percent of the urban population. Third, the average population density of the million-plus cities declined from 39000/sq. km to 26000/sq. km.[33]

The first two facts indicate that while there was a sharp increase in the number of new towns, the bulk of the increase in the urban population share was concentrated in existing cities, many of which grew rapidly to cross the million person mark. The third fact about the decline in the population density of the larger cities indicates that the growth in these cities was accommodated by an outward expansion wherein they assimilated neighboring/contiguous rural areas into their fold. Hence, urban sprawl was a key factor behind the growth in the urban labor force.[34] A number of factors could have contributed to such developments, including natural population growth and economic growth, both creating pressure on city land and causing the city limits to expand. In this work we do not seek to explain these causes, but instead take them as exogenous and analyze their consequences.

4.2 A Structural Explanation

We formalize a simple model with two sectors (agriculture and non-agriculture) and two types of labor (rural and urban). The goal of the exercise is to structurally identify the minimal features that can generate three key facts characterizing the Indian economy during 1983-2010 period, as outlined above: (i) a structural transformation; (ii) declining urban-rural wage gaps; and (iii) an improvement in the agricultural terms of trade. We then quantitatively examine the relative contributions of the identified factors to the observed wage convergence.

Consider a two-sector economy that is inhabited by two types of households: rural (R) of measure L_R and urban (U) of measure L_U. The total population is $L = L_U + L_R$. Preferences of agents are

$$V = \frac{c_i^{1-\rho}}{1-\rho}, \qquad i = R, U.$$

[33] These population figures and trends are taken from the Census of India (various rounds) and IIHS (2011).

[34] These developments are not specific to India. A recent report by the United Nations Human Settlements Programme (UN-HABITAT (2012)) shows that urban sprawl has become a remarkable characteristic of urban development worldwide in the last several decades.

Here $1/\rho$ is the elasticity of intertemporal substitution and c_i is the consumption aggregator which is given by

$$c_i = (c_{iA} - \bar{c})^{\theta} (c_{iS})^{1-\theta},$$

where $\bar{c}$ denotes minimum consumption needs of the agricultural good, c_A is consumption of the agricultural good and c_S consumption of the non-agricultural good, and θ is the consumption weight of agricultural goods.

Each household has one unit of labor time that can be used as either agricultural (A) or non-agricultural (S) labor. Hence,

$$1 = l_{iA} + l_{iS}$$

We assume that raw labor can be used directly in sector A but needs to be trained in order to make it productive in sector S. Using good A as the numeraire, the flow budget constraint facing the type-i household is

$$c_{iA} + pc_{iS} = w_{iA}l_{iA} + (w_{iS} - \tau_i) l_{iS} + \Omega_i/L_i \equiv y_i \ , \qquad i = R, U$$

where τ denotes the per unit labor time cost (in terms of the agricultural good) of converting raw labor time into productive labor time for sector S. p is the relative price of good S in terms of good A. Ω_i denotes the total dividend payments received by type-i households from agricultural and non-agricultural firms. w_{ij} is the wage rate received by type-i households for work in sector $j = A, S$. y_i denotes total income of household $i = U, R$.

Both sectors are assumed to be perfectly competitive. The representative firm in each sector produces output using the technology

$$\begin{aligned} Y_A &= AL_A \\ Y_S &= SL_S, \end{aligned}$$

where L_j denotes a sector-specific aggregator function that combines rural and urban labor, while A and S denote total factor productivities in sectors A and S. We shall assume that the sectoral labor aggregators are given by the constant elasticity of substitution functions

$$L_j = \left[\beta_j L_{Uj}^{\phi_j} + \left(1 - \beta_j\right) L_{Rj}^{\phi_j}\right]^{1/\phi_j}, \quad \phi_j \in (-\infty, 1] \ , \quad j = A, S \tag{4.3}$$

where the elasticity of substitution between the two types of labor in sector j is $\frac{1}{1-\phi_j}$. $\phi_j = 1$ corresponds to the linear aggregator where the two are perfect substitutes, while $\phi_j = -\infty$ is the Leontief case of zero substitutability between the two. In the special case of $\phi_j = 0$, we have the unit-elastic Cobb-Douglas case. β_j is the weight on urban labor in sector j.

The structure formalized above contains a few important features. The assumption of a minimum consumption need for the agricultural good is a common feature that is typically introduced in order to generate structural change in multi-sector models. The cost of training unskilled labor in order to make it productive for non-agricultural work is introduced in order to allow the model to generate a wage gap between sectors for the same type of labor. Our production specification of each good being produced by combining two different types of labor reflects our abstraction from migration and location issues in this model. A more elaborate economic environment would allow for multiple locations with comparative advantages in producing different goods and costs of migrating between locations. It is worth reiterating that our focus is on explaining the part of the rural-urban convergence that is not accounted for by education and migration. Hence, we abstract from these margins in the model. We believe that our more parsimonious specification here illustrates the key mechanisms at play without sacrificing analytical tractability.

Optimality for type-i households implies that

$$w_{iA} = w_{iS} - \tau_i \tag{4.4}$$

$$c_{iA} = (1-\theta)\,\bar{c} + \theta y_i \tag{4.5}$$

$$pc_{iS} = (1-\theta)\,(y_i - \bar{c})\,, \tag{4.6}$$

for $i = U, R$. Equation (4.4) makes clear that the cost of training τ is crucial for generating inter-sectoral wage gaps for each type of labor since labor is otherwise freely mobile across sectors.

Since both sectors are perfectly competitive, firms will hire labor till the going nominal wage of each type equals its marginal value product in that sector. This yields two equilibrium conditions from the firm side:

$$p = \frac{MPL_{RA} + \tau_R}{MPL_{RS}} = \frac{MPL_{UA} + \tau_U}{MPL_{US}} \tag{4.7}$$

where MPL_{ij} denotes the marginal product of labor type $i = U, R$ in sector $j = A, S$.

To complete the description of conditions that must be satisfied by all equilibrium allocations, market clearing in each sector dictates that

$$c_{UA}L_U + c_{RA}L_R = Y_A \tag{4.8}$$

$$c_{US}L_U + c_{RS}L_R = Y_S \tag{4.9}$$

Definition: *The Walrasian equilibrium for this economy is a vector of prices and wages* $\{p, w_{UA}, w_{US}, w_{RA}, w_{RS}\}$ *and quantities* $\{c_{UA}, c_{US}, c_{RA}, c_{RS}, l_{UA}, l_{US}, l_{RA}, l_{RS}, Y_A, Y_S\}$ *such that all worker-households and firms satisfy their optimality conditions, budget constraints are satisfied and all markets clear.*

4.2.1 Characterizing the Equilibrium

In order to characterize the equilibrium of this economy, it is convenient to use the following definitions:

$$k_A \equiv \frac{L_{UA}}{L_{RA}},\ k_S \equiv \frac{L_{US}}{L_{RS}},\quad k \equiv \frac{L_U}{L_R}$$

$$s_A \equiv \frac{L_{RA}}{L_R},\quad 1 - s_A = s_S \equiv \frac{L_{RS}}{L_R}$$

k_A and k_S denote the ratio of type U to type R labor in each sector, while k denotes the aggregate relative supply of type U to type R labor. Correspondingly, s_A and s_S denote the share of rural labor in sector A and S, respectively. Using this notation, the market clearing condition for type U labor can be written as

$$k_A s_A + k_S (1 - s_A) = k$$

Hence,

$$s_A = \frac{k - k_S}{k_A - k_S}$$

To solve the model recursively, note that we can use the firm optimality condition (equation (4.7)) to solve for k_A in terms of k_S. Under the general CES labor aggregator (equation (4.3)) with $\phi \neq 0$ this solution is derived by solving for k_A from the condition

$$\beta_S \left[A (1 - \beta_A) \left(\frac{L_A}{L_{RA}} \right)^{1-\phi_A} + \tau_R \right] = (1 - \beta_S) k_S^{1-\phi_S} \left[A \beta_A k_A^{\phi_A - 1} \left(\frac{L_A}{L_{RA}} \right)^{1-\phi_A} + \tau_U \right], \quad \phi \neq 0$$

where $\frac{L_A}{L_{RA}} = \left[\beta_A k_A^{\phi_A} + 1 - \beta_A \right]^{1/\phi_A}$. The solution for k_A can implicitly be defined as

$$k_A = \mu (k_S)$$

One can then use this solution for k_A to characterize the equilibrium for this economy by the system:

$$p = \frac{A \beta_A \{\mu (k_S)\}^{\phi_A - 1} \left[\beta_A \{\mu (k_S)\}^{\phi_A} + 1 - \beta_A \right]^{\frac{1-\phi_A}{\phi_A}} + \tau_U}{S \beta_S k_s^{\phi_S - 1} \left[\beta_S k_S^{\phi_S} + 1 - \beta_S \right]^{\frac{1-\phi_S}{\phi_S}}} \tag{4.10}$$

$$p = \left(\frac{1-\theta}{\theta} \right) \frac{A \left[\beta_A \{\mu (k_S)\}^{\phi_A} + 1 - \beta_A \right]^{\frac{1}{\phi_A}} \left[\frac{k - k_S}{\mu(k_S) - k_S} \right] - \bar{c} (1 + k) - (\tau_R + \tau_U k_S) \left[\frac{\mu(k_S) - k}{\mu(k_S) - k_S} \right]}{S \left[\beta_S k_S^{\phi_S} + 1 - \beta_S \right]^{\frac{1}{\phi_S}} \left[\frac{\mu(k_S) - k}{\mu(k_S) - k_S} \right]} \tag{4.11}$$

This is a two-equation system in two unknowns – k_S and p. Equilibrium solutions for the rest of the endogenous variables are derived recursively from the solutions for k_S and p. Note that equation (4.10) comes from combining the firm optimality conditions $pMPL_{US} = w_{US}$ and $MPL_{UA} = w_{UA}$

with the household optimality condition $w_{UA} = w_{US} - \tau_U$. Equation (4.11) arises from combining the household budget constraints with the market clearing conditions for the two goods.

Before proceeding a couple of observations about the properties of the equilibrium allocations in the model are worthwhile. First, under our specification, equilibrium quantities are going to be independent of S which is productivity in sector s. This can be seen from equations (4.10) and (4.11). Since S enters multiplicatively in the denominator of the right hand side of both equations, the equilibrium k_S will clearly be independent of S. The main role of the s-sector productivity is to affect the terms of trade p. The independence of k_S from S is due to the fact that the costs of training are incurred out of the agricultural good as well as the fact that there is no minimum consumption of the $s-$good. Relaxing these two assumptions would make allocations depend on S as well as A. In terms of the structural transformation dynamics though, these two changes are going to go in the same direction as in the current model. Increasing productivity in sector s will make the s good cheaper which in turn would make the training costs smaller. This would hasten the process of structural transformation. Moreover, the typical way of introducing minimum consumption in the s sector is to introduce it such that the lower income elasticity of agriculture is preserved. Hence, this would not change our structural transformation results. The specification we chose is possibly the most parsimonious one which generates the key facts for India that we outlined above.

4.2.2 A Special Case

In order to build intuition regarding the mechanisms at play in this model as well as the effects of exogenous shocks on factor allocations and prices, we now analytically examine a special case of the model described above by imposing the following two conditions:

Condition 1 *The labor aggregators in the two sectors are of the Cobb-Douglas form given by*

$$L_j = L_{Uj}^{\beta_j} L_{Rj}^{1-\beta_j} \ , \qquad j = A, S. \tag{4.12}$$

Condition 2 *There are no training costs of labor for working in the non–agricultural sector* S, *i.e.,* $\tau_U = \tau_R = 0$.

In this case the solution for k_A is given by

$$k_A = \gamma k_S \ , \quad \gamma \equiv \left(\frac{\beta_A}{1-\beta_A}\right)\left(\frac{1-\beta_S}{\beta_s}\right) \tag{4.13}$$

Moreover, the equilibrium system is given by

$$p = \frac{A\beta_A \gamma^{\beta_A - 1} k_S^{\beta_A}}{S\beta_S k_s^{\beta_S}} \tag{4.14}$$

$$p=\left(\frac{1-\theta}{\theta}\right)\left[\frac{A\gamma^{\beta_A}k_S^{\beta_A-1}\left(\frac{k-k_S}{\gamma-1}\right)-\bar{c}\left(1+k\right)}{Sk_S^{\beta_S-1}\left(\frac{\gamma k_S-k}{\gamma-1}\right)}\right] \tag{4.15}$$

Keeping in mind the empirical reality of rural labor being primarily employed in agriculture, we shall assume throughout the rest of the paper that the agricultural sector uses rural labor more intensively so that $\frac{L_{UA}}{L_{RA}}=k_A<k_S=\frac{L_{US}}{L_{RS}}$. To ensure this alignment of labor intensities we need $\gamma<1$. Hence we shall also impose the following condition:

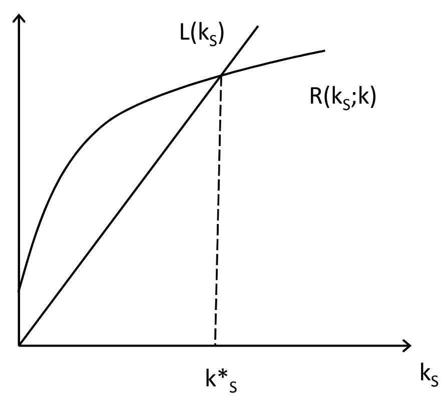
Condition 3 *The agricultural sector uses urban labor less intensively than the non-agricultural sector:* $\beta_A<\beta_S$.

The equilibrium solution can be characterized by using equations (4.14) and (4.15) to solve for k_S. They give

$$\left(1+\frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\right)k_S=\left(1+\frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\frac{1}{\gamma}\right)k+\bar{c}\left(1+k\right)\left(\frac{1-\gamma}{A\gamma^{\beta_A}}\right)k_S^{1-\beta_A} \tag{4.16}$$

The equilibrium is given by the k_S^* which solves this equation. The solution is graphically represented in Figure 9 where $L\left(k_S\right)=\left(1+\frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\right)k_S$ and $R\left(k_S;k,\bar{c},A\right)=\left(1+\frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\frac{1}{\gamma}\right)k+\bar{c}\left(1+k\right)\left(\frac{1-\gamma}{A\gamma^{\beta_A}}\right)k_S^{1-\beta_A}$. Note that $R\left(k_S;k,\bar{c},A\right)$ is increasing and concave in k_S and has a positive intercept term.

Figure 9: Characterizing the equilibrium k_S



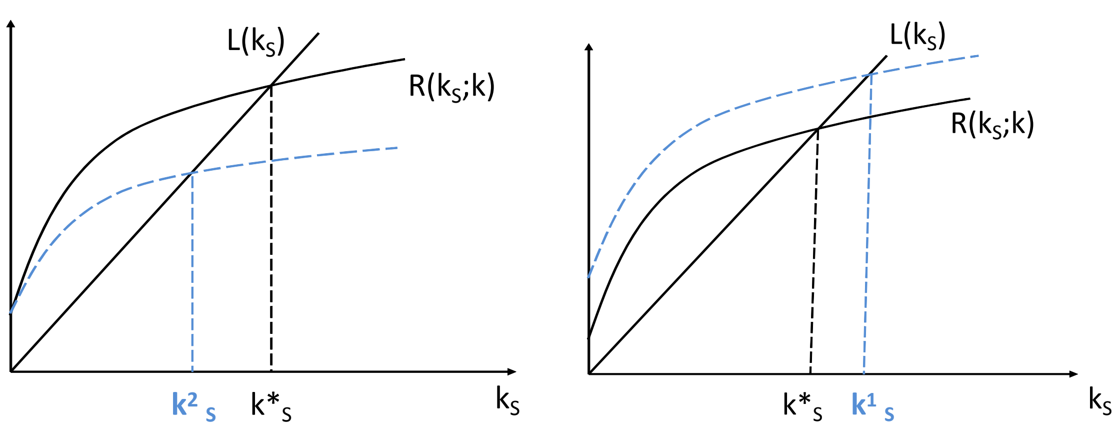
We are interested in analyzing the impact of two kinds of shocks to this economy, both of which are motivated by the data patterns that we documented above. First, we showed that there was an increase in agricultural productivity in India during this period along with an even faster increase in productivity in the non-agricultural sector. Second, we saw that there was an increase in the relative supply of urban to rural labor between 1983 and 2010. Our interest lies in examining the impact of these shocks on the wage gap, the structural transformation of the economy as well as the agricultural terms of trade.

Proposition 1 *Under Conditions 1, 2 and 3, an increase in agricultural productivity: (a) reduces the urban to rural labor ratios in both sectors; (b) raises the relative price p of good S; (c) reduces the rural wage while raising the urban wage; (d) reduces the allocation of rural labor to sector A; and (e) reduces the output share of good A.*

Proof. (a) From Panel (a) of Figure 10, an increase in A reduces the slope of the function $R(k_S; k, \bar{c})$ for all k_S while leaving the intercept unchanged. Hence, the equilibrium k_S falls as does $k_A = \gamma k_S$; (b) $k_S^{\beta_A - \beta_S}$ rises when k_S falls since $\beta_A < \beta_S$. Since k_S falls with A, $p = \frac{A\beta_A \gamma^{\beta_A - 1} k_S^{\beta_A - \beta_S}}{S\beta_S}$ must rise with A; (c) note that w_{UA} is decreasing in k_S while w_{RA} is rising in k_S. The result follows from the fact that $w_{UA} = w_{US}$ and $w_{RA} = w_{RS}$; (d) Using the solution for k_S in $s_A = \frac{k - k_S}{k_A - k_S}$ gives $s_A \equiv \frac{L_{RA}}{L_R} = \left(\frac{1}{1-\gamma}\right)\left(1 - \frac{k}{k_S}\right)$ which is clearly falling in k/k_S. The result follows from the fact that k_S falls when A rises; (e) Using the production functions and the solution for p, the agricultural share of output is $\lambda_A = \frac{1}{1+\left(\frac{1-s_A}{s_A}\right)\frac{\beta_A}{\beta_A}\frac{1}{\gamma}}$ which is rising in s_A. Since s_A declines when A rises, λ_A must also fall with A. ■

The logic underlying Proposition 1 is fairly standard given that this is a model with minimum consumption in the agricultural sector. This introduces differential income elasticities of the two goods. A rise in agricultural productivity A raises overall income which induces a larger increase in the demand for good S relative to the rise in demand for good A. Consequently, the price of the non-agricultural good p rises. As the economy shifts towards the non-agricultural sector, it begins to reallocate both urban and rural labor from agriculture to non-agriculture. Since agriculture is more rural labor intensive, it releases proportionately more rural labor which in turn reduces the urban to rural labor ratio in both sectors. The greater relative employment of rural labor in both sectors raises the returns to urban labor. Hence the urban wage rises while the rural wage falls.

Figure 10: Comparative static effects on k_S



(a) Rise in agricultural TFP

(b) Rise in relative urban labor supply

Proposition 2 *Under Conditions 1, 2 and 3, an increase in the stock of urban labor relative to rural labor, has the following effects: (a) it raises the urban to rural labor ratios in both sectors; (b)* *it reduces the relative price p of good S; (c) it raises the rural wage while reducing the urban wage; (d) it has an ambiguous effect on the allocation of rural labor to sector A; and (e) has an ambiguous effect on the output share of good A.*

Proof. (a) From Panel (b) of Figure 10, an increase in k shifts up the intercept of the function $R(k_S; k, \bar{c})$ while also making it's slope steeper at each point. Since $L(k_S)$ remains unchanged, the new equilibrium k_S is unambiguously higher. Hence, $k_A = \gamma k_S$ is higher as well; (b) It is easy to check that $p = \frac{A\beta_A \gamma^{\beta_A - 1} k_S^{\beta_A - \beta_S}}{S\beta_S}$ falls when k_S rises since $\beta_A < \beta_S$; (c) note that w_{UA} is decreasing in k_S while w_{RA} is rising in k_S. The result follows from the fact that $w_{UA} = w_{US}$ and $w_{RA} = w_{RS}$; (d) Using the solution for k_S in $s_A = \frac{k - k_S}{k_A - k_S}$ gives $s_A \equiv \frac{L_{RA}}{L_R} = \left(\frac{1}{1-\gamma}\right)\left(1 - \frac{k}{k_S}\right)$ which is clearly falling in k/k_S. The condition $L(k_S) = R(k_S; k, \bar{c})$ can be rewritten as

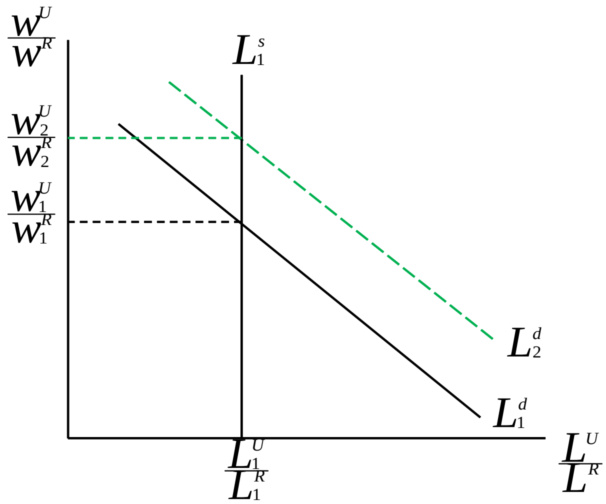
$$1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S} = \left(1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\frac{1}{\gamma}\right)\frac{k}{k_S} + \bar{c}(1+k)(1-\gamma)k_S^{-\beta_A}$$

Since k_S is rising in k, the effect of an increase in k_S on $\bar{c}(1+k)(1-\gamma)k_S^{-\beta_A}$ is ambiguous which implies that effect on k/k_S is also ambiguous; (e) Define the agricultural share of output as $\lambda_A = \frac{Y_A}{Y_A + pY_S}$. Using the production functions and the solution for p this can be written as $\lambda_A = \frac{1}{1+\left(\frac{1-s_A}{s_A}\right)\frac{\beta_A}{\beta_A}\frac{1}{\gamma}}$ which is rising in s_A. Since s_A responds ambiguously to a rise in k, the response of λ_A must also be ambiguous. ■

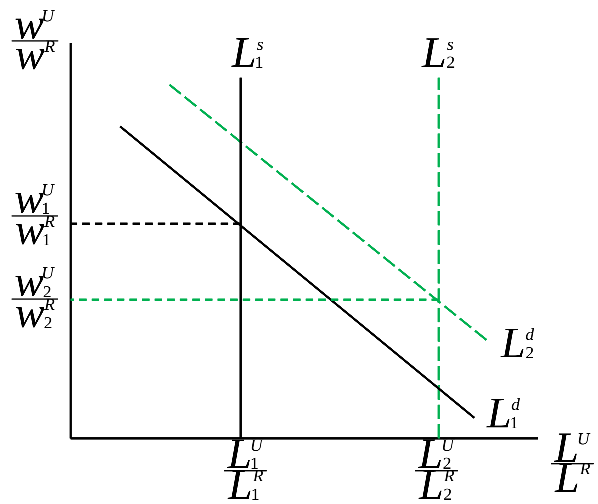
Intuitively, a rise in the urban to rural labor ratio k creates an excess supply of urban labor in both sectors thereby raising the urban to rural labor ratio in both sectors. Since the non-agricultural sector uses urban labor more intensively, it expands relatively more than the agricultural sector. Consequently, the relative price of the S good fall, i.e., p falls. The rise in the sectoral urban to rural labor ratios also cause rural wages to rise and urban wages to decline. This is similar to the Stolper-Samuelson effect in two-sector two-factor models. The effect on relative outputs of the two sectors is reminiscent of the Rybczynski effect of a rise in relative factor endowments with the caveat that the sectoral terms of trade are endogenous here as opposed to the exogenous terms of trade underlying the Rybczynski effect.

Proposition 2 makes clear the importance of the relative urban labor supply increase in generating the urban-rural wage convergence. The underlying mechanism is illustrated in Figure 11 which plots the relative demand and supply for urban labor against the relative urban to rural wage. Under our specification with no migration, the relative urban labor supply is exogenously given and hence independent of the relative urban wage. The relative urban labor demand, on the other hand, is a decreasing function of the relative urban wage. Panel (a) of Figure 11 shows that an increase in agricultural TFP shifts the relative demand schedule for urban labor to L_2^d which raises the relative urban wage to $\frac{w_2^U}{w_2^R}$. This is a key result in Proposition 1. Panel (b) of Figure 11 shows that if this labor demand shock is also accompanied by a large enough increase in relative urban labor supply to L_2^s then the relative urban wage can decline. Note though that Proposition 2 also suggests that an increase in relative urban labor supply alone may not induce a structural transformation of the economy even though it would induce a decline in the relative urban wage. Hence, we need both effects to be operative simultaneously.

Figure 11: Comparative static effects on wage gaps



(a) Rise in agricultural TFP



(b) Rise in relative urban labor supply

Figure 11 can also be used to illustrate the implication of our assumption of no urban-rural migration. Allowing for migration would, in effect, make the relative labor supply schedule an upward sloping function of the relative wage with the slope of the function depending on migration costs and how those costs change with increasing migration. Clearly, an increase in relative urban labor demand due to a positive shock to agricultural TFP would still have the effect of raising the relative urban wage. In the extreme case of a perfectly elastic relative urban labor supply schedule (under a zero cost of migration), the shift in relative urban labor demand would keep the relative urban wage unchanged. In order for the model to generate a *decline* in the relative urban wage, the relative urban labor supply schedule has to shift out. This is exactly what the process of urban agglomeration does in our model.

To summarize, Propositions 1 and 2 highlight three important features of our model economy. First, shocks to both productivity and the relative supply of urban to rural labor are necessary in order to jointly explain the observed changes in relative wages, agricultural terms of trade and the structural transformation. Increases in the relative endowment of urban labor gets the relative wage and terms of trade movements right but has ambiguous implications for the structural transformation of the economy. On the other hand, an increase in agricultural productivity generates the structural transformation but has counterfactual predictions for the wage gap as well as the terms of trade. Second, without the minimum consumption requirement the model cannot generate any structural transformation in this economy. This can be checked by setting $\bar{c} = 0$ in equation (4.16). Third, the sectoral urban to rural labor allocations are independent of the non-agricultural productivity parameter S. However, the relative price p does depend on S. In particular, suppose A and S both rise but A/S declines. In this case p could fall (i.e., the agricultural relative price could rise) in response to an increase in A as long as the fall in $\frac{A}{S}$ is large enough to offset the fall in k_S. This is easily ascertained from the expression $p = \frac{A\beta_A \gamma^{\beta_A - 1} k_S^{\beta_A}}{S\beta_S k_s^{\beta_S}}$, which must hold in equilibrium.

4.3 Quantitative Results

We now quantitatively assess the ability of the full model to explain the observed rural-urban wage dynamics along with the aggregate macroeconomic facts. We conduct the following experiment. First, we calibrate the key parameters of the model to match the urban-rural wage gaps, sectoral employment distribution in rural and urban areas, etc. in 1983. To be consistent with our empirical findings we control for education differences across sectors and migration flows in our calibration. We then perturb the model with two shocks: (a) shocks to rural and urban labor supplies; and (b) shocks to agricultural and non-agricultural productivity. These shocks are measured from the data. Keeping all other parameters unchanged, we examine the urban-rural gaps in 2010 that the model generates in response to these measured shocks. This enables us to quantify the contribution of these shocks to the observed convergence in urban-rural wages and employment between 1983 and 2010.

4.3.1 Calibration for 1983

To calibrate the urban and rural share in the labor force we use Census of India and NSS data. The Census is conducted every 10 years on the first year of each decade. Thus, in 1981 the total population of India was 683.3 million people, of which 525.6 million lived in rural areas and 157.7 million lived in urban areas. To obtain labor force numbers we multiply these population figures by the share of working age population in 1983 from the NSS equal to 0.54 in rural areas and 0.59 in urban areas; and by the labor force participation rate in 1983 from the NSS equal to 0.66 in rural areas and 0.59 in urban areas. These calculations give us the rural labor force share at 78 percent of total and urban labor force share at 22 percent of total in 1983.

The rest of the parameters are chosen to match a set of data moments. More precisely, we choose nine parameters that minimize the distance between nine moments in the data in 1983 and in the model. First, we match the sectoral distribution of the labor force in rural and urban areas. From panel (a) of Figure 4, in 1983, 78 percent of the rural labor force worked in agriculture while only 11 percent of the urban labor force was employed in agricultural jobs.

Our second set of targets are the four wage gaps. These are estimated from the 1983 NSS round and are summarized in Table 10 below. First, we match the within–agriculture and within–non-agriculture wage gaps between urban and rural areas. Those "within" gaps stood at -7 percent and 11 percent, respectively. "Between" gaps which capture the wage premium in non-agricultural jobs relative to agricultural jobs stood at 69 percent in rural areas and 87 percent in urban areas in 1983.

We also target the output share of agriculture in total GDP in India in 1983 at 36 percent. Lastly, we target two moments characterizing consumption expenditures in India. First is the share of agriculture in the total consumption of Indian households. In linking the model to the data above we followed the value-added approach to interpreting a sector.[35] Hence, to keep the model internally consistent we define the arguments in the utility functions in value added terms as well. To obtain such value-added equivalents of consumption we follow the literature and compute agricultural consumption in value added terms as the agricultural value added, while non-agricultural consumption is non-agricultural value added minus investment.[36] This gives us the share of agricultural value added in total consumption equal to 47 percent. Second, to help pin down the subsistence level we also target the *food based* poverty line in India in 1983. We estimate this poverty line to be 67 percent of per capita food consumption expenditures.[37]

Our free parameters are the preference parameters β_A and β_S along with with elasticity parameters ϕ_A and ϕ_S, the training costs τ_U and τ_R, the relative productivity level A/S, the agricultural consumption share θ and the minimum agricultural consumption parameter $\bar{c}$. These parameters are calibrated to jointly match the nine data moments described above.

Table 10: Data and model: 1983

	1983	
	data	model
employment shares:		
L_U	0.220	0.220
L_{RA}	0.780	0.853
L_{RS}	0.220	0.147
L_{UA}	0.110	0.105
L_{US}	0.890	0.895
wage gaps:		
within A	0.932	0.990
within S	1.079	1.027
R between	1.674	1.573
U between	1.815	1.631
overall mean	1.509	1.429
overall median	1.586	1.429
aggregates:		
S/A relative price	1.000	1.000
A share of Y	0.364	0.580
A share of C	0.474	0.501

In the empirical section above we showed the existence of significant differences in human capital between urban and rural areas. To make the model consistent with this, and to control for the initial

[35] See Herrendorf, Rogerson, and Valentinyi (2013b) for a careful discussion of value added and final expenditure approaches to interpreting the data.

[36] See appendix A.6 for data sources.

[37] This number is not too far from the 50 percent of food consumption used for subsistence assumed in Anand and Prasad (2010) based on a sample of 6 emerging economies, including India. Details on our computations are presented in Appendix A.6.

wage gap in 1983 accounted for by the differences in human capital, we adjust the labor input of each type in each sector by its respective human capital in 1983. We use years of education to proxy for human capital. Thus, in rural areas in 1983, labor employed in agriculture had 1.71 years of education, while those working in non-agriculture had 3.96 years of education. In the urban areas in 1983, the corresponding numbers were 2.63 in agriculture and 6.2 in non-agriculture. We will keep these values unchanged in our experiment below. The data targets and their values predicted by the model in 1983 are presented in Table 10.[38] All resulting parameter values are summarized in Table 11.

Table 11: Model parameters, 1983

	parameter	value
Share of U labor in total labor force	L_U	0.22
Urban labor weight in A sector	β_A	0.09
Urban labor weight in S sector	β_S	0.62
Training cost for U households	τ_U	12
Training cost for R households	τ_R	11
Elasticity of substitution between R and U labor in A	$1/(1-\phi_A)$	1.56
Elasticity of substitution between R and U labor in S	$1/(1-\phi_S)$	1.09
A consumption share	θ	0.3
Minimum consumption/Agri consumption	$\bar{c}/c_A$	0.57

4.3.2 Results

How much of the observed convergence in urban-rural wages is accounted for by changes in sectoral productivity and differential labor supply growth in the two sectors? To answer this question we re-calibrate the urban and rural labor force shares and TFP to their values in 2010.

Since we abstract from the rural-urban migration in the model, it is important to exclude these migration flows when computing the urban-rural labor force shares in 2010. For this purpose we compute *cumulative* net migration flows between rural and urban areas during 1983-2010 period from our NSS estimates and correct the Census population numbers in 2011 for these flows. From Table 8 net rural-to-urban migration flows in the five years preceding 1999-00 NSS survey round amounted to about 3 million people, while in five years preceding 2007-08 NSS survey round, they amounted to about 6 million people.[39] These numbers imply a cumulative net flow of about 21 million people from rural to urban areas between 1983 and 2010. We reverse this cumulative net flow by subtracting it from the urban population and adding it to the rural population in 2011. According to the 2011 Census, the urban population in India was 377.1 million while the rural population was 833.1 million people. Adjusting the resulting numbers by the share of working age population (equal

[38] The model overshoots the share of agriculture in India's GDP in 1983 since there is no investment, government spending or trade present in the model.

[39] Since migration data is not available in the 2009-10 NSS round we use the migration flow numbers for 2007-08 survey round.

to 0.62 in rural areas and to 0.68 in urban areas) and labor force participation rates (equal to 0.66 in rural areas and 0.59 in urban areas), gives us a *migration-adjusted* urban labor force share of 29 percent and rural labor force share of 71 percent in 2010.

To calibrate productivity shocks we focus on TFP, the dynamics of which during 1983-2010 period are presented in panel (b) of Figure 7. Specifically, agricultural TFP increased by 24 percent between 1983 and 2010, while non-agricultural TFP increased by 119 percent.

Table 12: Model and data: 1983 versus 2010

	1983		2010		2010-1983 change		
	data	model	data	model	data	model	explained share
employment shares:							
L_U	0.220	0.220	0.290	0.290	0.070	0.070	
L_{RA}	0.780	0.853	0.660	0.779	-0.120	-0.074	0.62
L_{RS}	0.220	0.147	0.340	0.221	0.120	0.074	0.62
L_{UA}	0.110	0.105	0.070	0.067	-0.040	-0.039	0.97
L_{US}	0.890	0.895	0.930	0.933	0.040	0.039	0.97
wage gaps:							
within A	0.932	0.990	1.000	0.976	0.068	-0.014	-0.21
within S	1.079	1.027	1.000	1.005	-0.079	-0.022	0.28
R between	1.674	1.573	1.518	1.337	-0.156	-0.236	1.51
U between	1.815	1.631	1.536	1.377	-0.279	-0.254	0.91
overall mean	1.509	1.429	1.270	1.228	-0.239	-0.202	0.84
overall median	1.586	1.429	1.126	1.228	-0.460	-0.202	0.44
aggregates:							
S/A relative price	1.000	1.000	0.752	0.782	-0.248	-0.218	0.88
A share of Y	0.364	0.580	0.160	0.497	-0.560	-0.144	0.26
A share of C	0.474	0.501	0.230	0.419	-0.515	-0.164	0.32

We feed the changes in labor force shares and sectoral productivity growth into the model while keeping all other parameters unchanged. The results are summarized in Table 12. First, the share of the workforce employed in agriculture declines by 7.4 percentage points for rural and 3.9 percentage points for urban workers. In the data, the agricultural share of rural jobs declined by 12 percentage points while the urban share of agricultural jobs fell by 4 percentage points between 1983 and 2010. Hence, shocks to labor force growth and sectoral productivity account for 62 percent of the observed decline in agricultural employment in rural areas and 97 percent of the observed decline in urban employment in agricultural jobs during the 1983-2010 period.

Second, wage gaps between urban and rural labor decline following the shocks. The "within" gaps between urban and rural wages fall by 0.014 in agriculture and by 0.022 in non-agriculture. The "between" non-agriculture and agriculture wage gaps also decline by 0.254 in urban areas and 0.236 in rural areas. The overall wage gap between urban and rural areas falls by 0.202 in response to the differential population growth in rural and urban areas and sectoral productivity growth. Given that the median wage gap in the data declined by 46 percentage points, these two shocks then account for about 44 percent of the observed decline in median urban-rural wage gaps. Since 60 percent of the median wage convergence was unexplained by standard covariate of wages, the two shocks account for 73 percent $\left(= \frac{0.202}{(0.6)(0.46)}\right)$ of the unexplained decline in the median wage gap.

Third, the model predicts a 22 percent decline in the relative price of non-agricultural goods, which is close to the 25 percent decline in the data. We conclude that the two aggregate shocks we emphasized account for about 4/5 of the observed fall in the price of non-agricultural goods in India during 1983-2010 period. The model also predicts a fall in the share of agriculture in output and consumption, but the declines are smaller than in the data.

We report the effects of each shock individually in Appendix A.7 and show that shocks to both urban-rural labor supply and to productivity are necessary to explain the data patterns numerically.

To assess the contribution of the declines in the "within" and "between" wage gaps to the rural-urban labor income convergence we perform the same decomposition as in equation (2.2) but using the wage and employment numbers predicted by the model for 1983 and 2010. As before, we only consider changes in sectoral productivity and differential population growth in rural and urban areas as the shocks. The results are presented in Table 13.

Table 13: Decomposition of labor income gap in the model

	wage component		labor reallocation	total
	within	between	component	
non-agri	-0.025	-0.158	-0.013	-0.197
agrarian	-0.005			-0.005
total	-0.030	-0.158	-0.013	-0.202
% contribution	14.7	78.6	6.6	100.0

Note: This table presents the decomposition of the change in urban-rural labor income gap between 1983 and 2010 predicted by the model. The decomposition is based on equation (2.2).

In line with the data decomposition in Table 2, convergence in wages is responsible for the majority of the labor income convergence in our experiment. Both the "between" and "within" wage components contribute substantially to the convergence in wages, with the "between" component contributing more. As in the data, the role played by labor reallocation component is very small.

Overall, our results suggest that aggregate factors have played an important role in urban-rural convergence in the past 27 years. Growth of agricultural productivity and an even faster growth of non-agricultural productivity and a relatively faster expansion of the urban labor force can jointly account for over 70 percent of the wage convergence left unexplained by standard covariates of wages, 4/5 of the observed decline in the relative price of non-agricultural goods, and a substantial part of sectoral employment convergence between urban and rural areas in India during this period. Furthermore, these factors induce both within- and between-wage convergence, in line with data.

5 Conclusion

The process of development tends to generate large scale structural transformations as economies shift from being primarily agrarian and rural towards becoming increasingly non-agricultural and urban. This transformation implies a reallocation and, possibly, re-training of the workforce. The capacity of markets and institutions in developing economies to cope with the demands of this restructuring is thus key to determining how smooth or disruptive this process is. Clearly, the greater the disruption, the more the likelihood of income redistributions through unemployment and wage losses due to incompatible skills.

We have examined this issue through the lens of the experience of India over the past three decades. India is particularly appropriate for two reasons. First, it has been undergoing precisely such a macroeconomic structural transformation during this period. Second, with over 800 million people still residing in rural India, the scale of the potential disruption due to the ongoing contraction of the agricultural sector is massive. We have found that the period 1983-2010 has been marked by a sharp and significant convergent trend in the labor income of the rural workforce towards the levels of their urban counterparts in India. A majority of this convergence is due to a decline in the wage gap between urban and rural areas. Thus, the median urban wage premium has declined from 59 percent in 1983 to 13 percent by 2010; similarly the mean wage gap has fallen from 51 percent to 27 percent. We find this rate of wage convergence to be very large and somewhat unexpected.

We evaluated two explanations for this wage convergence. First, we decomposed the urban-rural wage gap along the entire distribution into two components: differences in individual/household characteristics, and differences in returns to those characteristics. Surprisingly, we found that over 60 percent of the decline in the urban-rural wage gap was not due to convergence in individual characteristics such as demographics or education attainments, but rather is unexplained. Second, we examined the role of migration for the urban-rural wages gaps dynamics. While rural to urban migration has been happening, the overall flows have remained stable and small relative to the overall workforce. Rural migrants earn less than their urban counterpart, but the differences are not significant. However, the small size of the flows and the lack of a structural analysis of the issue in this paper suggests caution in drawing broader conclusions.

Given the lack of explanatory power of conventional worker characteristics, we then examined the possible role of aggregate shocks to the Indian economy during this period. Using a two-factor, two-sector model of structural transformation we showed both analytically and quantitatively that differential growth in urban and rural labor supply along with differential productivity shocks to agriculture and non-agriculture can potentially explain a large part of the observed convergence. In particular, our quantitative results suggest that around 70 percent of the unexplained wage convergence between rural and urban areas can be jointly accounted for by these two factors.

Our results highlight the key role of faster urban labor force growth in accounting for the wage convergence between urban and rural areas. Given that a large part of the faster growth in the urban labor force during this period was driven by urban agglomeration, we interpret our findings as indicating the importance of forming a better understanding of the process of urban sprawl in developing countries. We should point out that the importance of urban agglomeration in accounting for urban labor force growth is not unique to India. Lucas (1998) and UN-HABITAT (2012) document precisely this trend in a number of developing countries ranging from Latin America to Asia. Hence, we believe that the relevance of our results extends beyond just the Indian experience since 1983.

The empirical analysis in the paper also uncovered interesting distributional developments in India during this period. In particular, we found that the urban poor have become poorer relative to the rural poor while the urban rich did disproportionately better than the rural rich. While we have abstracted from both these issues in this paper, we intend to address them in future work.

References

Abler, D., G. Tolley, and G. Kripalani (1994): *Technical change and income distribution in Indian agriculture.* Westview Press, Inc.

Acemoglu, D., and V. Guerrieri (2008): "Capital Deepening and Nonbalanced Economic Growth," *Journal of Political Economy*, 116(3), 467–498.

Anand, R., and E. Prasad (2010): "Optimal Price Indices for Targeting Inflation under Incomplete Markets," IZA Discussion Papers 5137, Institute for the Study of Labor (IZA).

Baumol, W. J. (1967): "Macroeconomics of unbalanced growth: the anatomy of urban crisis," *The American Economic Review*, 57(3), 415–426.

Caselli, F., and J. Coleman (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109(3), 584–616.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), 1001–44.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

Fortin, N. M. (2006): "Greed, Altruism, and the Gender Wage Gap," Working papers, University of British Columbia.

Gollin, D., D. Lagakos, and M. Waugh (2012): "The Agricultural Productivity Gap in Developing Countries," Working paper, Arizona State University.

Gollin, D., S. Parente, and R. Rogerson (2002): "The Role of Agriculture in Development," *American Economic Review*, 92(2), 160–164.

Hall, R. E., and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *The Quarterly Journal of Economics*, 114(1), 83–116.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013a): "Growth and Structural Transformation," *Handbook of Economic Growth, forthcoming.*

——— (2013b): "Two Perspectives on Preferences and Structural Transformation," *American Economic Review, forthcoming.*

IIHS (2011): "Urban India 2011: Evidence," Report, Indian Institute of Human Settlement.

Kongsamut, P., S. Rebelo, and D. Xie (2001): "Beyond Balanced Growth," *Review of Economic Studies*, 68(4), 869–882.

Kuznets, S. (1966): *Modern Economic Growth.* Yale University Press.

Lagakos, D., and M. Waugh (2012): "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economics Review, forthcoming.*

Laitner, J. P. (2000): "Structural Change and Economic Growth," *Review of Economic Studies*, 67(3), 545–561.

Lucas, R. E. (1998): "Internal Migration and Urbanization: Recent Contributions and New Evidence," Ied discussion paper series number 91, Institute for Economic Development.

Michaels, G., F. Rauch, and S. J. Redding (2012): "Urbanization and Structural Transformation," *The Quarterly Journal of Economics*, 127(2), 535–586.

Ngai, L. R., and C. A. Pissarides (2007): "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97(1), 429–443.

Nguyen, B. T., J. W. Albrecht, S. B. Vroman, and M. D. Westbrook (2007): "A quantile regression decomposition of urban-rural inequality in Vietnam," *Journal of Development Economics*, 83(2), 466–490.

Qu, Z. F., and Z. Zhao (2008): "Urban-Rural Consumption Inequality in China from 1988 to 2002: Evidence from Quantile Regression Decomposition," IZA Discussion Papers 3659, Institute for the Study of Labor (IZA).

Rao, J. N. K., and C. F. J. Wu (1988): "Resampling inference with complex survey data," *Journal of the American Statistical Association*, 83, 231–241.

Rao, J. N. K., C. F. J. Wu, and K. Yue (1992): "Some recent work on resampling methods for complex surveys," *Survey Methodology*, 18, 209Ű217.

Smith, J. P., and F. R. Welch (1989): "Black Economic Progress after Myrdal," *Journal of Economic Literature*, 27(2), 519–64.

UN-HABITAT (2012): "State of the World's Cities 2010/2011 - Cities for All: Bridging the Urban Divide," Report, The United Nations Human Settlements Programme.

Wu, X., and J. M. Perloff (2005): "China's Income Distribution, 1985-2001," *The Review of Economics and Statistics*, 87(4), 763–775.

Young, A. (2012): "Inequality, the Urban-Rural Gap and Migration," Working paper, London School of Economics.

A Appendix: Not for publication

A.1 Data

The National Sample Survey Organization (NSSO), set up by the Government of India, conducts rounds of sample surveys to collect socioeconomic data. Each round is earmarked for particular subject coverage. We use the latest six large quinquennial rounds – 38(Jan-Dec 1983), 43(July 1987-June 1988), 50(July 1993-June 1994), 55(July 1999-June 2000), 61(July 2004-June 2005) and 66(July 2009-June 2010) on Employment and Unemployment (Schedule 10). Rounds 38 and 55 also contain migration particulars of individuals. We complement those rounds with a smaller 64th round(July 2007-June 2008) of the survey since migration information is not available in all other quinquennial survey rounds.

The survey covers the whole country except for a few remote and inaccessible pockets. The NSS follows multi-stage stratified sampling with villages or urban blocks as first stage units (FSU) and households as ultimate stage units. The field work in each round is conducted in several sub-rounds throughout the year so that seasonality is minimized. The sampling frame for the first stage unit is the list of villages (rural sector) or the NSS Urban Frame Survey blocks (urban sector) from the latest available census. The NSSO supplies household level multipliers with the unit record data for each round to help minimize estimation errors on the part of researchers. The coding of the data changes from round to round. We re-coded all changes to make variables uniform and consistent over the time.

In our data work, we only consider individuals that report their 3-digit occupation code and education attainment level. Occupation codes are drawn from the National Classification of Occupation (NCO) – 1968. We use the "usual" occupation code reported by an individual for the usual principal activity over the previous year (relative to the survey year). The dataset does not contain information on the years of schooling for the individuals. Instead it includes information on general education categories given as (i) not literate -01, literate without formal schooling: EGS/ NFEC/ AEC -02, TLC -03, others -04; (ii) literate: below primary -05, primary -06, middle -07, secondary -08, higher secondary -10, diploma/certificate course -11, graduate -12, postgraduate and above -13. We aggregate those into five similarly sized groups as discussed in the main text. We also convert these categories into years of education. The mapping we used is discussed in the main text.

The NSS only reports wages from activities undertaken by an individual over the previous week (relative to the survey week). Household members can undertake more than one activity in the reference week. For each activity we know the "weekly" occupation code, number of days spent working in that activity, and wage received from it. We identify the main activity for the individual as the one in which he spent maximum number of days in a week. If there are more than one activities with equal days worked, we consider the one with paid employment (wage is not zero or missing). Workers sometimes change the occupation due to seasonality or for other reasons. To minimize the effect of transitory occupations, we only consider wages for which the weekly occupation code coincides with usual occupation (one year reference). We calculate the daily wage by dividing total wage paid in that activity over the past week by days spent in that activity.

Lastly, we identify full time workers in our dataset. We assume that an individual is a full time worker if he is employed (based on daily status code) for at least two and half days combined in all activities during the reference week. We drop observations if total number of days worked in the reference week is more than seven.

A.2 Decomposition of labor income convergence

Equation (2.1) gives us average per capita labor income in urban (U) and rural (R) areas as

$$w_t^R = w_{1t}^R L_{1t}^R + w_{2t}^R L_{2t}^R \qquad \text{and} \qquad w_t^U = w_{1t}^U L_{1t}^U + w_{2t}^U L_{2t}^U,$$

where 1 and 2 refer to non-agricultural and agricultural jobs, respectively.

The relative labor income gap in period t is

$$\frac{w_t^U - w_t^R}{w_t^R} = \frac{\left(w_{1t}^U L_{1t}^U + w_{2t}^U L_{2t}^U\right) - \left(w_{1t}^R L_{1t}^R + w_{2t}^R L_{2t}^R\right)}{w_t^R}.$$

Adding and subtracting average labor income for each occupation (denoted by $w_{it}, i = 1, 2$), we can write the expression above as

$$\begin{aligned}\frac{w_t^U - w_t^R}{w_t^R} &= \frac{\left(w_{1t}^U - w_{1t}\right) L_{1t}^U + \left(w_{2t}^U - w_{2t}\right) L_{2t}^U}{w_t^R} - \frac{\left(w_{1t}^R - w_{1t}\right) L_{1t}^R + \left(w_{2t}^R - w_{2t}\right) L_{2t}^R}{w_t^R} \\ &\quad + \frac{w_{1t}\left(L_{1t}^U - L_{1t}^R\right) + w_{2t}\left(L_{2t}^U - L_{2t}^R\right)}{w_t^R}.\end{aligned}$$

Now we look at the change in the relative gap between periods t and $t-1$. To simplify the notation, let $\mu_{it}^j \equiv \left(w_{it}^j - w_{it}\right)/w_t^R$, with $i = 1, 2$; and $j = U, R$ and $\eta_{it} \equiv w_{it}/w_t^R$, $i = 1, 2$. Then the change in the relative gap can be written as

$$\begin{aligned}&\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} \\ =\; &\mu_{1t}^U L_{1t}^U + \mu_{2t}^U L_{2t}^U - \left(\mu_{1t}^R L_{1t}^R + \mu_{2t}^R L_{2t}^R\right) \\ &+\eta_{1t}\left(L_{1t}^U - L_{1t}^R\right) + \eta_{2t}\left(L_{2t}^U - L_{2t}^R\right) \\ &-\left(\mu_{1t-1}^U L_{1t-1}^U + \mu_{2t-1}^U L_{2t-1}^U\right) - \left(\mu_{1t-1}^R L_{1t-1}^R + \mu_{2t-1}^R L_{2t-1}^R\right) \\ &-\eta_{1t-1}\left(L_{1t-1}^U - L_{1t-1}^R\right) - \eta_{2t-1}\left(L_{2t-1}^U - L_{2t-1}^R\right).\end{aligned}$$

Define $\bar{x}_t = (x_t + x_{t-1})/2$, and $\Delta x_t = x_t - x_{t-1}$. Now, adding and subtracting $\left(\mu_{it}^j - \mu_{it-1}^j\right)\bar{L}_{it}^j$, where $\bar{L}_{it}^j = \left(L_{it}^j + L_{it-1}^j\right)/2$, and $i = 1, 2$ and $j = U, R$ and collecting the terms in the first and third lines above; adding and subtracting $\bar{\eta}_{it}\left[\left(L_{it}^U - L_{it-1}^U\right) - \left(L_{it}^R - L_{it-1}^R\right)\right]$, where $\bar{\eta}_{it} = \left(\eta_{it} + \eta_{it-1}\right)/2$ and $i = 1, 2$ and collecting the terms in the second and fourth lines above, we get

$$\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \Delta\mu_{1t}^U\bar{L}_{1t}^U + \Delta\mu_{2t}^U\bar{L}_{2t}^U - \Delta\mu_{1t}^R\bar{L}_{1t}^R - \Delta\mu_{2t}^R\bar{L}_{2t}^R$$
$$+\Delta L_{1t}^U\bar{\mu}_{1t}^U + \Delta L_{2t}^U\bar{\mu}_{2t}^U - \Delta L_{1t}^R\bar{\mu}_{1t}^R - \Delta L_{2t}^R\bar{\mu}_{2t}^R$$
$$+\bar{\eta}_{1t}\Delta\left(L_{1t}^U - L_{1t}^R\right) + \bar{\eta}_{2t}\Delta\left(L_{2t}^U - L_{2t}^R\right)$$
$$+\left(\overline{L_{1t}^U - L_{1t}^R}\right)\Delta\eta_{1t} + \left(\overline{L_{2t}^U - L_{2t}^R}\right)\Delta\eta_{2t}$$

Using the fact that $L_{2t}^j = 1 - L_{1t}^j$ we can rewrite the second row as

$$\Delta L_{1t}^U\left(\bar{\mu}_{1t}^U - \bar{\mu}_{2t}^U\right) - \Delta L_{1t}^R\left(\bar{\mu}_{1t}^R - \bar{\mu}_{2t}^R\right),$$

and the third row as

$$(\overline{\eta_{1t} - \eta_{2t}})\Delta\left(L_{1t}^U - L_{1t}^R\right),$$

and the fourth row as

$$\left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{2t}\right].$$

Thus, the change in the relative labor income gap becomes

$$\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \Delta\mu_{1t}^U\bar{L}_{1t}^U + \Delta\mu_{2t}^U\bar{L}_{2t}^U - \Delta\mu_{1t}^R\bar{L}_{1t}^R - \Delta\mu_{2t}^R\bar{L}_{2t}^R \tag{A1}$$
$$+\Delta L_{1t}^U\left(\bar{\mu}_{1t}^U - \bar{\mu}_{2t}^U\right) - \Delta L_{1t}^R\left(\bar{\mu}_{1t}^R - \bar{\mu}_{2t}^R\right) \tag{A2}$$
$$+(\overline{\eta_{1t} - \eta_{2t}})\Delta\left(L_{1t}^U - L_{1t}^R\right) \tag{A3}$$
$$+\left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{2t}\right] \tag{A4}$$

Row (A1) gives the within-occupation component of labor income convergence, rows (A2) and (A3) give the labor reallocation component of labor income convergence, while row (A4) gives the between-occupation component of labor income convergence.

A.3 Decomposition of the sectoral gaps in wages and consumption

We are interested in performing a time-series decomposition of rural-urban wage and consumption expenditure gaps between 1983 and 2004-05. We employ a two-fold Oaxaca-Blinder procedure where we use coefficients from a pooled regression with a group membership indicator (as in Fortin, 2006) as the reference coefficients. We use 1983 as the base year for the inter-temporal decomposition, so 1983 is the benchmark sample in our analysis.

Our econometric model for sector s and round t is given by

$$y_{st} = X'_{st}\beta_{ct} + e_{st}, \qquad s = 1, 2; \text{ and } t = 1, 2,$$

where y_{st} is a vector of outcomes (log wage) while X_{st} is the matrix of regressors for sector s in round t. Here β_{st} is a coefficient vector, and e_{st} is the vector of residuals. The differential in expected outcomes between urban and rural sectors in round t is then given by:

$$\Delta y_t^e = \Delta X_t' \tilde{\beta}_t + X'_{1t}(\beta_{1t} - \tilde{\beta}_t) + X'_{2t}(\tilde{\beta}_t - \beta_{2t}),$$

where $\tilde{\beta}_t$ is the vector of coefficients from the model with both groups pooled. The first term above is the explained part while the last two terms give the unexplained parts of the decomposition. Denote E_t to be the explained component of the decomposition, and U_t to be the unexplained part, then

$$\begin{aligned} E_t &= \Delta X_t' \tilde{\beta}_t, \qquad t = 1, 2, \\ U_t &= X'_{1t}(\beta_{1t} - \tilde{\beta}_t) + X'_{2t}(\tilde{\beta}_t - \beta_{2t}), \qquad t = 1, 2. \end{aligned}$$

The inter-temporal change in the outcome differentials can be written as the sum of changes in the explained, E and unexplained, U components:

$$\Delta y_2^e - \Delta y_1^e = (E_2 - E_1) + (U_2 - U_1) = \Delta E + \Delta U$$

These differentials are reported in Table 7.

A.4 Distributional effects of migration

Table A1 complements the results in Table 9 in the main text by presenting regression results from the RIF regressions for the 10th and 90th percentile of (log) wages. The regression specification is the same as in Section 3.2.

A.5 Measuring Total Factor Productivity

Following Hall and Jones (1999) we assume that output in each sector (agriculture, A, and non-agriculture, NA) is produced with Cobb-Douglas production function and that technological change is labor-augmenting:

$$Y_i = K_i^{\alpha_i}(Z_i H_i)^{\beta_i}, \qquad i = A, NA$$

where K_i denotes the stock of physical capital, H_i is the amount of human capital-augmented labor used in production, and Z_i is a labor-augmenting measure of productivity. We assume that each unit of homogeneous labor L_i has received E_i years of education. Therefore, $H_i = E_i L_i$. α_i is capital

Table A1: Wage gaps: Accounting for migration

	10th percentile			90th percentile		
	1983	1999-00	2007-08	1983	1999-00	2007-08
rural	-0.192***	0.006	0.122***	-0.511***	-0.679***	-0.900***
	(0.011)	(0.009)	(0.013)	(0.015)	(0.025)	(0.031)
rural-to-urban	0.086***	0.116***	0.180***	-0.147***	-0.220***	-0.453***
	(0.022)	(0.020)	(0.031)	(0.048)	(0.055)	(0.068)
urban-to-urban	0.149***	0.134***	0.237***	0.599***	1.242***	1.278***
	(0.016)	(0.019)	(0.028)	(0.057)	(0.112)	(0.132)
rural-to-rural	-0.175***	-0.046*	0.040	-0.155***	-0.080	-0.320***
	(0.031)	(0.026)	(0.041)	(0.033)	(0.058)	(0.072)
urban-to-rural	-0.029	0.141***	0.241***	0.875***	0.542***	0.601***
	(0.049)	(0.031)	(0.047)	(0.110)	(0.179)	(0.203)
N	63981	67322	69862	63981	67322	69862

Note: This table reports the estimates of coefficients on the rural dummy and dummies for rural-urban migration flows from the RIF regressions of log wages on a set of aforementioned dummies, age, age squared, and a constant for the 10th and 90th percentiles. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

income share, while β_i is labor income share in sector $i = A, NA$.

Sectoral real GDP is obtained from GDP by economic activity data from Statement 10 of National Accounts Statistics provided by the Ministry of Statistics and Programme Implementation (MOSPI) of Government of India. GDP is measured at factor cost. Real capital stock is obtained as net capital stock (equal to the sum of net fixed capital stock and inventories) by industry of use provided in Statement 22 of National Accounts Statistics by MOSPI. Both GDP and capital are measured in constant 1999-00 prices. Employment in each sector is computed from the NSS data using the employment shares in each sector and total labor force in India's economy in each survey year.

Based on the estimates by Abler, Tolley, and Kripalani (1994) we set capital and labor share in agriculture to be $\alpha_A = 0.25, \beta_A = 0.45$. The rest is returns to a fixed factor such as land. Note that under the assumption that the other input in agriculture is a fixed factor, our estimate of the change in the agricultural productivity over time is unaffected by the presence of this fixed factor. For the capital and labor shares in non-agriculture we used $\alpha_{NA} = 0.3$ and $\beta_{NA} = 0.7$, correspondingly.

A.6 Consumption moments: Data and calculations

A.6.1 Consumption value added

We used sectoral value added from GDP by economic activity data from Statement 10 of National Accounts Statistics provided by the Ministry of Statistics and Programme Implementation (MOSPI) of Government of India. Investment is measured as gross capital formation, and was obtained from Statement 20 of National Accounts Statistics provided by MOSPI. Both value added and investment is in constant 1999-00 prices and can be accessed from
http://mospi.nic.in/Mospi_New/site/India_Statistics.aspx?status=1&menu_id=43.

A.6.2 Measuring food-based poverty line

In the model parameter $\bar{c}$ denotes minimum consumption needs of the agricultural good. To obtain a target for this parameter we used the *food-based* poverty line in India in 1983. The data for poverty lines was obtained from "Report of The Expert Group on Estimation of Proportion and Number of Poor" prepared by Perspective Planning Division of Planning Commission, Government of India in July of 1993. It reports poverty lines in rural and urban areas of India in 1983 using 1980-81 prices. These poverty lines are Rs. 1073.4 per capita per year in rural areas and Rs. 1411.7 per capita per year in urban areas. The report also contains the distribution of household consumption expenditures by food and non-food items. For rural households, the food share is 79 percent for those below the poverty line and 66 percent for those above. Using the proportion of households below the poverty line we get the average share of food expenditures as 71 percent in the rural areas. The corresponding numbers for urban households are 74 percent, 57 percent and 62 percent. To compute the food-based poverty lines we use the average food share of those above the poverty line in order to estimate the unconstrained spending on food. This gives us food-based poverty lines at Rs. 765.8 per person per year in rural and Rs. 877.6 per person per year in urban areas.

The last step is to get household consumption expenditure in rural and urban areas separately. According to the National Income Statistics of India, the per capita consumption expenditure in India in 1983 was Rs. 1591.4. The Planning Commission reports the rural and urban per capita consumption expenditure in 1983 at 1993-94 prices. Deflating these numbers using the Agricultural Labor price index for rural workers and the Industrial Workers index for urban workers to derive the corresponding levels at 1980-81 prices gives rural per capita consumption to be 73 percent of urban per capita consumption. These historical price deflators are available from the Reserve Bank of India. Using the 73 percent ratio to decompose the national per capita average of Rs. 1591.4 into its rural and urban components gives rural per capita consumption to be Rs. 1471.6 and urban consumption of Rs. 2015.9. To compute the food share of consumption in rural and urban households we use the average share of food in all rural and all urban households, which is 71 percent and 62 percent, respectively. Using these ratios along with the respective per capita consumptions gives rural and urban agricultural consumption. The ratio of the food-based poverty line to agricultural consumption is then readily computed as 0.68 for rural and 0.65 for urban workers, with the average being 67 percent. These computations are summarized in Table A2.

A.7 The effects of individual shocks

Table A3 presents changes in various variables triggered by both urban-to-rural labor supply shock and TFP shocks (column "full model"), as well as by each shock individually: column "LU/LR" is for relative labor supply shock alone and column "S/A" is for TFP shocks alone. Column "data" reports the changes in those variables in the data during 1983-2010 period. The table makes clear

Table A2: Computations of food-based poverty line, 1983

		share of C	share of Agri C
Poverty line, rural	Rs. 1073.4		
Poverty line, urban	Rs. 1411.7		
Food share of poverty line, rural	0.66		
Food share of poverty line, urban	0.57		
Food-based poverty line, rural	Rs. 765.8	0.48	0.68
Food-based poverty line, urban	Rs. 877.6	0.40	0.65
Food-based poverty line, average			0.67

that each shock alone is not sufficient to reproduce the patterns in the data. For instance, relative TFP shock leads to the correct pattern of structural transformation, with employment in agriculture declining in both urban and rural areas, but also leads to an increase in urban-rural wage gap and non-agriculture relative prices, both of which are counterfactual. In contrast, an increase in urban labor relative to rural labor leads to the correct dynamics of relative wages and prices, but misfires on the labor force movements across sectors. A combination of the two shocks is thus necessary to account for all data facts, as we argued in the main text.

Table A3: Contribution of individual shocks, 1983-2010

	2010-1983 change			
	data	full model	LU/LR	S/A
employment shares:				
L_U	0.070	0.070	0.070	0.000
L_{RA}	-0.120	-0.074	0.017	-0.094
L_{RS}	0.120	0.074	-0.017	0.094
L_{UA}	-0.040	-0.039	0.101	-0.064
L_{US}	0.040	0.039	-0.101	0.064
wage gaps:				
within A	0.068	-0.014	-0.482	0.679
within S	-0.079	-0.022	-0.310	0.495
R between	-0.156	-0.236	-0.015	-0.231
U between	-0.279	-0.254	0.567	-0.407
overall median	-0.460	-0.202	-0.505	0.444
aggregates:				
S/A relative price	-0.248	-0.218	-0.192	0.006
A share of Y	-0.560	-0.144	0.060	-0.189
A share of C	-0.515	-0.164	0.054	-0.193