IGC International Growth Centre

The Rural-Urban Divide in India

Viktoria Hnatkovska[a] and **Amartya Lahiri**[b]

[a] University of British Columbia (contact: hnatkovs@mail.ubc.ca) and [b] University of British Columbia (contact: amartya.lahiri@ubc.ca)

Working Paper
February 2013

International Growth Centre
London School of Economics and Political Science
4th Floor,
32 Lincoln's Inn Fields
Houghton Street
London WC2A 2AE
United Kingdom

For media or communications enquiries, please contact Mazida Khatun
mazida.khatun@theigc.org

Directed and Organised by
LSE
University of Oxford

www.theigc.org

The Rural-Urban Divide in India*

Viktoria Hnatkovska† and Amartya Lahiri‡

August 2012

Abstract

We examine the gaps between rural and urban India in terms of the education attainment, occupation choices, consumption and wages. The study covers the period 1983-2010 and uses household survey data from successive rounds of the National Sample Survey. We find a significant narrowing of the differences in education, occupation distribution, and wages between individuals in rural India and their urban counterparts. However, individual characteristics do not appear to account for much of this convergence. We also examine the effects of the targeted rural employment program NREGA that was introduced in 2005. We find that NREGA's effect on the rural-urban wage and consumption gaps have been negligible. Migration did not play an important role either. These results suggest that the astounding urban-rural convergence in India remains a puzzle.

JEL Classification: J6, R2

Keywords: Rural urban disparity, education gaps, wage gaps

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

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

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

1 Introduction

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

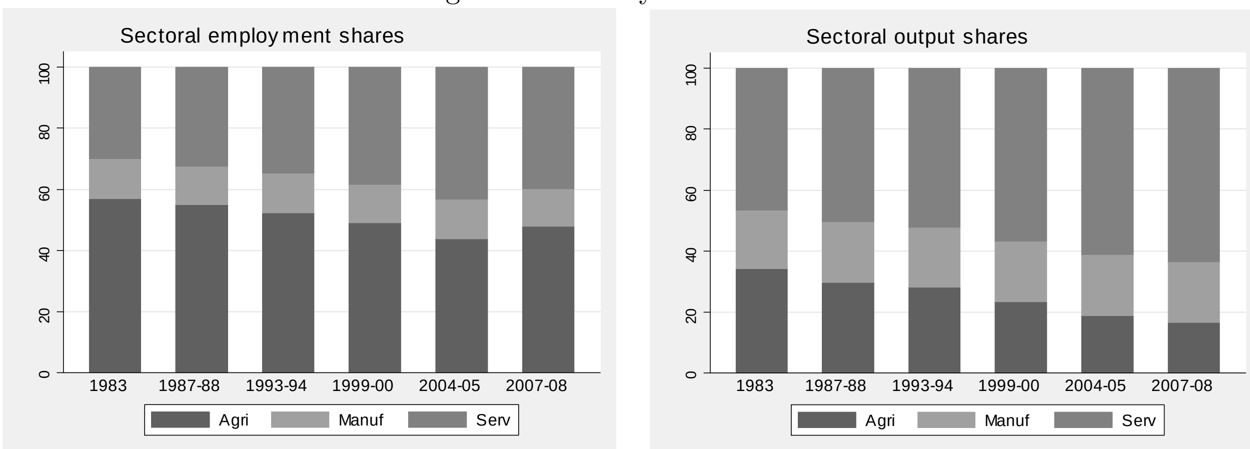
A parallel literature has addressed the issue of the redistributionary effects associated with these structural transformations, both in terms of theory and data. The main focus of this research is on the relationship between development and inequality.[1] This work is related to the issue of rural-urban dynamics since the process of structural transformation implies contracting and expanding sectors which, in turn, implies a reallocation and, possibly, re-training of the workforce. The capacity of institutions in such transforming economies to cope with these demands is a fundamental factor that determines how smooth or disruptive this process is. Clearly, the greater the disruption, the more the likelihood of income redistributions through unemployment and wage losses due to incompatible skills.

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

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

output and employment shares. The big expansion has occurred in the service sector. The industrial sector has also expanded but at a far lower pace. These patterns of structural transformation since 1983 are shown in Figure 1.

Figure 1: Industry distribution



(a) (b)

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

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

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

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

We also examine the potential effect of an important rural employment program introduced in 2006 called National Rural Employment Guarantee Act (NREGA) on the rural-urban wage and consumption gaps. In order to examine the effect of the program we use a state level analysis. Our results indicate that the state-level wage and consumption gaps between rural and urban areas did not change disproportionately in the 2009-10 survey round, relative to their trend during the entire period 1983-2010. We also find that states that were more rural, and hence more exposed to the policy, did not exhibit differential responses of the percentile gaps in wages and consumption in 2009-10, relative to trend. We conclude that the effect of this program on the gaps was, at best, very muted.

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

Our broad conclusion from these results is that the incentives generated by the institutional structure of the country have provided useful signals to the workforce in guiding their choices during this period. As a result, there has been significant churning at the micro levels of the economy. Some of the resulting changes have been truly striking with the median wage premium of urban workers relative to rural workers having declined from 101 percent in 1983 to just 11 percent in 2009-10. This is a welcome sign. Moreover, these results for India stand in sharp relief to the experience of China where Qu and Zhao (2008) report that rural-urban consumption and income gaps actually widened between 1988 and 2002.

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

The rest of the paper is organized as follows: the next section presents the data and some sample statistics. Section 3 presents the main results on changes in the rural-urban gaps as well as the analysis of the rural employment guarantee reform introduced in India in 2005. 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.[2] These restrictions leave us with, on average, 140,000 to 180,000 individuals per survey round.

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

The sample statistics across the rounds are given in Table 1. The table breaks down the overall patterns by individuals and households and by rural and urban locations. Clearly, the sample is overwhelmingly rural with about 73 percent of households on average being resident in rural areas. Rural residents are sightly less likely to be male, more likely to be married, and belong to larger households than their urban counterparts. Lastly, rural areas have more members of backward castes as measured by the proportion of scheduled castes and tribes (SC/STs).

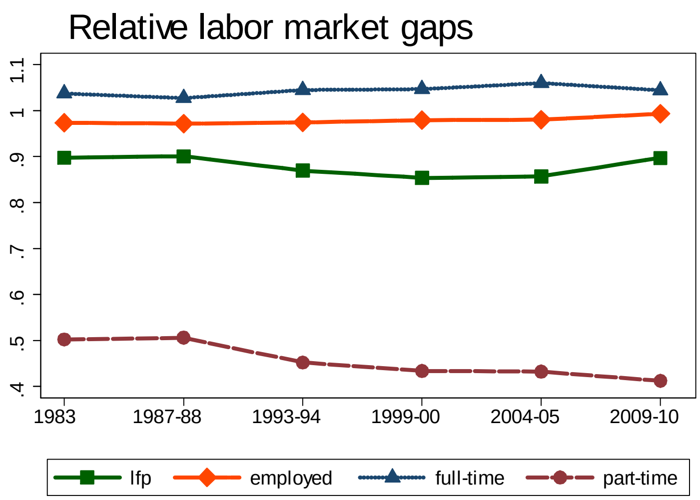
Table 1: Sample summary statistics

	(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.

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 2 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.

Figure 2: 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.

3 Rural-Urban Gaps

We now turn to our central goal of uncovering the gaps in the characteristics of the workforce between rural and urban areas. There are four indicators of primary interest: education attainments levels of the workforce, the occupation distribution of the workforce, the wage levels of workers and their consumption levels. We investigate each of them in turn.

3.1 Education

Our first indicator of interest is the education levels of the rural and urban workforce. Education in the NSS data is presented as a category variable with the survey listing the highest education attainment level in terms of categories such as primary, middle etc. In order to ease the presentation we proceed in two ways. First, we construct a variable for the years of education. We do so by assigning years of education to each category based on a simple mapping: not-literate = 0 years; literate but below primary = 2 years; primary = 5 years; middle = 8 years; secondary and higher secondary = 10 years; graduate = 15 years; post-graduate = 17 years. Diplomas are treated similarly depending on the specifics of the attainment level.[3] Second, we use the reported education categories but aggregate them into five broad groups: 1 for illiterates, 2 for some but below primary school, 3 for primary school, 4 for middle, and 5 for secondary and above. The results from the two approaches are similar. While we use the second method for our econometric specifications since these are the actually reported data (as opposed to the years series that was constructed by us), we also show results from the first approach below.

Table 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.

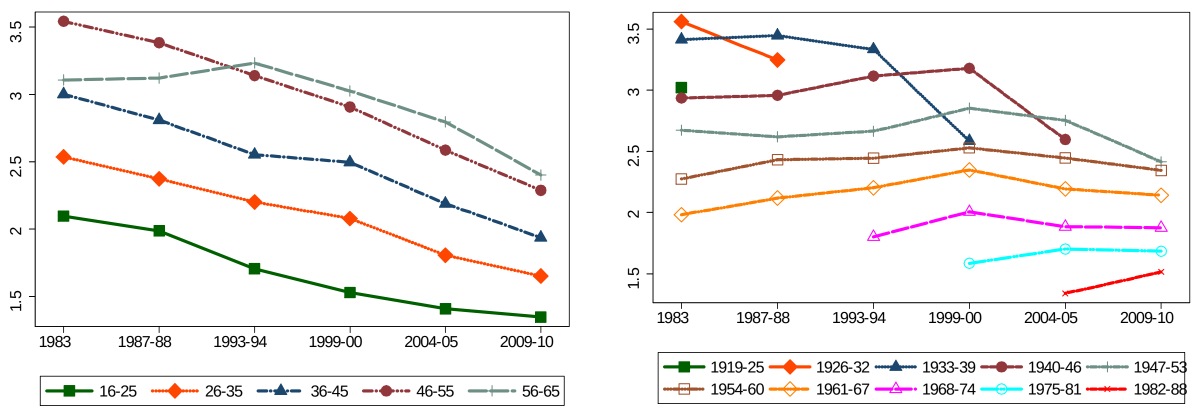
Table 2, while revealing an improving trend for the average worker, nevertheless masks potentially important underlying heterogeneity in education attainment by cohort, i.e., variation by the age of

[3] 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.

the respondent. Panel (a) of Figure 3 shows the relative gap in years of education between the typical urban and rural worker by age group. There are two key results to note from panel (a): (i) the gaps have been getting smaller over time for all age groups; (ii) the gaps are smaller for the younger age groups.

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

Figure 3: Education gaps by age groups and birth cohorts

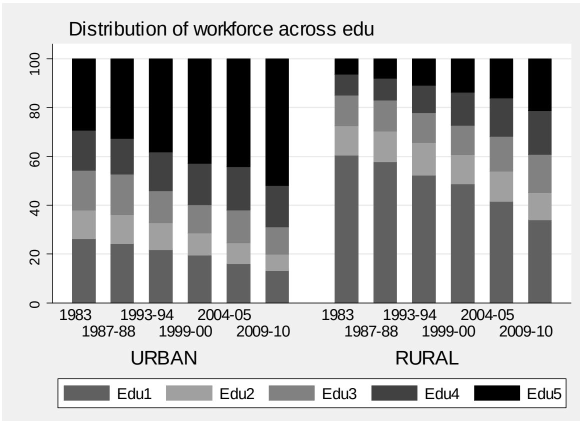


(a) (b)

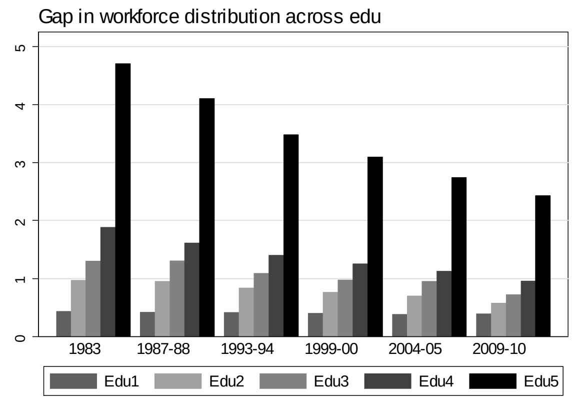
Notes: The figures show the relative gap in the average years of education between the urban and rural workforce over time for different for different age groups and birth cohorts.

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.

Figure 4: Education distribution



(a)



(b)

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

Panel (a) of Figure 4 shows the distribution of the urban and rural workforce by education category. Recall that education categories 1, 2 and 3 are "illiterate", "some but below primary education" and "primary", respectively. Hence in 1983, 55 percent of the urban labor force and over 85 percent of the rural labor force had primary or below education, reflecting the abysmal delivery of public services in education in the first 35 years of post-independence India. By 2010, the primary and below category had come down to 30 percent for urban workers and 60 percent for rural workers. Simultaneously, the other notable trend during this period is the perceptible increase in the secondary and above category for workers in both sectors. For the urban sector, this category expanded from about 30 percent in 1983 to over 50 percent in 2010. Correspondingly, the share of the secondary and higher educated rural worker rose from just around 5 percent of the rural workforce in 1983 to above 20 percent in 2010. This, along with the decline in the proportion of rural illiterate workers from 60 percent to around 30 percent, represent the sharpest and most promising changes in the past 27 years.

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

While the visual impressions suggest convergence in education, are these trends statistically significant? We turn to this issue next by estimating ordered multinomial probit regressions of education categories 1 to 5 on a constant and the rural dummy. The aim is to ascertain the significance of the difference between rural and urban areas in the probability of a worker belonging to each category as well as the significance of changes over time in these differences. Table 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 4. The figure shows trends in the relative gaps while the probit regressions show trends in the absolute gaps.

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

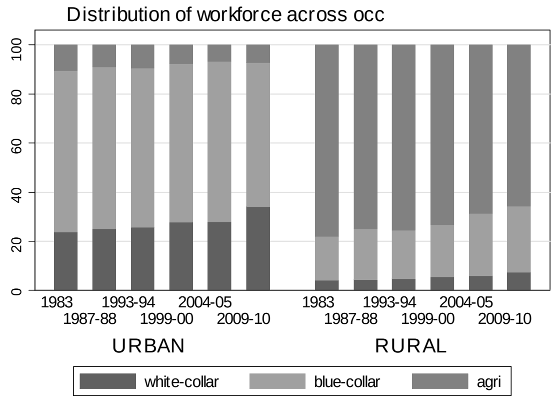
3.2 Occupation Choices

We now turn to our second measure of interest: the occupation choices being made by the workforce in urban and rural areas. Our interest lies in determining whether the occupation choices being made in the two sectors are showing some signs of convergence. Clearly, there are some fundamental differences in the sectoral compositions of rural and urban areas making it unlikely/impossible for the occupation distributions to converge. However, the country as a whole has been undergoing a structural transformation with an increasing share of output accruing to services with a corresponding decline in the output share of agriculture. Are these trends translating into symmetric changes in the rural and urban occupation distributions? Or, is the expansion of the non-agricultural sector (broadly defined) restricted to urban areas only?

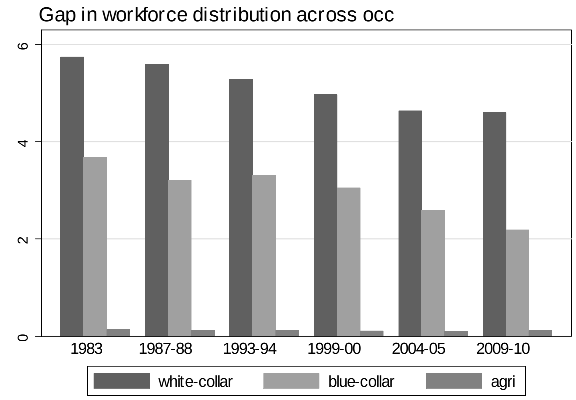
To examine this issue, we aggregate the reported 3-digit occupation categories in the survey into three broad occupation categories: white-collar occupations like administrators, executives, managers, professionals, technical and clerical workers; blue-collar occupations such as sales workers, service workers and production workers; agricultural occupations collecting farmers, fishermen, loggers, hunters etc.. Figure 5 shows the distribution of these occupations in urban and rural India across the survey rounds (Panel (a)) as well as the gap in these distributions between the sectors (Panel (b)).

The urban and rural occupation distributions have the obvious feature that urban areas have a much smaller fraction of the workforce in agrarian occupations while rural areas have a minuscule 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

Figure 5: Occupation distribution



(a)



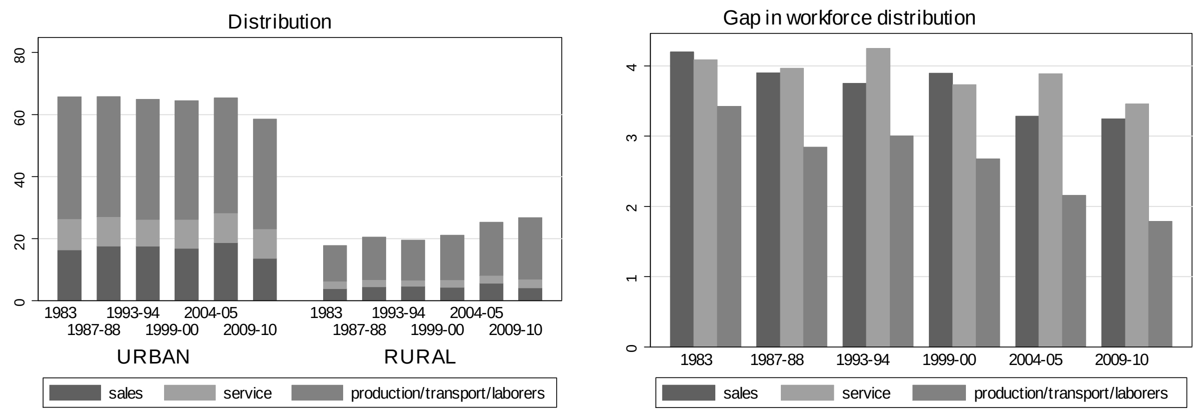
(b)

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

dominance of these occupations. Importantly though, the share of blue-collar jobs has been rising in rural areas. In fact, as Panel (b) of Figure 5 shows, the share of both white collar and blue collar jobs in rural areas are rising faster than their corresponding shares in urban areas.

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

Figure 6: Occupation distribution within blue-collar jobs



(a) (b)

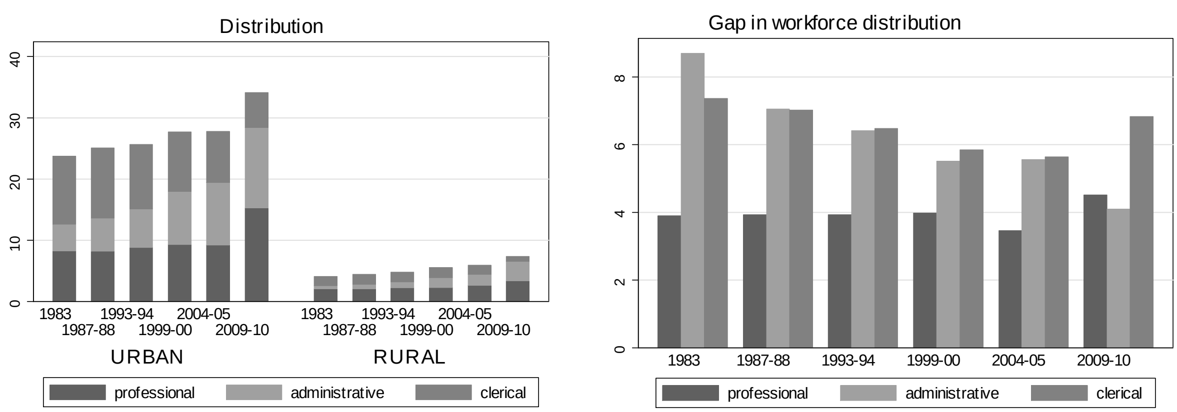
Notes: Panel (a) of this figure presents the distribution of workforce within blue-collar jobs for different NSS rounds. The left set of bars refers to urban workers, while the right set is for rural workers. Panel (b) presents relative gaps in the distribution of urban relative to rural workers across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

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

Next, we turn to white-collar jobs. Panel (a) of Figure 7 presents the distribution of all white-collar jobs in each sector into three types of occupations. The first is *professional, technical and related workers.* This group includes, for instance, chemists, engineers, agronomists, doctors and veterinarians, accountants, lawyers and teachers. The second is *administrative, executive and managerial workers*, which include, for example, officials at various levels of the government, as well as proprietors, directors and managers in various business and financial institutions. The third type of occupations consists of *clerical and related workers.* These are, for instance, village officials, book keepers, cashiers, various clerks, transport conductors and supervisors, mail distributors and communications operators. The figure shows that administrative jobs is the fastest growing occupation within the white-collar group in both rural and urban areas. It was the smallest category among all white-collar jobs in both sectors in 1983, but has expanded dramatically ever since to overtake clerical jobs as the second most popular occupation among white-collar jobs after professional occupations. Lastly, the share of professional jobs has also increased while the share of clerical and related jobs has shrunk in both the rural and urban sectors during the same time.

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

Figure 7: Occupation distribution within white-collar jobs



(a) (b)

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

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

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

[4]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.

the probability of being in white-collar and blue-collar jobs, while having significantly positive effects on the probability of being in agrarian jobs. However, as Panel (b) of the Table indicates, between 1983 and 2010 the negative effect of the rural dummy in blue-collar occupations has declined (the marginal effect has become less negative) while the positive effect on being in agrarian occupations has become smaller, with both changes being highly significant. Since there was an initial under-representation of blue-collar occupations and over-representation of agrarian occupations in rural sector, these results as indicate an ongoing process of convergence across rural and urban areas in these two occupation. At the same time, the gap in the share of the workforce in white-collar jobs between urban and rural areas has widened.

Table 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.

3.3 Wages

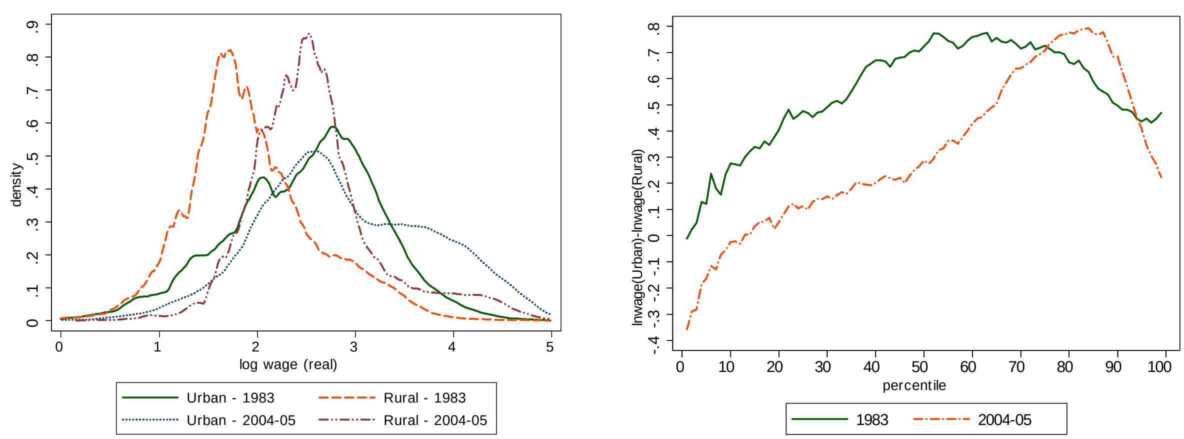
The next point of interest is the behavior of wages in urban and rural India. Wages are obtained as the daily wage/salaried income received for the work done by respondents during the previous week (relative to the survey week). Wages can be paid in cash or kind, where the latter are evaluated at the current retail prices. We convert wages into real terms using state-level poverty lines that differ for rural and urban sectors. We express all wages in 1983 rural Maharashtra poverty lines.[5] Importantly, we are interested not just in the mean or median wage gaps, but rather in the behavior of the wage gap across the entire wage distribution.

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

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 8 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2004-05 survey rounds. The plot shows a very clear rightward shift of the wage density function during this period for rural workers. The shift in the wage distribution for urban workers is much more muted. In fact, the mean almost did not change, and most of the changes in the distribution took place at the two ends. Specifically, a fat left tail in the urban wage distribution in 1983, indicating a large mass of urban labor having low real wages, has disappeared and was replaced by a fat right tail.

Figure 8: The log wage distributions of urban and rural workers for 1983 and 2004-05



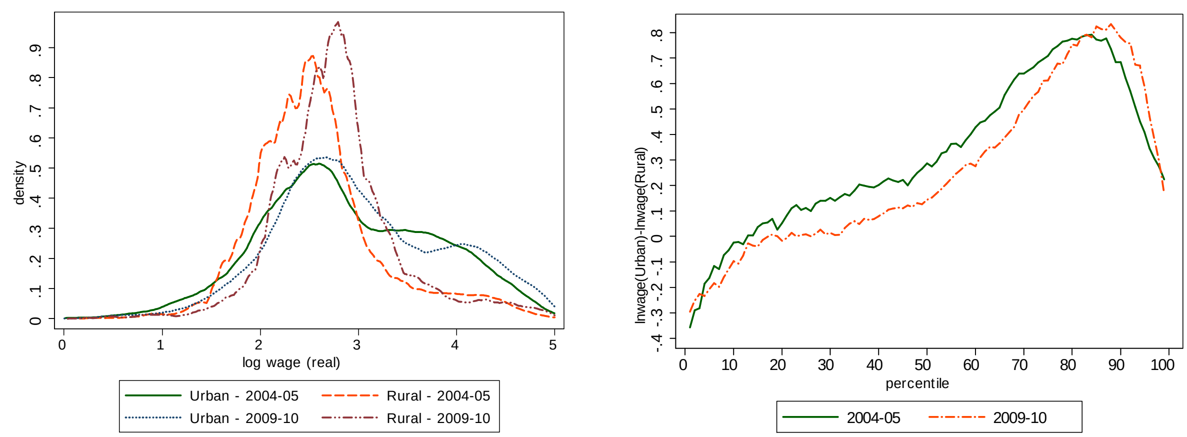
(a) (b)

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

Panel (b) of Figure 8 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 rising 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 9 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.

Figure 9: The log wage distributions of urban and rural workers for 2004-05 and 2009-10



(a) (b)

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

Figures 8 and 9 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.[6]

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.[7] 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 8: the wage gap between rural and urban areas fell between 1983 and 2005 for all but the richest wage groups.

While the wage convergence for most of the distribution is interesting, what were the factors driving this convergence? We turn to this issue next. Our focus is on two aspects of the wage gaps: Was the wage convergence documented above driven by a convergence of measured covariates of wages; or was it due to changes in unmeasured factors? We proceed with two approaches. Our first

[6] 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 regression methods to determine the unconditional effect of rural status on wages for different quantiles. The RIF methodology essentially gets around this problem for quantiles. Details regarding this method can be found in Firpo, Fortin, and Lemieux (2009).

[7] 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.

approach is to use the procedure developed by DiNardo, Fortin, and Lemieux (1996) (DFL from hereon) to decompose the overall difference in the observed wage distributions of urban and rural labor within a sample round into two components – the part that is explained by differences in attributes and the part that is explained by differences in the wage structure of the two groups. To obtain the explained part, for each set of attributes we construct a counterfactual density for urban workers by assigning them the rural distribution of the attributes.[8,9]

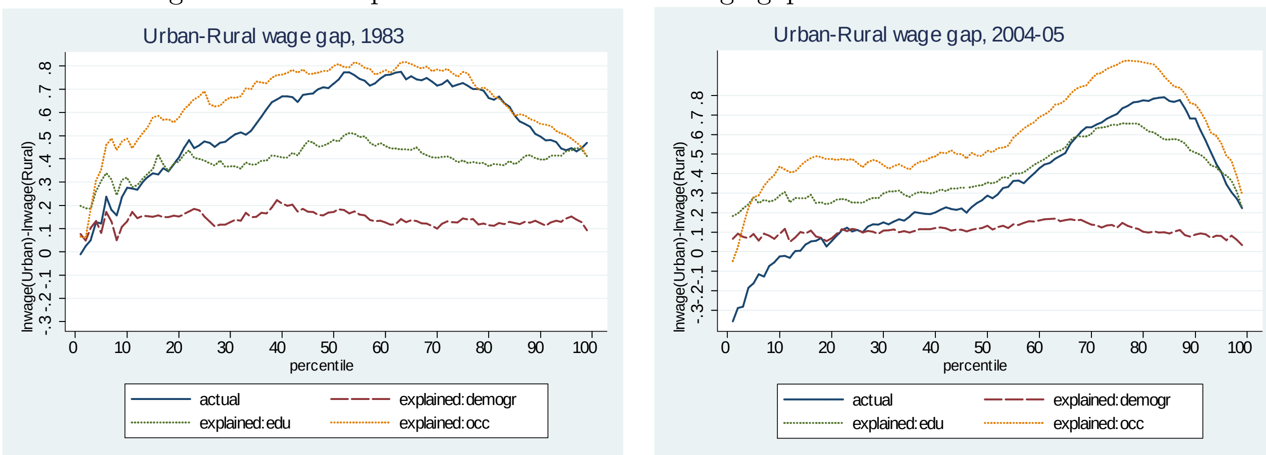
We consider several sets of attributes. First, we evaluate the role of individual demographic characteristics such as age, age squared, a dummy for the caste group (SC/ST or not) and a geographic zone of residence. The latter are constructed by grouping all Indian states into six regions – North, South, East, West, Central and North-East. Note that we control for caste by including a dummy for whether or not the individual is an SC/ST in order to account for the fact that SC/STs tend to be disproportionately rural. Given that they are also disproportionately poor and have little education, controlling for SC/ST status is important in order to determine the independent effect of rural status on wages. Second, we add education to the set of attributes and obtain the incremental contribution of education to the observed wage convergence. Lastly, we evaluate the role played by occupation differences in the urban-rural wage convergence. Figure 10 presents our findings for the pre-NREGA period – for 1983 (panel (a)) and 2004-05 (panel (b)). The solid line shows the actual urban-rural (log) wage gaps for the entire wage distribution, while the broken lines show the gaps explained by differences in attributes of the two groups, where we introduced the attributes sequentially.

Figure 10 shows that demographic characteristics explain a small fraction of the urban-rural wage gap. Moreover, this fraction remains stable at around 0.1 along the entire distribution in both 1983 and 2004-05. For the 1983 wage distribution gap, differences in education account for almost the entire wage gap at the bottom of the distribution, while differences in occupation explain the wage gap for the upper 50 percent of the distribution. Turning to 2004-05 however, the picture is different. Here differences in education attainments between urban and rural workers explain a large fraction of the gap at the top end of the distribution (median and above). However, for the bottom end of the distribution the education differences suggest that there should exist a large gap in *favor* of urban

[8] The DFL method involves first constructing a counterfactual wage density function for urban individuals by giving them the attributes of rural households. This is done by a suitable reweighing of the estimated wage density function of urban households. The counterfactual density is then compared with the actual wage density to assess the contribution of the measured attributes to the observed wage gap.

[9] We choose to do the reweighing this way to avoid a common support problem, i.e., there may not be enough rural workers at the top end of the distribution to be able to mimic the urban distribution.

Figure 10: Decomposition of Urban-Rural wage gaps for 1983 and 2004-05



(a) (b)

Notes: Each panel shows the actual log wage gap between urban and rural workers for each percentile, and the counterfactual percentile log wage gaps when urban workers are sequentially given rural attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus occupations ("occ"). The left panel shows the decomposition for 1983 while the right panel is for 2004-05.

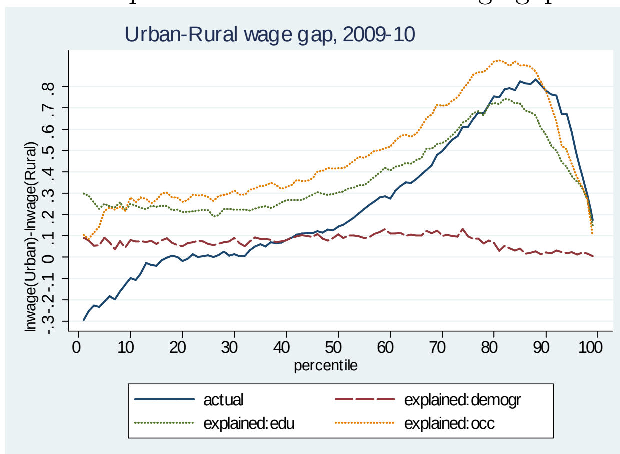
workers. This finding stands in stark contrast to the actual data which shows that wages of rural workers are in fact *higher* than wages of urban workers for the bottom 15 percent of distribution. Clearly, the data wage gap for the bottom 15 percent is the opposite of what their demographic characteristics and education endowments predict. Adding occupations deepens the puzzle further. Based on differences in occupations, the urban-rural gap should be more than 20 percent higher than the actual gap in the data. These results suggest that differences in the wage structure of the urban and rural workers play an important role in our data.

We conduct an analogous decomposition of the urban-rural wage gap in the post-NREGA period of 2009-10. As can be seen from Figure 11 the contribution of various factors in that round resembles closely our findings for 2004-05. We assess the role played by NREGA using state-level variation in wage gaps in Section 3.6.

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

[10] 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

Figure 11: Decomposition of Urban-Rural wage gaps for 2009-10



Notes: This figure shows the actual log wage gap between urban and rural workers for each percentile in 2009-10, and the counterfactual percentile log wage gaps when urban workers are sequentially given rural attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus occupations ("occ").

Our set of explained factors, as before, includes demographic characteristics such as individual's age, age squared, caste, and geographic region of residence. Additionally, we control for the education level of the individual by including dummies for education categories 1-5.[11]

Table 6: Decomposing changes in rural-urban wage gaps over time

(a). Change 1983 to 2009-10				explained
	(i) measured gap	(ii) explained	(iii) unexplained	(iv) education
10th quantile	-0.371***	-0.096***	-0.275***	-0.059***
	(0.036)	(0.016)	(0.040)	(0.013)
50th quantile	-0.568***	-0.202***	-0.366***	-0.166***
	(0.022)	(0.014)	(0.019)	(0.012)
90th quantile	0.332***	0.229***	0.103***	0.284***
	(0.041)	(0.046)	(0.045)	(0.044)
mean	-0.263***	-0.115***	-0.148***	-0.078***
	(0.019)	(0.014)	(0.017)	(0.012)
(b). Change in explained component				
10th quantile	-0.096***	-0.060***	-0.036***	-0.049***
	(0.016)	(0.008)	(0.013)	(0.006)
50th quantile	-0.202***	-0.064***	-0.137***	-0.052***
	(0.014)	(0.012)	(0.014)	(0.009)
90th quantile	0.229***	0.060***	0.169***	0.084***
	(0.046)	(0.021)	(0.040)	(0.020)
mean	-0.115***	-0.032***	-0.083***	-0.015
	(0.014)	(0.012)	(0.008)	(0.010)

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

Table 6 shows the results of the decomposition exercise. Panel (a) shows the decomposition of

as the benchmark sample. Details of the decomposition method can be found in the Appendix A.1.

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

the measured gap (column (i)) into the explained and unexplained components (columns (ii) and (iii)), as well as the part of the gap that is explained by education alone (column (iv)). The results indicate that the part of the wage gap that is explained by the included covariates varies from 25 percent for the bottom 10 percent to about 90 percent for the top 10 percent. Based on the explained component of the mean and median urban-rural wage gaps, about 50 percent of the gap is explained by the included covariates. Importantly, education alone accounts for the majority of the explained component along every point of the distribution.

If the explained component of a regression is βX, then changes in that component has two components: the change in X and the change β, which is the measured return to X. Since X is measured in the data, the part of the change in the explained component that is due to X is "explained" by the data while the part due to β is not directly explained. Panel (b) of the Table 6 decomposes changes in the explained component itself into the explained and unexplained parts. For the 10th percentile, most of the change in the measured component of the gap was due to changes in the explained part (or X). For the median and the 90th percentile however, most of the change in the explained component was due to changes in returns rather than changes in the component itself.

Overall, our conclusion from the wage data is that wages have converged significantly between rural and urban India during since 1983 for all except the very top of the income distribution. Education has been an important contributor to these convergent patterns. However, a large fraction of the trend is due to unmeasured factors, especially for the left tail of distribution. This is particularly puzzling since the actual wage gaps for the bottom 10 percent of the urban and rural wage distributions are in favor of rural workers while the covariates predict the opposite!

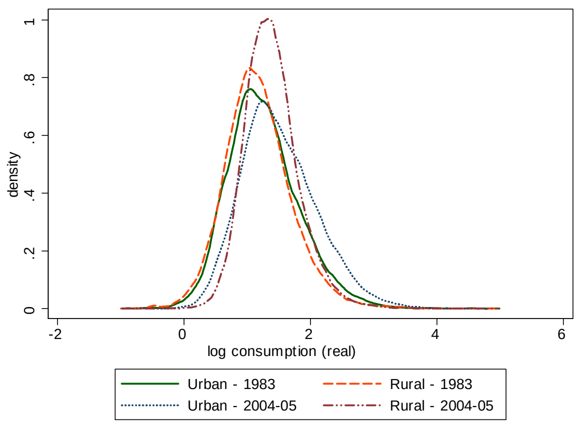
3.4 Consumption

Our last indicator of interest is the household expenditure on consumption in rural and urban India. This measure is the one that is often used in studies on poverty and inequality. The variable we use is monthly per capita household consumption expenditure, or "mpce". This variable is collected at the level of the household. We convert consumption expenditures into real terms using official state-poverty lines, with rural Maharashtra in 1983 as the base – same as we did for wages. In order to make real consumption results comparable with the wage data, we convert consumption into per-capita daily value terms.

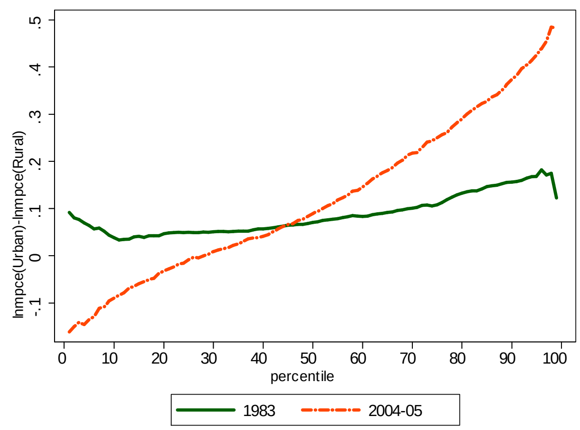
We start by plotting the kernel densities of the (log) per capita real consumption (we will use mpce and consumption interchangeably from hereon) for rural and urban households for 1983 and

2004-05 in Panel (a) of Figure 12. Panel (b) of the Figure plots the percentile gaps in (log) per capita log household consumption between urban and rural households for these two survey rounds.

Figure 12: The log consumption distributions of urban and rural households for 1983 and 2004-05



(a)



(b)

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

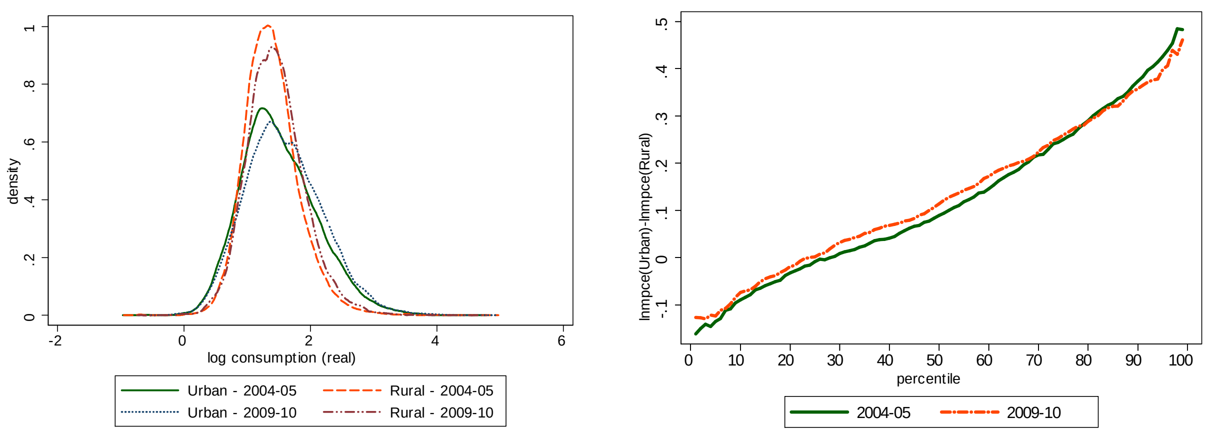
Panel (a) of Figure 12 shows that the consumption distribution shifted to the right for both rural and urban households during this period. Panel (b) of the figure, which plots the gaps for each percentile however shows several interesting features. First, consumption gaps in both 1983 and 2004-05 are positive and upward sloping. Second, the plot for 2004-05 is significantly steeper than the schedule for 1983. Effectively, the consumption gaps between urban and rural areas declined for percentiles below the median of the consumption distribution but widened above the median. Interestingly, rural households till around the 30th percentile actually consumed more than their counterparts in urban areas in 2004-05 (the gaps were negative) whereas the gaps were positive in 1983. Thus, the rural poor have clearly done better than the urban poor during this period, as measured by their household consumption levels.[12] This is an interesting finding in its own right and may indicate the effects of weaker social and family insurance mechanisms in urban areas for the poor which leads them to possibly consume less than they would otherwise. This channel has been suggested as a rationalization for low inequality and mobility in rural India by Munshi and Rosenzweig (2009).

Next, we consider the post-NREGA period. Panel (a) of Figure 13 contrasts densities of the (log)

[12] Note that these findings are similar in nature to the results we obtained for wages though the specifics are different. In particular, for wages the gap shrank for a much larger portion of the distribution than the corresponding narrowing of expenditure gaps in 2004-05.

per capita real consumption expenditures in the rural and urban sectors for 2004-05 and 2009-10; while panel (b) reports the percentile gaps in the two periods. Clearly, the trends we uncovered in the pre-reform periods continued on into the 2009-10. Both rural and urban consumption distributions have shifted to the right, although the shifts were relatively small. This led to a small increase in the consumption gaps during this period for all percentiles except at the very top of the distribution.

Figure 13: The log consumption distributions of urban and rural households for 2004-05 and 2009-10



(a) (b)

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

Are the measured consumption gaps shown in Panel (b) of Figures 12 and 13 significant? Are the changes in the gaps over time significant? We address this by estimating RIF regressions of log consumption on a constant and a rural dummy for all the survey rounds. We estimate these regressions for the 10th, 50th and 90th percentiles. We also run a standard OLS specification to determine the effect of rural status on the unconditional mean of log consumption. The results are reported in Table 7.

For all except the 10th percentile of households, rural consumption expenditure was significantly lower than the corresponding urban consumption across the survey rounds. The rural dummy for the 10th percentile regression was significantly negative in 1983 but turned positive and significant for all subsequent rounds indicating that rural consumption became higher than urban consumption for this group. Panel (b) of Table 7 shows that between 1983 and 2010 there was significant convergence in urban and rural consumption expenditures for the 10th percentile, but a significant divergence in consumption levels of the mean, median and 90th percentile, i.e., the gap has widened for them

Table 7: Consumption gaps and changes

	1983	1987-88	1993-94	1999-00	2004-05	2009-10
Panel (a): Rural dummy coefficient						
10th quantile	-0.039***	0.049***	0.009	0.020**	0.080***	0.070***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.009)	(0.011)
50th quantile	-0.066***	-0.039***	-0.088***	-0.096***	-0.072***	-0.091***
	(0.007)	(0.006)	(0.005)	(0.006)	(0.007)	(0.009)
90th quantile	-0.164***	-0.179***	-0.290***	-0.355***	-0.413***	-0.411***
	(0.011)	(0.010)	(0.011)	(0.012)	(0.015)	(0.018)
mean	-0.085***	-0.054***	-0.115***	-0.134***	-0.119***	-0.131***
	(0.006)	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)
N	87335	93701	87098	88620	90838	75123

	1983 to 1993-94	1993-94 to 2009-10	1983 to 2009-10
Panel (b): Changes			
10th quantile	0.048***	0.061***	0.109***
	(0.011)	(0.013)	(0.014)
50th quantile	-0.022***	-0.003	-0.025***
	(0.009)	(0.010)	(0.011)
90th quantile	-0.126***	-0.121***	-0.247***
	(0.016)	(0.021)	(0.021)
mean	-0.030***	-0.016*	-0.046***
	(0.008)	(0.009)	(0.010)

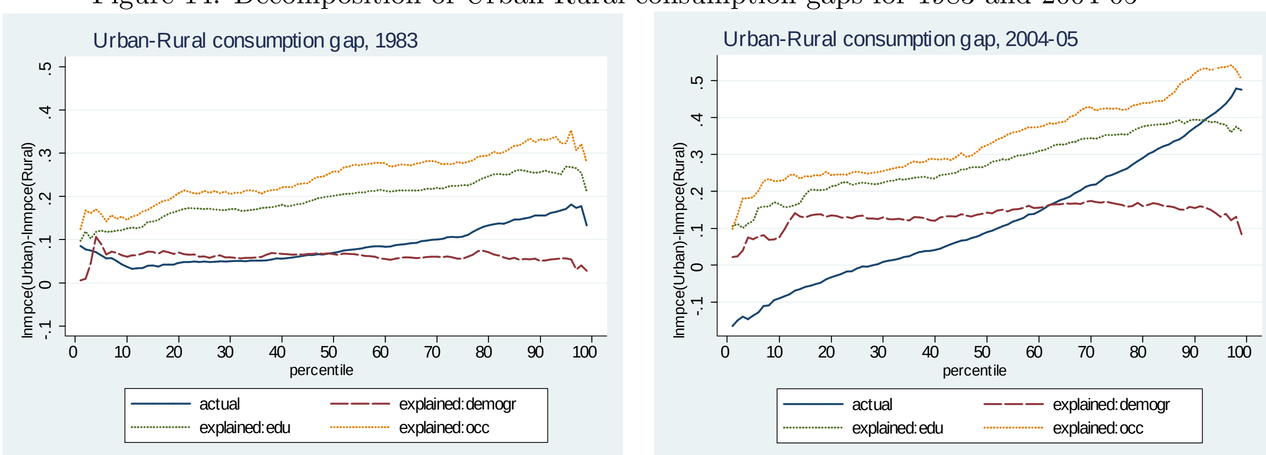
Note: Panel (a) reports the estimates of the coefficient on the rural dummy from RIF regressions of log consumption expenditures on rural dummy and a constant. Panel (b) reports the changes in the estimated coefficients over successive decades and the entire sample period. N refers to the number of observations. Standard errors are in parenthesis. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

during this period. This pattern is in contrast to the wage gap patterns which showed convergence for all except the 90th percentile.

Are most of the changes in consumption gaps due to changes in attributes or are they due to changes in the consumption structure, i.e., are they due to changes in the $X's$ or the $\beta's$? We address this question with DFL decompositions. Figure 14 reports the actual log consumption gap for urban relative to rural households computed for every percentile (solid line); and the explained gaps, computed as the difference between the actual and reweighted urban consumption percentiles, where we reweighed the consumption distribution of urban workers by assigning them rural attributes. As with wages, we examine the influence of individual covariates by introducing them sequentially. We first consider demographic characteristics, which include household size, the number of earning members of the household, caste dummy and regional dummies. Then we introduce education attainments, which consist of the education attainments of the household head and the highest level of education attained in the household. Lastly, we add occupation dummies for the household head.

Panel (a) shows the results for 1983, while panel (b) reports them for the pre-NREGA period of 2004-05. Our findings are similar of those for wages. Demographic characteristics explain almost the entire gap at the bottom end of the distribution (up to the median) in 1983. In 2004-05 however, demographics lose their explanatory power. In fact, the gap predicted by the demographics remains flat for the entire distribution, while the actual consumption gap is clearly upward sloping. Furthermore, the actual gap is negative at the bottom end of the distribution, while demographics predict a positive gap in the order of about 10 percent. Adding education and occupations to the set of attributes deepens the puzzle: based on the differences in education attainments and occupation distributions, the consumption gap between urban and rural workers should be in the order of 15 to 20 percent at the bottom end of the distribution and in the order of 30 percent at the top end in 1983. The actual gap is significantly smaller. Similarly, in 2004-05 the actual gaps are orders of magnitude smaller than the gaps implied by the differences in attributes, especially at the lower end of the distribution. Clearly, there is a large unexplained component that is responsible for the observed consumption gaps, especially for the poorer households.

Figure 14: Decomposition of Urban-Rural consumption gaps for 1983 and 2004-05



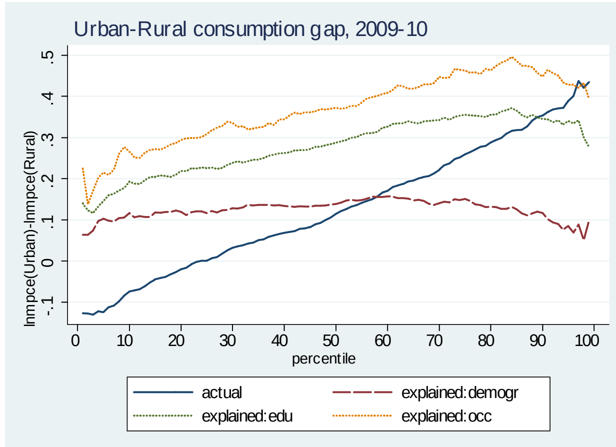
(a) (b)

Notes: Each panel shows the actual log consumption gap between urban and rural workers for each percentile, and the counterfactual percentile log consumption gaps when urban workers are sequentially given rural attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus occupations ("occ"). The left panel shows the decomposition for 1983 while the right panel is for 2004-05.

We also perform an analogous decomposition of consumption gaps in the post-reform 2009-10 period and find a pattern similar to 2004-05, with a large unexplained component behind the observed gaps. In section 3.6 we ask whether the NREGA policy is driving part of this unexplained component.

Were the consumption gaps changing over time due to changes in measured covariates of consumption or were the changes due to unexplained factors? In order to determine this we conduct a Oaxaca-Blinder type time series decomposition of the measured changes into their explained and

Figure 15: Decomposition of Urban-Rural consumption gaps for 2009-10



Notes: This figure shows the actual log consumption gap between urban and rural workers for each percentile in 2009-10, and the counterfactual percentile log consumption gaps when urban workers are sequentially given rural attributes. Three sets of attributes are considered: demographic (denoted by "demogr"), demographics plus education ("edu"), and all of the above plus occupations ("occ").

unexplained components. We do this using the OLS regression for the mean and the RIF regressions for different quantiles of the consumption distributions in rural and urban areas. As attributes of consumption we introduce controls for household size, the number of earning members of the household, region dummies, and a caste dummy for SC/ST status. We also control for the education characteristics of the household by adding the education attainment level of the household head and the highest level of education attained in the household. Table 7 shows the results of the decomposition.

In summary, three aspects of the results for the consumption gaps noteworthy. First, rural-urban consumption gaps have appear to have widened between 1983 and 2010 for most groups except the bottom 10 percent of households. This is in sharp contrast to the wage gaps which appear to have narrowed for all but the richest households. Second, the explained part of the changes in the consumption gaps accounted for by household attributes comprises around zero percent of the total change for the 10th percentile and 30 percent for the mean, median and the 90th percentile. Hence, the majority of the changes in the quantile gaps were driven by unexplained or unmeasured factors. This is slightly lower than the explained component for changes in the wage gaps. Third, changes in the education level of households explain a small fraction of the actual change in the rural-urban consumption gap. This is in contrast to changes in the rural-urban wage gaps where education played a larger and significant role.

Table 8: Decomposing changes in rural-urban consumption expenditure gaps over time

(a). Change (1983 to 2004-05)	(i) measured gap	(ii) explained	(iii) unexplained	explained (iv) education
10th quantile	-0.111***	0.001	-0.112***	-0.031***
	(0.019)	(0.010)	(0.018)	(0.007)
50th quantile	0.044***	0.015*	0.028**	-0.034***
	(0.017)	(0.010)	(0.014)	(0.007)
90th quantile	0.194***	0.043***	0.151***	-0.030**
	(0.022)	(0.016)	(0.023)	(0.014)
mean	0.043***	0.017**	0.027***	-0.033***
	(0.015)	(0.009)	(0.012)	(0.007)
(b). Change in explained component				
10th quantile	0.001	-0.004	0.004	-0.031***
	(0.010)	(0.005)	(0.008)	(0.003)
50th quantile	0.015*	0.014**	0.001	-0.022***
	(0.010)	(0.006)	(0.008)	(0.004)
90th quantile	0.043***	0.049***	-0.006	-0.001
	(0.016)	(0.010)	(0.014)	(0.008)
mean	0.017**	0.018***	-0.001	-0.018***
	(0.009)	(0.006)	(0.005)	(0.005)

Note: Panel (a) presents the change in the urban-rural consumption gap between 1983 and 2004-05. Panel (b) reports the decomposition of the time-series change in the explained component of the change in the consumption gap over 1983–2004-05 period. All gaps are decomposed into explained and unexplained components using RIF regression approach of Firpo, Fortin, and Lemieux (2009). 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.

3.5 The Role of Migration

A natural explanation for the narrowing of the wage gaps that we have documented above is migration from rural to urban areas. Indeed, two of the older theories of structural transformation – the Lewis and Harris-Todaro models – both formalize the process through which rural/agricultural workers migrate to urban areas in search of higher wages. Even from a neoclassical perspective, i.e., from a non-dualistic economy view of the world, rural migration to urban areas would tend to raise rural wages as long as the marginal product of labor in agriculture is positive while simultaneously putting downward pressure on urban wages. This would induce a narrowing of the rural-urban wage gaps.

In order to assess the contribution of migration to wage gaps, we examined the migration data contained in the NSS surveys. Unfortunately, migration particulars are not available in all the survey rounds that we study as questions on migration were not asked at all in most of them. Specifically, we have information on whether a surveyed individual migrated during the previous five years leading up to the survey date for the 38th round (1983) and 55th round (1999-00). We also have this information for the smaller 64th survey round conducted by the NSS in 2007-08.[13] We use information from these three rounds to form an assessment of the role of migration.

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

Table 9 shows the main patterns of migration for these three rounds. The first feature to note is that the number of recent migrants (those who migrated during the preceding five years) as a share of the total workforce has declined from 7.2 percent in 1983 to 6.2 percent in 2007-08.[14] Of these migrants, the largest single group were those who moved between rural areas, although the share of rural-to-rural migration in overall migration flows has declined from about 50 percent in 1983 to just above 38 percent in 2007-08. The share of urban migrants to rural areas has stayed relatively unchanged around 10 percent during this period. In contrast, urban areas have experienced an increase in migration inflows from both rural and urban areas. Thus, the share of rural-to-urban migration in total migration flows has increased from 22 percent in 1983 to about 30 percent in 2007-08. Urban-to-urban migration, which stood at 19 percent in 1983, rose to 23 percent in 2007-08, thereby failing to keep pace with the rise in the rural-to-urban flows. Interestingly, majority of the increase in migration to urban areas took place in the latter half of our sample – since 1999-00. To put these flows in perspective, the rural-to-urban migrants account for around 7 percent of the urban workforce. This share has remained stable over the period. Note that the net flow of workers from rural to urban areas is lower as there is some reverse flow as well. While clearly not insignificant, the share of migrant workers from rural areas in the urban workforce is relatively small given the overall size of the urban workforce.

Table 9: Migration trends: 1983-2008

	migrant	migrants				rural-to-urban	for job
	total	rural-to-urban	urban-to-urban	rural-to-rural	urban-to-rural	urban	rural-to-urban
1983	0.072	0.224	0.185	0.496	0.087	0.072	0.778
	(0.001)	(0.005)	(0.005)	(0.006)	(0.003)	(0.002)	(0.010)
1999-00	0.068	0.230	0.182	0.468	0.106	0.067	0.740
	(0.001)	(0.006)	(0.005)	(0.007)	(0.004)	(0.002)	(0.012)
2007-08	0.062	0.301	0.227	0.379	0.084	0.072	0.810
	(0.001)	(0.007)	(0.007)	(0.008)	(0.004)	(0.002)	(0.011)

The last column of Table 9 also shows that the majority of the rural-to-urban migration is job related. The rest is mostly for marriage reasons. Same is true about urban-to-urban migration flows. Interestingly, job related migration from rural to urban areas appears to have increased in 2007-08 relative to 1999-2000 despite the introduction of the rural employment program NREGA in 2005. Migration to rural areas is in equal proportion for job, marriage and other reasons.[15]

What do the wage and consumption profiles of these recently migrated workers look like? We perform a simple evaluation of migrant workers wages and consumption and their effect on urban-rural wage and consumption convergence by amending our regression specifications in Sections 3.3

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

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

and 3.4 to include four additional dummy variables, each identifying a migration flow between rural and urban areas. We also re-define the rural dummy to identify rural non-migrant workers only. If migration flows contribute significantly to the urban-rural gaps, we should see the coefficient on rural dummy change in value and/or significance after migration flow dummies are introduced.

Table 10 reports our results for (log) wages. We find that migration flows from urban areas have coefficients that are positive and significant, suggesting that urban migrants earn more than the benchmark group – urban non-migrants. Migrants from rural areas, in contrast, earn less than urban non-migrants, but the difference is significant mainly for rural-to-rural migration flows. Note also that the negative effects on wages for this group is declining over time, in line with the aggregate wage convergence. Wages of migrants who moved from rural to urban areas are no different than the wages of urban non-migrants.[16] These results apply to both mean and median wages. Do these migration flows contribute to the urban-rural wages gap convergence? A comparison of regression coefficients on the rural dummy in Table 10 and in the benchmark specification without migration flows dummies in Table 5 reveals that they are practically the same. We find that this result also holds for individuals at the two ends of the wage distribution (see Table A1 in Appendix A.2).[17] This suggests to us that the wage gap between urban and rural non-migrants has been narrowing at the same rate as the overall urban-rural gap.

Table 10: Wage gaps: Accounting for migration

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

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

We repeat the same analysis for consumption in Table 11. Migration flows dummies are positive

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

[17] The regression results for migrants in the 10th percentile of the distribution reported in A1 should be treated with caution due to the small number of observations. This might be reflecting problems in sample coverage of poorer migrants to urban areas.

and significant, suggesting that migrants consume more per capita than urban non-migrants. This is true even for rural-to-rural and rural-to-urban migrants. The coefficients on the rural dummy are negative, significant and very similar in magnitude to the rural dummy coefficients in our benchmark regression specification in Table 7 without migration flows dummies. Same results hold for the 10th and 90th percentiles of consumption distribution (see Table A2 in Appendix A.2). We interpret this result as an indication that migration flows contributed little to the observed urban-rural relative consumption.

Table 11: Consumption gaps: Accounting for migration

	mean			median		
	1983	1999-00	2007-08	1983	1999-00	2007-08
rural	-0.061***	-0.116***	-0.133***	-0.047***	-0.083***	-0.091***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
rural-to-urban	0.205***	0.106***	0.128***	0.180***	0.084***	0.115***
	(0.019)	(0.020)	(0.026)	(0.023)	(0.028)	(0.024)
urban-to-urban	0.394***	0.401***	0.451***	0.327***	0.293***	0.292***
	(0.023)	(0.028)	(0.027)	(0.022)	(0.019)	(0.020)
rural-to-rural	0.114***	0.120***	0.200***	0.119***	0.134***	0.177***
	(0.020)	(0.026)	(0.025)	(0.024)	(0.028)	(0.024)
urban-to-rural	0.355***	0.155***	0.191***	0.319***	0.101***	0.126***
	(0.037)	(0.032)	(0.041)	(0.032)	(0.038)	(0.039)
N	87335	88620	85916	87335	88620	85916

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

Overall, we do not find significant evidence that migration may have contributed to the shrinking wage gaps between rural and urban areas. Of course this conclusion is subject to an obvious caveat that migration decision itself is endogenous to wage gaps between rural and urban areas. Thus, how much of the overall wage convergence as well as its distributional differences can be explained by migration requires a more structural analysis which is beyond the scope of this paper. We hope to return to this issue in future work.

3.6 Effects of National Rural Employment Guarantee Act

In 2005 the government of India enacted the National Rural Employment Guarantee Act (NREGA, since renamed the The Mahatma Gandhi National Rural Employment Guarantee Act). The objective of the Act was to provide an economic safety net to the rural poor by providing rural households with 100 days of guaranteed employment every year for at least one adult member for doing casual manual labour at the rate of Rupees 60 per day (approximately US $1.30 per day). The employment has to be productive, projects must not involve any machines or contractors, and the identification of projects is based on the economic, social and environmental benefits of different types of works, their contribution to social equity, and their ability to create permanent assets. In addition, the employment is generally provided within a radius of 5 kilometers of the village where the applicant resides at the time of applying. The question that arises is whether any of the rural-urban convergence between 2004-05 and 2009-10 can be attributed to the NREGA reform? To answer this question we turn to a state-level analysis.

We are interested in whether there was a break in wage and consumption gap dynamics after the introduction of NREGA in 2006. Our strategy to study this is to examine whether the 2009-10 round exhibits a disproportionate change in the size of the gaps relative to the previous five rounds. We proceed by estimating fixed effects regressions on the mean, median, 10th and 90th percentile *state wage gaps*, where in each regression we control for the time-invariant state-level fixed effects. We include a trend variable ("trend") to obtain the estimate of the average change between rounds in the state wage gaps during the sample period. To account for the potential break in the size of the gaps in 2009-10 (the period associated with NREGA), we include a dummy variable "2009-10 dummy" which is equal to one for observations in the 2009-10 round.

The shortcoming of this approach, of course, is that the 2009-10 dummy may be significant for reasons that are unrelated to NREGA. To separate the effects of NREGA from other potential factors, we exploit the cross-state variation in the exposure of states to NREGA. Specifically, we introduce a control for the share of the rural labor force in each state (variable "rural share") and its interaction with the 2009-10 time dummy. Recall that NREGA applies only to rural employment. If NREGA had a significant negative effect on the state wage gaps (i.e., it reduced the wage gap), then the coefficients on the 2009-10 dummy and the interaction between the rural share and the 2009-10 dummy should both be negative. Our sample includes the 17 major states in India. The results are presented in Table 12.

Consistent with our earlier findings for individual data, the median and 10th percentile *state-level* wage gaps have significantly declined during 1983-2010 period. The wage gap for the 90th percentile has widened while the changes in the mean gap were not significant. At the same time, the coefficient on "2009-10 dummy" is not statistically different from zero for all but the 90th percentile where the sign is positive, i.e. the gap at the top end of thee distribution appears to have widened relative to the trend. Overall, these results suggest that the decline in the wage gap during 2009-10 period was not different from the earlier rounds for those moments. The coefficients on both the rural share as well as the interaction term between the rural share and the 2009-10 dummy are consistently insignificant (again except for the 90th percentile where it enters with a positive coefficient) indicating that state level exposure to the NREGA program had no significant effect on the state-level wage gaps. [18]

Table 12: Estimating the effects of NREGA reform on the urban-rural log wage gaps

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

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

We consider an analogous exercise for consumption gaps in Table 13. The results suggest that the state-level consumption gaps have declined for the 10th percentile, and widened for the 90th percentile. The changes in the mean and median consumption gaps were both insignificant. The coefficients on the 2009-10 dummy, the rural share as well as the interactive term between the rural share and the 2009-10 dummy were all insignificant except for the 10th percentile where the average state-level consumption gaps have fallen and have fallen by more than the trend in states with larger rural population shares. Clearly, the NREGA policy does not appear to have had any differential effect on the rural-urban consumption gaps beyond the average effect across states.

Table 13: Estimating the effects of NREGA reform on the urban-rural consumption gaps

	mean	median	10th percentile	90th percentile
trend	0.007	0.002	-0.004***	0.008***
	(0.005)	(0.001)	(0.001)	(0.001)
2009-10 dummy	0.387	0.025	0.020	0.157
	(0.649)	(0.160)	(0.176)	(0.201)
rural share	0.405	-0.140	-0.833*	0.576
	(1.661)	(0.409)	(0.451)	(0.515)
rural share x 2009-10 dummy	-0.619	-0.093	-0.095	-0.311
	(0.872)	(0.215)	(0.237)	(0.270)
N	96	96	96	96

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

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

[18] The results reported here are based on state-level moments. Hence they are not directly comparable to the results in Table 5 which report estimates on moments constructed from the All-India sample.

4 Conclusion

This paper has examined and contrasted the patterns of economic change in rural and urban India over the past three decades. We have found this period has been marked by a sharp and significant convergent trend in the education attainment levels of the rural workforce towards the levels of their urban counterparts. This process has also been accompanied by some convergence in the occupation choices being made in the two sectors. Specifically, the contraction in agrarian jobs in rural areas that has accompanied the ongoing structural transformation of the Indian economy away from agriculture has been met by an expansion of blue-collar occupations in rural areas at a significantly faster rate than the corresponding expansion of blue-collar occupations in urban areas. As a result there appears to have set in a process of convergence between the rural and urban occupation distributions (even though the absolute differences between the two sectors continues to be large).

We have also found a significant convergent trend in rural wages towards urban levels over this period with the median urban wage premium having declined from 101 percent in 1983 to 11 percent by 2010. We find this rate of wage convergence to be very large and somewhat unexpected. Even more surprisingly, we find that the majority of the decline in the rural-urban wage gap is not due to convergence in individual characteristics such as education attainments or policy changes, but rather is unexplained. Furthermore, if in 1983 the majority of the urban-rural wage gap was accounted for by demographics, education and occupation attributes of individuals, by 2010, most of the wage gap was due to urban-rural differences in wage structure rather than differences in their characteristics.

The convergence in consumption between rural and urban households has been more muted than that for some of the other indicators we examined. However, there were some shared features of the consumption and wage dynamics in that the rural poor did better over time than the corresponding urban poor (10th percentile) in terms of both indicators so that by 2004-05, the 10th percentile wage and consumption in rural areas exceeded that of their urban counterparts. There was however divergence in the fortunes of the 90th percentile in both wages and consumption where the pre-existing urban advantage became more pronounced over time.

The effect of the National Rural Employment Guarantee Act (NREGA) on the fortunes of the rural population has been a keenly debated issue in recent years. Using a state-level analysis, we find that neither did the gaps in 2009-10 narrow faster relative to trend, nor was there a disproportionate reduction in the gaps in states that had a larger rural share. Given that NREGA was introduced in 2005, we interpret these findings as indicating that the effect of the program on the state-level rural-urban wage and consumption gaps was, at best, very muted. We should reiterate though that this conclusion is based on just state-level wage and consumption gaps which has its own limitations, primarily the limited number of data points. This is an issue that we hope to return to in greater detail in future work.

We have also examined the role of migration in this process. While rural to urban migration has been happening, the overall flows have remained stable and small relative to the overall workforce. Rural migrants earn less than their urban counterpart, but the differences are not significant. However, the small size of the flows and the lack of a structural analysis of the issue in this paper suggests caution in drawing broader conclusions.

We believe these results to be suggestive of the fact that the massive macroeconomic changes that have been underway in India during this period have led to a healthy churning of the labor force in the country. The results we have obtained here for the rural-urban gaps are similar in spirit to those in Hnatkovska, Lahiri, and Paul (2012a) and Hnatkovska, Lahiri, and Paul (2012b) for the gaps between scheduled castes/tribes and others in the Indian workforce. There too we found significant convergence across the two groups in education, occupation choices, wages and consumption. Clearly, some of the market incentives that were unleashed by the economic reforms have been providing the right signals to economic agents to make the appropriate market based private choices in terms of their schooling and employment decisions. In fact, the wage convergence and some degree of consumption convergence between rural and urban India between 1983 and 2010 stands in contrast to the experience of China where Qu and Zhao (2008) report that rural-urban consumption and income gaps actually widened between 1988 and 2002.

References

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

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

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

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

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

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

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

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

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

Hnatkovska, V., A. Lahiri, and S. Paul (2012a): "Breaking the Caste Barrier: Intergenerational Mobility in India," *Journal of Human Resources, forthcoming.*

——— (2012b): "Castes and Labor Mobility," *American Economic Journal: Applied Economics*, 4(2), 274–307.

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

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

Munshi, K., and M. Rosenzweig (2009): "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth," Working papers, Brown University, Department of Economics.

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

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

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).

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

A Appendix

A.1 Decompositions 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 or log consumption expenditure) 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:

$$\Delta y^e_2 - \Delta y^e_1 = (E_2 - E_1) + (U_2 - U_1) = \Delta E + \Delta U$$

These differentials are reported in Panel (a) of Tables 6 and 8. 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 Tables 6 for wages and 8 for consumption.

A.2 Distributional effects of migration

Table A1 presents regression results from the RIF regression of (log) wages on a several dummies: a rural non-migrant dummy, and a set of four migration flows dummies between rural and urban areas. RIF regressions are estimated for the 10th and 90th percentile of (log) wages.

Table A1: Wage gaps: Accounting for migration

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

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

Table A2 presents the corresponding results for per capital consumption expenditures.

Table A2: Consumption gaps: Accounting for migration

	10th percentile			90th percentile		
	1983	1999-00	2007-08	1983	1999-00	2007-08
rural	-0.029***	0.034***	0.053***	-0.124***	-0.328***	-0.406***
	(0.009)	(0.008)	(0.010)	(0.012)	(0.013)	(0.015)
rural-to-urban	0.096***	0.128***	0.117***	0.298***	0.105**	0.174***
	(0.024)	(0.020)	(0.032)	(0.046)	(0.047)	(0.059)
urban-to-urban	0.165***	0.203***	0.195***	0.748***	0.706***	0.919***
	(0.020)	(0.014)	(0.019)	(0.061)	(0.064)	(0.082)
rural-to-rural	0.023	0.070**	0.176***	0.275***	0.142**	0.247***
	(0.029)	(0.033)	(0.023)	(0.046)	(0.057)	(0.067)
urban-to-rural	0.154***	0.160***	0.179***	0.648***	0.192**	0.349***
	(0.028)	(0.030)	(0.029)	(0.096)	(0.078)	(0.106)
N	87335	88620	85916	87335	88620	85916

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