Structural Transformation and the Rural-Urban Divide*

Viktoria Hnatkovska† and Amartya Lahiri‡

November 2012

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 this using the Indian experience between 1983 and 2010, a period when India has been undergoing such a transformation. We examine the gaps between rural and urban India in terms of the education attainment, occupation choices, consumption and wages during this period. We find a significant narrowing of the differences in education, occupation distribution, and wages between individuals in rural India and their urban counterparts. However, individual characteristics do not appear to account for much of this convergence. Migration did not play an important role either. We use a simple two sector model of structural transformation to rationalize the rural-urban convergence in India as the consequence of (a) higher labor supply growth in urban areas; and (b) aggregate productivity growth.

JEL Classification: J6, R2

Keywords: Rural urban disparity, education gaps, wage gaps

---

*We would like to thank IGC for a grant funding this research, seminar participants at UBC and the IGC-India 2012 conference in Delhi.

†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

A topic of long running interest to social scientists has been the processes that surround the transformation of economies along the development path. As is well documented, the process of development tends to generate large scale structural transformations of economies as they shift from being primarily agrarian towards more industrial and service oriented activities. A related aspect of this transformation is how the workforce in such economies adjusts to the changing macroeconomic structure in terms of their labor market choices such as investments in skills, choices of occupations, location and industry of employment. Indeed, some of the more widely cited contributions to development economics have tended to focus precisely on these aspects. The well known Harris-Todaro model of Harris and Todaro (1970) was focused on the process through which rural labor migrates to urban areas in response to wage differentials while the equally venerated Lewis model, formalized in Lewis (1954), addressed the issue of shifting incentives for employment between rural agriculture and urban industry.

A parallel literature has addressed the issue of the redistributionary effects associated with these structural transformations, both in terms of theory and data. The main focus of this research is on the relationship between development and inequality.[1] This work is related to the issue of rural-urban dynamics since the process of structural transformation implies contracting and expanding sectors which, in turn, implies a reallocation and, possibly, re-training of the workforce. The capacity of institutions in such transforming economies to cope with these demands is a fundamental factor that determines 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.

India over the past three decades has been on exactly such a path of structural transformation. Prodded by a sequence of reforms starting in the mid 1980s, the country is now averaging annual growth rates routinely is excess of 8 percent. This is in sharp contrast to the first 40 years since 1947 (when India became an independent country) during which period the average annual output growth hovered around the 3 percent mark, a rate that barely kept pace with population growth. This latter phase has also been marked by a significant transformation in the output composition of the country with the agricultural sector gradually contracting both in terms of its output and

[1] Perhaps the best known example of this line of work is the "Kuznets curve" idea that inequality follows an inverse-U shape with development or income (see Kuznets (1955)). More recent work on this topic explores the relationship between inequality and growth (see, for example, Persson and Tabellini (1994) and Alesina and Rodrik (1994) for illustrative evidence regarding this relationship in the cross-country data).

employment shares.

How has the workforce in rural and urban India responded to these shifting aggregate sectoral patterns? Have these changes been accompanied by widening rural-urban disparities or have the disparities between them been shrinking over time? In this paper we address these issues by studying the evolution of education attainment levels, the occupation choices, the wage and consumption expenditures of rural and urban workers in India between 1983 and 2010. We do this by using data from six rounds of the National Sample Survey (NSS) of households in India from 1983 to 2009-10.

We find, reassuringly, that this period has been marked by significant narrowing of the gaps between rural and urban areas in all of these measures. The shrinking of the rural-urban gaps have been the sharpest in education attainment and wages, but there have also been important convergent trends in occupation choices. There has been a significantly faster expansion of blue-collar jobs (primarily production and service workers) in rural areas, which is surprising given the usual priors that blue and white collar occupations are mostly centered around urban locations. We also find striking wage convergence with the median wage premium of urban workers relative to rural workers having declined from 101 percent in 1983 to just 11 percent in 2009-10. We also find some interesting distributional features of the changes in wages and consumption during this period. Specifically, the rural poor (10th percentile) appeared to have gained relative to the urban poor whereas the rural rich (the 90th percentile) failed to keep pace with the urban rich.

A key feature of our findings is that most of the changes in the wage and consumption gaps between rural and urban areas cannot be explained by standard demographic and individual characteristics such as education and age. Changing occupation choices though appear to have played a significant role in inducing the shrinking gaps. The tepid contribution of education to the rural-urban gaps stands in sharp contrast to their contribution to gaps between backward castes and others, which too declined during this period. Hnatkovska, Lahiri, and Paul (2012) show that the declining caste gaps in wages and consumption were mostly accounted for by education. The rural-urban gaps, in contrast, changed primarily due to changes in the occupation distribution and due to changes in the returns to the covariates of the gaps rather than due to changes in the covariates themselves. It bears repetition that this does not suggest that the covariates did not change. Indeed, a central finding of the paper is the declining education attainment gaps between rural and urban workers during this period.[2]

[2] We also examine the potential effect of an important rural employment program introduced in 2005 called National Rural Employment Guarantee Act (NREGA) on the rural-urban wage and consumption gaps. In order to examine the effect of the program we use a state level analysis. Our results indicate 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

Using data on migration from the NSS surveys, we also relate the convergence trends to migration of workers from rural to urban areas. We find that annual migration flows have declined from 1.2 percent of the workforce in 1983 to 0.9 percent in 2007-08. Around a quarter of these flows was from rural to urban areas. Consequently, while the gross flow of workers from rural to urban areas is significant, it is also small relative to the overall urban workforce. We find these migrant workers do earn lower wages than their urban non-migrant counterparts, but the difference is not statistically significant. Overall our results indicate that migration did not play an important role in inducing convergent dynamics between urban and rural areas. However, since the migration decision is likely to be endogenous to the wage gap, a concrete conclusion regarding this issue requires more structural work than the current study.

If individual covariates do not explain much of the convergence, how does one explain it? We write down a simple model with two factors – urban and rural labor – and two sectors – agriculture and non-agriculture. We introduce the possibility of structural transformation of the economy by introducing a minimum consumption need of the agricultural good. We use this model to show both analytically and quantitatively that the wage and occupational convergence between rural and urban areas along with the structural transformation of the economy arecan be jointly explained by two factors: (a) a faster increase in urban labor force growth; and (b) differential productivity growth between the agriculture and non-agricultural sectors. Our broad conclusion from these results is that the incentives generated by the institutional structure of the country have provided useful signals to the workforce in guiding their choices during this period of massive aggregate changes. As a result, there has been significant churning at the micro levels of the economy.

There is a large body of work on inequality and poverty in India. A sample of this work can be found in Banerjee and Piketty (2001), Bhalla (2003), Deaton and Dreze (2002) and Sen and Himanshu (2005). While some of these studies do examine inequality and poverty in the context of the rural and urban sectors separately (see Deaton and Dreze (2002) in particular), most of this work is centered on either measuring inequality (through Gini coefficients) or poverty, focused either on consumption or income alone, and restricted to a few rounds of the NSS data at best. An overview of this work can be found in Pal and Ghosh (2007). Our study is distinct from this body of work in that we examine multiple indicators of economic achievement over a 27 year period. This gives us both a broader view of developments as well as a time-series perspective on post-reform India.

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 and consumption in 2009-10, relative to trend. We conclude that the effect of this program on the gaps was, at best, very muted.

The rest of the paper is organized as follows: the next section presents the data and some sample statistics. Section 3 presents the main results on changes in the rural-urban gaps as well as the analysis of the rural employment guarantee reform introduced in India in 2005. Section 5 presents the model and the results while the last section contains concluding thoughts.

2 Empirical motivation

How did the structural transformation in the Indian economy affect the urban-rural inequality? We focus on differences in labor income between urban and rural areas to address this question.[3] Our data comes from successive rounds of the National Sample Survey (NSS) of households in India for employment and consumption. The survey rounds that we include 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 focus is on determining the trends in occupations and wages, amongst other things, 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 be 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.[4] These restrictions leave us with, on average, 140,000 to 180,000 individuals per survey round.

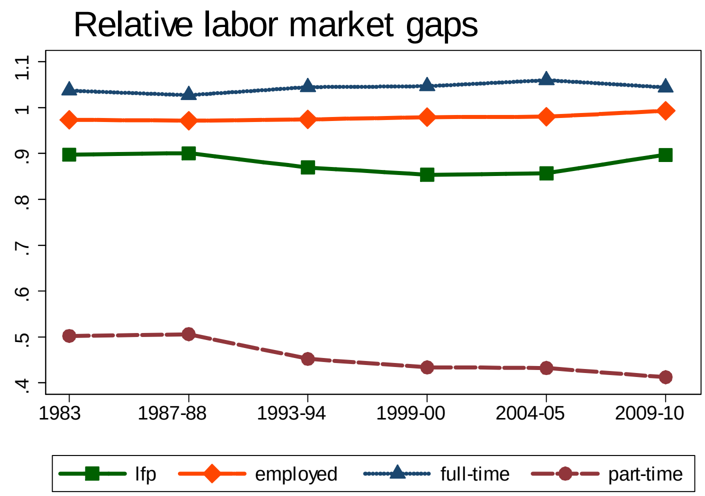
Our focus on full time workers may potentially lead to mistaken inference if there have been significant differential changes in the patterns of part-time work and/or labor force participation patterns in rural and urban areas. To check this, Figure 1 plots the urban to rural ratios in labor force participation 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. Details on our data are provided in Appendix A.1.

To obtain a measure of labor income we need wages and occupation distribution of the labor force. Wages are obtained as the daily wage/salaried income received for the work done by respondents during the previous week (relative to the survey week). Wages can be paid in cash or kind, where the latter are evaluated at the current retail prices. We convert wages into real terms using state-level poverty lines that differ for rural and urban sectors. We express all wages in 1983 rural Maharashtra

[3]We also consider per capita consumption expenditures, and find that our findings are robust. These results are presented in the online appendix available at http://faculty.arts.ubc.ca/vhnatkovska/research.htm.

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

Figure 1: Labor force participation and employment gaps



Note: "lfp" refers to the ratio of labor force participation rate of urban to rural sectors. "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.

poverty lines.[5] When it comes to occupations, there are some fundamental differences in the sectoral compositions of rural and urban areas making it unlikely/impossible for the occupation distributions to converge. However, the country as a whole has been undergoing a structural transformation with an increasing share of output accruing to services with a corresponding decline in the output share of agriculture. This could have led to urban-rural convergence in labor income as well. To assess the role played by labor reallocation across jobs, we aggregate the reported 3-digit occupation categories in the survey into three broad occupation categories: *white-collar* occupations like administrators, executives, managers, professionals, technical and clerical workers; *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 three occupations in each location: white-collar jobs (occ 1), blue collar jobs (occ 2),

[5] In 2004-05 the Planning Commission of India has 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 2009-10 poverty lines obtained from the Planning Commission under the new methodology to the old basket using 2004-05 adjustment factor. That factor was obtained from the poverty lines under the old and new methodologies available for 2004-05 survey year. As a test, we used the same adjustment factor to obtain the implied "old" poverty lines for 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.

and agrarian jobs (occ 3):

$$w_t^j = w_{1t}^j L_{1t}^j + w_{2t}^j L_{2t}^j + w_{3t}^j L_{3t}^j, \tag{2.1}$$

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 $i = 1, 2, 3$ and $j = U, R$. Also $L_{1t}^j + L_{2t}^j + L_{3t}^j = 1$. The labor income gap between urban and rural areas can then be expressed as

$$\begin{aligned}
\frac{w_t^U - w_t^R}{w_t^R} \quad = \quad & \frac{\left(w_{1t}^U - w_{1t}\right) L_{1t}^U + \left(w_{2t}^U - w_{2t}\right) L_{2t}^U + \left(w_{3t}^U - w_{3t}\right) L_{3t}^U}{w_t^R} \\
& - \frac{\left(w_{1t}^R - w_{1t}\right) L_{1t}^R + \left(w_{2t}^R - w_{2t}\right) L_{2t}^R + \left(w_{3t}^R - w_{3t}\right) L_{3t}^R}{w_t^R} \\
& + \frac{\left(w_{1t} - w_{3t}\right) \left(L_{1t}^U - L_{1t}^R\right) + \left(w_{2t} - w_{3t}\right) \left(L_{2t}^U - L_{2t}^R\right)}{w_t^R},
\end{aligned}$$

where w_{it} is the economy-wide average daily wage in occupation $i = 1, 2, 3$. The decomposition above show that the urban-rural labor income gap can arise due to two channels. First, the gap may occur if urban and rural wages and employment within each occupation are different (rows 1 and 2 on the right in the expression above). We refer to this channel as *within-occupation* channel. Second, the gap may arise if there is cross-occupation inequality in wages and employment shares (last row in the expression above). This is the *between-occupation* channel.

Expression above allows us establish the link between structural transformation and convergence in labor income between rural and urban areas through a simple decomposition of the change in labor income gap between period t and $t-1$:

$$\begin{aligned}
& \frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} \\
= \quad & \Delta\mu_{1t}^U \bar{L}_{1t}^U + \Delta\mu_{2t}^U \bar{L}_{2t}^U + \Delta\mu_{3t}^U \bar{L}_{3t}^U - \Delta\mu_{1t}^R \bar{L}_{1t}^R - \Delta\mu_{2t}^R \bar{L}_{2t}^R - \Delta\mu_{3t}^R \bar{L}_{3t}^R \\
& + \left(\overline{L_{1t}^U - L_{1t}^R}\right) \left[\Delta\eta_{1t} - \Delta\eta_{3t}\right] + \left(\overline{L_{2t}^U - L_{2t}^R}\right) \left[\Delta\eta_{2t} - \Delta\eta_{3t}\right] \\
& + \Delta L_{1t}^U \left(\bar{\mu}_{1t}^U - \bar{\mu}_{3t}^U\right) + \Delta L_{2t}^U \left(\bar{\mu}_{2t}^U - \bar{\mu}_{3t}^U\right) - \Delta L_{1t}^R \left(\bar{\mu}_{1t}^R - \bar{\mu}_{3t}^R\right) - \Delta L_{2t}^R \left(\bar{\mu}_{2t}^R - \bar{\mu}_{3t}^R\right) \\
& + (\overline{\eta_{1t} - \eta_{3t}}) \Delta \left(L_{1t}^U - L_{1t}^R\right) + (\overline{\eta_{2t} - \eta_{3t}}) \Delta \left(L_{2t}^U - L_{2t}^R\right)
\end{aligned} \tag{2.2}$$

Appendix A.2 presents detailed derivations of this decomposition. Here $\mu_{it}^j \equiv \left(w_{it}^j - w_{it}\right)/w_t^R, \eta_{it} \equiv w_{it}/w_t^R, \bar{x}_t = (x_t + x_{t-1})/2$, and $\Delta x_t = x_t - x_{t-1}$. This decomposition breaks up the change in labor income gap over time into two basic components: changes in wages and changes in employment. In addition, wage component is further split up into within-occupation component and 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 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 income may arise if wages in different occupations are converging – between-occupation wage convergence component. Lastly, rows three and four 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 1: Decomposition of labor income gap, 1983-2010

	wage component		labor reallocation component	total
	within	between		
white-collar	-0.003	-0.056	0.148	0.089
blue-collar	-0.136	-0.120	-0.068	-0.324
agrarian	0.010			0.010
total	-0.130	-0.177	0.080	-0.226
% contribution	0.574	0.782	-0.356	1.000

Table 1 presents the results of the decomposition by occupations and components. During 1983-2010 period, the aggregate labor income gap between urban and rural areas have declined by 22.6 percent. All of this decline is due to convergence of wages during this time, with roughly equal contribution of within-occupation and between-occupation components. More precisely, convergence of rural and 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 additional 0.18 (or 78 percent) decline in labor income gap. The majority of these changes were driven by blue-collar occupations. White-collar jobs also saw wage convergence both within occupations and between occupations, although the convergence was smaller than in blue-collar jobs.

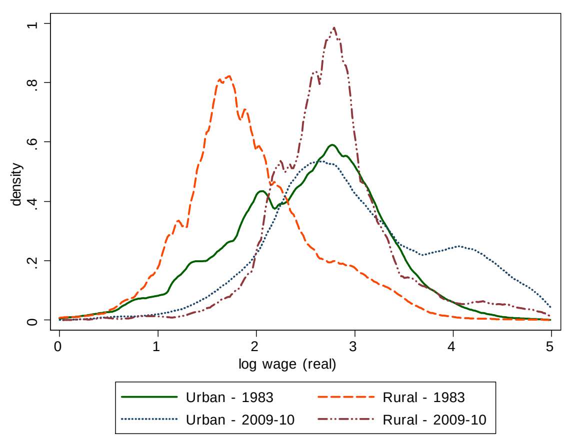
This convergence driven by wages was somewhat offset by reallocation of workers across occupations. The latter has led to an increase of the labor income gap by 0.08 (or 36 percent). All of this divergence in employment shares was accounted for by white-collar jobs, where employment shares in urban and rural areas have diverged and thus led to a divergence of the labor income gap by 0.15. Employment shares in blue-collar jobs, on the other hand, have converged and thus helped to offset some of the divergence brought on by white-collar jobs.

Clearly, convergence between urban and rural *wages* 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 more details and consider some factors behind this convergence.

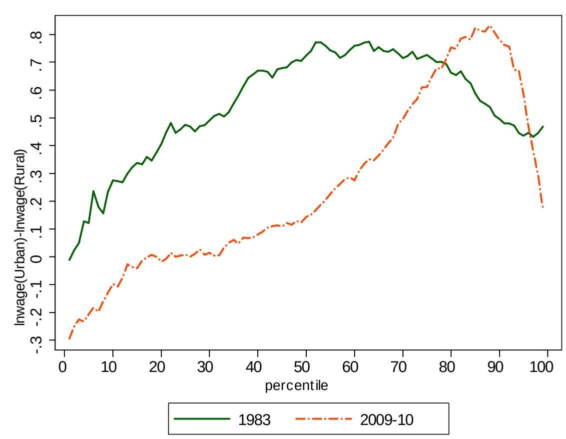
3 Rural-Urban Wage Gaps

In studying urban-rural wage convergence we are interested not just in the mean or median wage gaps, but rather in the behavior of the wage gap across the entire wage distribution. Thus, we start by taking a look at the distribution of log wages for rural and urban workers in our sample. Panel (a) of Figure A1 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2009-10 survey rounds.[6] The plot shows a very clear rightward shift of the wage density function during this period for rural workers. The shift in the wage distribution for urban workers is much more muted. In fact, the mean 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, has disappeared and was replaced by a fat right tail.

Figure 2: The log wage distributions of urban and rural workers for 1983 and 2004-05



(a)



(b)

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

[6] 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 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 1983-2010 period. The results for pre- and post-NREGA subsamples are provided in the Appendix A.3.

Panel (b) of Figure A1 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 gap 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. Indeed, it is easy to see from Panel (b) that the median log wage gap between urban and rural wages fell from around 0.7 to around 0.1. Hence, the median wage premium of urban workers declined from around 101 percent to 11 percent. 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 in the two sectors, rural wages were actually higher than urban wages. This was in stark contrast to the picture in 1983 when urban wages were higher than rural wages for all percentiles.

Figures A1 suggest wage convergence between rural and urban areas. But is this borne out statistically? To test for this, we estimate Recentered Influence Function (RIF) regressions developed by Firpo, Fortin, and Lemieux (2009) 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. Our interest is in the coefficient on rural dummy. The controls for age are intended to flexibly control for the fact that wages are likely to vary with age and experience. We perform the analysis for different unconditional quantiles as well as the mean of the wage distribution.[7]

Panel (a) of Table 2 reports the estimated coefficient on the rural dummy for the 10th, 50th and 90th percentiles as well as the mean for different survey rounds.[8] 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 over the entire period as well all sub-periods within (see Panel (b)) with the largest convergence having occurred for the median. In fact, 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

[7] We use the RIF approach (developed by Firpo, Fortin, and Lemieux (2009)) because we are interested in estimating the effect of the rural dummy for different points of the distribution, not just the mean. However, since the law of iterated expectations does not go through for quantiles, we cannot use standard mean regression methods to determine the unconditional effect of rural status on wages for different quantiles. The RIF methodology gets around this problem for quantiles. Details regarding this method can be found in Firpo, Fortin, and Lemieux (2009).

[8] Due to an anomalous feature of missing rural wage data for 1987-88, we chose to drop 1987-88 from the study of wages in order to avoid spurious results.

Table 2: 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 conditional mean 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.

time, for the 90th percentile the wage gap actually increased over time. These results corroborate the visual impression from Figure A1: the wage gap between rural and urban areas fell between 1983 and 2010 for all but the richest wage groups.

We now turn to our central goal of uncovering the factors behind converging wage gaps in rural and urban areas. We consider three explanations. First, wage convergence may have arisen due to convergence of individual and household characteristics. Second, the wage levels of urban and rural workers may have been brought closer together through workers migration between urban and rural areas. Third, economic reform targeting workers in rural areas may have played a role. We investigate each of these explanations in turn.

3.1 Individual and household attributes

We begin by summarizing demographic characteristics, education attainment levels and occupation distribution of rural and urban workforce in our sample. We then evaluate the contribution of each of these attributes to the urban-rural wage gap using DiNardo, Fortin, and Lemieux (1996) technique. This technique allows us to examine urban-rural inequality in wages across the entire distribution. We show that this is important as urban-rural gap is much larger at the top of the wage distribution. Furthermore, the technique allows for attributes to have marginal effects that vary along the distribution.

3.1.1 Demographics

We summarize demographic characteristics across the rounds in Table 3. 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 73 percent of households 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).

Table 3: Sample summary statistics

	(a) Individuals			(b) Households		
Urban	age	male	married	proportion	SC/ST	hh size
1983	35.03	0.87	0.78	0.26	0.16	5.01
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1987-88	35.45	0.87	0.79	0.24	0.15	4.89
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1993-94	35.83	0.87	0.79	0.26	0.16	4.64
	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1999-00	36.06	0.86	0.79	0.28	0.18	4.65
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	36.18	0.86	0.77	0.27	0.18	4.47
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2009-10	36.96	0.86	0.79	0.29	0.17	4.27
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Rural						
1983	35.20	0.77	0.81	0.74	0.30	5.42
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1987-88	35.36	0.77	0.82	0.76	0.31	5.30
	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1993-94	35.78	0.77	0.81	0.74	0.32	5.08
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
1999-00	36.01	0.73	0.82	0.72	0.34	5.17
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
2004-05	36.56	0.76	0.82	0.73	0.33	5.05
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
2009-10	37.66	0.77	0.83	0.71	0.34	4.77
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Difference						
1983	-0.17***	0.11***	-0.04***	-0.48***	-0.15***	-0.41***
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
1987-88	0.09	0.10***	-0.03***	-0.51***	-0.16***	-0.40***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1993-94	0.04	0.10***	-0.02***	-0.47***	-0.16***	-0.44***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
1999-00	0.05	0.13***	-0.04***	-0.45***	-0.16***	-0.52***
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
2004-05	-0.39***	0.10***	-0.05***	-0.45***	-0.15***	-0.58***
	(0.10)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
2009-10	-0.70***	0.09***	-0.04***	-0.42***	-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.

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 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 in 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.

3.1.2 Education

Next we turn to the education levels of the rural and urban workforce. Education in the NSS data is presented as a category variable with the survey listing the highest education attainment level in terms of categories such as primary, middle etc. 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.[9] 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. While we use the second method for our econometric specifications since these are the actually reported data (as opposed to the years series that was constructed by us), we also show results from the first approach below.

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 years of education of the urban worker was 164 percent higher than 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 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 the rural workers improving faster.

[9] 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.

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)

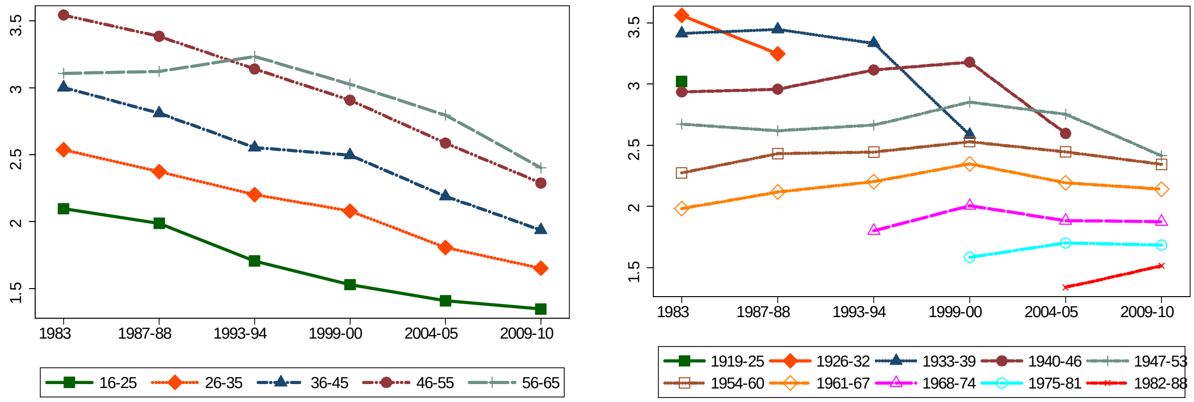
Notes: This table presents the average years of education for the overall sample and separately for the urban and rural workforce; as well as the relative gap in the years of education obtained as the ratio of urban to rural education years. Standard errors are in parenthesis.

Table 4, while revealing an improving trend for the average worker, nevertheless masks potentially important underlying heterogeneity in education attainment by cohort, i.e., variation by the age of the respondent. Panel (a) of Figure 3 shows the relative gap in years of education between the typical urban and rural worker by age group. There are two key results to note from panel (a): (i) the gaps have been getting smaller over time for all age groups; (ii) the gaps are smaller for the younger age groups.

Is the education convergence taking place uniformly across all birth cohorts, or are the changes mainly being driven by ageing effects? To disentangle the two we compute relative education gaps for different birth cohorts for every survey year. Those are plotted in panel (b) of Figure 3. Clearly, almost all of the convergence in education attainments takes place through cross-cohort improvements, with the younger cohorts showing the smallest gaps. Ageing effects are symmetric across all cohorts, except the very oldest. Most strikingly, the average gap in 2009-10 between urban and rural workers from the youngest birth cohort (born between 1982 and 1988) has almost disappeared while the corresponding gap for those born between 1954 and 1960 stood at 150 percent. Clearly, the declining rural-urban gaps are being driven by declining education gaps amongst the younger workers in the two sectors.

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 movement in the category of

Figure 3: Education gaps by age groups and birth cohorts

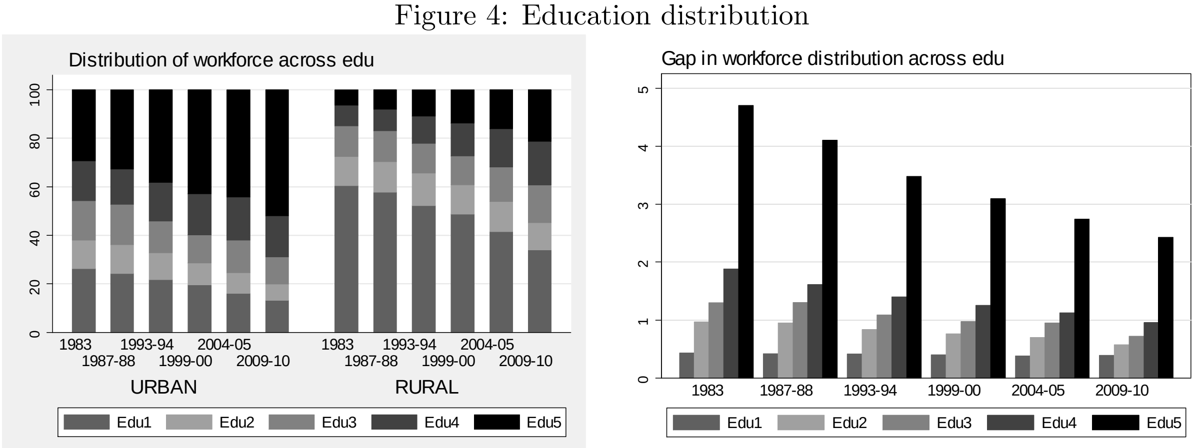


(a) (b)

Notes: The figures show the relative gap in the average years of education between the urban and rural workforce over time for different for different age groups and birth cohorts.

education.

Figure 4: 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).

Panel (a) of Figure 4 shows the distribution of the urban and rural workforce by education category. Recall that education categories 1, 2 and 3 are "illiterate", "some 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.

Simultaneously, 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 percent to around 30 percent, represent the sharpest and most promising changes in the past 27 years.

Panel (b) of Figure 4 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 the 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 while the further away from one they are, the more skewed the distribution is. As the Figure indicates, the biggest convergence in the education distribution 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 significance of 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 ("illiterate" and "some but below primary education", respectively) 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 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

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.

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 4. The figure shows trends in the relative gaps while the probit regressions show trends in the absolute gaps.

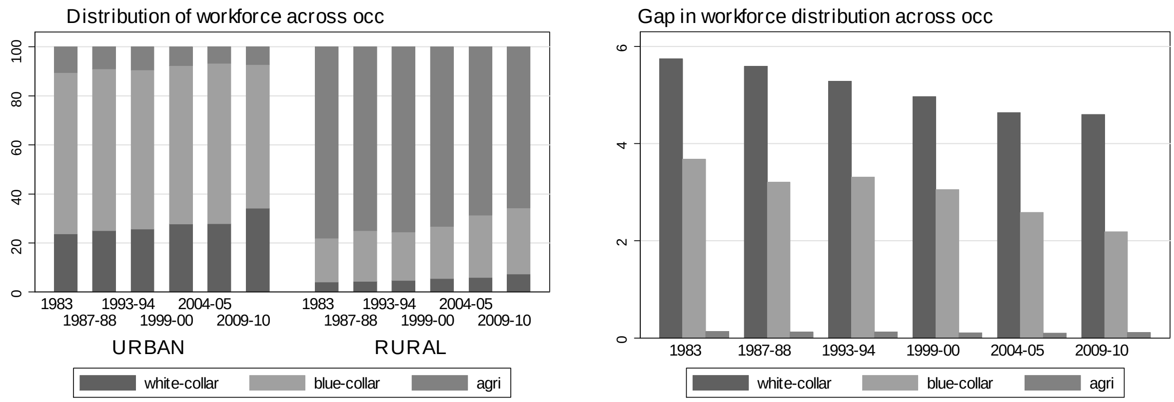
In summary, the overwhelming feature of the data on education attainment gaps suggests a strong and significant trend toward education convergence between the urban and rural workforce. This is evident when comparing average years of education, the relative gaps by education category as well as the absolute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

3.1.3 Occupation Choices

We now turn to the occupation choices being made by the workforce in urban and rural areas. To examine this issue, we we consider three occupation categories: white-collar occupations, blue-collar occupations, and agricultural occupations. Figure 5 shows the distribution of these occupations in urban and rural India across the survey rounds (Panel (a)) as well as the gap in these distributions between the sectors (Panel (b)).

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

Figure 5: Occupation distribution



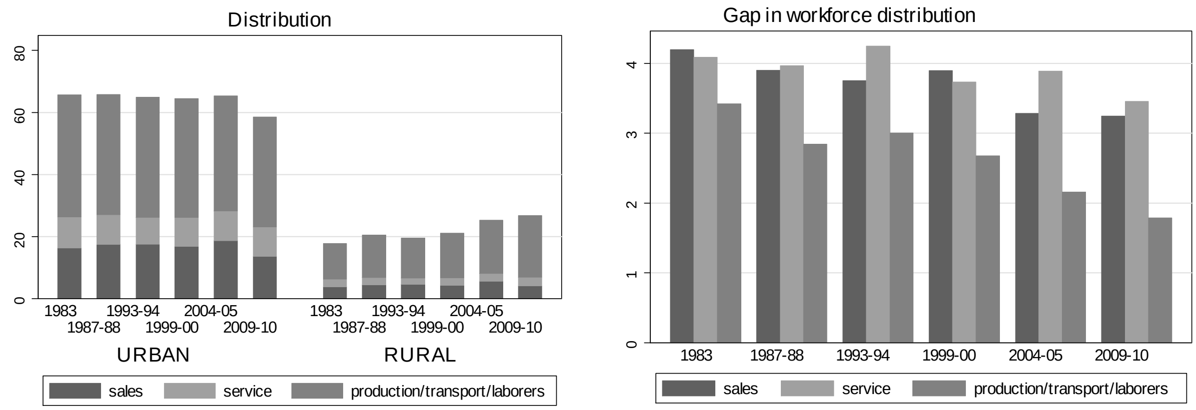
(a) (b)

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.

share of people working in white collar jobs. The crucial aspect though is the share of the workforce in blue collar jobs that pertain to both services and manufacturing. The urban sector clearly has a dominance of these occupations. Importantly though, the share of blue-collar jobs has been rising in rural areas. In fact, as Panel (b) of Figure 5 shows, the share of both white collar and blue collar jobs in rural areas are rising faster than their corresponding shares in urban areas.

What are the non-farm occupations that are driving the convergence between rural and urban areas? We answer this question by considering disaggregated occupation categories within the white-collar and blue-collar jobs. We start with the blue-collar jobs that have shown the most pronounced increase in rural areas. Panel (a) of Figure 6 presents the break-down of all blue-collar jobs into three types of occupations. The first group are *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. The second group are *service workers*, including hotel and restaurant staff, maintenance workers, barbers, policemen, firefighters, etc. The third group consists of *production and transportation workers and laborers*. This group includes among others miners, quarrymen, and various manufacturing workers. The main result that jumps out of panel (a) of Figure 6 is the rapid expansion of blue-collar jobs in the rural sector. The share of rural population employed in blue-collar jobs has increased from under 18 percent to 27 percent between 1983 and 2010. This increase is in sharp contrast with the urban sector where the population share of blue-collar jobs remained roughly unchanged at around 65 percent during this period. Most of the increase in blue-collar jobs in the rural sector was accounted for by a two-fold expansion in the share of production jobs (from 11 percent in 1983 to 20 percent in 2010). While sales and service jobs in the rural areas expanded as well, the increase was much less dramatic. In the urban sector however, the trends have been quite different: While sales and service jobs have remained relatively unchanged, the share of production jobs has actually declined.

Figure 6: Occupation distribution within blue-collar jobs



(a) (b)

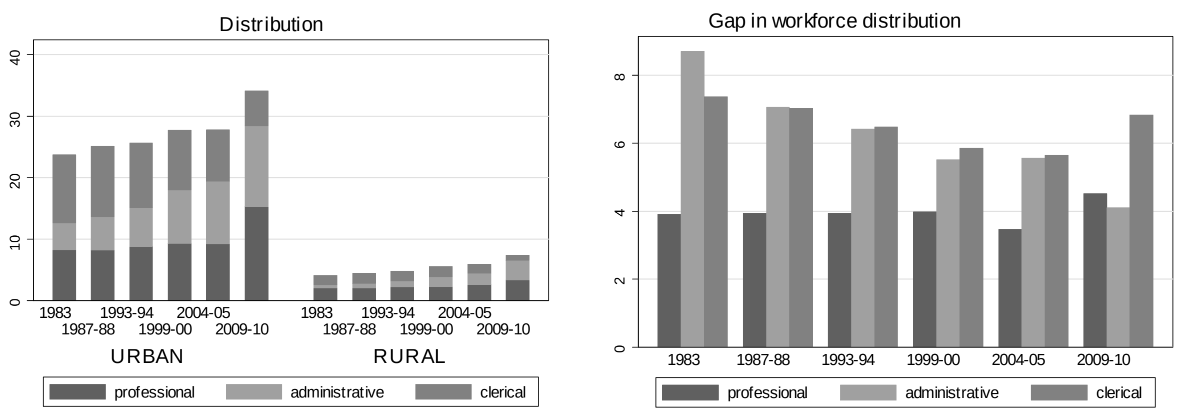
Notes: Panel (a) of this figure presents the distribution of workforce within blue-collar jobs for different NSS rounds. The left set of bars refers to urban workers, while the right set is for rural workers. Panel (b) presents relative gaps in the distribution of urban relative to rural workers across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

Clearly, such distributional changes should have led to a convergence in the rural and urban occupation distributions. To illustrate this, panel (b) of Figure 6 presents the relative gaps in the workforce distribution across various blue-collar occupations. The largest gaps in the sectoral employment shares were observed in sales and service jobs, where the gap was 4 times in 1983. The distributional changes discussed above have led to a decline in the urban-rural gaps in these jobs. The more pronounced decline in the relative gap was in production occupations: from 3.5 in 1983 to less than 2 in 2010.

Next, we turn to white-collar jobs. Panel (a) of Figure 7 presents the distribution of all white-collar jobs in each sector into three types of occupations. The first is *professional, technical and related workers*. This group includes, 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 figure shows that administrative jobs is the fastest growing occupation within the white-collar group in both rural and urban areas. It was the smallest category among all white-collar jobs in both sectors in 1983, but has expanded dramatically ever since to overtake clerical jobs as the second most popular occupation among white-collar jobs after professional occupations. Lastly, the share of professional jobs has also increased while the share of clerical and related jobs has shrunk in both the rural and urban sectors during the same time.

Have the expansions and contractions in various jobs been symmetric across rural and urban sectors? Panel (b) of Figure 7 presents relative gaps in the workforce distribution across various white-collar occupations. The biggest difference in occupation distribution between urban and rural sectors was in administrative jobs, but the gap has declined more than two-fold between 1983 and 2010. Similarly, the relative gap in clerical jobs has fallen, although the decline was more muted.[10] The gap in professional jobs remained relatively unchanged at 4 during the same period.

Figure 7: Occupation distribution within white-collar jobs



(a) (b)

Notes: Panel (a) of this figure presents the distribution of workforce within white-collar jobs 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 different occupation categories.

Overall, these results suggest that the expansion of rural non-farm sector has led to rural-urban occupation convergence, contrary to a popular belief that urban growth was deepening the rural-urban divide in India.

Is this visual image of sharp changes in the occupation distribution and convergent trends statistically significant? To examine this we estimate a multinomial probit regression of occupation choices

[10] There is a jump in the urban-rural gap in clerical occupations in 2010 which we believe may be driven by the small number of observations for these jobs in rural areas.

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 significantly negative marginal effect on the probability of being in white-collar and blue-collar jobs, while having significantly 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 sector, these results as indicate an ongoing process of convergence across rural and urban areas in these two occupation. At the same time, the gap in the share of the workforce in white-collar jobs between urban and rural areas has widened.

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 present 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. N refers to the number of observations. Agrarian jobs is the reference group in the regressions. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Note that these results are consistent with the labor income decomposition results reported in section 2. There we showed that labor reallocation channel in white-collar jobs have contributed to a widening of the labor income gap between urban and rural areas. This was because the employment distribution was becoming more uneven in these jobs, in terms of *absolute differences*, in line with the evidence in Table 6. In terms of the *relative differences*, however, the occupation distribution between urban and rural areas was converging in white-collar jobs, as Figure 5 shows. Blue-collar and agrarian jobs have shown convergence over time in both absolute and relative terms.

3.1.4 Decomposition of wage gaps

How much of the wage convergence documented above driven by a convergence of measured covariates? Or was it due to changes in unmeasured factors? We proceed with two approaches. 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.[11,12]

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. Note that 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 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.[13] By endowing urban workers with the occupation distribution of rural workforce allows us to effectively control for the between-occupation wage structure effects driving urban-rural wage gap. The unexplained part, therefore, must be predominantly due to differences in the within-occupation wage structure between urban and rural workers.

Figure 8 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.

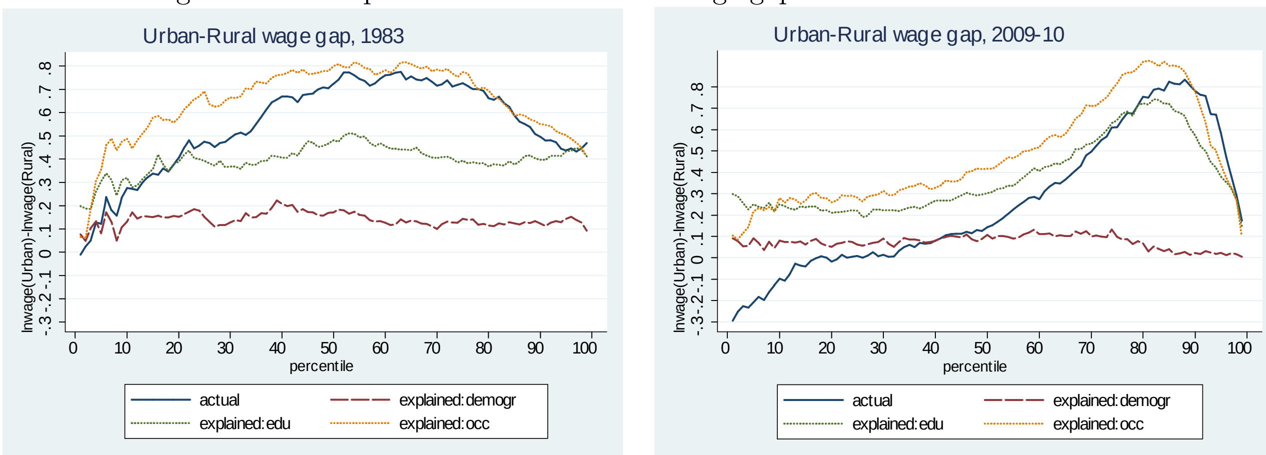
Figure 8 shows that demographic characteristics explain a small fraction of the urban-rural wage gap. Moreover, this fraction remains stable at around 0.1 along the entire distribution in both 1983 and 2009-10. For the 1983 wage distribution gap, differences in education account for almost the entire wage gap at the bottom of the distribution, while differences in occupation explain the

[11] 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 reweighing of the estimated wage density function of urban households. The counterfactual density is then compared with the actual wage density to assess the contribution of the measured attributes to the observed wage gap.

[12] We choose to do the reweighing 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 be able to mimic the urban distribution.

[13] Our occupation controls include more disaggregated occupation categories discussed in Section 3.1.3.

Figure 8: 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.

wage gap for the upper 50 percent of the distribution. Turning to 2009-10 however, the picture is different. Here differences in education attainments between urban and rural workers explain a large fraction of the gap at the top end of the distribution (median and above). However, for the bottom end of the distribution the education differences suggest that there should exist a large gap in *favor* of urban workers. This finding stands in stark contrast to the actual data which shows that wages of rural workers are in fact *higher* than wages of urban workers for the bottom 15 percent of distribution. Clearly, the data wage gap for the bottom 15 percent is the opposite of what their demographic characteristics and education endowments predict. Adding occupations deepens the puzzle further. Based on differences in occupations, the urban-rural gap should be more than 20 percent higher than the actual gap in the data. These results suggest that differences in the wage structure of the urban and rural workers play an important role in our data. The unexplained component remains large when we consider the within-occupation wage gaps for each occupation separately. The unexplained component is particularly pronounced in blue-collar and argarian jobs. Similarly, we find the unexplained component of between-occupation wage gaps to be large as well. Therefore, both between- and within-occupation components of urban-rural wage gaps contribute to our finding of large wage structure effects.[14]

Our second approach aims to understand the time-series evolution of wage gaps between urban

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

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 OLS regressions for the decomposition at the mean, and Recentered Influence Function (RIF) regressions for decompositions at the 10th, 50th, and 90th quantiles.[15]

Our set of explained 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 1-5.[16]

Table 7: Decomposing changes in rural-urban wage gaps over time

(a). Change 1983 to 2009-10				explained
	(i) measured gap	(ii) explained	(iii) unexplained	(iv) 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)
(b). Change in explained component				
10th quantile	-0.096***	-0.060***	-0.036***	-0.049***
	(0.016)	(0.008)	(0.013)	(0.006)
50th quantile	-0.202***	-0.064***	-0.137***	-0.052***
	(0.014)	(0.012)	(0.014)	(0.009)
90th quantile	0.229***	0.060***	0.169***	0.084***
	(0.046)	(0.021)	(0.040)	(0.020)
mean	-0.115***	-0.032***	-0.083***	-0.015
	(0.014)	(0.012)	(0.008)	(0.010)

Note: Panel (a) presents the change in the rural-urban wage gap between 1983 and 2009-10. Panel (b) reports the decomposition of the time-series change in the explained component of the change in the wage gap over 1983-2010 period. All gaps are decomposed into explained and unexplained components using the RIF regression approach of Firpo, Fortin, and Lemieux (2009) for the 10th, 50th and 90th quantiles. Both panels also report the contribution of education to the explained gaps. Bootstrapped standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Table 7 shows the results of the decomposition exercise. Panel (a) shows the decomposition of the 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 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, about 50 percent of 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.

[15] 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.4.

[16] We do not include occupation amongst the explanatory variables since it is likely to be endogenous to wages.

If the explained component of a regression is βX, then changes in that component has two components: the change in X and the change β, which is the measured return to X. Since X is measured in the data, the part of the change in the explained component that is due to X is "explained" by the data while the part due to β is not directly explained. Panel (b) of the Table 7 decomposes changes in the explained component itself into the explained and unexplained parts. For the 10th percentile, most of the change in the measured component of the gap was due to changes in the explained part (or X). For the median and the 90th percentile however, most of the change in the explained component was due to changes in returns rather than changes in the component itself.

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, a large fraction of the trend is due to unmeasured factors, especially for the left tail of distribution. This is particularly puzzling since the actual wage gaps for the bottom 10 percent of the urban and rural wage distributions are in favor of rural workers while the covariates predict the opposite!

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. Indeed, two of the older theories of structural transformation – the Lewis and Harris-Todaro models – both formalize the process through which rural/agricultural workers migrate to urban areas in search of higher wages. Even from a neoclassical perspective, i.e., from a non-dualistic economy view of the world, 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.[17] We use information from these three rounds to form an assessment of the role of migration.

[17] 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.

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 of the total workforce has declined from 7.2 percent in 1983 to 6.2 percent in 2007-08.[18] 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 above 38 percent in 2007-08. The share of urban migrants to rural areas has stayed relatively unchanged around 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, thereby failing to keep pace with the rise in the rural-to-urban flows. Interestingly, 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 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. While clearly not insignificant, the share of migrant workers from rural areas in the urban workforce is relatively small given the overall size of the urban workforce.

Table 8: Migration trends: 1983-2008

	migrant	migrants				rural-to-urban	for job
	total	rural-to-urban	urban-to-urban	rural-to-rural	urban-to-rural	urban	rural-to-urban
1983	0.072	0.224	0.185	0.496	0.087	0.072	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	0.067	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	0.072	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. Same is true about 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 2005. Migration to rural areas is in equal proportion for job, marriage and other reasons.[19]

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 regression specifications in Section 3.1.4 to include four additional dummy variables, each identi-

[18] These numbers imply annual migration flows of about 1 percent of total workforce.

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

fying 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 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 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.[20] 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 2 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 A2 in Appendix A.5).[21] This suggests to us 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

[20] 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.

[21] The regression results for migrants in the 10th percentile of the distribution reported in A2 should be treated with caution due to the small number of observations. This might be reflecting problems in sample coverage of poorer migrants to urban areas.

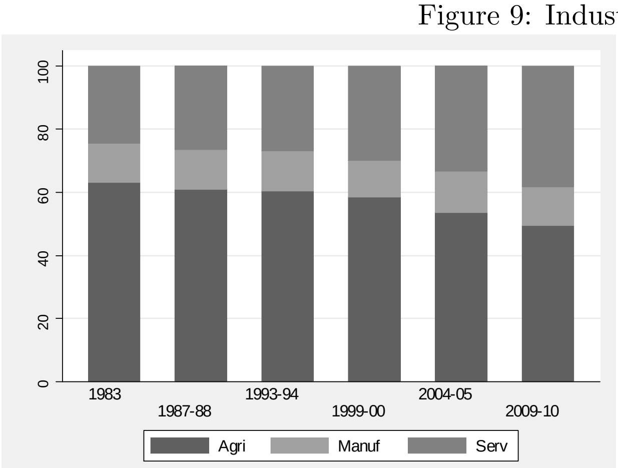
wage gaps between rural and urban areas. Of course this conclusion is subject to an obvious caveat that migration decision itself is endogenous to wage gaps between rural and urban areas. Thus, how much of the overall wage convergence as well as its distributional differences can be explained by migration requires a more structural analysis which is beyond the scope of this paper. We hope to return to this issue in future work.

4 Aggregate facts

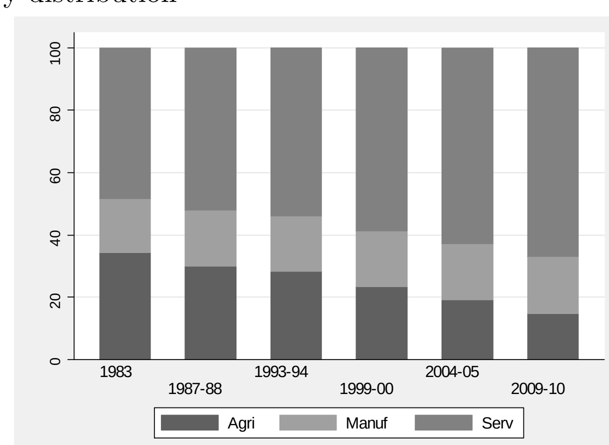
The previous results suggest that a majority of the convergence between rural and urban India cannot be accounted for by convergence in the individual characteristics of the two groups. This then leads us to examining alternative explanations for the convergent trends. One possibility is that aggregate developments dduring this period may have accounted for the trends. To examine this possibility, we start by presenting some of the key aggregate developments in India during 1983-2010 period. In particular, we show how employment and output shares in the three sectors of India's economy have evolved during this time. We also characterize the evolution of sectoral productivity and relative prices.

Figure 9 shows employment shares (panel (a)) and output shares (panel (b)) in agriculture, manufacturing and services during 1983-2010 period. As is easy to see, the agriculture has been releasing workers, and its share of output is declining over time, while service sector is expanding both as a share of employment and as a share of output.

Figure 9: Industry distribution



(a) sectoral employment shares

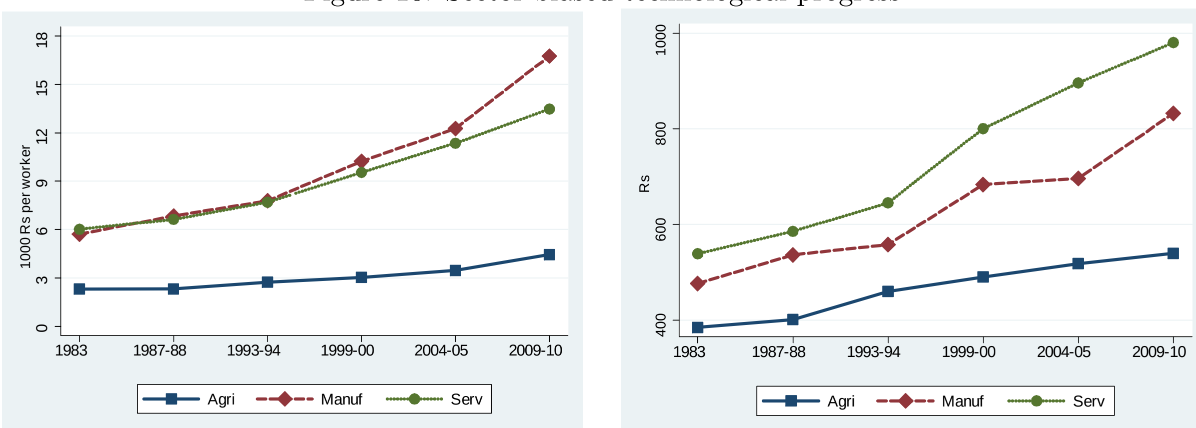


(b) sectoral output shares

Notes: Panel (a) of this Figure presents the distribution of workforce across three industry categories for different NSS rounds. Panel (b) presents distribution of output (measured in constant 1980-81 prices) across three industry categories. The source for the figure is Hnatkovska and Lahiri (2011).

Underlying this process of sectoral reallocation were the changing patterns of productivity. Figure 10 presents labor productivity and total factor productivity (TFP) in agriculture, manufacturing, and services during 1983-2010 period.

Figure 10: Sector-biased technological progress

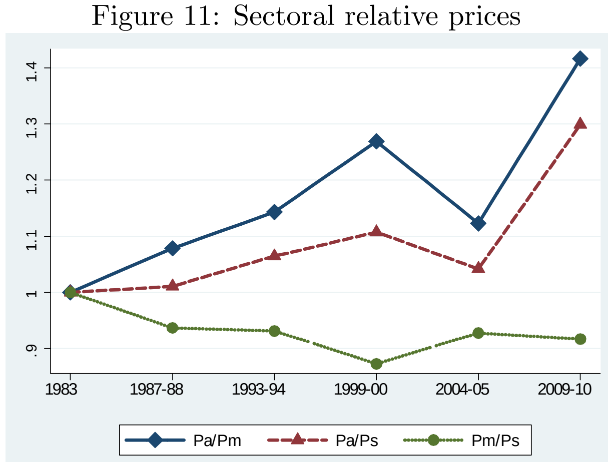


(a) Sectoral labor productivity, per worker (b) Sectoral TFP

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.

Lastly, Figure 11 presents the evolution of sectoral relative prices during the same time period. Thus, the period of 1983-2010 was characterized by an increase in prices of agricultural goods relative to both manufacturing and service goods. At the same time the price of manufacturing relative to service goods somewhat declined.

Figure 11: Sectoral relative prices



Notes: This figure shows the relative sectoral prices. Pa, Pm and Ps refer to price indices in agriculture, manufacturing and services, respectively.

5 A Structural Explanation

A key feature of the empirical results described above is that measured correlates such as demographic patterns and human capital attainment rates explain at best about 50 percent of the convergence in average wages between rural and urban areas. What accounts for the rest of the observed convergence? We turn to this issue now by formalizing 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 some minimal features that can generate a structural transformation (a relative decline of the agricultural sector accompanied by a reallocation of both factors towards the non-agricultural sector) as well as declining wage gaps between rural and urban labor. We then want to quantitatively examine the relative contributions of the identified factors to the observed wage convergence.

Consider a static, 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

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

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.

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

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

where ψ 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.

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

$$Y_A = AL_A$$

$$Y_S = SL_S$$

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

$$L_j = \left[\beta_j L_{Uj}^{\phi_j} + \left(1 - \beta_j\right) L_{Rj}^{\phi_j}\right]^{1/\phi_j}, \quad \phi_j \in (-\infty, 1], \quad j = A, S \tag{5.3}$$

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.

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 of 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. 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

$$w_{iA} = w_{iS} - \psi_i \tag{5.4}$$

$$c_{iA} = (1 - \theta)\,\bar{c} + \theta y_i \tag{5.5}$$

$$pc_{iS} = (1 - \theta)\,(y_i - \bar{c}) \tag{5.6}$$

where $y_i = w_{iA} l_{iA} + (w_{iS} - \gamma_i) l_{iS} + \Omega_i / L_i$ denotes total income of household $i = U, R$. Equation (5.4) makes clear that the cost of training ψ 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:

$$p = \frac{MPL_{RA}}{MPL_{RS}} = \frac{MPL_{UA}}{MPL_{US}} \tag{5.7}$$

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, note that market clearing in each sector dictates that

$$c_{UA} L_U + c_{RA} L_R = Y_A \tag{5.8}$$

$$c_{US} L_U + c_{RS} L_R = Y_s \tag{5.9}$$

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 satisfy their optimality conditions, budget constraints are satisfied and all markets clear.*

5.1 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 (5.7)) to solve for k_A in terms of k_S. Under the general CES labor aggregator (equation (5.3)) with $\phi \neq 0$ this solution is derived by solving for k_A from the condition

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

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

$$k_A = \mu\left(k_S\right)$$

In the general CES case, one can use this solution for k_A to characterize the equilibrium for this economy by the system:

$$p = \frac{A\beta_A \left\{\mu\left(k_S\right)\right\}^{\phi_A - 1} \left[\beta_A \left\{\mu\left(k_S\right)\right\}^{\phi_A} + 1 - \beta_A\right]^{\frac{1-\phi_A}{\phi_A}} + \tau_U}{S\beta_S k_s^{\phi_S - 1} \left[\beta_S k_S^{\phi_S} + 1 - \beta_S\right]^{\frac{1-\phi_S}{\phi_S}}} \tag{5.10}$$

$$p = \left(\frac{1-\theta}{\theta}\right) \frac{A\left[\beta_A \left\{\mu\left(k_S\right)\right\}^{\phi_A} + 1 - \beta_A\right]^{\frac{1}{\phi_A}} \left[\frac{k - k_S}{\mu(k_S) - k_S}\right] - \bar{c}\left(1+k\right) - \left(\tau_R + \tau_U k_S\right)\left[\frac{\mu(k_S) - k}{\mu(k_S) - k_S}\right]}{S\left[\beta_S k_S^{\phi_S} + 1 - \beta_S\right]^{\frac{1}{\phi_S}} \left[\frac{\mu(k_S) - k}{\mu(k_S) - k_S}\right]} \tag{5.11}$$

This is a two-equation system with 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 (5.10) comes from combining the firm optimality conditions $pMPK_{US} = w_{US}$ and $MPK_{UA} = w_{UA}$ with the household optimality condition $w_{UA} = w_{US} - \tau_U$. Equation (5.11) arises from combining the household budget constraints with the market clearing conditions for the two goods.

5.2 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 5.1 *The labor aggregators in the two sectors are of the Cobb-Douglas form given by*

$$L_j = L_{Uj}^{\beta_j} L_{Rj}^{1-\beta_j} \ , \qquad j = A, S. \tag{5.12}$$

Condition 5.2 *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

$$k_A = \gamma k_S \ , \quad \gamma \equiv \frac{\beta_A}{1-\beta_A} \frac{1-\beta_S}{\beta_s} \tag{5.13}$$

Moreover, the equilibrium system is given by

$$p = \frac{A\beta_A \gamma^{\beta_A - 1} k_S^{\beta_A}}{S\beta_S k_s^{\beta_S}} \tag{5.14}$$

$$p = \left(\frac{1-\theta}{\theta}\right) \left[\frac{A\gamma^{\beta_A} k_S^{\beta_A - 1}\left(\frac{k-k_S}{\gamma-1}\right) - \bar{c}\,(1+k)}{S k_S^{\beta_S - 1}\left(\frac{\gamma k_S - k}{\gamma-1}\right)}\right] \tag{5.15}$$

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 $\frac{L_{UA}}{L_{RA}} = k_A < k_S = \frac{L_{US}}{L_{RS}}$. Hence, $\beta_A < \beta_S$ and $\gamma < 1$.

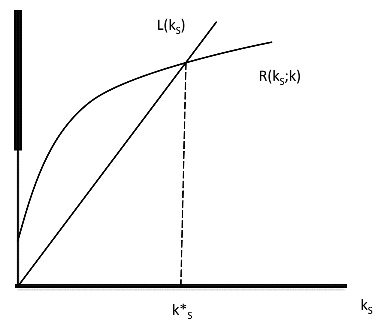
The equilibrium solution can be characterized by solving for k_S by equating these two equations. This gives

$$\left(1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\right) k_S = \left(1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\frac{1}{\gamma}\right) k + \bar{c}\,(1+k)\left(\frac{1-\gamma}{A\gamma^{\beta_A}}\right) k_S^{1-\beta_A} \tag{5.16}$$

The equilibrium is given by the k_S^* which solves this equation. The solution is graphically represented in Figure 12 where $L(k_S) = \left(1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\right) k_S$ and $R(k_S; k, \bar{c}, A) = \left(1 + \frac{\theta}{1-\theta}\frac{\beta_A}{\beta_S}\frac{1}{\gamma}\right) k + \bar{c}\,(1+k)\left(\frac{1-\gamma}{A\gamma^{\beta_A}}\right) k_S^{1-\beta_A}$. Note that $R(k_S; k, \bar{c}, A)$ is increasing and concave in k_S and has a positive intercept term.

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 saw that there was an increase in the relative supply of urban to rural labor between 1983 and 2010. Second, we also saw 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 sectors. Our interest lies in examining the impact of these shocks on the wage gap, the structural transformation of the economy as well as

Figure 12: Characterizing the equilibrium k_S



the agricultral terms of trade.

Proposition 5.3 *Under Conditions 5.1, 5.2, an increase in k, 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 (a) of Figure 13, 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. Hence, s_A must fall when k rises; (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 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 rises, i.e., p rises. The rise in the sectoral urban to rural labor ratios also cause rural wages to rise and urban wages to decline. 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 5.4 *Under Conditions 5.1, 5.2, an increase in agricultural productivity A: (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 (b) of Figure 13, 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 5.4 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 agricultural good falls, p rises. As the economy tries to shift 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.

Propositions 5.3 and 5.4 highlight two important features of our model economy. First, we need shocks to both productivity and the relative supply of urban to rural labor in order to jointly explain

Figure 13: Comparative static effects on k_S

(a) Rise in relative urban labor supply

(b) Rise in Agricultural TFP

the observed changes in relative wages, agricultural terms of trade and the structural transformation. Introducing increases in the relative endowment of urban labor by itself 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 only generates the observed 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 5.16.

5.3 A Calibrated Model

We now quantitatively assess the ability of the model to explain the observed rural-urban wage dynamics along with the aggregate macroeconomic facts. As we saw in the previous subsection, we need movements in both supply of urban labor relative to rural labor as well as shocks to agricultural productivity in order for the model to have the ability to generate both the relative wage dynamics along with the sectoral terms of trade and structural transformation of the economy that were observed during the period 1983-2010. We should also note that without a cost ψ of acquiring skills to work in sector S, this model cannot generate any inter-sectoral wage gap for either type of labor. Since this is a robust feature of the data, we shall introduce this training cost in the calibrated model below.

Our strategy is to calibrate the key parameters of the model so as to match the rural-urban gaps in wages and employment share in 1983. We then perturb the model with two shocks: (a) shocks to labor supply of the two groups; and (b) shocks to agricultural and non-agricultural productivity.

These shocks are measured from the data. Keeping all other parameters fixed, we examine the rural-urban gaps in 2010 that the model generates in response to these measured shocks.

6 Conclusion

This paper has examined the patterns of labor income changes in rural and urban India over the past three decades. We have found that this period 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. Majority of this convergence is accounted for by 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. Furthermore, urban to rural wage convergence has taken place both within each occupation and across occupations.

We evaluated two explanations for this wage convergence. First, we decomposed the urban-rural log wage gap along the distribution into two components: one due to differences in individual and household characteristics and one due to differences in returns to those characteristics. Surprisingly, we found that the majority of the decline in the rural-urban wage gap was not due to convergence in individual characteristics such as demographics or education attainments, but rather is unexplained. Furthermore, if in 1983 the majority of the urban-rural wage gap was accounted for by demographics, education and occupation attributes of individuals, by 2010, most of the wage gap was due to urban-rural differences in wage structure rather than differences in their characteristics.

Second, we have 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 individual-level 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 strucctural transformation we sshowed 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. We believe this mayy be a common phenomenon in other countries as well. We intend to examine the cross-country evidence on this in future work.

References

Alesina, A., and D. Rodrik (1994): "Distributive Politics and Economic Growth," *The Quarterly Journal of Economics*, 109(2), 465–90.

Banerjee, A., and T. Piketty (2001): "Are the rich growing richer: Evidence from Indian tax data," Working papers, MIT and CEPREMAP.

Bhalla, S. S. (2003): "Recounting the poor: Poverty in India, 1983-99," *Economic and Political Weekly*, pp. 338–349.

Deaton, A., and J. Dreze (2002): "Poverty and inequality in India: A re-examination," *Economic and Political Weekly*, pp. 3729–3748.

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.

Harris, J. R., and M. P. Todaro (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60(1), 126–142.

Hnatkovska, V., and A. Lahiri (2011): "Convergence Across Castes," Igc working papers, International Growth Center.

Hnatkovska, V., A. Lahiri, and S. Paul (2012): "Castes and Labor Mobility," *American Economic Journal: Applied Economics*, 4(2), 274–307.

Kuznets, S. (1955): "Economic growth and income inequality," *American Economic Review*, 45(1), 1–28.

Lewis, A. (1954): "Economic development with unlimited supply of labour," *The Manchester School*, 22(2), 139–191.

Pal, P., and J. Ghosh (2007): "Inequality in India: A survey of recent trends," DESA Working Papers 45, UN.

Persson, T., and G. Tabellini (1994): "Is Inequality Harmful for Growth?," *American Economic Review*, 84(3), 600–621.

Sen, A., and Himanshu (2005): "Poverty and inequality in India: Getting closer to the truth," in *Data and Dogma: The Great Indian Poverty Debate*, ed. by A. Deaton, and V. Kozel. Macmillan, New Delhi.

A Appendix

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 recoded 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

$$\begin{aligned} w_t^R &= w_{1t}^R L_{1t}^R + w_{2t}^R L_{2t}^R + w_{3t}^R L_{3t}^R, \\ w_t^U &= w_{1t}^U L_{1t}^U + w_{2t}^U L_{2t}^U + w_{3t}^U L_{3t}^U, \end{aligned}$$

where 1,2,3 refer to while-collar, blue-collar and agricultural jobs, respectively.

The relative labor income gap in period t is

$$\frac{w_t^U - w_t^R}{w_t^R} = \frac{\left(w_{1t}^U L_{1t}^U + w_{2t}^U L_{2t}^U + w_{3t}^U L_{3t}^U\right) - \left(w_{1t}^R L_{1t}^R + w_{2t}^R L_{2t}^R + w_{3t}^R L_{3t}^R\right)}{w_t^R}.$$

Adding and subtracting average labor income for each occupation (denoted by $w_{it}, i = 1, 2, 3$), we can write the expression above as

$$\begin{aligned} \frac{w_t^U - w_t^R}{w_t^R} &= \frac{\left(w_{1t}^U - w_{1t}\right) L_{1t}^U + \left(w_{2t}^U - w_{2t}\right) L_{2t}^U + \left(w_{3t}^U - w_{3t}\right) L_{3t}^U}{w_t^R} \\ &- \frac{\left(w_{1t}^R - w_{1t}\right) L_{1t}^R + \left(w_{2t}^R - w_{2t}\right) L_{2t}^R + \left(w_{3t}^R - w_{3t}\right) L_{3t}^R}{w_t^R} \end{aligned}$$

$$+\frac{w_{1t}\left(L_{1t}^{U}-L_{1t}^{R}\right)+w_{2t}\left(L_{2t}^{U}-L_{2t}^{R}\right)+w_{3t}\left(L_{3t}^{U}-L_{3t}^{R}\right)}{w_t^R}.$$

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 $i=1,2,3$; and $j=U,R$ and $\eta_{it} \equiv w_{it}/w_t^R$, $i=1,2,3$. Then the change in the relative gap can be written as

$$\begin{aligned}
& \frac{w_t^U-w_t^R}{w_t^R}-\frac{w_{t-1}^U-w_{t-1}^R}{w_{t-1}^R} \\
= \quad & \mu_{1t}^{U}L_{1t}^{U}+\mu_{2t}^{U}L_{2t}^{U}+\mu_{3t}^{U}L_{3t}^{U}-\left(\mu_{1t}^{R}L_{1t}^{R}+\mu_{2t}^{R}L_{2t}^{R}+\mu_{3t}^{R}L_{3t}^{R}\right) \\
& +\eta_{1t}\left(L_{1t}^{U}-L_{1t}^{R}\right)+\eta_{2t}\left(L_{2t}^{U}-L_{2t}^{R}\right)+\eta_{3t}\left(L_{3t}^{U}-L_{3t}^{R}\right) \\
& -\left(\mu_{1t-1}^{U}L_{1t-1}^{U}+\mu_{2t-1}^{U}L_{2t-1}^{U}+\mu_{3t-1}^{U}L_{3t-1}^{U}\right)-\left(\mu_{1t-1}^{R}L_{1t-1}^{R}+\mu_{2t-1}^{R}L_{2t-1}^{R}+\mu_{3t-1}^{R}L_{3t-1}^{R}\right) \\
& -\eta_{1t-1}\left(L_{1t-1}^{U}-L_{1t-1}^{R}\right)-\eta_{2t-1}\left(L_{2t-1}^{U}-L_{2t-1}^{R}\right)-\eta_{3t-1}\left(L_{3t-1}^{U}-L_{3t-1}^{R}\right).
\end{aligned}$$

Define $\bar{x}_t=\left(x_t+x_{t-1}\right)/2$, 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 $i=1,2,3$ 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 $i=1,2,3$ and collecting the terms in the second and fourth lines above, we get

$$\begin{aligned}
& \frac{w_t^U-w_t^R}{w_t^R}-\frac{w_{t-1}^U-w_{t-1}^R}{w_{t-1}^R} \\
= \quad & \Delta\mu_{1t}^{U}\bar{L}_{1t}^{U}+\Delta\mu_{2t}^{U}\bar{L}_{2t}^{U}+\Delta\mu_{3t}^{U}\bar{L}_{3t}^{U}-\Delta\mu_{1t}^{R}\bar{L}_{1t}^{R}-\Delta\mu_{2t}^{R}\bar{L}_{2t}^{R}-\Delta\mu_{3t}^{R}\bar{L}_{3t}^{R} \\
& +\Delta L_{1t}^{U}\bar{\mu}_{1t}^{U}+\Delta L_{2t}^{U}\bar{\mu}_{2t}^{U}+\Delta L_{3t}^{U}\bar{\mu}_{3t}^{U}-\Delta L_{1t}^{R}\bar{\mu}_{1t}^{R}-\Delta L_{2t}^{R}\bar{\mu}_{2t}^{R}-\Delta L_{3t}^{R}\bar{\mu}_{3t}^{R} \\
& +\bar{\eta}_{1t}\Delta\left(L_{1t}^{U}-L_{1t}^{R}\right)+\bar{\eta}_{2t}\Delta\left(L_{2t}^{U}-L_{2t}^{R}\right)+\bar{\eta}_{3t}\Delta\left(L_{3t}^{U}-L_{3t}^{R}\right) \\
& +\left(\overline{L_{1t}^{U}-L_{1t}^{R}}\right)\Delta\eta_{1t}+\left(\overline{L_{2t}^{U}-L_{2t}^{R}}\right)\Delta\eta_{2t}+\left(\overline{L_{3t}^{U}-L_{3t}^{R}}\right)\Delta\eta_{3t}
\end{aligned}$$

Using the fact that $L_{3t}^{j}=1-L_{1t}^{j}-L_{2t}^{j}$ we can rewrite the second row as

$$\Delta L_{1t}^{U}\left(\bar{\mu}_{1t}^{U}-\bar{\mu}_{3t}^{U}\right)+\Delta L_{2t}^{U}\left(\bar{\mu}_{2t}^{U}-\bar{\mu}_{3t}^{U}\right)-\Delta L_{1t}^{R}\left(\bar{\mu}_{1t}^{R}-\bar{\mu}_{3t}^{R}\right)-\Delta L_{2t}^{R}\left(\bar{\mu}_{2t}^{R}-\bar{\mu}_{3t}^{R}\right),$$

and the third row as

$$(\bar{\eta}_{1t}-\bar{\eta}_{3t})\Delta\left(L_{1t}^{U}-L_{1t}^{R}\right)+\left(\bar{\eta}_{2t}-\bar{\eta}_{3t}\right)\Delta\left(L_{2t}^{U}-L_{2t}^{R}\right),$$

and the fourth row as

$$\left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{3t}\right] + \left(\overline{L_{2t}^U - L_{2t}^R}\right)\left[\Delta\eta_{2t} - \Delta\eta_{3t}\right].$$

Thus the change in the relative labor income gap becomes

$$\begin{aligned}
&\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} \\
&= \Delta\mu_{1t}^U \bar{L}_{1t}^U + \Delta\mu_{2t}^U \bar{L}_{2t}^U + \Delta\mu_{3t}^U \bar{L}_{3t}^U - \Delta\mu_{1t}^R \bar{L}_{1t}^R - \Delta\mu_{2t}^R \bar{L}_{2t}^R - \Delta\mu_{3t}^R \bar{L}_{3t}^R && \text{(A1)} \\
&+ \Delta L_{1t}^U \left(\bar{\mu}_{1t}^U - \bar{\mu}_{3t}^U\right) + \Delta L_{2t}^U \left(\bar{\mu}_{2t}^U - \bar{\mu}_{3t}^U\right) - \Delta L_{1t}^R \left(\bar{\mu}_{1t}^R - \bar{\mu}_{3t}^R\right) - \Delta L_{2t}^R \left(\bar{\mu}_{2t}^R - \bar{\mu}_{3t}^R\right) && \text{(A2)} \\
&+ (\overline{\eta_{1t} - \eta_{3t}})\Delta\left(L_{1t}^U - L_{1t}^R\right) + (\overline{\eta_{2t} - \eta_{3t}})\,\Delta\left(L_{2t}^U - L_{2t}^R\right) && \text{(A3)} \\
&+ \left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{3t}\right] + \left(\overline{L_{2t}^U - L_{2t}^R}\right)\left[\Delta\eta_{2t} - \Delta\eta_{3t}\right] && \text{(A4)}
\end{aligned}$$

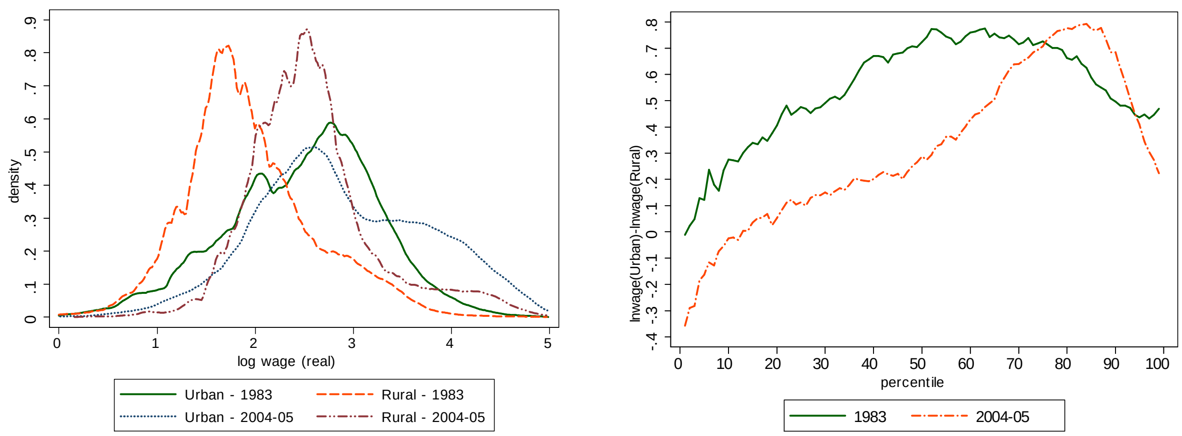
Row (A1) gives the within-occupation component of labor income convergence, rows (A2) and (A3) give the labor reallocation component of labor income convergence, while row (A4) gives the between-occupation component of labor income convergence.

A.3 National Rural Employment Guarantee Act and wage gaps

In 2005 the government of India enacted the National Rural Employment Guarantee Act (NREGA, since renamed the The Mahatma Gandhi National Rural Employment Guarantee Act). The objective of the Act was to provide an economic safety net to the rural poor by providing rural households with 100 days of guaranteed employment every year for at least one adult member for doing casual manual labour at the rate of Rupees 60 per day (approximately US $1.30 per day). The employment has to be productive, projects must not involve any machines or contractors, and the identification of projects is based on the economic, social and environmental benefits of different types of works, their contribution to social equity, and their ability to create permanent assets. In addition, the employment is generally provided within a radius of 5 kilometers of the village where the applicant resides at the time of applying. This Act could clearly have affected rural and urban wages. To control for the effects of this policy on real wages, we split our sample period into the pre-NREGA (1983 to 2004-05) and post-NREGA (2004-05 to 2009-10) periods. We first present the wage distributions and gaps for the two sub-periods, the DFL decomposition for the immediate pre-NREGA period, and then conduct a more detailed evaluation of NREGA using state-level variation in the intensity of the policy.

We begin with the pre-NREGA period of 1983 to 2004-05. Panel (a) of Figure A1 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2004-05 survey rounds. In line with our findings for the 1983-2010 period, the plot shows a very clear rightward shift of the wage density function during this period for rural workers, while the shift in the wage distribution for urban workers is much more muted.

Figure A1: The log wage distributions of urban and rural workers for 1983 and 2004-05



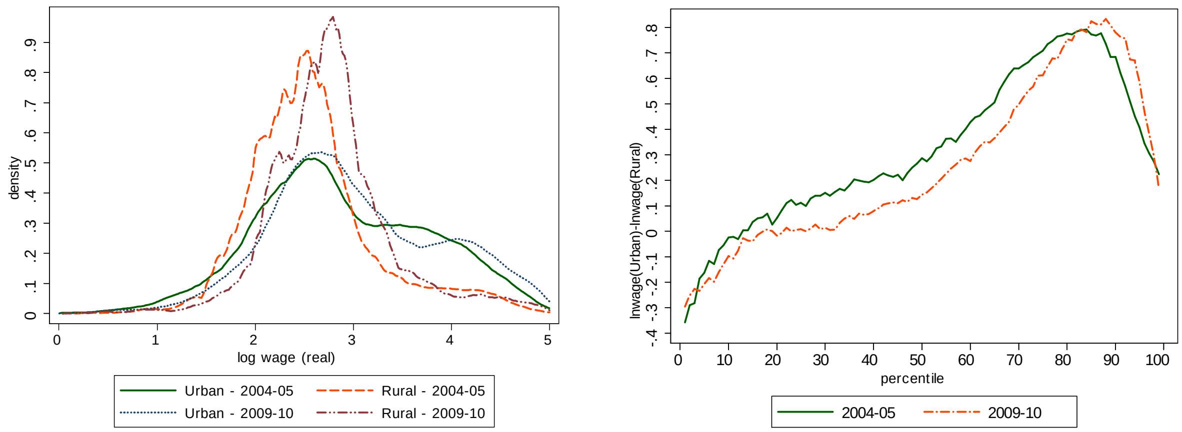
(a) (b)

Notes: Panel (a) shows the estimated kernel densities of log real wages for urban and rural workers, while panel (b) shows the difference in percentiles of log-wages between urban and rural workers plotted against the percentile. The plots are for 1983 and 2004-05 NSS rounds.

Panel (b) of Figure A1 presents the percentile (log) wage gaps between urban and rural workers for 1983 and 2004-05. In line with our earlier findings, the gap schedule for 2004-05 lies to the right of that for 1983 till the 70th percentile indicating that for most of the wage distribution, the gap between urban and rural wages has declined over this period. Next consider the post-NREGA wage distributions. Indeed, it is easy to see from Panel (b) that the median log wage gap between urban and rural wages fell from around 0.7 to around 0.2. Hence, the median wage premium of urban workers declined from around 101 percent to 22 percent. Figure A2 contrasts the real wage densities of rural and urban workers in 2004-05 and 2009-10. The figure shows that the urban-rural wage convergence we uncovered for 1983-2005 period continued in the post-reform period as well. Panel (a) indicates a clear rightward shift in the urban wage distribution, while panel (b) shows that the percentile gaps in 2009-10 lie to the right and below the gaps for 2004-05 period for up to 80th percentile. In fact, the median wage premium of the urban worker has declined from 22 percent to 11 percent during this period.

We conduct the DFL decomposition of the urban-rural wage gap in the immediate pre-NREGA

Figure A2: The log wage distributions of urban and rural workers for 2004-05 and 2009-10

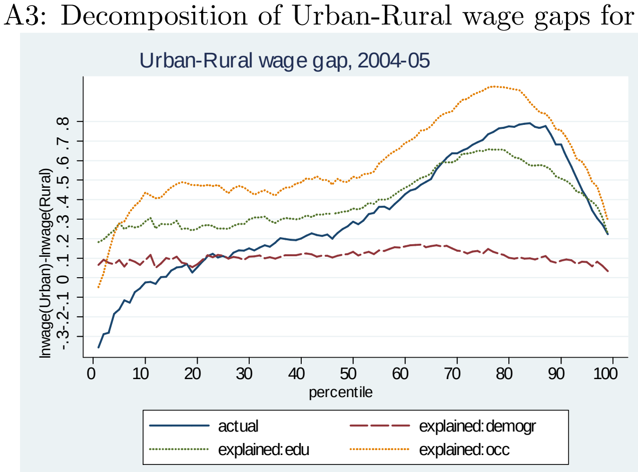


(a) (b)

Notes: Panel (a) shows the estimated kernel densities of log real wages for urban and rural workers, while panel (b) shows the difference in percentiles of log-wages between urban and rural workers plotted against the percentile. The plots are for 2004-05 and 2009-10 NSS rounds.

period of 2004-05. As can be seen from Figure A3 the contribution of various factors in that round resembles closely our findings for 2009-10. Overall these results suggest that urban-rural wage gaps have been steadily declining since 1983.

Figure A3: Decomposition of Urban-Rural wage gaps for 2004-05



Notes: This figure shows the actual log wage gap between urban and rural workers for each percentile in 2009-10, 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").

Next, we study the role played by NREGA in more details. In particular, we want to assess whether any of the rural-urban convergence between 2004-05 and 2009-10 can be attributed to the NREGA reform? To answer this question we turn to a state-level analysis and ask whether there was a break in wage gap dynamics after the introduction of NREGA in 2005. Our strategy is to examine whether the 2009-10 round exhibits a disproportionate change in the size of the gaps relative to the previous five rounds. We proceed by estimating fixed effects regressions on the mean, median, 10th and 90th percentile *state wage gaps*, where in each regression we control for the time-invariant state-level fixed effects. We include a trend variable ("trend") to obtain the estimate of the average change between rounds in the state wage gaps during the sample period. To account for the potential break in the size of the gaps in 2009-10 (the period associated with NREGA), we include a dummy variable "2009-10 dummy" which is equal to one for observations in the 2009-10 round.

The shortcoming of this approach, of course, is that the 2009-10 dummy may be significant for reasons that are unrelated to NREGA. To separate the effects of NREGA from other potential factors, we exploit the cross-state variation in the exposure of states to NREGA. Specifically, we introduce a control for the share of the rural labor force in each state (variable "rural share") and its interaction with the 2009-10 time dummy. Recall that NREGA applies only to rural employment. If NREGA had a significant negative effect on the state wage gaps (i.e., it reduced the wage gap), then the coefficients on the 2009-10 dummy and the interaction between the rural share and the 2009-10 dummy should both be negative. Our sample includes the 17 major states in India. The results are presented in Table A1.

Consistent with our earlier findings for individual data, the median and 10th percentile *state-level* wage gaps have significantly declined during 1983-2010 period. The wage gap for the 90th percentile has widened while the changes in the mean gap were not significant. At the same time, the coefficient on "2009-10 dummy" is not statistically different from zero for all but the 90th percentile where the sign is positive, i.e. the gap at the top end of thee distribution appears to have widened relative to the trend in 2009-10 round. Overall, these results suggest that the decline in the wage gap during 2009-10 period was not different from the earlier rounds for majority of the wage distribution. The coefficients on both the rural share as well as the interaction term between the rural share and the 2009-10 dummy are consistently insignificant (again except for the 90th percentile where it enters with a positive coefficient) indicating that state level exposure to the NREGA program had no significant effect on the state-level wage gaps.[22]

Overall, our results suggest that NREGA had little effect on the wage convergence between the urban and rural sectors.

[22] The results reported here are based on state-level moments. Hence they are not directly comparable to the results in Table 2 which report estimates on moments constructed from the All-India sample.

Table A1: Estimating the effects of NREGA reform on the urban-rural log wage gaps

	mean	median	10th percentile	90th percentile
trend	0.000	-0.014***	-0.009***	0.010***
	(0.002)	(0.003)	(0.002)	(0.003)
2009-10 dummy	0.139	0.008	-0.182	0.784*
	(0.333)	(0.392)	(0.310)	(0.409)
rural share	0.540	-0.486	0.149	1.956*
	(0.941)	(1.110)	(0.877)	(1.159)
rural share x 2009-10 dummy	-0.164	-0.115	0.229	-0.870
	(0.446)	(0.526)	(0.415)	(0.548)
N	80	80	80	80

Note: The table presents the fixed effects regression results of the (log) wage gaps between urban and rural workers on the trend ("round"), dummy for 2009-10 round ("2009-10 dummy"), the share of rural population ("rural share") and the interactive term between rural share and 2009-10 dummy. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

A.4 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 Panel (a) of Table 7. Note, however, that inter-temporal changes in the explained and unexplained components may be due to changes in either the attribute gaps or in the returns to those attributes. Since the unexplained part is typically small in our decompositions, we focus on the inter-temporal decomposition of the explained part, ΔE, in the main text. $\Delta E = \Delta X_2' \tilde{\beta}_2 - \Delta X_1' \tilde{\beta}_1$ can be broken down as

$$\Delta E = \Delta X_2' \left(\tilde{\beta}_2 - \tilde{\beta}_1 \right) + \left(\Delta X_2' - \Delta X_1' \right) \tilde{\beta}_1,$$

where the first term is the unexplained part of ΔE, while the second term is the explained part of ΔE. This decomposition is presented in Panel (b) of Table 7 .

A.5 Distributional effects of migration

Table A2 presents regression results from the RIF regression of (log) wages on a several dummies: a rural non-migrant dummy, and a set of four migration flows dummies between rural and urban areas. RIF regressions are estimated for the 10th and 90th percentile of (log) wages.

Table A2: 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.