Urban Sprawl and Rural Development: Theory and Evidence from India*

Viktoria Hnatkovska† and Amartya Lahiri‡

September 2017

Abstract

We examine the evolution of the fortunes of rural and urban workers in India between 1983 and 2010, a period of rapid growth in India. We find evidence of a significant convergence of education attainments, occupation distribution, and wages of rural workers towards those of urban workers. However, individual worker characteristics account for *at most* 40 percent of the wage convergence. We develop a two-sector model of structural transformation to rationalize the rest of the rural-urban wage convergence in India as the consequence of urbanization through land reclassification induced by productivity growth.

JEL Classification: E2, O1, R2

Keywords: Rural urban disparity, wage gaps, urbanization

*This research was funded by a grant from IGC. We would like to thank, without implicating, seminar participants at UBC, the IGC-India 2012 conference in Delhi, and numerous universities for helpful comments and suggestions. The online appendix to the paper is available at http://faculty.arts.ubc.ca/vhnatkovska/research.htm.

†University of British Columbia. E-mail address: hnatkovs@mail.ubc.ca.

‡University of British Columbia and CAFRAL. E-mail address: amartyalahiri@gmail.com.

1 Introduction

A typical pattern observed in countries as they develop is a contraction in the agricultural sector accompanied by an expansion of the non-agricultural sectors. Since the contracting agricultural sector is primarily rural while the expanding sectors are mostly urban, this structural transformation process has potentially important implications for the evolution of economic inequality within such developing economies. The process induces potentially costly reallocation of workers across sectors and locations. Not surprisingly, in a recent cross-country study on a sample of 65 countries, **?** finds that around 40 percent of the average inequality in consumption is due to urban-rural gaps.

In this paper we examine the consequences of structural transformation for the fortunes of rural and urban workers by focusing on the experience of India between 1983 and 2010. Our paper makes three key contributions. First, we believe this is the first paper that provides a comprehensive empirical documentation of the trends in rural and urban disparities in India since 1983 in education, occupation distributions, and wages, as well as an econometric attribution of the changes in the rural-urban wage gaps to measured and unmeasured factors. Second, we provide a structural explanation for the observed wage convergence using a unique formulation that relies on land reclassification from rural to urban. Third, our results suggest a common driving process behind both structural transformation and changes in rural-urban wage inequality. This latter connection has been largely omitted in the literature.

Two features of India during this period make it a particularly relevant case. First, India has had a very well publicized take-off in macroeconomic growth during this period. As we will show below, this growth take-off has also been accompanied by a structural transformation of the Indian economy along the lines described above. Second, the size of the rural sector in India is huge with upwards of 800 million people still residing in the primarily agrarian rural India in 2011. Hence, the scale of the potential disruption and reallocation unleashed by this process is massive.

Using six rounds of the National Sample Survey (NSS) of households in India between 1983 and 2010, we analyze the evolution over time of education attainment, occupation choices, and labor income of rural and urban workers. Our analysis yields several key results. First, while the educational attainments of both rural and urban individuals have been rising, the gap between them has been shrinking dramatically over time both in terms of years of schooling as well as in the relative distribution of workers in different education categories. In terms of occupations, we show that the share of non-farm jobs (both white- and blue-collar) has expanded dramatically in rural areas, leading to a rural-urban occupation convergence.

Second, there has been a significant decline in labor income differences between rural and urban India with almost all of the measured convergence being due to shrinking wage gaps, both between and within occupations. Specifically, the mean wage premium (in logs) of the urban worker over the rural worker fell from 51% to 27% while the corresponding median wage premium (in logs) declined from 59% to 13% between 1983 and 2010. We also find that the urban wage premia declined for all income groups up to the 75th percentile with the urban wage premium at the bottom end of the wage distribution (till the 15th percentile) having actually turned negative during our sample period. Examining the convergence patterns between rural and urban workers along the entire wage distribution is an important feature of the empirical part of our study.

Third, we show that converging individual characteristics can explain *at most* 40 percent of the observed wage convergence between rural and urban areas. Hence, most of the convergence remains unexplained.

The large unexplained residual wage convergence between urban and rural workers presents a puzzle: what factors could have induced the remaining convergence? One possibility is that the relative supply of urban labor may have increased during this period. For a given relative demand for urban labor, an increase in the relative supply of urban labor would cause the urban wage premium to fall. As we show below, this was indeed the case in India with the share of urban workers in the total workforce in India rising between 1983 and 2010 by 8 percentage points. Amongst others, one way in which urban labor supply could rise is through migration from rural to urban areas. Indeed, most of the literature on development economics tends to focus on the economic forces that drive such migration, a classic example being the Harris-Todaro model. However, for India Hnatkovska and Lahiri (2015) find that net flows of workers from rural to urban areas has remained small and relatively stable since 1983. This makes it unlikely that migration, by itself, could have driven the urban-rural wage convergence in India.

A separate driver of relative urban labor supply growth is reclassification of rural locations into urban areas. As we show below, this force was very strong in India with a number of rural areas getting reclassified as urban due to growth or assimilation into contiguous urban areas. This reclassification process caused previously rural workers to become urban workers without having changed their physical location.[1]

[1] Note that the definition of "rural" and "urban" settlements remains invariant in the dataset. To be precise, in accordance with the Census, NSS Organization of India defines an "urban" area as all places with a Municipality, Corporation or Cantonment and places notified as town area; or all other places which satisfied the following criteria: (i) a minimum population of 5000; (ii) at least 75% of the male working population are non-agriculturists; (iii) a density of population of at least 1000 per sq. mile (390 per sq. km.).

We formalize this alternative channel of relative urban labor supply growth by developing a model of locational reclassification. Our model incorporates two locations, rural and urban, into a standard two-sector, non-homothetic model of structural transformation. Crucially, we allow for rural locations to be reclassified as urban at a cost. We show that our model can jointly generate urban-rural wage convergence, increased urbanization through land reclassification, as well as structural transformation of the economy in response to total factor productivity growth.

Intuitively, under non-homothetic preferences, a rise in agricultural productivity releases labor from agriculture which induces the structural transformation of the economy. This process however also raises the relative attractiveness of urban locations which induces a reclassification of some rural land to urban. This land relassification converts the affected rural workers into m urban workers. The consequent increase in the *relative supply* of urban labor tends to lower the urban-rural wage gap while inducing an expansion in the output share of the non-agricultural sector and a fall in the relative price of the non-agricultural good. Both of these are key features of the Indian data.

We view our model of land reclassification as an independent contribution of the paper since most of the literature on structural transformation and urbanization tends to focus on migration of workers from rural to urban areas. More generally, we believe the model to be illustrative of the mechanisms at play in the data whenever land reclassification plays an important role in increasing the urban labor supply.

Our interest in rural-urban gaps probably is closest in spirit to the work of **?** who examined rural-urban consumption expenditure gaps in 65 countries. Like us, he finds that only a small fraction of the rural-urban inequality can be accounted for by individual characteristics, such as education differences. He attributes the remaining gaps to competitive sorting of workers to rural and urban areas based on their unobserved skills.[2]

Our work is also related to an empirical literature studying rural-urban gaps in different countries (see, for instance, Nguyen, Albrecht, Vroman, and Westbrook (2007) for Vietnam, Wu and Perloff (2005) and Qu and Zhao (2008) for China and others). These papers generally employ household survey data and relate changes in urban-rural inequality to individual and household characteristics. Our study is the first to conduct a similar analysis for India and for multiple years, as well as extend the analysis to consider aggregate factors.

The rest of the paper is organized as follows: the next section presents the data and some

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

motivating statistics. Section 3.1 presents the main results on evolution of the rural-urban gaps as well as the analysis of the extent to which these changes were due to changes in individual characteristics of workers. Section 5 presents our model and examines the role of aggregate shocks in explaining the patterns. The last section contains concluding thoughts.

2 Data

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.[3] These restrictions leave us with, on average, 140,000 to 180,000 individuals per survey round.

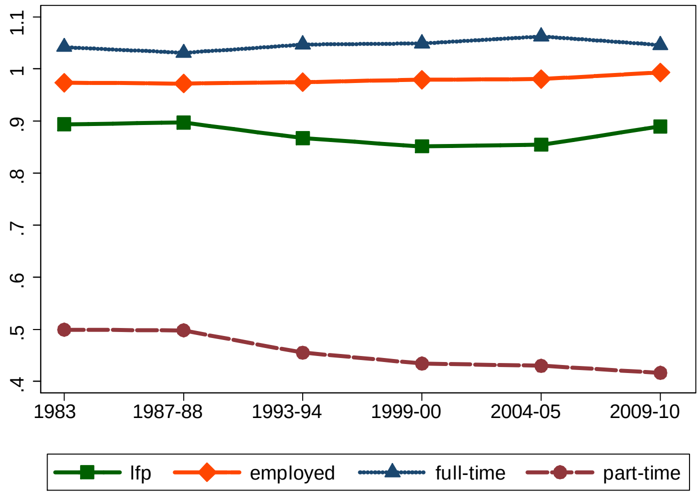
Our focus on full time workers may potentially lead to mistaken inference if there have been significant differential changes in the patterns of part-time work and/or labor force participation patterns in rural and urban areas. To check this, Figure 1 plots the urban to rural ratios in labor force participation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

We summarize demographic characteristics in our sample across the rounds in Table 1. The table breaks down the overall patterns by individuals and households and by rural and urban locations. Clearly, the sample is overwhelmingly rural with about 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).

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

[3] 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.

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 locations, 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 Empirical findings

How did urban and rural workers fare during our sample period? We characterize differences in education attainments, occupations, labor income and wages of rural and urban workforce to answer this question.[4]

3.0.1 Education

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 =

[4]We also consider per capita consumption expenditures, and find that our findings are generally robust.

Table 1: Sample summary statistics

	(a) Individuals			(b) Households		
Urban	age	male	married	proportion	SC/ST	hh size
1983	35.03	0.87	0.78	0.26	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.

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.[5] 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

[5] 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.

also show results from the first approach below.

Table 2 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.

Table 2: Education Gap: Years of Schooling

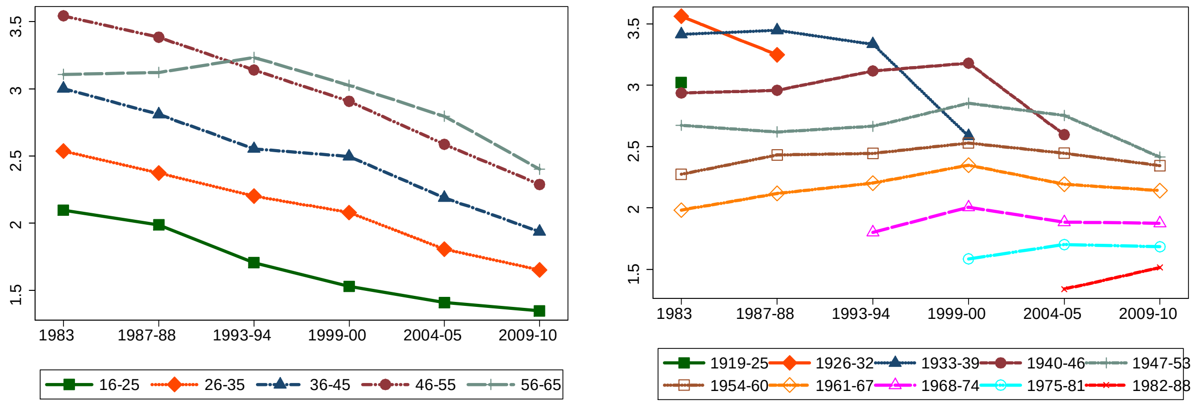
	Average years of education			Relative education gap
	Overall	**Urban**	**Rural**	**Urban/Rural**
1983	3.02	5.83	2.20	2.64***
	(0.01)	(0.03)	(0.01)	(0.02)
1987-88	3.21	6.12	2.43	2.51***
	(0.01)	(0.03)	(0.01)	(0.02)
1993-94	3.86	6.85	2.98	2.30***
	(0.01)	(0.03)	(0.02)	(0.02)
1999-2000	4.36	7.40	3.43	2.16***
	(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 2, 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 2 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 2. 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.

Figure 2: Education gaps by age groups and birth cohorts



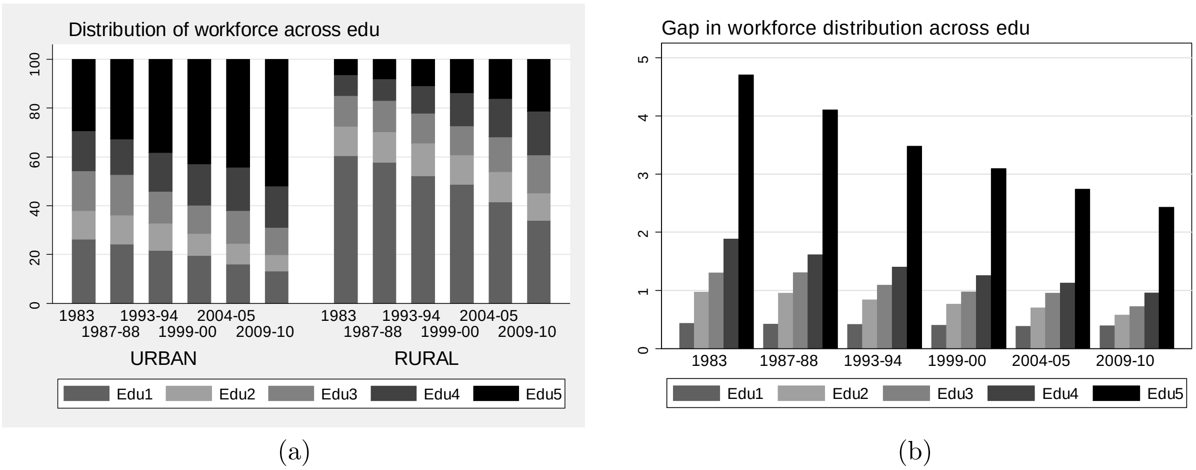
(a) (b)

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.

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 education.

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

Figure 3: Education distribution



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).

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 3 shows the changes in the relative education distributions of the urban and rural workforce. For each survey year, the Figure shows the fraction of urban workers in each education category relative to the fraction of rural workers in that category. Thus, in 1983 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 3 shows the results.

Table 3: 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.

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.

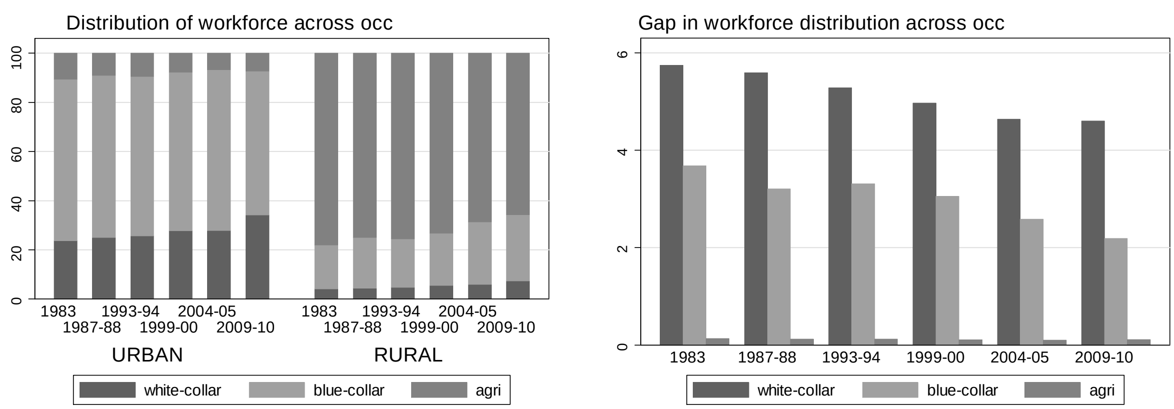
Panel (b) of Table 3 shows that the changes over time in these marginal effects were also significant for all rounds and all categories. The trends though are interesting. There are clearly significant convergent trends for education categories 1, 3 and 4. Category 1, where rural workers were over-represented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

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

3.0.2 Occupation Choices

We now turn to the occupation choices being made by the workforce in urban and rural areas. To examine this issue, we 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. Figure 4 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)).

Figure 4: Occupation distribution



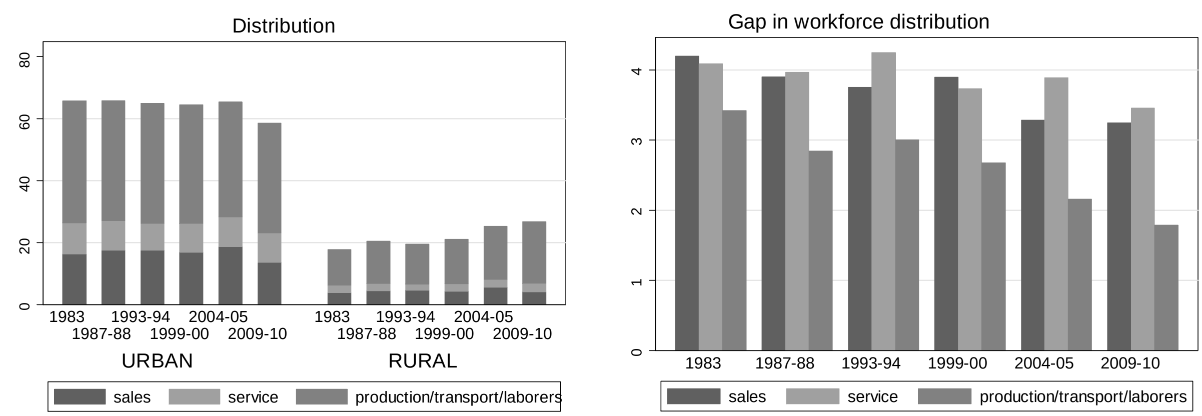
(a) (b)

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

The urban and rural occupation distributions have the obvious feature that urban areas have a much smaller fraction of the workforce in agrarian occupations while rural areas have a minuscule share of people working in white collar jobs. 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 4 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 5 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, quarry men, and various manufacturing workers. The main result that jumps out of panel (a) of Figure 5 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 5: 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 5 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 6 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.

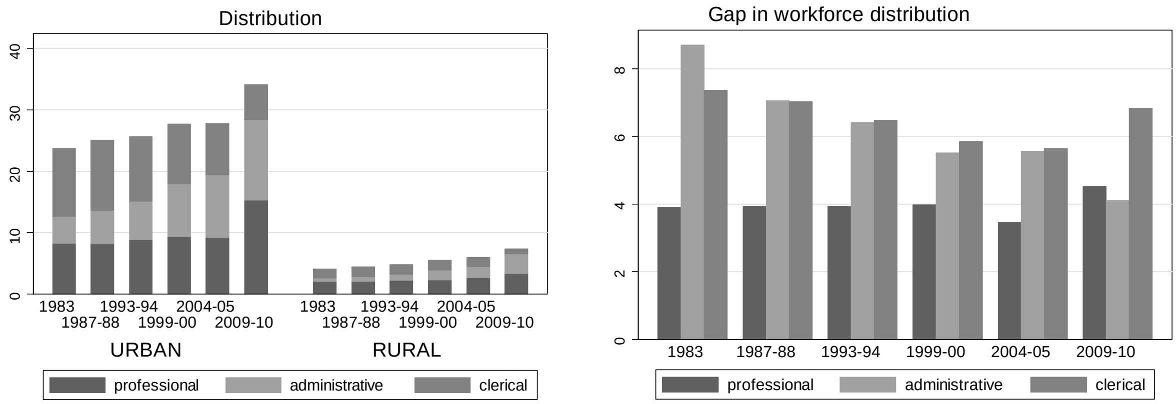
Have the expansions and contractions in various jobs been symmetric across rural and urban sectors? Panel (b) of Figure 6 presents relative gaps in the workforce distribution across various white-collar occupations. The biggest difference in occupation distribution between urban and rural sectors was 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.[6] The gap in professional jobs remained relatively unchanged at 4 during the same period.

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 statis-

[6] 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.

Figure 6: 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.

tically significant? To examine this we estimate a multinomial probit regression of occupation choices on a rural dummy and a constant for each survey round. The results for the marginal effects of the rural dummy are shown in Table 4. 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 occupations. At the same time, the gap in the share of the workforce in white-collar jobs between urban and rural areas has widened. Note that this result is not inconsistent with Figure 4, which indicates convergence in the workforce distribution in white-collar jobs. The key difference is that Table 4 reports *absolute differences* in workforce distribution between rural and urban workforce, while Figure 4 reports *relative differences* in that distribution. At the same time, blue-collar and agrarian jobs have shown convergence over time in both absolute and relative terms.

Table 4: 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.

3.1 Wages

We obtain wages 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 poverty lines.[7]

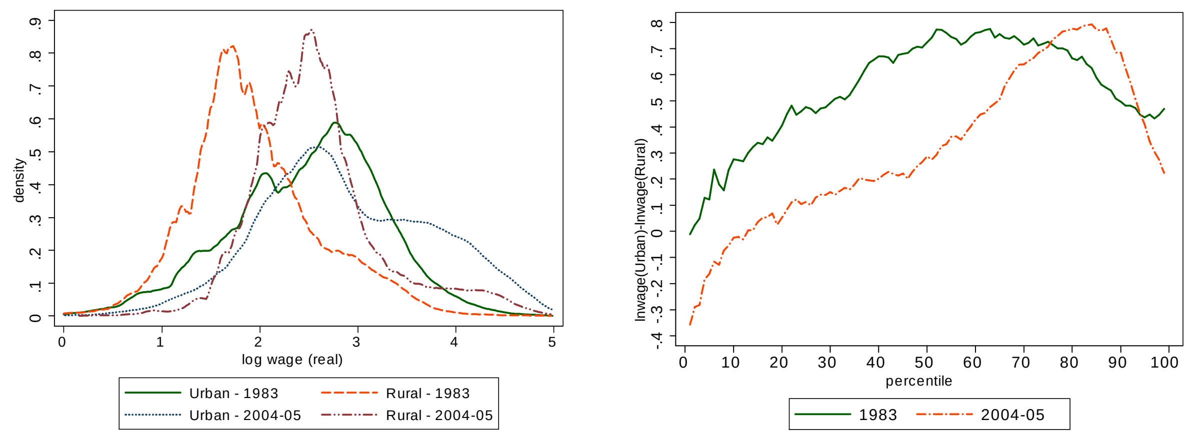
In studying urban-rural real wage convergence we are interested not just in the mean or median wage gaps, but rather in the behavior of the real wage gap across the entire wage distribution. Thus, we start by taking a look at the distribution of log real wages for rural and urban workers in our sample. In order to present the results, we break up our sample into two sub-periods: 1983 to 2004-05 and 2004-05 to 2009-10. We do this to distinguish long run trends since 1983 from the potential effects of The Mahatma Gandhi National Rural Employment Guarantee Act (NREGA) that was introduced 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 split our sample period into the pre- and post-NREGA periods.

We begin with the pre-NREGA period of 1983 to 2004-05. Panel (a) of Figure 7 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2004-05 survey rounds. The plot

[7] 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.

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



(a) wage densities

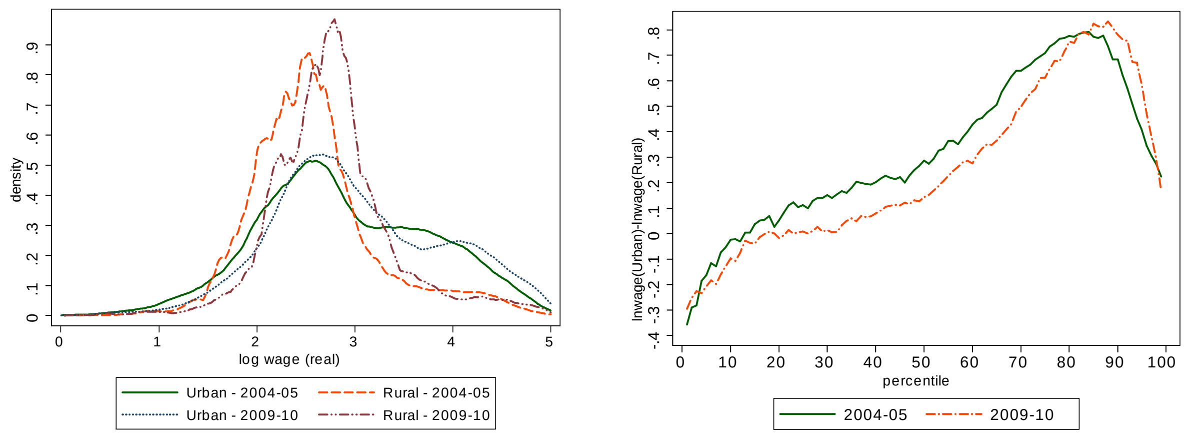
(b) wage gaps

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 7 presents the percentile (log) wage gaps between urban and rural workers for 1983 and 2004-05. 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 higher for richer wage groups. A rightward shift in the schedule over time implies that the wage gap has shrunk. The plot 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. 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. Between the 70th and 90th percentiles however, the wage gaps are larger in 2004-05 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 2004-05 period, as we discussed above. A last noteworthy feature is that in 2004-05, for the bottom 15 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.

Next we turn to the analysis of the post-NREGA wage distributions. Figure 8 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.[8]

Figure 8: The log wage distributions of urban and rural workers for 2004-05 and 2009-10



(a) wage densities

(b) wage gaps

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.

Figures 7 and 8 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.[9]

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

[9] 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

Table 5: 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.

Panel (a) of Table 5 reports the estimated coefficient on the rural dummy for the 10th, 50th and 90th percentiles as well as the mean for different survey rounds.[10] 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 time, for the 90th percentile the wage gap actually increased over time. These results corroborate the visual impression from Figure 7: the wage gap between rural and urban areas fell between 1983 and 2005 for all but the richest wage groups.

3.2 Labor income

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), and agrarian jobs (occ 3):

$$w_t^j = w_{1t}^j L_{1t}^j + w_{2t}^j L_{2t}^j + w_{3t}^j L_{3t}^j, \tag{3.1}$$

where L_{it}^j is employment share of occupation i in location j, and w_{it}^j is average daily real 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

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).

[10] 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.

gap between urban and rural areas can then be expressed as

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

where w_{it} is the economy-wide average daily real wage in occupation $i = 1, 2, 3$. The decomposition above shows 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 as the *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.

The 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 $t-1$:

$$\begin{aligned}&\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} \\ = \;& \Delta\mu_{1t}^U \bar{L}_{1t}^U + \Delta\mu_{2t}^U \bar{L}_{2t}^U + \Delta\mu_{3t}^U \bar{L}_{3t}^U - \Delta\mu_{1t}^R \bar{L}_{1t}^R - \Delta\mu_{2t}^R \bar{L}_{2t}^R - \Delta\mu_{3t}^R \bar{L}_{3t}^R \\ & + \left(\overline{L_{1t}^U - L_{1t}^R}\right)\left[\Delta\eta_{1t} - \Delta\eta_{3t}\right] + \left(\overline{L_{2t}^U - L_{2t}^R}\right)\left[\Delta\eta_{2t} - \Delta\eta_{3t}\right] \\ & + \Delta L_{1t}^U\left(\bar{\mu}_{1t}^U - \bar{\mu}_{3t}^U\right) + \Delta L_{2t}^U\left(\bar{\mu}_{2t}^U - \bar{\mu}_{3t}^U\right) - \Delta L_{1t}^R\left(\bar{\mu}_{1t}^R - \bar{\mu}_{3t}^R\right) - \Delta L_{2t}^R\left(\bar{\mu}_{2t}^R - \bar{\mu}_{3t}^R\right) \\ & + \left(\overline{\eta_{1t} - \eta_{3t}}\right)\Delta\left(L_{1t}^U - L_{1t}^R\right) + \left(\overline{\eta_{2t} - \eta_{3t}}\right)\Delta\left(L_{2t}^U - L_{2t}^R\right) \qquad (3.2)\end{aligned}$$

The derivation of this decomposition is described in the online Appendix to this paper. Here $\mu_{it}^j \equiv \left(w_{it}^j - w_{it}\right)/w_t^R, \eta_{it} \equiv w_{it}/w_t^R, \bar{x}_t = \left(x_t + x_{t-1}\right)/2$, and $\Delta x_t = x_t - x_{t-1}$. This decomposition breaks up the change in labor income gap over time into two basic components: changes in wages and changes in employment. In addition, the wage component is further split up into the within-occupation and between-occupation components. These are, respectively, the first and second rows of equation (3.2).

The first row of equation (3.2) summarizes the change in the labor income gap attributable to changes in rural and urban real wages in each occupation for a given level of employment. 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

(3.2) gives the convergence in labor income due to convergence of wages in different occupations – the between-occupation component. Lastly, rows three and four give 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 6: 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 6 presents the results of the decomposition by occupations and components. During the 1983-2010 period, the aggregate labor income gap between urban and rural areas declined by 22.6 percent. All of this decline is due to convergence of wages with roughly equal contributions of the 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 an additional 0.18 (or 78 percent) decline in the 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.

We now turn to our central goal of uncovering the factors behind converging wage gaps in rural and urban areas. We consider several explanations. First, wage convergence may have arisen due to convergence of individual and household characteristics. Second, aggregate shocks, such as productivity changes, may have played a role. We investigate each of these explanations in turn.

4 Decomposition of wage gaps

How much of the wage convergence documented above is driven by the convergence of measured covariates? Or was it due to changes in unmeasured factors? 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. Additionally, we control for the education level of the individual by including dummies for education categories 1-5.[11] 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.[12]

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 40 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.

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

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

[12] 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.2.

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

(a). Change 1983 to 2009-10	(i) measured gap	(ii) explained	(iii) unexplained	explained (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.

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 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!

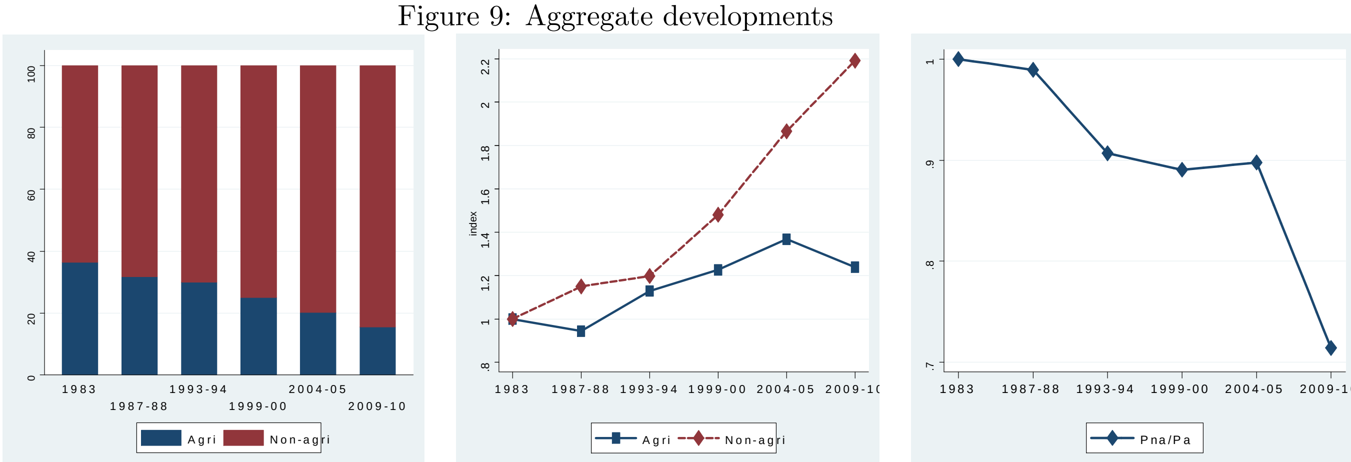
5 An explanation for wage convergence

The empirical results suggest that a majority of the convergence between rural and urban India cannot be accounted for by convergence in the individual characteristics of the two groups. What then explains the convergent trends? One possible explanation is the increase in relative urban labor supply during this period. In the NSS data, the proportion of the urban employment grew from

22 percent to 30 percent of the overall employment between 1983 and 2010. This increase in urban employment also finds an echo in the Census figures for India for the overall population where the urban population share rose from 23 to 31 percent between 1981 and 2011. Clearly, an expansion in relative urban labor supply could reduce the relative urban-rural wage

At this stage it is worthwhile noting that during the period between 1983 and 2010, India also underwent some very stark and rapid macreconomic changes. Specifically, the period was marked by a sharp increase in aggregate productivity growth, as well as a structural transformation of the economy and associated changes in the agricultural terms of trade. Figure 9 shows the basic facts on these variables. Panel (a) shows that the output share of agriculture fell from 36 percent in 1983 to 16 percent in 2010. Panel (b) shows the sectoral TFP growth with agricultural TFP growing by 24 percent between 1983 and 2010 while non-agricultural TFP grew by a remarkable 119 percent during the same period. Lastly, Panel (c) of Figure 9 shows that this period was characterized by a 29 percent decline in the relative price of non-agricultural output. These patterns jointly illustrate the ongoing structural transformation of the economy, its productivity growth and the accompanying improvement in the agricultural terms of trade during this period.[13,14]

Figure 9: Aggregate developments



(a) output shares (b) sectoral TFP (c) sectoral relative price

Notes: Panel (a) of this Figure presents the distribution of output across agricultural and non-agricultural sectors for different NSS rounds. Panel (b) shows sectoral TFP while Panel (c) plots the evolution of the relative price of the non-agricultural good.

It is worth pointing out that the literature on structural transformation typically generates these

[13] Note that below we present aggregate facts for industries rather than occupations. This is innocuous since we will only distinguish between agriculture and non-agriculture based activities, and because the vast majority of agricultural jobs are in the agriculture industry. This guarantees a tight mapping between occupations and industries. We should note that the employment share of agriculture also fell during this perod, but at a slower rate.

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

sectoral dynamics associated with economic development by introducing non-homothetic preferences through features such as minimum consumption requirements of agriculture. An increase in agricultural productivity in such an environment raises the relative demand for the non-agricultural good since the demand for agricultural goods expands less than proportionately due to the non-homotheticity. Note that if urban areas have a comparative advantage in producng the non-agricultural good, the increased relative demand for the non-agricutural good will act like an increase in the demand for resources required for producing the non-agricultural good. This would typically raie the demand for both urban labor and urban land.

In general, urban employment growth can occur due to three factors: natural growth due to fertility and death rate differentials; migration; and agglomeration of rural areas into urban areas. In India, natural growth was, and still is, higher in rural areas. Increasing rural to urban migration could potentially be an important factor underlying changes in the urban-rural wage gap. However, using the NSS surveys, Hnatkovska and Lahiri (2015) find that 5-year net flow of workers from rural to urban areas in India is small and has remained relatively stable at around 1 percent of all full-time employed workforce. They also find that migrants from rural to urban areas do not earn significantly lower wages than their urban non-migrant counterparts. Moreover, the wage differential between rural and urban non-migrant workers has been narrowing at the same rate as the overall wage gap between rural and urban workers. These results suggest that migration, by itself, could not have driven the convergent dynamics between urban and rural areas in India.[15]

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

[15] However, in Hnatkovska and Lahiri (2016) we explore the power of the contrasting strengths of the migration channel in China and India in explaining the different trends in rural-urban wage gaps in the two countries.

declined from 39000/sq. km to 26000/sq. km.[16]

The first two facts indicate that while there was a sharp increase in the number of new towns, the bulk of the increase in the urban population share was concentrated in existing cities, many of which grew rapidly to cross the million person mark. The third fact about the decline in the population density of the larger cities indicates that the growth in these cities was accommodated by an outward expansion wherein they assimilated neighboring/contiguous rural areas into their fold. Hence, urban sprawl was a key factor behind the growth in the urban labor force.[17]

Could these aggregate changes be related to the changing rural-urban gaps? In the remainder of this section we present a simple two-sector, two-location model where productivity shocks alone can jointly cause rising urbanization, structural transformation of the economy, fall in the relative price of non-agricultural output, as well as a declining urban-rural wage gap. Crucially, the model aims to reproduce the Indian pattern of urbanization occurring through reclassification of rural land into urban through urban sprawl into surrounding rural areas with the urban workforce expanding primarily due to the conversion of workers in these reclassified areas.

5.1 The model

Consider a two-good economy with two locations: R and U. The initial distribution of the workforce and land across the two locations are:

$$\rho + \omega = 1 \tag{5.3}$$

$$T_R + T_U = T \tag{5.4}$$

where ρ denotes the measure of workers initially in rural areas while ω is the initial measure of urban workers. The normalization of total labor to unity is with no loss off generality. T_R and T_U are the endowments of land in the two locations. The total land and labor endowments for the economy are assumed to be fixed.

Land can be reclassified across locations subject to a cost of δ per unit of re-classified land. We assume that once a location gets reclassified it also acquires the productivity of the new location. The reclassification cost is a stand-in for all the costs of urban and municipal infrastructure building that is needed by urban areas. Indeed, this infrastructure is also what gives the city its comparative advantage in producing the non-agricultural good. We assume that the re-classification cost is paid

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

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

in units of the final good. Using r to denote reclassification of land, the post-reclassification land endowment distribution is

$$\hat{T}_U = T_U + r$$

$$\hat{T}_R = T_R - r$$

We assume that labor cannot migrate across locations. However, labor on re-classified land gets re-classified as well, in proportion to the labor/land ratio in the reclassified location. Hence,

$$\hat{\rho} = \rho - \frac{\rho}{T_R} r = \rho \left(1 - \frac{r}{T_R}\right)$$

$$\hat{\omega} = \omega + \frac{\rho}{T_R} r = \omega \left(1 + \frac{\rho}{\omega} \frac{r}{T_R}\right),$$

where $\hat{\rho}$ and $\hat{\omega}$ are the measures of workers in rural and urban areas, respectively, post re-classification.

Note that if urban land is reclassified as rural then the labor force adjustment would be given by

$$\hat{\rho} = \rho + \frac{\omega}{T_u} r$$

$$\hat{\omega} = \omega - \frac{\omega}{T_U} r$$

We shall conduct the analysis here under the assumption that it is rural land that gets re-classified as urban and not the other way around.

Let the production technologies in per worker terms be given by the Leontief functions:

$$Y_{iA} = A_i \min[1, xX_i], \qquad i = R, U$$

$$Y_{iN} = N_i \min[1, zZ_i], \qquad i = R, U$$

where Y_{iA} and Y_{iN} are output per worker in location $i = R, U$ in agriculture and non-agriculture, respectively, while x and z are constant production function coefficients that are independent of location. A and N denote total factor productivity in agriculture and non-agriculture, respectively, while X and Z denote land per worker in the two sectors. Notice that this specification makes the technologies for producing the two goods specific to the location since productivity is location specific.

The Leontief specification implies that

$$X_i = \frac{1}{x}, \qquad i = R, U$$

$$Z_i = \frac{1}{z}, \qquad i = R, U$$

Hence, land/labor ratios in both sectors are identical across locations due to the technological specification. We shall assume throughout that $x < z$ so that the land per worker is higher in the agricultural sector.

Let α_i denote the measure of workers in agriculture in location $i = R, U$. Similarly, let η_i be the measure of workers in non-agriculture in location $i = R, U$. The technology implies that output of the two goods in each location is given by

$$Y_{iA} = \alpha_i A_i$$

$$Y_{iN} = \eta_i N_i$$

The total use of land in the two locations cannot exceed their supply. Hence,

$$\frac{\alpha_R}{x} + \frac{\eta_R}{z} \leq T_R - r \tag{5.5}$$

$$\frac{\alpha_U}{x} + \frac{\eta_U}{z} \leq T_U + r \tag{5.6}$$

Moreover, the use of labor in the two locations cannot exceed the total supply of labor in those locations:

$$\alpha_R + \eta_R \leq \left(1 - \frac{r}{T_R}\right)\rho \tag{5.7}$$

$$\alpha_U + \eta_U \leq \left(1 + \frac{\rho}{\omega}\frac{r}{T_R}\right)\omega \tag{5.8}$$

Lastly, the total economy-wide production of the two goods are given by

$$Y_A = \alpha_R A_R + \alpha_U A_U$$

$$Y_N = \eta_R N_R + \eta_U N_U$$

We solve for the optimal allocations of labor and land across the two locations that would be chosen by a planner who wants to maximize the value of final output net of the re-classification cost. Specifically, the planner maximizes

$$(Y_A - a)^{\gamma} Y_N^{1-\gamma} - \delta r.$$

Note that the objective function can be interpreted as a utility function of the representative household (which can pool resources for consumption across its members located in the two locations) where utility is linear in consumption of the final good where the final good is produced using a Cobb-Douglas aggregator over the agricultural and non-agricultural goods:

$$Y = (Y_A - a)^{\gamma} Y_N^{1-\gamma}$$

The term a captures the non-homotheticity due to a minimum consumption requirement of the agricultural good. This is a standard method used to generate the structural transformation.

The endogenous variables for this problem are the four labor allocations and the land reclassification: $(\alpha_R, \alpha_U, \eta_R, \eta_U, r)$. The exogenous variables in our environment are $(A_R, A_U, N_R, N_U, \rho, \omega, T_R, T_U)$ while the model parameters are (γ, δ, x, z). The planner's objective is to choose $\alpha_R, \alpha_U, \eta_R, \eta_U$ and r to maximize

$$W = (\alpha_R A_R + \alpha_U A_U - a)^{\gamma} (\eta_R N_R + \eta_U N_U)^{1-\gamma} - \delta r \tag{5.9}$$

subject to equations (5.3), (5.4), (5.5), (5.6), (5.7) and (5.8).

In the analysis below we shall maintain the following assumptions:

Assumption 1. $x < z$

Assumption 2. $A_R = g\bar{A}_R, A_U = g\bar{A}_U, N_R = g\bar{N}_R, N_U = g\bar{N}_U$

Assumption 3. $\bar{A}_R \geq \bar{A}_U, \quad \bar{N}_R \leq \bar{N}_U$

Assumption 4. $\frac{\rho}{xT_R} \geq 1 \geq \frac{\rho}{zT_R}$

Assumption 1 ensures that the land per unit worker in agriculture is higher than in non-agriculture. Assumption 2 specifies the productivity processes as potentially having a common aggregate component g. Assumption 3 implies that rural agricultural productivity is at least as high as urban agricultural productivity while urban non-agricultural productivity in urban areas is at least as high as non-agricultural productivity in rural areas. These two assumption, we believe, capture the conventional perspective on the geographic distribution of sectoral productivities. Assumption 4 ensures that the non-negativity conditions on sectoral labor allocations are satisfied, i.e., $\alpha_R \geq 0$ and $\eta_R \geq 0$.

Assuming no wastage of land and labor resources, one can reduce this planning problem of maximizing social welfare (given by equation 5.9 above) to just choosing the share of rural land to reclassify, r. Note that we can first use equations (5.7) and (5.8) to eliminate η_R and η_U and then use equations (5.5) and (5.6) to solve for α_R and α_U in terms of r alone. The first order condition for this problem is

$$0 \leq \frac{dW}{dr} = \gamma \left(\frac{Y_A - a}{Y_N}\right)^{\gamma-1} \left(\frac{1}{x} - \frac{1}{z}\right)^{-1} \left(1 - \frac{\rho}{zT_R}\right)(A_U - A_R)$$
$$+ (1-\gamma)\left(\frac{Y_A - a}{Y_N}\right)^{\gamma} \left(\frac{1}{x} - \frac{1}{z}\right)^{-1} \left(\frac{\rho}{xT_R} - 1\right)(N_U - N_R) - \delta$$

Note that along all interior solutions we have $\alpha_R > 0$ and $\eta_R > 0$. It is easy to check that these imply that $1 > \frac{\rho}{zT_R}$ and $\frac{\rho}{xT_R} > 1$ which are satisfied by Assumption 3.[18,19]

The first order condition illustrates the mechanisms at play in the model. Consider, for illustrative purposes, a situation where $A_U = A_R$, i.e., there are no differences in agricultural productivity across locations. The cost of reclassifying an additional unit of rural land is δ. The gain from this reclassification is the increase in the final output due to the higher productivity of urban land in non-agricultural use, $N_U > N_R$. However, this gain is proportional to the additional final consumption good produced by the incremental unit of the non-agricultural good (or, the marginal product of non-agricultural output). Since this marginal product is decreasing in the amount of the non-agricultural good, the solution is typically interior. For a given productivity level„ a fall in the reclassification cost would induce an increase in the amount of rural land reclassified in order to balance the gains with the lower cost. Similarly, an increase in urban productivity of the non-agricultural sector N_U would induce an increase in the proportion of rural land that is reclassified, i.e., an increase in r until the additional gains were exactly offset by the cost.

5.2 Effects of TFP growth

We want to examine the effect of productivity growth in our model economy. We shall focus on TFP growth as captured by increases in the parameter g (see Assumption 2 above). Our first focus of interest is on the effect of productivity growth on the land and labor distribution across rural and

[18] Note that if $\frac{dW}{dr} < 0$ then the incentive is for urban land to be re-classified as rural though the labor constraints in that case would be different. Hence, one would have to solve a different problem. For our exercise here we shall restrict parameter ranges such that $\frac{dW}{dr}\Big|_{r=0} > 0$. This is sufficient for $r > 0$ in equilibrium.

[19] It is straightforward to verify that $\frac{d\ln\left(\frac{Y_A - a}{Y_N}\right)}{dr} = \left(\frac{1}{Y_A - a}\right)\left(\frac{\frac{\rho}{zT_R} - 1}{\frac{1}{x} - \frac{1}{z}}\right)(A_R - A_U) + \frac{1}{Y_N}\left(\frac{1 - \frac{\rho}{xT_R}}{\frac{1}{x} - \frac{1}{z}}\right)(N_U - N_R) < 0$ under assumptions 1, 3 and 4. This implies that $\frac{d^2W}{dr^2} < 0$. Hence, the second order condition for a maximum is satisfied.

urban locations in this economy. The following proposition summarizes the key result:

Proposition 5.1 *Under Assumptions 1-4, an increase in total factor productivity g induces rising urbanization with an increase in the share of both urban land and urban labor.*

Proof. Using the productivity specification in Assumption 2, we can write

$$\frac{Y_A - a}{Y_N} = \frac{\alpha_R \bar{A}_R + \alpha_U \bar{A}_U - \frac{a}{g}}{\eta_R \bar{N}_R + \eta_U \bar{N}_U}$$

Since $\alpha_R, \alpha_U, \eta_R$ and η_U are not direct functions of the TFP parameter g (see equations (5.5)-(5.8)), it is easy to see that $\frac{\partial\left(\frac{Y_A - a}{Y_N}\right)}{\partial g} > 0$. Hence,$\frac{\partial^2 W}{\partial r \partial g} > 0$. Since the second order condition holds, i.e., $\frac{\partial^2 W}{\partial r^2} < 0$, the optimal r must rise. Consequently, the urban population share must also rise since $\hat{\omega} = \omega\left(1 + \frac{\rho}{\omega}\frac{r}{T_R}\right)$. ■

Intuitively, an increase in productivity raises income which raises the relative marginal utility value of non-agricultural goods due to the non-homotheticity in agricultural goods demand. Since urban land is more productive for non-agricultural goods, this causes an increased demand for urban land and thus an increase in the share of rural land that is reclassified as urban. Note that in our specification, the cost of land reclassification, δ, is assumed to be independent of the productivity parameter g. If the reclassification cost also increased with g then the optimal r would rise by less than in our benchmark case.

Having established that a rise in productivity induces an increase in urbanization, we now switch attention to its effect on the urban-rural wage gap and the relative price of non-agriculture. To get the equivalent of the relative price of non-agriculture to agriculture from the model we use

$$p = \frac{(1-\gamma)}{\gamma}\frac{(Y_A - a)}{Y_N} \tag{5.10}$$

which, in the decentralized case, would correspond to equating the relative price to the marginal rate of substitution or, $p = \frac{MU(c_N)}{MU(c_A)}$.

It is clear from equation (5.10) that the effect on the relative price of non-agriculture depends on the response of $\frac{Y_A - a}{Y_N}$ to the increase in g. The direct effect of the rise in g is to raise $\frac{Y_A - a}{Y_N}$ due to the non-homothetic parameter a which is independent of g. However, as we saw in Proposition 5.1, the rise in g also induces a rise in urbanization r. The rise in r, in turn, reduces $\frac{Y_A - a}{Y_N}$. The equilibrium effect on p depends on the net impact of these two offsetting effects.

In terms of the urban-rural wage gaps, the sectoral wages in a decentralized version of the model would be equal to the respective value marginal products of labor:

$$\begin{aligned} w_{AR} &= A_R \\ w_{AU} &= A_U \\ w_{NR} &= \frac{(1-\gamma)}{\gamma}\left(\frac{Y_A - a}{Y_N}\right) N_R \\ w_{NU} &= \frac{(1-\gamma)}{\gamma}\left(\frac{Y_A - a}{Y_N}\right) N_U \end{aligned}$$

In deriving the above we have used the expression for the relative price of the non-agricultural good from equation (5.10) above.

The urban to rural wage ratio is the employment share-weighted average of the sectoral wages in the two locations: $\frac{w_U}{w_R} = \frac{\frac{\alpha_U}{\hat{\omega}} A_U + \frac{\eta_U}{\hat{\omega}} N_U}{\frac{\alpha_R}{\hat{\rho}} A_R + \frac{\eta_R}{\hat{\rho}} N_R}$. This can be written as

$$\frac{w_U}{w_R} = \frac{\rho\left(1 - \frac{r}{T_R}\right)}{\omega\left(1 + \frac{r}{T_R}\frac{\rho}{\omega}\right)} \left[\frac{\alpha_U A_U + \eta_U \frac{(1-\gamma)}{\gamma}\left(\frac{Y_A - a}{Y_N}\right) N_U}{\alpha_R A_R + \eta_R \frac{(1-\gamma)}{\gamma}\left(\frac{Y_A - a}{Y_N}\right) N_R}\right]. \tag{5.11}$$

Equation (5.11) makes clear that the effect of a rise in productivity on the urban-rural wage gap is also ambiguous. The increase in the urban labor share reduces the urban-rural wage gap. However, the relative price effect has an independent effect on the wage gap. Given the ambiguous relative price effect, the effect of productivity increases on the urban-rural wage gap also remains ambiguous.

Could a rise in productivity jointly induce a rise in urbanization and structural transformation of the economy, a fall in the urban-rural wage gap and a fall in the relative price of non-agricultural goods? Given the theoretical ambiguity noted above, we illustrate this possibility by simulating the model. The parameter configuration underlying the simulated model is given in the Table 8 below.

Table 8: Model parameters

	Parameter	Value
Agricultural output share	γ	0.5
Per unit land reclassification cost	δ	0.05
Worker per unit land in agri	x	0.5
Worker per unit land in non-agri	z	50
Initial measure of rural workers	ρ	0.7
Initial measure of urban workers	ω	0.3
Initial measure of rural land	T_R	0.6
Initial measure of rural land	T_U	0.4
Minimum consumption of agri	a	0.08

While this is not a calibration exercise, we have chosen most of the parameters to reflect some well known features of the Indian economy. Thus, the agricultural output share γ is set to 0.5 which reflects the high agricultural output share in 1983 in India. Similarly, the parameters x and z are set to reflect the fact that agriculture is primarily rural and rural population density is much smaller than urban population density.[20] The parameters ρ and ω which give the initial rural and urban population shares are picked to match the 70 percent share of the rural population. We also assume that the 60 percent of India's land is rural which corresponds to the share of agricultural land in India as reported by the World Bank.[21] We should reiterate though the simulations below are intended solely to illustrate the price effects of productivity changes in the model.

For our simulation exercise we solve the model under the baseline parameterization reported above in Table 8 and an initial sector-location productivity configuration given in Table 9 below. The first four rows of the Table give the baseline levels of sectoral productivities in each location. We then successively solve the model 65 times by increasing all the productivities by a common proportional factor each time. We set the size of the proportional TFP increase at each step to $g = 0.002$, as shown in the last row of Table 9

Table 9: Productivity parameterization

	Parameter	Value
Rural agricultural productivity	A_R	0.29
Urban agricultural productivity	A_U	1.145
Rural non-agri productivity	N_R	2.29
Urban non-agri productivity	N_U	4.58
TFP step increase	g	0.002

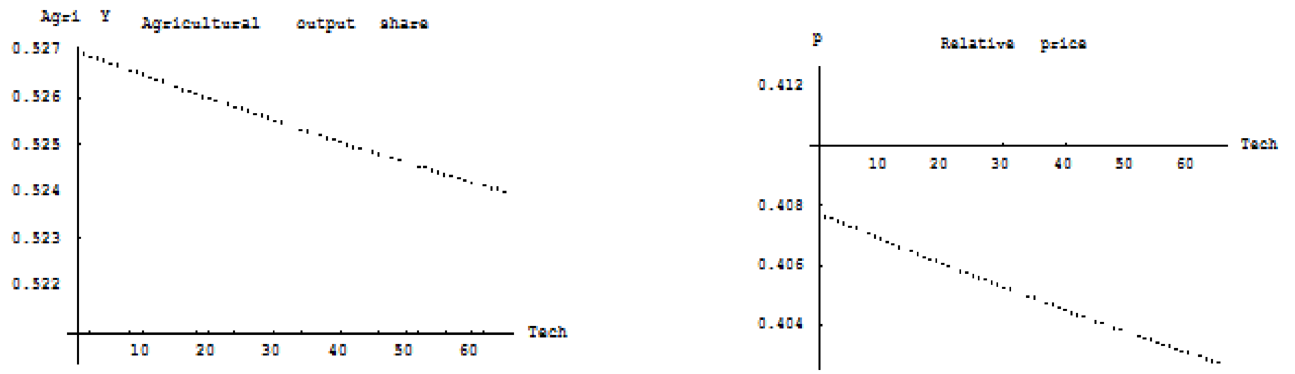
Figure 10 shows the effect of total factor productivity (TFP) shocks on the agricultural share of output and the relative price of the non-agricultural good p. As TFP rises, the economy becomes less agricultural while the agricultural terms of trade improves, i.e., p declines. This is precisely the pattern we saw in India between 1983 and 2010.

Figure 11 depicts the effect of improvements in TFP on the degree of urbanization and the urban-rural wage gap. Clearly, as productivity rises, the economy becomes urbanized while the urban-rural wage gap declines. This too is precisely what one saw in India between 1983 and 2010.

[20] As an example, in 2011 the population density in the state of Delhi (which is almost completely urban) was 11,297 people per squatre kilometer while the population density in the mostly rural state of Arunachal Pradesh was 17. Our parameters imply a 100-fold higher population density in urban areas.

[21] For a country-wise breakdown of the share of agricultural land in the total land area, see http://data.worldbank.org/indicator/AG.LND.AGRI.ZS

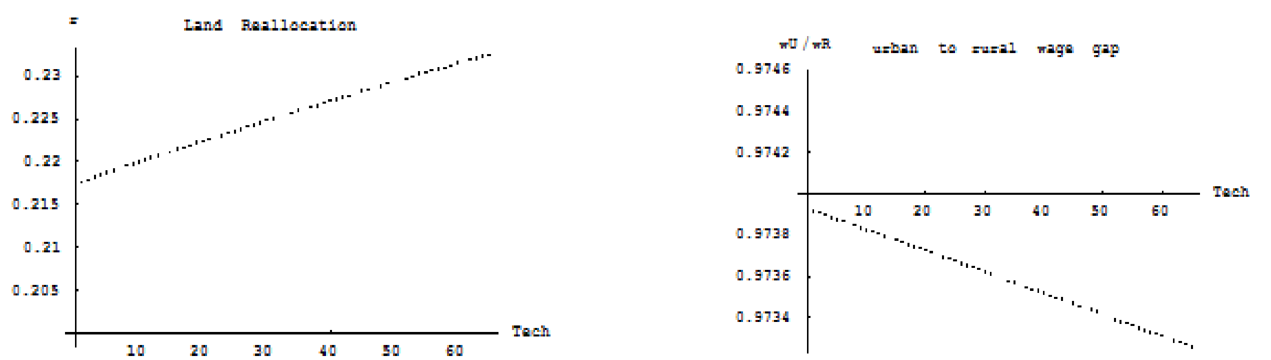
Figure 10: Sectoral transformation and technological progress



(a) Agricultural output share (b) Relative price of non-agriculture

Notes: Panel (a) shows the output share of agriculture in response to TFP shocks, while panel (b) shows the response of the relative price of the non-agricultural good to TFP shocks.

Figure 11: Urbanization and wage gaps under technological progress



(a) Urbanization (b) Urban-rural wage gap

Notes: Panel (a) shows the degree of urbanization in response to TFP shocks, while panel (b) shows the response of the urban-rural wage gap in response to TFP shocks.

These results clearly illustrate that productivity growth in our model can jointly generate structural transformation of the economy, an improvement in the agricultural terms of trade and a decline in the urban-rural wage gap. In standard models of structural transformation with non-homothetic preferences, TFP increases induce a fall in the relative price of agriculture due to the decline in the relative demand for it. The key difference in our model is the increase in the relative supply of urban labor due in response to the shock. The resultant supply response of the non-agricultural good causes the rise in the relative price of agriculture as also the declining urban-rural wage gap.

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 with the majority of this convergence being accounted for by a decline in the wage gap between urban and rural areas. The median urban wage premium in India declined from 59 percent in 1983 to 13 percent by 2010 while the mean wage gap fell from 51 percent to 27 percent. Importantly, we found that a majority of this fall in the urban-rural wage gap cannot be unexplained by standard individual worker and demographic covariates of wages.

Given the lack of explanatory power of conventional worker characteristics, we then examined the possible role of aggregate shocks to the Indian economy during this period. Using a two-factor, two-sector model we showed how aggregate productivity shocks can generate rising urbanization due to land reclassification, a structural transformation of the economy from rural-agrarian towards an urban-non-agrarian economy, and declining urban-rural wage gaps. This depiction fits the dynamic evolution of the Indian economy between 1983 and 2010 quite well.

Our results highlight the key role of faster urban labor force growth in accounting for the wage convergence between urban and rural areas. Given that a large part of the faster growth in the urban labor force during this period was driven by urban agglomeration, we interpret our findings as indicating the importance of forming a better understanding of the process of urban sprawl in developing countries. The importance of urban agglomeration in accounting for urban labor force growth is not unique to India. Lucas (1998) and UN-HABITAT (2012) document precisely this trend in a number of developing countries ranging from Latin America to Asia. Hence, we believe that the relevance of our results extends beyond just the Indian experience since 1983. Indeed, in Hnatkovska and Lahiri (2016) we explore the consequences of differences in the costs to urbanize (including the costs of migration) between China and India for understanding the differences in the evolution of the urban-rural wage gaps in the two countries.

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

References

FIRPO, S., N. M. FORTIN, AND T. LEMIEUX (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953–973.

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

GOLLIN, D., D. LAGAKOS, AND M. WAUGH (2012): "The Agricultural Productivity Gap in Developing Countries," Working paper, Arizona State University.

HNATKOVSKA, V., AND A. LAHIRI (2015): "Rural and Urban Migrants in India: 1983-2008," *World Bank Economic Review*, 29, S257–S270.

——— (2016): "Urbanization, Structural Transformation and Rural-Urban Disparities," Working paper, University of British Columbia.

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

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

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

NGUYEN, B. T., J. W. ALBRECHT, S. B. VROMAN, AND M. D. WESTBROOK (2007): "A quantile regression decomposition of urban-rural inequality in Vietnam," *Journal of Development Economics*, 83(2), 466–490.

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

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

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

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 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:

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

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 .