Structural Transformation and the Rural-Urban Divide*

Viktoria Hnatkovska† and Amartya Lahiri‡

December 2012

Abstract

Development of an economy typically goes hand-in-hand with a declining importance of agriculture in output and employment. Given the primarily rural population in developing countries and their concentration in agrarian activities, this has potentially large implications for inequality along the development path. We examine the Indian experience between 1983 and 2010, a period when India has been undergoing such a transformation. We find a significant decline in the wage differences between individuals in rural and urban India during this period. However, individual characteristics such as education, occupation choices and migration account for at most 40 percent of the wage convergence. We use a two-sector model of structural transformation to rationalize the rest of the rural-urban convergence in India as the consequence of two observed features during this period: (a) higher labor supply growth in urban areas; and (b) differential sectoral productivity growth. Quantitative results suggest that 90 percent of the unexplained wage convergence between rural and urban areas can be jointly explained by these two factors.

JEL Classification: J6, R2

Keywords: Rural urban disparity, education gaps, wage gaps

---

*We would like to thank IGC for a grant funding this research, seminar participants at UBC, Wharton and the IGC-India 2012 conference in Delhi. An online Appendix to this paper is available from the authors’ websites.
†Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada and Wharton School, University of Pennsylvania. E-mail address: hnatkova@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

Development of an economy typically goes hand-in-hand with a declining importance of agriculture in output and employment. Given the primarily rural population in developing countries and their concentration in agrarian activities, this has potentially large implications for inequality along the development path. This paper examines the issue within the context of the experience of India between 1983 and 2010. With upwards of 800 million people still residing in rural areas, India provides a potentially dramatic illustration of the importance of this issue particularly in light of the large macroeconomic changes that it has been undergoing during this period.

We document that the labor income gap between urban and rural areas in India has declined sharply during this period. Upwards of 60 percent of this convergence, however, cannot be explained by changing individual or household characteristics of rural and urban workers. Moreover, migration flows between rural and urban areas are small and remained stable, thus contributing little to the observed convergence. Instead, we propose a mechanism that relies on aggregate changes and thus present a joint explanation of rural-urban convergence and the ongoing structural transformation of the economy.

Our explanation builds upon traditional theories of structural transformation but with a key difference. Traditional approaches emphasize the demand-side reasons for structural change by relying on aggregate productivity growth and non-homothetic preferences (see Chenery and Srinivasan (1988), Matsuyama (1991), Laitner (2000), Kongsamut, Rebelo, and Xie (2001), Gollin, Parente, and Rogerson (2002)). These theories imply that as an economy grows the relative demand for agricultural goods and, therefore, farm labor declines. While these models can potentially match the declining share of agriculture in employment and output, they also imply a decline in the price of agricultural goods and farm relative wages, both of which are inconsistent with the factual movements in sectoral relative prices and wages in India. We augment this standard mechanism for structural change with two key supply-side factors: (i) a rise in relative non-farm productivity; and (ii) higher labor force growth in urban relative to rural areas. Both are robust features of Indian data. By reducing relative labor supply to agricultural activities, these two factors can jointly account for structural transformation, rural-urban wage convergence, and an improvement in the agricultural terms of trade. A quantitative evaluation of the model suggests that around 90 percent of the unexplained wage convergence is due to these two forces.

The two channels we put forward are complementary to the skill acquisition story proposed by Caselli and Coleman (2001) in their study of regional convergence between North and South of the United States.
USA. Like our two factors, in their model a fall in the cost of acquiring skills to work in the non-agricultural sector can induce a fall in farm labor supply and lead to an increase in farm wages and relative prices.\(^1\)

We document the developmental process for India during the 1983-2010 period using six rounds of the National Sample Survey (NSS) of households. We start by showing that there has been a significant decline in labor income differences between rural and urban India during this period. Using a simple decomposition exercise we show that almost all of the measured convergence is due to shrinking wage gaps, both between and within occupations. We also find that labor reallocations across occupations have played a minor role at best in accounting for the observed convergence. In terms of actual wage gaps, we find that the mean wage premium of the urban worker over the rural worker fell significantly from 51 percent to 27 percent while the corresponding median wage premium declined from 59 percent to 13 percent between 1983 and 2010.

What accounts for the wage convergence between rural and urban India? The natural candidates are individual characteristics of workers such as their education levels and occupation choices. We find evidence of significant convergent trends in both education attainment rates as well as the occupation choices of rural workers toward those of urban workers. However, using the decomposition methods of DiNardo, Fortin, and Lemieux (1996) and Firpo, Fortin, and Lemieux (2009) for the entire wage distribution, we show that individual characteristics including education and occupation choices can only explain at most 40 percent of the observed wage convergence between rural and urban areas. Most of the convergence remains unexplained. It bears repetition that this does not suggest that the covariates did not change. Indeed, a central finding of the paper is the decline in the education gaps between rural and urban workers during this period.\(^2\)

A related narrative in the structural transformation literature suggests an important role for migration of workers from rural to urban areas in the process of moving from agriculture to industrial activities. To investigate the role of this margin we examine the migration data contained in the NSS surveys. We find that 5-year gross migration flows have declined marginally from 7.2 percent of the workforce in 1983 to 6.2 percent in 2007-08. Around a quarter of these flows were from rural to

---

1While likely an important additional factor behind the rural-urban wage convergence, the fall in skill acquisition costs is prohibitively difficult to quantify in the Indian context. Therefore, we choose to abstract from it in this study.

2We also examine the potential effect of an important rural employment program introduced in 2005 called National Rural Employment Guarantee Act (NREGA) on the rural-urban wage gaps. We use a state level analysis and find that the state-level wage and consumption gaps between rural and urban areas did not change disproportionately in the 2009-10 survey round, relative to their trend during the entire period 1983-2010. We also find that states that were more rural, and hence more exposed to the policy, did not exhibit differential responses of the percentile gaps in wages in 2009-10, relative to trend. We conclude that the effect of this program on the gaps was very muted. These results are available in an online appendix.
urban areas, while around 10 percent were from urban to rural areas. Consequently, the net flow of workers from rural to urban areas is small and has remained relatively stable at around 1 percent of all full-time employed workforce. We also find that migrants from rural to urban areas do not earn significantly lower wages than their urban non-migrant counterparts. Moreover, the wage difference between rural and urban non-migrant workers has been narrowing at the same rate as the overall wage gap between rural and urban workers. These results indicate to us that migration did not play an important role in inducing convergent dynamics between urban and rural areas.

If individual covariates do not explain much of the convergence, how does one explain it? Motivated by the fact that India has been undergoing a massive macroeconomic transformation during the period under study, we develop a simple model with two factors (urban labor and rural labor) and two sectors (agriculture and non-agriculture) to examine the potential contribution of aggregate shocks. We introduce the possibility of structural transformation of the economy by having a minimum consumption need of the agricultural good. We use this model to show that the wage convergence between rural and urban areas along with the structural transformation of the economy can be jointly explained by two factors: (i) a relatively faster increase in the urban labor force; and (ii) faster productivity growth in the non-agricultural sectors relative to agricultural productivity growth.\footnote{Our focus on these two channels is in part motivated by the recent work on non-balanced growth which emphasizes sectoral developments behind structural transformation of economies. For instance, \cite{Ngai2007} focus on exogenous Total Factor Productivity differences across sectors, while \cite{Acemoglu2008} emphasize differences in sectoral factor proportions and capital deepening.} Using a calibrated version of the model we show that these two factors can account for 90 percent of the unexplained wage convergence between rural and urban areas. Crucially, the model can also explain about 2/3 of the observed improvement in the agricultural terms of trade along with the structural transformation of the economy during this period.

Overall, our paper makes two key contributions. First, we believe this is the first paper that provides a comprehensive empirical documentation of the trends in rural and urban disparities in India since 1983 in wages, education and occupation distributions as well as an econometric attribution of the changes in the rural-urban wage gaps to measured and unmeasured factors. Second, we also provide a structural explanation for the observed wage convergence which is largely unexplained by the standard covariates of wages.

The interest in rural-urban inequality dates back to the classic theories in development by \cite{Lewis1954} and \cite{HarrisTodaro1970}. They recognized that the process of development tends to generate large scale structural transformation as economies shift from being primarily agrarian towards more industrial and service oriented activities. More recently, \cite{Young2012} has examined...
the cross-country evidence from 65 countries on urban-rural inequality. Strikingly, he finds that around 40 percent of the average inequality in countries included in the sample is due to urban-rural gaps.

There is a large body of work on inequality and poverty in India. While some of these studies do examine inequality and poverty in the context of 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, and is 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.

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

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

2 Empirical motivation

We start by focusing on differences in labor income between urban and rural areas and trends therein since 1983. Our data comes from successive rounds of the Employment & Unemployment surveys of the National Sample Survey (NSS) of households in India. The survey rounds that we include in the study are 1983 (round 38), 1987-88 (round 43), 1993-94 (round 50), 1999-2000 (round 55), 2004-05 (round 61), and 2009-10 (round 66). Since our interest is in determining the trends in wages and determinants of wages such as education and occupation, we choose to restrict the sample to individuals in the working age group 16-65, who are working full time (defined as those who worked

\[4\text{Since a large fraction of rural workers in India may be self-employed and thus do not report wage income, we also consider per capita consumption expenditures, and find that our findings are generally robust, especially for the lower percentiles of the consumption distribution. These results are presented in the online appendix.}\]
at least 2.5 days in the week prior to being sampled), who are not enrolled in any educational institution, and for whom we have both education and occupation information. We further restrict the sample to individuals who belong to male-led households. These restrictions leave us with, on average, 140,000 to 180,000 individuals per survey round. Details on our data are provided in Appendix [A.1].

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

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

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

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

5 This avoids households with special conditions since male-led households are the norm in India.
6 Using poverty lines that differ between urban and rural areas may generate real wage convergence if urban prices are growing faster than rural prices. This is indeed the case in India during our study period. However, only a small fraction of the observed real wage convergence is driven by the price dynamics. In the online appendix we show that
<table>
<thead>
<tr>
<th></th>
<th>(a) Individuals</th>
<th>(b) Households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban</td>
<td>Rural</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>male</td>
</tr>
<tr>
<td>1983</td>
<td>35.03</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1987-88</td>
<td>35.45</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1993-94</td>
<td>35.83</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1999-00</td>
<td>36.06</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>2004-05</td>
<td>36.18</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>2009-10</td>
<td>36.96</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

### Difference

<table>
<thead>
<tr>
<th></th>
<th>(a) Individuals</th>
<th>(b) Households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban</td>
<td>Rural</td>
</tr>
<tr>
<td>1983</td>
<td>-0.17***</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1987-88</td>
<td>0.09</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1993-94</td>
<td>0.04</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>1999-00</td>
<td>0.05</td>
<td>0.13***</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>2004-05</td>
<td>-0.39***</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>2009-10</td>
<td>-0.70***</td>
<td>0.09***</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

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.

Maharashtra poverty lines:\(^7\) To assess the role played by labor reallocation across jobs, we aggregate the reported 3-digit occupation categories in the survey into three broad occupation categories: white-collar occupations like administrators, executives, managers, professionals, technical and clerical nominal wages are converging as fast as real wages (except at the mean) during 1983-2010 period.

\(^7\)In 2004-05 the Planning Commission of India changed the methodology for estimation of poverty lines. Among other changes, they switched from anchoring the poverty lines to a calorie intake norm towards consumer expenditures more generally. This led to a change in the consumption basket underlying poverty lines calculations. To retain comparability across rounds we convert the 2009-10 poverty lines obtained from the Planning Commission under the new methodology to the old basket using a 2004-05 adjustment factor. That factor was obtained from the poverty lines under the old and new methodologies available for the 2004-05 survey year. As a test, we used the same adjustment factor to obtain the implied "old" poverty lines for the 1993-94 survey round for which the two sets of poverty lines are also available from the Planning Commission. We find that the actual old poverty lines and the implied "old" poverty lines are very similar, giving us confidence that our adjustment is valid.
workers; blue-collar occupations such as sales workers, service workers and production workers; and agrarian occupations collecting farmers, fishermen, loggers, hunters etc..

We define labor income per worker in Rural (R) or Urban (U) location as the sum of labor income in the three occupations in each location – white-collar jobs (occ 1), blue collar jobs (occ 2), and agrarian jobs (occ 3):

\[
\omega^j_t = \omega^1_{1t} L^{1t}_j + \omega^2_{2t} L^{2t}_j + \omega^3_{3t} L^{3t}_j,
\]

where \( L^{i}_{jt} \) is employment share of occupation \( i \) in location \( j \), and \( \omega^j_{it} \) is average daily wage in occupation \( i \) in location \( j \), with \( i = 1, 2, 3 \) and \( j = U, R \). Also \( L^{j}_{1t} + L^{j}_{2t} + L^{j}_{3t} = 1 \). The labor income gap between urban and rural areas can then be expressed as

\[
\frac{\omega^U_t - \omega^R_t}{\omega^R_t} = \frac{(\omega^U_{1t} - \omega_{1t})}{\omega^R_t} L^{1t}_U + \frac{(\omega^U_{2t} - \omega_{2t})}{\omega^R_t} L^{2t}_U + \frac{(\omega^U_{3t} - \omega_{3t})}{\omega^R_t} L^{3t}_U
- \frac{(\omega^R_{1t} - \omega_{1t})}{\omega^R_t} L^{1t}_R + \frac{(\omega^R_{2t} - \omega_{2t})}{\omega^R_t} L^{2t}_R + \frac{(\omega^R_{3t} - \omega_{3t})}{\omega^R_t} L^{3t}_R
+ \frac{(\omega_{1t} - \omega_{3t})}{\omega^R_t} (L^{1t}_U - L^{1t}_R) + \frac{(\omega_{2t} - \omega_{3t})}{\omega^R_t} (L^{2t}_U - L^{2t}_R),
\]

where \( \omega_{it} \) is the economy-wide average daily wage in occupation \( i = 1, 2, 3 \). The decomposition above shows that the urban-rural labor income gap can arise due to two channels. First, the gap may occur if wages and employment within each occupation are different across urban and rural areas (rows...
1 and 2 on the right in the expression above). We refer to this channel as the within-occupation channel. Second, the gap may arise if there is cross-occupation inequality in wages and employment shares (last row in the expression above). This is the between-occupation channel.\footnote{This decomposition is similar in spirit to that used by Caselli and Coleman (2001) for industries.}

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

$$
\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \Delta \mu_{1t}^U \bar{L}_{1t} + \Delta \mu_{2t}^U \bar{L}_{2t} + \Delta \mu_{3t}^U \bar{L}_{3t} - \Delta \mu_{1t}^R \bar{L}_{1t} - \Delta \mu_{2t}^R \bar{L}_{2t} - \Delta \mu_{3t}^R \bar{L}_{3t} + (\bar{L}_{1t}^U - \bar{L}_{1t}^R) [\Delta \eta_{1t} - \Delta \eta_{3t}] + (\bar{L}_{2t}^U - \bar{L}_{2t}^R) [\Delta \eta_{2t} - \Delta \eta_{3t}] + \Delta L_{1t}^U (\bar{\mu}_{1t}^U - \bar{\mu}_{3t}^U) + \Delta L_{2t}^U (\bar{\mu}_{2t}^U - \bar{\mu}_{3t}^U) - \Delta L_{1t}^R (\bar{\mu}_{1t}^R - \bar{\mu}_{3t}^R) - \Delta L_{2t}^R (\bar{\mu}_{2t}^R - \bar{\mu}_{3t}^R) + (\eta_{1t} - \eta_{3t}) \Delta (L_{1t}^U - L_{1t}^R) + (\eta_{2t} - \eta_{3t}) \Delta (L_{2t}^U - L_{2t}^R) \tag{2.2}
$$

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

Table 2 presents the results of the decomposition by occupations and components. During the 1983-2010 period, the aggregate labor income gap between urban and rural areas declined by 0.226. All of this decline is due to a convergence of wages, with roughly equal contributions of the within and between-occupation components. More precisely, convergence of rural and urban wages within each occupation has led to a 0.13 (or 57 percent) decline in the labor income gap between the two sectors. The between-occupation wage convergence in urban and rural areas produced an additional 0.18 (or 78 percent) decline in labor income gap. The majority of these changes were driven by
Table 2: Decomposition of labor income gap, 1983-2010

<table>
<thead>
<tr>
<th></th>
<th>wage component</th>
<th>labor reallocation</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>within</td>
<td>between</td>
<td>component</td>
</tr>
<tr>
<td>white-collar</td>
<td>-0.003</td>
<td>-0.056</td>
<td>0.148</td>
</tr>
<tr>
<td>blue-collar</td>
<td>-0.136</td>
<td>-0.120</td>
<td>-0.068</td>
</tr>
<tr>
<td>agrarian</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>-0.130</td>
<td>-0.177</td>
<td>0.080</td>
</tr>
<tr>
<td>% contribution</td>
<td>57.4</td>
<td>78.2</td>
<td>-35.6</td>
</tr>
</tbody>
</table>

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

blue-collar occupations. White-collar jobs also saw wage convergence both within occupations and between occupations, although the convergence was smaller than in blue-collar jobs.

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

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

3 Rural-Urban Wage Gaps

We first examine the distribution of log wages for rural and urban workers in our sample. Panel (a) of Figure 2 plots the kernel densities of log wages for rural and urban workers for the 1983 and 2009-10 survey rounds. The plot shows a very clear rightward shift of the wage density function for rural workers during this period. The shift in the wage distribution for urban workers is much more

9The Mahatma Gandhi National Rural Employment Guarantee Act (NREGA) was enacted in 2005. NREGA provides a government guarantee of a hundred days of wage employment in a financial year to all rural household whose adult members volunteer to do unskilled manual work. This Act could clearly have affected rural and urban wages. To control for the effects of this policy on real wages, we perform all evaluations on two subsamples: the pre-NREGA and post-NREGA periods. We find that the introduction of NREGA did not change the trends in real wages. Therefore, we proceed by presenting the results for the 1983-2010 period. The results for the pre- and post-NREGA subsamples are provided in an online Appendix.
muted. In fact, the mode almost did not change, and most of the changes in the distribution took place at the two ends. Specifically, a fat left tail in the urban wage distribution in 1983, indicating a large mass of urban labor having low real wages, disappeared. Instead a fat right tail has emerged.

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

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

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

Panel (b) of Figure 2 presents the percentile (log) wage gaps between urban and rural workers for 1983 and 2009-10. The plots give a sense of the distance between the urban and rural wage densities functions in those two survey rounds. An upward sloping schedule indicates that wage gaps are rising for richer wage groups. A rightward shift in the schedule over time implies that the wage gap has shrunk. The plot for 2009-10 lies to the right of that for 1983 till the 75th percentile indicating that for most of the wage distribution, the gap between urban and rural wages has declined over this period. Panel (b) shows that the median unconditional log wage gap between urban and rural wages fell from around 0.7 to around 0.1. Between the 75th and 90th percentiles however, the wage gaps are larger in 2009-10 as compared to 1983. This is driven by the emergence of a large mass of people in the right tail of the urban wage distribution in 2009-10 period, as we discussed above. A last noteworthy feature is that in 2009-10, for the bottom 20 percentiles of the wage distribution, rural wages were actually higher than urban wages. This was in stark contrast to 1983 when urban wages were higher than rural wages for all percentiles.

Figure 2 suggests wage convergence between rural and urban areas. To test whether this is statistically significant, we estimate regressions of the log real wages of individuals in our sample on a constant, controls for age (we include age and age squared of each individual) and a rural dummy
for each survey round. The controls for age are intended to account for potential life-cycle differences between urban and rural workers. We perform the analysis for different unconditional quantiles as well as the mean of the wage distribution.\(^{10}\)

### Table 3: Wage gaps and changes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10th quantile</td>
<td>-0.208***</td>
<td>-0.031***</td>
<td>-0.013</td>
<td>0.017</td>
<td>0.087***</td>
<td>0.177***</td>
<td>0.113***</td>
<td>0.295***</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td></td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>50th quantile</td>
<td>-0.586***</td>
<td>-0.405***</td>
<td>-0.371***</td>
<td>-0.235***</td>
<td>-0.126***</td>
<td>0.181***</td>
<td>0.279***</td>
<td>0.460***</td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>90th quantile</td>
<td>-0.504***</td>
<td>-0.548***</td>
<td>-0.700***</td>
<td>-0.725***</td>
<td>-1.135***</td>
<td>-0.044***</td>
<td>-0.587***</td>
<td>-0.631***</td>
</tr>
<tr>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.024)</td>
<td>(0.028)</td>
<td>(0.038)</td>
<td></td>
<td>(0.022)</td>
<td>(0.042)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>mean</td>
<td>-0.509***</td>
<td>-0.394***</td>
<td>-0.414***</td>
<td>-0.303***</td>
<td>-0.270***</td>
<td>0.115***</td>
<td>0.124***</td>
<td>0.239***</td>
</tr>
<tr>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td></td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
</tbody>
</table>

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 3 reports the estimated coefficient on the rural dummy for the 10th, 50th and 90th percentiles as well as the mean for different survey rounds.\(^{11}\) Clearly, rural status significantly reduced wages for almost all percentiles of the distribution across the rounds. However, the size of the negative rural effect has become significantly smaller over time for the 10th and 50th percentiles as well as the mean (see Panel (b)).\(^{12}\) The largest convergence occurred for the median. Furthermore, the coefficient on the rural dummy for the 10th percentile has turned positive, indicating a gap in favor of the rural poor. At the same time, the wage gap actually increased over time for the 90th percentile. These results corroborate the visual impression from Figure 2: the wage gap between rural and urban areas fell between 1983 and 2010 for all but the richest wage groups.

### 3.1 The role of education and occupation

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

\(^{10}\)We use the Recentered Influence Function (RIF) regressions developed by Firpo, Fortin, and Lemieux (2009) to estimate the effect of the rural dummy for different points of the wage distribution.

\(^{11}\)Due to widespread missing rural wage data for 1987-88, we chose to drop that round from the study of wages.

\(^{12}\)The decline in the mean wage gap reported in Table 3 is slightly higher than the decline in Table 2. This is because we report conditional wage gaps (with controls for age and age squared) in Table 3 and unconditional wage gaps in Table 2.
3.1.1 Education trends

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.\(^{13}\) Second, we use the reported education categories but aggregate them into five broad groups: 1 for illiterates, 2 for some but below primary school, 3 for primary school, 4 for middle, and 5 for secondary and above. The results from the two approaches are similar.

Table 4 shows the average years of education of the urban and rural workforce across the six rounds in our sample. The two features that emerge from the table are: (a) education attainment rates as measured by years of education were rising in both urban and rural sectors during this period; and (b) the rural-urban education gap shrank monotonically over this period. The average 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 rural workers improving faster.\(^{14}\)

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

\(^{13}\)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.

\(^{14}\)We have also examined rural-urban gaps in years of education by age and birth cohorts. While we don’t report those results here, our principal findings are (i) the gaps have been narrowing over time for all cohorts; and (ii) the gaps are smaller for younger and newer cohorts.
### Table 4: Education Gap: Years of Schooling

<table>
<thead>
<tr>
<th>Year</th>
<th>Average years of education</th>
<th>Relative education gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall Urban Rural</td>
<td>Urban/Rural</td>
</tr>
<tr>
<td>1983</td>
<td>3.02 5.83</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>(0.01) (0.03)</td>
<td>(0.01) (0.02)</td>
</tr>
<tr>
<td>1987-88</td>
<td>3.21 6.12</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>(0.01) (0.03)</td>
<td>(0.01) (0.02)</td>
</tr>
<tr>
<td>1993-94</td>
<td>3.86 6.85</td>
<td>2.98</td>
</tr>
<tr>
<td></td>
<td>(0.01) (0.03)</td>
<td>(0.02) (0.02)</td>
</tr>
<tr>
<td>1999-2000</td>
<td>4.36 7.40</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>(0.02) (0.04)</td>
<td>(0.02) (0.02)</td>
</tr>
<tr>
<td>2004-05</td>
<td>4.87 7.66</td>
<td>3.96</td>
</tr>
<tr>
<td></td>
<td>(0.02) (0.04)</td>
<td>(0.02) (0.02)</td>
</tr>
<tr>
<td>2009-10</td>
<td>5.70 8.42</td>
<td>4.72</td>
</tr>
<tr>
<td></td>
<td>(0.03) (0.04)</td>
<td>(0.03) (0.01)</td>
</tr>
</tbody>
</table>

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. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

movements in urban and rural areas are being driven by very different categories of education.

Figure 3: Education distribution

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

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

Panel (b) of Figure 3 shows the changes in the relative education distributions of the urban and rural workforce. For each survey year, the Figure shows the fraction of urban workers in each education category relative to the fraction of rural workers in that category. Thus, in 1983 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 shrunk rapidly. The trends in the other three categories were more muted as compared to the convergence in categories 4 and 5.

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

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Edu 1</td>
<td>0.352*** (0.003)</td>
<td>0.340*** (0.002)</td>
<td>0.317*** (0.002)</td>
<td>0.303*** (0.003)</td>
<td>0.263*** (0.003)</td>
<td>0.229*** (0.003)</td>
<td>-0.035*** (0.004)</td>
<td>-0.088*** (0.004)</td>
<td>-0.123*** (0.004)</td>
</tr>
<tr>
<td>Edu 2</td>
<td>0.003*** (0.001)</td>
<td>0.010*** (0.000)</td>
<td>0.021*** (0.001)</td>
<td>0.028*** (0.001)</td>
<td>0.037*** (0.001)</td>
<td>0.044*** (0.001)</td>
<td>0.018*** (0.001)</td>
<td>0.023*** (0.001)</td>
<td>0.041*** (0.001)</td>
</tr>
<tr>
<td>Edu 3</td>
<td>-0.047*** (0.001)</td>
<td>-0.038*** (0.000)</td>
<td>-0.016*** (0.000)</td>
<td>-0.001* (0.000)</td>
<td>0.012*** (0.001)</td>
<td>0.031*** (0.001)</td>
<td>0.031*** (0.001)</td>
<td>0.047*** (0.001)</td>
<td>0.078*** (0.001)</td>
</tr>
<tr>
<td>Edu 4</td>
<td>-0.092*** (0.001)</td>
<td>-0.078*** (0.001)</td>
<td>-0.065*** (0.001)</td>
<td>-0.054*** (0.001)</td>
<td>-0.041*** (0.001)</td>
<td>-0.020*** (0.001)</td>
<td>0.027*** (0.001)</td>
<td>0.045*** (0.001)</td>
<td>0.072*** (0.001)</td>
</tr>
<tr>
<td>Edu 5</td>
<td>-0.216*** (0.003)</td>
<td>-0.204*** (0.003)</td>
<td>-0.257*** (0.003)</td>
<td>-0.276*** (0.003)</td>
<td>-0.268*** (0.003)</td>
<td>-0.284*** (0.003)</td>
<td>-0.041*** (0.004)</td>
<td>-0.027*** (0.005)</td>
<td>-0.068*** (0.005)</td>
</tr>
</tbody>
</table>

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 while it significantly reduced the probability of belonging to categories 4-5. In category 3 the sign on the rural dummy had switched from negative to positive in 2004-05 and stayed that way in 2009-10.

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

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

3.1.2 Occupation Choices

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

The urban and rural occupation distributions have the obvious feature that urban areas have a much smaller fraction of the workforce in agrarian occupations while rural areas have a minuscule share of people working in white-collar jobs. Moreover, the urban sector clearly has a dominance in the share of the workforce in blue-collar jobs that pertain to both services and manufacturing. Importantly though, the share of blue-collar jobs has been rising in rural areas. In fact, as Panel (b)
Figure 4: Occupation distribution

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

The figure shows, the shares of both white-collar and blue-collar jobs in rural areas are rising faster than their corresponding shares in urban areas. Overall, these results suggest that the expansion of the rural non-farm sector has led to rural-urban occupation convergence.

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

Note that these results are consistent with the labor income decomposition results reported in section 2. There we showed that labor reallocation channel in white-collar jobs has contributed to a

---

15 Most of the relative increase in rural blue-collar jobs is accounted for by a two-fold expansion in the share of rural production and transportation jobs. While sales and service jobs in the rural areas expanded as well, the increase was much less dramatic. The relative expansion of rural white collar jobs was spread across most categories of white-collar jobs though the sharpest change was in administrative jobs.
Table 6: Marginal effect of rural dummy in multinomial probit regressions for occupations

<table>
<thead>
<tr>
<th></th>
<th>Panel (a): Marginal effects, unconditional</th>
<th>Panel (b): Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>white-collar</td>
<td>-0.196***</td>
<td>-0.206***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>blue-collar</td>
<td>-0.479***</td>
<td>-0.453***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>agri</td>
<td>0.675***</td>
<td>0.659***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

N

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

The widening of the labor income gap between urban and rural areas. This was because the employment distribution was becoming more uneven in these jobs, in terms of absolute differences, in line with the evidence in Table 6. In terms of the relative differences, however, the occupation distribution between urban and rural areas was converging in white-collar jobs, as Figure 4 shows. Blue-collar and agrarian jobs have shown convergence over time in both absolute and relative terms.

3.1.3 Decomposition of wage gaps

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

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

We consider several sets of attributes. First, we evaluate the role of individual demographic characteristics such as age, age squared, a dummy for the caste group (SC/ST or not) and a geographic zone of residence. The latter are constructed by grouping all Indian states into six regions – North, South, East, West, Central and North-East. We control for caste by including a dummy for

16The DFL method involves first constructing a counterfactual wage density function for urban individuals by giving them the attributes of rural households. This is done by a suitable reweighting of the estimated wage density function of urban households. The counterfactual density is then compared with the actual wage density to assess the contribution of the measured attributes to the observed wage gap.

17We choose to do the reweighting this way to avoid a common support problem, i.e., there may not be enough rural workers at the top end of the distribution to mimic the urban distribution.
whether or not the individual is an SC/ST in order to account for the fact that SC/STs tend to be disproportionately rural. Given that they are also disproportionately poor and have little education, controlling for SC/ST status is important in order to determine the independent effect of rural status on wages. Second, we add education to the set of attributes and obtain the incremental contribution of education to the observed wage convergence. Lastly, we evaluate the role played by differences in the occupation distribution for the urban-rural wage gaps.

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

Figure 5: Decomposition of Urban-Rural wage gaps for 1983 and 2009-10

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

---

18 Our occupation controls include 7 disaggregated occupation categories. Within the blue-collar jobs we distinguish sales workers, which include manufacturer’s agents, retail and wholesales merchants and shopkeepers, salesmen working in trade, insurance, real estate, and securities; as well as various money lenders; service workers, including hotel and restaurant staff, maintenance workers, barbers, policemen, firefighters; and production and transportation workers and laborers, which include among others miners, quarrymen, and various manufacturing workers. The white-collar group is disaggregated into three categories of workers as well. First group consists of professional, technical and related workers who include, for instance, chemists, engineers, agronomists, doctors and veterinarians, accountants, lawyers and teachers. The second is administrative, executive and managerial workers, which include, for example, officials at various levels of the government, as well as proprietors, directors and managers in various business and financial institutions. The third type of occupations consists of clerical and related workers. These are, for instance, village officials, book keepers, cashiers, various clerks, transport conductors and supervisors, mail distributors and communications operators. The seventh group is agricultural workers.
Figure 5 shows that demographic characteristics explain a small fraction of the urban-rural wage gap. Moreover, this fraction remains stable at around 0.1 along the entire distribution in both 1983 and 2009-10. In 1983 differences in education account for almost the entire wage gap at the bottom of the distribution, while differences in occupation explain the wage gap for the upper 50 percent of the distribution. Put differently, education and occupation choices can jointly account for almost the entire wage gap distribution in 1983. Turning to 2009-10 however, the picture is different. Here differences in education attainments between urban and rural workers explain a large fraction of the gap at the top end of the distribution (70th percentile and above). However, for those below the 60th percentile, covariates such as demographic characteristics, education and occupation choices systematically over-predict the actual wage gaps. This is particularly stark for the bottom 15 percent where the actual wage gap is negative while the demographic characteristics, education endowments and differences in occupations predict that the urban-rural gap should be positive 30 percent.

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

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

Our set of explained factors, as before, includes demographic characteristics such as individual’s

---

19 These results are not presented, but are available in the online appendix.
20 The inter-temporal decomposition at the mean is in the spirit of Smith and Welch (1989).
21 All decompositions are performed using a pooled model across rural and urban sectors as the reference model. Following Fortin (2000), we allow for a group membership indicator in the pooled regressions. We also used 1983 round as the benchmark sample. Details of the decomposition method can be found in the Appendix A.3.
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.\(^{22}\)

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

<table>
<thead>
<tr>
<th>(a). Change 1983 to 2009-10</th>
<th>(i) measured gap</th>
<th>(ii) explained</th>
<th>(iii) unexplained</th>
<th>(iv) education</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th quantile</td>
<td>-0.371***</td>
<td>-0.096***</td>
<td>-0.275***</td>
<td>-0.050***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.016)</td>
<td>(0.040)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>50th quantile</td>
<td>-0.568***</td>
<td>-0.292***</td>
<td>-0.366***</td>
<td>-0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.014)</td>
<td>(0.019)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>90th quantile</td>
<td>0.332***</td>
<td>0.229***</td>
<td>0.103***</td>
<td>0.284***</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.046)</td>
<td>(0.045)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>mean</td>
<td>-0.263***</td>
<td>-0.115***</td>
<td>-0.148***</td>
<td>-0.078***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.012)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b). Change in explained component</th>
<th>(i) explained</th>
<th>(ii) explained</th>
<th>(iii) unexplained</th>
<th>(iv) education</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th quantile</td>
<td>-0.096***</td>
<td>-0.060***</td>
<td>-0.036***</td>
<td>-0.049***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>50th quantile</td>
<td>-0.202***</td>
<td>-0.064***</td>
<td>-0.137***</td>
<td>-0.052***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>90th quantile</td>
<td>0.229***</td>
<td>0.060***</td>
<td>0.169***</td>
<td>0.084***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.021)</td>
<td>(0.040)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>mean</td>
<td>-0.115***</td>
<td>-0.032***</td>
<td>-0.083***</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.008)</td>
<td>(0.010)</td>
</tr>
</tbody>
</table>

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

Table 7 shows the results of the decomposition exercise. Bootstrapped standard errors are in parenthesis.\(^{23}\) Panel (a) shows the decomposition of the change in measured gap (column (i)) into the explained and unexplained components (columns (ii) and (iii)), as well as the part of the gap that is explained by education alone (column (iv)). The results indicate that the part of the wage gap that is explained by the included covariates varies from 25 percent for the bottom 10 percent to about 90 percent for the top 10 percent. Based on the explained component of the mean and median urban-rural wage gaps, at most 40 percent of the decrease in the gap is explained by the included covariates. Importantly, education alone accounts for the majority of the explained component along every point of the distribution.

Overall, our conclusion from the wage data is that wages have converged significantly between rural and urban India during since 1983 for all except the very top of the income distribution.

---

\(^{22}\)We do not include occupation amongst the explanatory variables since it is likely to be endogenous to wages. This is a problem for the RIF and OLS regressions, since they impute occupations based on the estimated coefficients, but less so for the DFL decomposition which uses reported individual occupations. In fact, in the original application of the DFL method, DiNardo, Fortin, and Lemieux (1996) do include occupation dummies in their estimation of the effects of unionization on the wage distribution in the USA. We followed their approach for the DFL decompositions.

\(^{23}\)In the computations we accounted for the complex survey design of the NSS data. We also use adjusted sampling weights that account for the pooled sampling (over rounds) in our decompositions. The variance is estimated using the resulting replicated point estimates (see Rao and Wu (1988) and Rao, Wu, and Yue (1992)).
Education has been an important contributor to these convergent patterns. However, on average over 60 percent of the convergence is due to unmeasured factors.

### 3.2 The Role of Migration

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

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

Table 8 shows the main patterns of migration for these three rounds. The first feature to note is that the number of recent migrants (those who migrated during the preceding five years) as a share of all full-time employed workers has declined from 7.2 percent in 1983 to 6.2 percent in 2007-08. Of these migrants, the largest single group were those who moved between rural areas, although the share of rural-to-rural migration in overall migration flows has declined from about 50 percent in 1983 to just below 38 percent in 2007-08. The share of urban migrants to rural areas has stayed relatively unchanged around 9-10 percent during this period. In contrast, urban areas have experienced an increase in migration inflows from both rural and urban areas. Thus, the share of rural-to-urban migration in total migration flows has increased from 22 percent in 1983 to about 30 percent in 2007-08. Urban-to-urban migration, which stood at 19 percent in 1983, rose to 23 percent in 2007-08. Interestingly, the majority of the increase in migration to urban areas took place in the latter half of our sample – since 1999-00.

To put these flows in perspective, the rural-to-urban migrants account for around 7 percent of the urban full-time workforce. This share has remained stable over the period. Note that the net flow

---

24 We identify migrants as individuals who reported that their place of enumeration is different from the last usual residence and who left their last usual place of residence within the previous five years. These variables are available on a consistent basis across the three survey rounds. For these individuals we also know the reason for leaving the last usual residence and its location.
of workers from rural to urban areas is lower as there is some reverse flow as well\textsuperscript{25} In particular, the net inflow of migrants from rural to urban areas in the five years preceding 1983 was about 4.5 percent of all urban full-time employed workers, while in 2007-08 the corresponding number was 5 percent. Expressed as a share of all full-time employed workers, net migration flows from rural to urban areas were about 1 percent in 1983 and 1.3 percent in 2010. While clearly not insignificant, the share of migrant workers from rural areas in the urban workforce is relatively small. Overall, between 1978 and 1983, about 4 million people moved from rural to urban locations, on net. Between 2003 and 2008, the net inflow of migrants into urban areas from rural locations was about 10 million people\textsuperscript{26}

### Table 8: Migration trends: 1983-2008

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

The last column of Table\textsuperscript{8} also shows that the majority of the rural-to-urban migration is job related. The rest is mostly for marriage reasons. The same is true for urban-to-urban migration flows. Interestingly, job related migration from rural to urban areas appears to have increased in 2007-08 relative to 1999-2000 despite the introduction of the rural employment program NREGA in 2005. Migration to rural areas is in equal proportion for job, marriage and other reasons\textsuperscript{27}

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

\textsuperscript{25} These bidirectional migration flows were emphasized in Young (2012).

\textsuperscript{26} These numbers were obtained by multiplying the net flow as a share of full-time employed workers by the share of full time employment in the labor force in that year equal to 0.89; the resulting number was multiplied by the labor force participation in 1983 equal to 0.64. All these shares were computed using 1983 NSS survey. Lastly, the resulting number was multiplied by the population in India which was equal to 683.3 million people according to 1981 Census. The corresponding numbers for 2007-08 were: the share of full time employment in the labor force \(0.90\); labor force participation \(0.64\); population in 2011 Census \(1210.2\) million.

\textsuperscript{27} Other reasons include natural disaster, social problems, displacement, housing based movement, health care, etc..
Table 9 reports our results for (log) wages. We find that migration flows from urban areas have coefficients that are positive and significant, suggesting that urban migrants earn more than the benchmark group – urban non-migrants. Migrants from rural areas, in contrast, earn less than urban non-migrants, but the difference is significant mainly for rural-to-rural migration flows. Note also that the negative effects on wages for this group is declining over time, in line with the aggregate wage convergence. Wages of migrants who moved from rural to urban areas are no different than the wages of urban non-migrants. These results apply to both mean and median wages. Do these migration flows contribute to the urban-rural wages gap convergence? A comparison of regression coefficients on the rural dummy in Table 9 and in the benchmark specification without migration flows dummies in Table 3 reveals that they are practically the same. We find that this result also holds for individuals at the two ends of the wage distribution (see Table A1 in Appendix A.4). This suggests that the wage gap between urban and rural non-migrants has been narrowing at the same rate as the overall urban-rural gap.

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 the migration decision itself is endogenous to wage gaps between rural and urban areas. Such an analysis is left for future research.

---

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.
4 The Role of Aggregate Shocks

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

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

To account for the data we augment the standard structural transformation model with two key supply-side effects. First, in conformity with the data, we allow for a rise in non-farm productivity relative to farm productivity. Second, we allow for differential labor force growth in urban and rural areas. This too is a key feature of Indian data for the period 1983-2010. Both these features induce an increase in the relative supply of labor to non-farm activities and can, therefore, potentially overturn the counterfactual movements in the agricultural terms of trade and wages implied by the standard demand side channels of structural transformation.

4.1 Key aggregate facts

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

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

The ongoing process of structural transformation of the Indian economy during the 1983-2010 period can be seen through Figure 6 which shows employment shares (panel (a)) and output shares (panel (b)) in agriculture, and non-agriculture. As is easy to see, agriculture has been releasing workers, and its share of output has been declining over time. The non-agricultural sector, on the other hand, has expanded both as a share of employment and as a share of output. These are the textbook features of structural transformation. More precisely, the share of agriculture in total employment has declined from 63 percent in 1983 to 49 percent in 2010. This mirrors almost exactly the structural transformation in occupations, where the share of agricultural jobs has declined from 63 percent in 1983 to 50 percent in 2010. The decline of agriculture in total output was even more pronounced with its share falling from 33 percent in 1983 to 18 percent in 2010.

Figure 6: Employment and output distribution

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

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

Data description and details on how TFP was computed can be found in Appendix A.5.
(a) Labor productivity per worker
Notes: Panel (a) shows sectoral labor productivity during 1983-2010 period, while panel (b) shows sectoral total factor productivity (TFP) during the same time period.

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

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

Another fact that we already discussed but highlight here again was the gradual increase in the share of urban labor in the overall Indian labor force. As we showed in Table 1 using the NSS data, the proportion of the urban full time employed labor grew from 22 percent of all full time employed workers to 27 percent between 1983 and 2010. This increase in urban labor finds an echo in the Census figures for India for the overall population where the urban population share rose from 23 to 31 percent between 1981 and 2011. Based on this we conclude that labor supply in urban India

---

30 These numbers were obtained using nominal and real output series from the National Accounts Statistics provided by the Ministry of Statistics and Programme Implementation (MOSPI) of Government of India.
grew at a faster rate than in rural India during this period.

### 4.2 A Structural Explanation

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

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

\[
V = \frac{c_i^{1-\rho}}{1-\rho}, \quad i = R, U
\]

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

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

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

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

\[
1 = l_{iA} + l_{iS}
\]

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

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

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

\( y_i \) denotes total income of household \( i = U, R \).

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

\[
Y_A = AL_A \\
Y_S = SL_S,
\]

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

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

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

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

Optimality for type-\( i \) households implies that

\[
w_{iA} = w_{iS} - \tau_i
\]
\[ c_{iA} = (1 - \theta) \bar{c} + \theta y_i \]  
\[ pc_{iS} = (1 - \theta) (y_i - \bar{c}) , \]

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

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

\[ p = \frac{MPL_{RA}}{MPL_{RS}} = \frac{MPL_{UA}}{MPL_{US}} \]  

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

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

\[ c_{UA} L_U + c_{RA} L_R = Y_A \]  
\[ c_{US} L_U + c_{RS} L_R = Y_S \]  

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

### 4.2.1 Characterizing the Equilibrium

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

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

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

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

\[ k_A s_A + k_S (1 - s_A) = k \]
Hence,
\[ s_A = \frac{k - k_S}{k_A - k_S} \]

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

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

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

\[ k_A = \mu(k_S) \]

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

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

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

\[
S \left[ \beta_S k_S^{\phi_S} + 1 - \beta_S \right]^{\frac{1}{\phi_S}} \left[ \frac{\mu(k_S) - k}{\mu(k_S) - k_S} \right] \tag{4.11}
\]

This is a two-equation system in two unknowns – \( k_S \) and \( p \). Equilibrium solutions for the rest of the endogenous variables are derived recursively from the solutions for \( k_S \) and \( p \). Note that equation (4.10) comes from combining the firm optimality conditions \( pMPL_{US} = w_{US} \) and \( MPL_{UA} = w_{UA} \) with the household optimality condition \( w_{UA} = w_{US} - \tau_U \). Equation (4.11) arises from combining the household budget constraints with the market clearing conditions for the two goods.

### 4.2.2 A Special Case

In order to build intuition regarding the mechanisms at play in this model as well as the effects of exogenous shocks on factor allocations and prices, we now analytically examine a special case of the model described above by imposing the following two conditions:
**Condition 4.1** The labor aggregators in the two sectors are of the Cobb-Douglas form given by

\[ L_j = L_j^{\beta_j} L_j^{1-\beta_j}, \quad j = A, S. \]  

**Condition 4.2** There are no training costs of labor for working in the non-agricultural sector \( S \), i.e., \( \tau_U = \tau_R = 0 \).

In this case the solution for \( k_A \) is given by

\[ k_A = \gamma k_S, \quad \gamma = \frac{\beta_A}{1-\beta_A} \frac{1-\beta_S}{\beta_S} \]  

Moreover, the equilibrium system is given by

\[ p = \frac{A \gamma A^\gamma A^{-1} k_S^A}{S \beta Sk_S^S} \]  

\[ p = \left( \frac{1-\theta}{\theta} \right) \left[ \frac{A \gamma A^\gamma A^{-1} \frac{k - k_S}{\gamma - 1} - \bar{c}(1+k)}{S \beta Sk_S^S}\right] \]  

Keeping in mind the empirical reality of rural labor being primarily employed in agriculture, we shall assume throughout the rest of the paper that the agricultural sector uses rural labor more intensively so that \( \frac{L_{UA}}{L_{RA}} = k_A < k_S = \frac{L_{US}}{L_{RS}} \). Hence, we shall assume that \( \beta_A < \beta_S \) and \( \gamma < 1 \).

The equilibrium solution can be characterized by using equations (4.14) and (4.15) to solve for \( k_S \). This gives

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

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

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

**Proposition 4.3** Under Conditions 4.1, 4.2, an increase in \( k \), the stock of urban labor relative to rural labor, has the following effects: (a) it raises the urban to rural labor ratios in both sectors; (b) it reduces the relative price \( p \) of good \( S \); (c) it raises the rural wage while reducing the urban wage; (d) it has an ambiguous effect on the allocation of rural labor to sector \( A \); and (e) has an ambiguous effect on the output share of good \( A \).

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

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

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

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

**Proposition 4.4** Under Conditions 4.1, 4.2, an increase in agricultural productivity \( A \): (a) reduces the urban to rural labor ratios in both sectors; (b) raises the relative price \( p \) of good \( S \); (c) reduces the rural wage while raising the urban wage; (d) reduces the allocation of rural labor to sector \( A \); and (e) reduces the output share of good \( A \).

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

The logic underlying Proposition 4.4 is fairly standard given that this is a model with minimum consumption in the agricultural sector. This introduces differential income elasticities of the two goods. A rise in agricultural productivity \( A \) raises overall income which induces a larger increase in the demand for good \( S \) relative to the rise in demand for good \( A \). Consequently, the price of the non-agricultural good \( p \) rises. As the economy shifts towards the non-agricultural sector, it begins to reallocate both urban and rural labor from agriculture to non-agriculture. Since agriculture is more rural labor intensive, it releases proportionately more rural labor which in turn reduces the urban to rural labor ratio in both sectors. The greater relative employment of rural labor in both sectors
raises the returns to urban labor. Hence the urban wage rises while the rural wage falls.

Figure 10: Comparative static effects on $k_S$

(a) Rise in relative urban labor supply  
(b) Rise in Agricultural TFP

Propositions 4.3 and 4.4 highlight three important features of our model economy. First, we need shocks to both productivity and the relative supply of urban to rural labor in order to jointly explain the observed changes in relative wages, agricultural terms of trade and the structural transformation. Increases in the relative endowment of urban labor gets the relative wage and terms of trade movements right but has ambiguous implications for the structural transformation of the economy. On the other hand, an increase in agricultural productivity generates the structural transformation but has counterfactual predictions for the wage gap as well as the terms of trade. Second, without the minimum consumption requirement the model cannot generate any structural transformation in this economy. This can be checked by setting $\bar{c} = 0$ in equation (4.16). Third, the sectoral urban to rural labor allocations are independent of the non-agricultural productivity parameter $S$. However, the relative price $p$ does depend on $S$. In particular, suppose $A$ and $S$ both rise but $A/S$ declines. In this case $p$ could fall (i.e., the agricultural relative price could rise) in response to an increase in $A$ as long as the fall in $\frac{A}{S}$ is large enough to offset the fall in $k_S$. This is easily ascertained from the expression $p = \frac{A^\frac{\beta}{\gamma} - 1}{S^\frac{\beta}{\gamma} k_S^s}$, which must hold in equilibrium. Thus, differential sectoral productivity growth is important in getting the relative price movements right.

4.3 Quantitative Results

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

4.3.1 Calibration for 1983

To calibrate the urban and rural share in the labor force we use Census of India and NSS data. The Census is conducted every 10 years on the first year of each decade. Thus, in 1981 the total population of India was 683.3 million people, of which 525.6 million lived in rural areas and 157.7 million lived in urban areas. The urban population included 8 million migrants who moved to urban areas from rural locations in the preceding 10 years, on net. Since we abstract from rural-urban migration in the model, we adjust the population numbers from the Census for the net migration flow between the two areas. Namely, we subtract the net migration flow from the urban population and add it to the rural population. To obtain the number of people in the labor force we multiply the resulting numbers by the share of working age population, which in rural areas was equal to 0.53 and in urban areas equal to 0.59 in 1983, according to the 1981 Census. Lastly, we multiply the resulting numbers by the labor force participation rate computed from the 1983 NSS round, which was equal to 0.66 in rural areas and 0.59 in urban areas. These calculations give us migration-adjusted labor force in 1983 equal to 185.57 million people in rural areas and 51.80 million people in urban areas. The resulting rural labor force share is 78 percent and urban labor force share is 22 percent in 1983.

The sectoral labor force shares in each location are obtained using the NSS data and are borrowed from panel (a) of Figure 4. We find that 78 percent of the rural labor force worked in agriculture, while only 11 percent of the urban labor force was employed in agricultural jobs in 1983. Parameters $\beta_A$ and $\beta_S$ are calibrated jointly with elasticity parameters $\phi_A$ and $\phi_S$ to match this distribution. These parameters also affect the wages earned by each labor type in each sector. The targets for wages are estimated from the 1983 NSS round and are summarized in Table 10 below.

First, we match the within-agriculture and within-non-agriculture wage gaps between urban and
rural areas. Those "within" gaps stood at -7 percent and 11 percent, respectively. "Between" gaps which capture the wage premium in non-agricultural jobs relative to agricultural jobs stood at 69 percent in rural areas and 87 percent in urban areas in 1983. We calibrate training costs $\tau_U$ and $\tau_R$ to match these "between" wage gaps. Of course, other parameters affect the "between" wage gaps as well, thus the training costs are calibrated jointly with the other parameters.

Lastly, we calibrate the productivity levels in the agricultural and non-agricultural sectors to match the output share of agriculture in total GDP in India in 1983 at 33 percent. The agricultural consumption share $\theta$ together with the minimum consumption parameter $\bar{c}$ are calibrated to match the share of food expenditures in the total consumption expenditures of Indian households equal to 45 percent. This number was computed as the ratio of the official poverty line in India in 1983 to the average per capita consumption expenditure of Indian households in the same year.

<table>
<thead>
<tr>
<th>Table 10: Data and model: 1983</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
</tr>
<tr>
<td>employment shares:</td>
</tr>
<tr>
<td>$L_U$</td>
</tr>
<tr>
<td>$L_{RA}$</td>
</tr>
<tr>
<td>$L_{RS}$</td>
</tr>
<tr>
<td>$L_{UA}$</td>
</tr>
<tr>
<td>$L_{US}$</td>
</tr>
<tr>
<td>wage gaps:</td>
</tr>
<tr>
<td>within $A$</td>
</tr>
<tr>
<td>within $S$</td>
</tr>
<tr>
<td>$R$ between</td>
</tr>
<tr>
<td>$U$ between</td>
</tr>
<tr>
<td>overall mean</td>
</tr>
<tr>
<td>overall median</td>
</tr>
<tr>
<td>aggregates:</td>
</tr>
<tr>
<td>$S/A$ relative price</td>
</tr>
<tr>
<td>$Y$ share of $A$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of U labor in total labor force</td>
<td>$L_U$</td>
</tr>
<tr>
<td>Urban labor weight in A sector</td>
<td>$\beta_A$</td>
</tr>
<tr>
<td>Urban labor weight in S sector</td>
<td>$\beta_S$</td>
</tr>
<tr>
<td>Training cost for U households</td>
<td>$\tau_U$</td>
</tr>
<tr>
<td>Training cost for R households</td>
<td>$\tau_R$</td>
</tr>
<tr>
<td>Elasticity of substitution between R and U labor in A</td>
<td>$\phi_A$</td>
</tr>
<tr>
<td>Elasticity of substitution between R and U labor in S</td>
<td>$\phi_S$</td>
</tr>
<tr>
<td>A consumption share</td>
<td>$\tilde{\theta}$</td>
</tr>
<tr>
<td>Minimum consumption</td>
<td>$\bar{c}$</td>
</tr>
</tbody>
</table>

### 4.3.2 Results

How much of the observed convergence in urban-rural wages is accounted for by changes in sectoral productivity and differential labor supply growth in the two sectors? To answer this question we re-calibrate the urban and rural labor force shares to their values in 2011 using the Census data. As before we exclude any growth in urban population arising from net inflow of migrants from rural areas. According to the 2011 Census, the urban population in India was 377.1 million while the rural population was 833.1 million. The net migration flow from rural to urban areas in the 10 years preceding 2011 was 20 million. Adjusting the resulting numbers by the share of working age population (equal to 0.6 in rural areas and to 0.659 in urban areas in 2011 Census) and labor force participation rates (equal to 0.66 in rural areas and 0.59 in urban areas in 2010 NSS data), gives us an urban population share of 29 percent and rural population share of 71 percent.

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

We feed the changes in labor force shares and sectoral productivity growth into the model while keeping all other parameters unchanged. The results are summarized in Table 12. First, the share of the workforce employed in agriculture declines by 5.7 percentage points for rural and 3 percentage points for urban workers. In the data, the agricultural share of rural jobs declined by 12 percentage points while the urban share of agricultural jobs fell by 4 percentage points between 1983 and 2010. Hence, shocks to labor force growth and sectoral productivity account for 47 percent of the observed convergence.
Table 12: Model and data: 1983 versus 2010

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>employment shares:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_U$</td>
<td>0.220</td>
<td>0.220</td>
<td>0.290</td>
<td>0.290</td>
<td>0.070</td>
<td>0.070</td>
</tr>
<tr>
<td>$L_R$</td>
<td>0.780</td>
<td>0.853</td>
<td>0.660</td>
<td>0.796</td>
<td>-0.120</td>
<td>-0.057</td>
</tr>
<tr>
<td>$L_R$</td>
<td>0.220</td>
<td>0.147</td>
<td>0.340</td>
<td>0.204</td>
<td>0.120</td>
<td>0.057</td>
</tr>
<tr>
<td>$L_U$</td>
<td>0.110</td>
<td>0.109</td>
<td>0.070</td>
<td>0.079</td>
<td>-0.040</td>
<td>-0.030</td>
</tr>
<tr>
<td>$L_U$</td>
<td>0.890</td>
<td>0.891</td>
<td>0.930</td>
<td>0.921</td>
<td>0.040</td>
<td>0.030</td>
</tr>
<tr>
<td><strong>wage gaps:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within $A$</td>
<td>0.932</td>
<td>0.999</td>
<td>1.000</td>
<td>0.927</td>
<td>0.068</td>
<td>-0.072</td>
</tr>
<tr>
<td>within $S$</td>
<td>1.079</td>
<td>1.032</td>
<td>1.000</td>
<td>0.962</td>
<td>-0.079</td>
<td>-0.070</td>
</tr>
<tr>
<td>$R$ between</td>
<td>1.674</td>
<td>1.482</td>
<td>1.518</td>
<td>1.253</td>
<td>-0.156</td>
<td>-0.228</td>
</tr>
<tr>
<td>$U$ between</td>
<td>1.815</td>
<td>1.531</td>
<td>1.536</td>
<td>1.301</td>
<td>-0.279</td>
<td>-0.230</td>
</tr>
<tr>
<td>overall mean</td>
<td>1.509</td>
<td>1.374</td>
<td>1.270</td>
<td>1.126</td>
<td>-0.239</td>
<td>-0.248</td>
</tr>
<tr>
<td>overall median</td>
<td>1.586</td>
<td>1.374</td>
<td>1.126</td>
<td>1.126</td>
<td>-0.460</td>
<td>-0.248</td>
</tr>
<tr>
<td><strong>aggregates:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S/A$ relative price</td>
<td>1.000</td>
<td>7.007</td>
<td>0.752</td>
<td>5.938</td>
<td>-0.248</td>
<td>-0.153</td>
</tr>
<tr>
<td>$Y$ share of $A$</td>
<td>0.330</td>
<td>0.595</td>
<td>0.180</td>
<td>0.538</td>
<td>-0.455</td>
<td>-0.095</td>
</tr>
</tbody>
</table>

Decline in agricultural employment in rural areas and 75 percent of the observed decline in urban employment in agricultural jobs during the 1983-2010 period.

Second, wage gaps between urban and rural labor decline following the shocks. The "within" gaps between urban and rural wages fall by 0.072 in agriculture and by 0.07 in non-agriculture. The "between" non-agriculture and agriculture wage gaps also decline by 0.23 in urban areas and 0.228 in rural areas. The overall wage gap between urban and rural areas falls by 0.25 in response to the differential population growth in rural and urban areas and sectoral productivity growth. Given that the median wage gap in the data declined by 46 percentage points, these two shocks then account for about 54 percent of the observed decline in median urban-rural wage gaps. Since 60 percent of the median wage convergence was unexplained by standard covariate of wages, the two shocks account for 92 percent ($\frac{0.25}{0.6}(0.46)$) of the unexplained decline in the median wage gap.

Third, the model predicts a 15 percent decline in the relative price of non-agricultural goods, which is qualitatively consistent with the data but is smaller than the 25 percent decline in the data. We conclude that the two aggregate shocks we emphasized account for about 2/3 of the observed fall in the price of non-agricultural goods in India during 1983-2010 period.

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

Table 13: Decomposition of labor income gap in the model

<table>
<thead>
<tr>
<th></th>
<th>wage component</th>
<th>labor reallocation</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>within</td>
<td>between</td>
<td>component</td>
</tr>
<tr>
<td>non-agri</td>
<td>-0.064</td>
<td>-0.155</td>
<td>-0.008</td>
</tr>
<tr>
<td>agrarian</td>
<td>-0.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>-0.085</td>
<td>-0.155</td>
<td>-0.008</td>
</tr>
<tr>
<td>% contribution</td>
<td>34.3</td>
<td>62.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>

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

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

Overall, our results suggest that aggregate factors have played an important role in urban-rural convergence in the past 27 years. Faster growth of non-agricultural productivity and a relatively faster expansion of the urban labor force can jointly account for over 90 percent of the wage convergence left unexplained by standard covariates of wages, 2/3 of the observed decline in the relative price of non-agricultural goods, and a substantial part of sectoral employment convergence between urban and rural areas in India during this period. Furthermore, these factors induce both within- and between-wage convergence, in line with data.

5 Conclusion

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

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

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

Given the lack of explanatory power of conventional worker characteristics, we then examined the possible role of aggregate shocks to the Indian economy during this period. Using a two-factor, two-sector model of structural transformation we showed both analytically and quantitatively that differential growth in urban and rural labor supply along with differential productivity shocks to agriculture and non-agriculture can potentially explain a large part of the observed convergence. In particular, our quantitative results suggest that around 90 percent of the unexplained wage convergence between rural and urban areas can be jointly accounted for by these two factors.

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


A Appendix

A.1 Data

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

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

In our data work, we only consider individuals that report their 3-digit occupation code and education attainment level. Occupation codes are drawn from the National Classification of Occupation (NCO) – 1968. We use the "usual" occupation code reported by an individual for the usual principal activity over the previous year (relative to the survey year). The dataset does not contain information on the years of schooling for the individuals. Instead it includes information on general education categories given as (i) not literate -01, literate without formal schooling: EGS/ NFEC/ AEC -02, TLC -03, others -04; (ii) literate: below primary -05, primary -06, middle -07, secondary -08, higher secondary -10, diploma/certificate course -11, graduate -12, postgraduate and above -13. We aggregate those into five similarly sized groups as discussed in the main text. We also convert these categories into years of education. The mapping we used is discussed in the main text.
The NSS only reports wages from activities undertaken by an individual over the previous week (relative to the survey week). Household members can undertake more than one activity in the reference week. For each activity we know the "weekly" occupation code, number of days spent working in that activity, and wage received from it. We identify the main activity for the individual as the one in which he spent maximum number of days in a week. If there are more than one activities with equal days worked, we consider the one with paid employment (wage is not zero or missing). Workers sometimes change the occupation due to seasonality or for other reasons. To minimize the effect of transitory occupations, we only consider wages for which the weekly occupation code coincides with usual occupation (one year reference). We calculate the daily wage by dividing total wage paid in that activity over the past week by days spent in that activity.

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

A.2 Decomposition of labor income convergence

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

\[ w^R_t = w^R_{1t} L^R_{1t} + w^R_{2t} L^R_{2t} + w^R_{3t} L^R_{3t}, \]
\[ w^U_t = w^U_{1t} L^U_{1t} + w^U_{2t} L^U_{2t} + w^U_{3t} L^U_{3t}, \]

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

The relative labor income gap in period t is

\[ \frac{w^U_t - w^R_t}{w^R_t} = \frac{(w^U_{1t} L^U_{1t} + w^U_{2t} L^U_{2t} + w^U_{3t} L^U_{3t}) - (w^R_{1t} L^R_{1t} + w^R_{2t} L^R_{2t} + w^R_{3t} L^R_{3t})}{w^R_t}. \]

Adding and subtracting average labor income for each occupation (denoted by \(w^i_t, i = 1, 2, 3\)), we can write the expression above as

\[ \frac{w^U_t - w^R_t}{w^R_t} = \frac{(w^U_{1t} - w^R_{1t}) L^U_{1t} + (w^U_{2t} - w^R_{2t}) L^U_{2t} + (w^U_{3t} - w^R_{3t}) L^U_{3t}}{w^R_t} - \frac{(w^R_{1t} - w^U_{1t}) L^R_{1t} + (w^R_{2t} - w^U_{2t}) L^R_{2t} + (w^R_{3t} - w^U_{3t}) L^R_{3t}}{w^R_t}. \]
\[
+ \frac{w_{1t}(L_{1t}^U - L_{1t}^R) + w_{2t}(L_{2t}^U - L_{2t}^R) + w_{3t}(L_{3t}^U - L_{3t}^R)}{w_t^R}.
\]

Now we look at the change in the relative gap between periods \(t\) and \(t - 1\). To simplify the notation, let \(\mu_{it}^j = \left(\frac{w_{jt}^L - w_{jt}^R}{w_t^R}\right)\), with \(i = 1, 2, 3\); and \(j = U, R\) and \(\eta_{it} = w_{it}/w_t^R\), \(i = 1, 2, 3\).

Then the change in the relative gap can be written as

\[
\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \mu_{1t}^U L_{1t}^U + \mu_{2t}^U L_{2t}^U + \mu_{3t}^U L_{3t}^U - (\mu_{1t}^R L_{1t}^R + \mu_{2t}^R L_{2t}^R + \mu_{3t}^R L_{3t}^R) + \eta_{1t}(L_{1t}^U - L_{1t}^R) + \eta_{2t}(L_{2t}^U - L_{2t}^R) + \eta_{3t}(L_{3t}^U - L_{3t}^R) - (\mu_{1t-1}^U L_{1t}^U - \mu_{1t-1}^R L_{1t-1}^R + \mu_{2t-1}^U L_{2t-1}^U + \mu_{2t-1}^R L_{2t-1}^R + \mu_{3t-1}^U L_{3t-1}^U - \mu_{3t-1}^R L_{3t-1}^R) - \eta_{1t-1}(L_{1t-1}^U - L_{1t-1}^R) - \eta_{2t-1}(L_{2t-1}^U - L_{2t-1}^R) - \eta_{3t-1}(L_{3t-1}^U - L_{3t-1}^R) .
\]

Define \(\bar{x}_t = (x_t + x_{t-1})/2\), and \(\Delta x_t = x_t - x_{t-1}\). Now, adding and subtracting \(\left(\mu_{it}^j - \mu_{it-1}^j\right)\bar{L}^j_{it}\), where \(\bar{L}^j_{it} = \left(L^j_{it} + L^j_{it-1}\right)/2\), and \(i = 1, 2, 3\) and \(j = U, R\) and collecting the terms in the first and third lines above; adding and subtracting \(\bar{\eta}_{it} \left[(L_{it}^U - L_{it}^R) - (L_{it}^R - L_{it}^U)\right]\), where \(\bar{\eta}_{it} = (\eta_{it} + \eta_{it-1})/2\) and \(i = 1, 2, 3\) and collecting the terms in the second and fourth lines above, we get

\[
\frac{w_t^U - w_t^R}{w_t^R} - \frac{w_{t-1}^U - w_{t-1}^R}{w_{t-1}^R} = \Delta \mu_{1t}^U \bar{L}_{1t}^U + \Delta \mu_{2t}^U \bar{L}_{2t}^U + \Delta \mu_{3t}^U \bar{L}_{3t}^U - \Delta \mu_{1t}^R \bar{L}_{1t}^R - \Delta \mu_{2t}^R \bar{L}_{2t}^R - \Delta \mu_{3t}^R \bar{L}_{3t}^R + \Delta \mu_{1t}^U \bar{L}_{1t}^U + \Delta \mu_{2t}^R \bar{L}_{2t}^R + \Delta \mu_{3t}^U \bar{L}_{3t}^U - \Delta \mu_{2t}^R \bar{L}_{2t}^R + \Delta \mu_{3t}^U \bar{L}_{3t}^U + \Delta \mu_{1t}^R \bar{L}_{1t}^R + \Delta \mu_{2t}^R \bar{L}_{2t}^R + \Delta \mu_{3t}^R \bar{L}_{3t}^R + \bar{\eta}_{1t} \Delta \left(L_{1t}^U - L_{1t}^R\right) + \bar{\eta}_{2t} \Delta \left(L_{2t}^U - L_{2t}^R\right) + \bar{\eta}_{3t} \Delta \left(L_{3t}^U - L_{3t}^R\right) + \left(L_{1t}^U - L_{1t}^R\right) \Delta \eta_{1t} + \left(L_{2t}^U - L_{2t}^R\right) \Delta \eta_{2t} + \left(L_{3t}^U - L_{3t}^R\right) \Delta \eta_{3t}.
\]

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

\[
\Delta \mu_{1t}^U \left(\bar{L}_{1t}^U - \bar{L}_{1t}^U\right) + \Delta \mu_{2t}^U \left(\bar{L}_{2t}^U - \bar{L}_{2t}^U\right) + \Delta \mu_{3t}^U \left(\bar{L}_{3t}^U - \bar{L}_{3t}^U\right) - \Delta L_{1t}^R \left(\bar{L}_{1t}^R - \bar{L}_{1t}^R\right) - \Delta L_{2t}^R \left(\bar{L}_{2t}^R - \bar{L}_{2t}^R\right) ,
\]

and the third row as

\[
(\bar{\eta}_{1t} - \bar{\eta}_{3t}) \Delta \left(L_{1t}^U - L_{1t}^R\right) + (\bar{\eta}_{2t} - \bar{\eta}_{3t}) \Delta \left(L_{2t}^U - L_{2t}^R\right) ,
\]
and the fourth row as
\[
\left( L_{1t}^U - L_{1t}^R \right) [\Delta \eta_{1t} - \Delta \eta_{3t}] + \left( L_{2t}^U - L_{2t}^R \right) [\Delta \eta_{2t} - \Delta \eta_{3t}] .
\]

Thus, the change in the relative labor income gap becomes
\[
\frac{w_t^U - w_t^R}{w_t^R - w_{t-1}^R} = \Delta \mu_{1t}^U L_{1t}^U + \Delta \mu_{2t}^U L_{2t}^U + \Delta \mu_{3t}^U L_{3t}^U - \Delta \mu_{1t}^R L_{1t}^R - \Delta \mu_{2t}^R L_{2t}^R - \Delta \mu_{3t}^R L_{3t}^R \\
\quad + \Delta L_{1t}^U (\tilde{\mu}_{1t}^U - \tilde{\mu}_{3t}^U) + \Delta L_{2t}^U (\tilde{\mu}_{2t}^U - \tilde{\mu}_{3t}^U) - \Delta L_{1t}^R (\tilde{\mu}_{1t}^R - \tilde{\mu}_{3t}^R) - \Delta L_{2t}^R (\tilde{\mu}_{2t}^R - \tilde{\mu}_{3t}^R) \\
\quad + (\eta_{1t} - \eta_{3t}) \Delta (L_{1t}^U - L_{1t}^R) + (\eta_{2t} - \eta_{3t}) \Delta (L_{2t}^U - L_{2t}^R) + \left( L_{1t}^U - L_{1t}^R \right) [\Delta \eta_{1t} - \Delta \eta_{3t}] + \left( L_{2t}^U - L_{2t}^R \right) [\Delta \eta_{2t} - \Delta \eta_{3t}] 
\]
(A1)

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

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

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

Our econometric model for sector s and round t is given by
\[
y_{st} = X_{st}' \beta_{st} + e_{st}, \quad s = 1, 2; \text{ and } t = 1, 2,
\]
where \( y_{st} \) is a vector of outcomes (log wage) while \( X_{st} \) is the matrix of regressors for sector s in round t. Here \( \beta_{st} \) is a coefficient vector, and \( e_{st} \) is the vector of residuals. The differential in expected outcomes between urban and rural sectors in round t is then given by:
\[
\Delta y_t^c = \Delta X_t' \tilde{\beta}_t + X_{1t}' (\beta_{1t} - \tilde{\beta}_t) + X_{2t}' (\tilde{\beta}_t - \beta_{2t}),
\]
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

$$
E_t = \Delta X_t' \tilde{\beta}_t, \quad t = 1, 2, \\
U_t = X_{1t}'(\beta_{1t} - \tilde{\beta}_t) + X_{2t}'(\tilde{\beta}_t - \beta_{2t}), \quad t = 1, 2.
$$

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

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

These differentials are reported in Panel (a) of Table 7. Note, however, that inter-temporal changes in the explained and unexplained components may be due to changes in either the attribute gaps or in the returns to those attributes. We focus on the inter-temporal decomposition of the explained part, $\Delta E$, in the main text. $\Delta E = \Delta X_2' \tilde{\beta}_2 - \Delta X_1' \tilde{\beta}_1$ can be broken down as

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

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

### A.4 Distributional effects of migration

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

### A.5 Measuring Total Factor Productivity

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

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

where $K_i$ denotes the stock of physical capital, $H_i$ is the amount of human capital-augmented labor used in production, and $Z_i$ is a labor-augmenting measure of productivity. We assume that each
### Table A1: Wage gaps: Accounting for migration

<table>
<thead>
<tr>
<th></th>
<th>10th percentile</th>
<th>90th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1983</td>
<td>1999-00</td>
</tr>
<tr>
<td>rural</td>
<td>-0.192***</td>
<td>0.006</td>
</tr>
<tr>
<td>rural-to-urban</td>
<td>0.086***</td>
<td>0.116***</td>
</tr>
<tr>
<td>urban-to-urban</td>
<td>0.149***</td>
<td>0.134***</td>
</tr>
<tr>
<td>rural-to-rural</td>
<td>-0.175***</td>
<td>-0.046*</td>
</tr>
<tr>
<td>urban-to-rural</td>
<td>-0.029</td>
<td>0.141***</td>
</tr>
</tbody>
</table>

|                | (0.011)         | (0.009)         | (0.013)         | (0.015)         | (0.025)         | (0.031)         |
| rural-to-urban | (0.022)         | (0.020)         | (0.031)         | (0.048)         | (0.055)         | (0.068)         |
| urban-to-urban | (0.016)         | (0.019)         | (0.028)         | (0.057)         | (0.112)         | (0.132)         |
| rural-to-rural | (0.031)         | (0.026)         | (0.041)         | (0.033)         | (0.058)         | (0.072)         |
| urban-to-rural | (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.

A unit of homogeneous labor $L_i$ has received $E_i$ years of education. Therefore, $H_i = E_iL_i$. $\alpha_i$ is capital income share, while $\beta_i$ is labor income share in sector $i = A, NA$.

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

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