Empir Econ (2010) 39:371–394
DOI 10.1007/s00181-009-0308-4

ORIGINAL PAPER

Economic liberalization and urban–rural inequality in India: a quantile regression analysis

Rubiana Chamarbagwala

Received: 24 September 2008 / Accepted: 5 May 2009 / Published online: 22 August 2009

Abstract We examine India's urban–rural inequality in welfare in 1993–1994 and 2004, a period which coincides with the country's economic liberalization reforms and rapid economic growth. Using real monthly per capita household consumption expenditure as our measure of welfare, we estimate quantile regressions to analyze the urban–rural welfare gap across the entire welfare distribution. While the urban–rural welfare gap was fairly convex across the welfare distribution in 1993–1994, it became more concave in 2004, with the gap narrowing for the lowest and highest quintiles and widening for the middle three quintiles. The urban–rural gap in returns to all levels of education widened substantially for the bottom four quintiles but became increasingly negative for the top quintile. Applying the Machado and Mata (J Appl Econom 20:445–465, 2005) decomposition technique to decompose the urban–rural welfare gap at each percentile, we find that for the bottom 40% of the distribution, differences in the distribution of covariates became less important while differences in the distribution of returns to covariates became more important in explaining the gap. The opposite was true for the top 40% of the distribution. Our analysis suggests that while the rural poor appear to be catching up with their urban counterparts in terms of labor market characteristics, ten years of economic reforms have intensified the urban–rural gap in returns to these characteristics. On the other hand, the rural rich lag even further behind the urban rich with respect to their labor market characteristics even though the urban–rural gap in the returns to these characteristics has diminished during the reform period. Future efforts to generate urban–rural equality may require policies that seek to equalize returns to labor market characteristics between the two sectors at

R. Chamarbagwala (✉)
Department of Economics, Indiana University, Wylie Hall,
Room 105, 100 S. Woodlawn, Bloomington, IN 47405, USA
e-mail: rchamarb@indiana.edu

the lower half of the distribution and improve rural labor market characteristics at the top half of the distribution.

Keywords Urban–rural inequality · Economic liberalization · Quantile regression decomposition · India

1 Introduction

In this paper, we explore urban–rural inequality in India during 1993–1994 and 2004, a period during which the country implemented substantial trade and investment liberalization reforms. We use the natural logarithm of real monthly per capita household consumption expenditure (log RMPCE) as our measure of welfare and focus on the urban–rural difference in the distribution of log RMPCE.[1] We first estimate quantile regressions for the full sample in order to examine urban–rural differences in the returns to covariates across the two years. We then use the Machado and Mata (2005) technique to decompose the urban–rural gap in log RMPCE into two parts: the part that is explained by the urban–rural difference in the distributions of observed household characteristics (covariate effect) and the part that is explained by the distributions of returns to these characteristics (returns effect).

Urban–rural inequality remains a critical policy concern in India, especially since the rural sector accounts for the vast majority (72%) of the country's population.[2] As shown in Fig. 1, our data reveal a large urban–rural gap in welfare in both years, 1993–1994 and 2004. The positive slopes of both lines show that the urban–rural gap was larger for higher percentiles of the welfare distribution in both years—the gap between the urban and rural rich was much larger than that between the urban and rural poor. However, the welfare gap was fairly convex across the distribution in 1993–1994 but became concave in 2004, with the gap narrowing for the bottom and top quintiles and widening for the middle three quintiles of the distribution. In this paper, we examine how the distributions of labor market characteristics and the returns to these characteristics contributed to urban–rural inequality during this period.

Analyzing India's urban–rural gap has two important implications for policy makers. First, a narrowing urban–rural welfare gap at the bottom and top quintiles of the distribution suggests that the benefits of India's economic reforms have indeed trickled down from the urban to the rural sector among households in these two quintiles. Second, understanding the urban–rural gap has implications for migration within India. A narrowing urban–rural welfare differential could lead to less rural-to-urban migration, thus reducing urbanization which has important consequences for population density, congestion, pollution, as well as poverty in urban India.

There is a vast literature (discussed in Sect. 2) that measures poverty and inequality separately for rural and urban sectors in India and for the country's states and union

---

[1] The use of expenditure rather than income or wages is used as a measure of welfare in several studies on less developed countries (LDCs). This is because income or wages of agricultural workers and laborers in the informal sector may be highly volatile and fluctuate according to seasonal factors (Deaton 1997).

[2] According to the 2001 Census of India, the rural population is 742,490,639 (72.2%) and the urban population is 286,119,689 (27.8%).

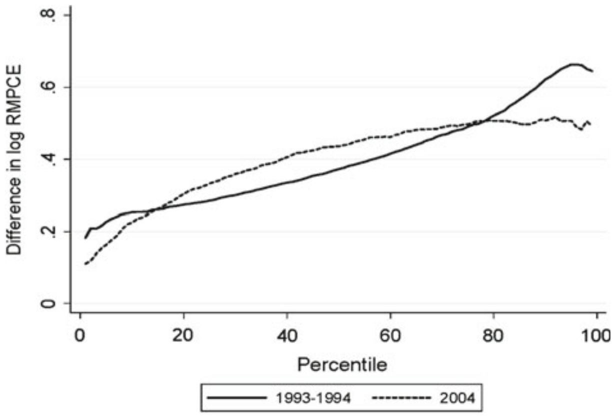

Fig. 1 Urban–rural gap in log real monthly per capita expenditure

territories. In general, these studies show for India's post-reform period during the 1990s, increasing inequality in urban and rural sectors (Deaton 2005; Deaton and Dreze 2002; Sen and Himanshu 2003; Sundaram and Tendulkar 2003a,b) as well as a sharp increase in inter-state inequality (Ghosh and Chandrasekhar 2003; Ahluwalia 2002; Singh et al. 2003).

Our analysis contributes to the literature on poverty and inequality in India in several ways. First, we examine urban–rural inequality in India between 1993–1994 and 2004, a period that marks significant economic reforms and growth in the country and for which little research exists on the urban–rural gap in poverty and inequality. Second, we investigate urban–rural inequality in welfare, as measured by log RMPCE, across the entire welfare distribution. Third, we estimate returns to covariates for each percentile of the welfare distribution. This provides additional information with respect to the returns to education as well as non-agricultural employment, which cannot be obtained from a simple mean regression. Fourth, we allow not only for differential intercepts but also for different returns to covariates in rural and urban sectors. Finally, we decompose the urban–rural gap in welfare to estimate the contribution of covariate and returns effects to urban–rural inequality in both years.

In order to estimate covariate and returns effects of the urban–rural welfare gap, we apply the Machado and Mata (2005) technique. This decomposition technique has been used in similar analyses by several authors (Albrecht et al. 2006, 2003; Autor et al. 2006; Ganguli and Terrell 2005; Machado et al. 2006) and allows us to decompose the urban–rural gap across the entire welfare distribution. While this paper does not represent a novel application of the Machado–Mata technique, it uses this methodology to contribute to the literature on urban–rural inequality in India. We follow Albrecht et al. (2006) closely, who conduct such a study on urban–rural inequality in Vietnam.

The Oaxaca-Blinder decomposition method, which decomposes only the difference in means, cannot explain the factors that contribute to the urban–rural gap across all percentiles of the welfare distribution. The Machado and Mata (2005) technique involves estimating and comparing three different distributions of welfare (as measured by log RMPCE)—the urban, rural, and counterfactual distributions. The urban and rural distributions are estimated via quantile regressions on log RMPCE for the urban and rural sectors, respectively. The counterfactual distribution is estimated from quantile regressions on log RMPCE using rural returns to covariates but the urban distribution of covariates. Comparing the urban and counterfactual distributions provides an estimate of the contribution of the returns effect to the urban–rural gap. On the other hand, a comparison of the rural and counterfactual distributions provides an estimate of the contribution of the covariate effect to urban–rural inequality.

Our decomposition reveals very different patterns in the urban–rural gap across the welfare distribution in 1993–1994 and 2004, as shown in Fig. 5. The narrowing urban–rural welfare gap below the 20th percentile of the distribution can be attributed mostly to rural labor acquiring labor market characteristics that were more similar to their urban counterparts—i.e. a smaller covariate effect.[3] On the other hand, higher returns to rural compared to urban labor market characteristics—i.e. a smaller returns effect—were primarily responsible for the narrowing urban–rural welfare gap above the 80th percentile of the distribution. The widening urban–rural gap between the 20th and 50th percentiles was driven by a higher returns effect (lower returns to rural compared to urban labor market characteristics) whereas a higher covariate effect (worse rural compared to urban labor market characteristics) seems to have driven greater urban–rural inequality between the 50th and 80th percentiles.

Our results suggest that future efforts to generate urban–rural equality may require different policies for poorer and richer sections of the Indian economy. Greater urban–rural equality at the lower half of the welfare distribution may require policies that seek to equalize returns to labor market characteristics—such as education, skilled occupations, and non-farm jobs—between the two sectors. Such policies could focus on generating increased demand for labor with superior characteristics—such as education and skill—in rural India, thereby increasing rural returns to these characteristics. On the other hand, programs that improve the education and skills of rural labor may be necessary to reduce urban–rural inequality at the top half of the welfare distribution, where the welfare gap is the largest.

The paper is organized as follows. Section 2 briefly describes India's economic reforms and discusses the related literature. Section 3 describes the data and variables used in the analysis and Sect. 4 presents aggregate inequality measures and changes in inequality between 1993–1994 and 2004. Our quantile regression results are presented in Sect. 5 whereas Sect. 6 provides results of the decomposition. Finally, Sect. 7 provides concluding remarks.

2 Background: economic reforms, poverty, and inequality in India

Three broad phases can be identified in the course of India's economic development since the country's independence in 1947. During the first phase, which lasted from 1951 to 1963, the mean growth rate of GDP was 3.8%, industry was in its infancy, and

[3] Differences in education and types of employment between urban and rural households in both years are presented in Table 1.

Table 1 Within-quintile means of variables: rural and urban India

Variable	Rural expenditure quintile ($N = 69,184$)					Urban expenditure quintile ($N = 46,123$)				
	First	Second	Third	Fourth	Fifth	First	Second	Third	Fourth	Fifth
1993–1994										
log RMPCE	4.08	4.45	4.69	4.95	5.49	4.34	4.79	5.10	5.46	6.15
hhsize	5.68	5.47	5.23	5.05	4.48	5.75	5.13	4.64	3.95	3.10
Age	42.64	43.86	44.55	45.45	46.11	43.02	43.55	43.38	42.71	42.83
Female	0.09	0.08	0.09	0.10	0.11	0.12	0.10	0.10	0.10	0.10
Sc	0.28	0.23	0.19	0.15	0.10	0.20	0.14	0.11	0.08	0.04
St	0.19	0.16	0.15	0.14	0.12	0.05	0.06	0.09	0.09	0.05
Primary	0.09	0.12	0.13	0.15	0.15	0.15	0.17	0.15	0.12	0.06
Middle	0.06	0.08	0.10	0.13	0.15	0.13	0.18	0.18	0.16	0.09
Secondary	0.02	0.03	0.05	0.07	0.13	0.07	0.13	0.17	0.20	0.19
Higher	0.00	0.01	0.02	0.03	0.07	0.03	0.06	0.09	0.12	0.14
College	0.00	0.01	0.02	0.03	0.08	0.02	0.05	0.10	0.20	0.43
Self employed	0.37	0.51	0.57	0.62	0.62	0.40	0.39	0.37	0.32	0.28
Agriculture	0.83	0.79	0.75	0.71	0.63	0.23	0.18	0.16	0.14	0.16
Manufactures	0.05	0.06	0.06	0.06	0.05	0.19	0.19	0.19	0.19	0.17
Services	0.12	0.15	0.19	0.23	0.32	0.58	0.63	0.65	0.67	0.67
Variable	Rural expenditure quintile ($N = 18{,}975$)					Urban expenditure quintile ($N = 10{,}656$)				
	First	Second	Third	Fourth	Fifth	First	Second	Third	Fourth	Fifth
2004										
log RMPCE	4.30	4.65	4.87	5.13	5.68	4.55	5.05	5.37	5.69	6.26
hhsize	6.57	6.22	5.80	5.36	4.61	5.71	5.04	4.21	3.62	2.98
Age	45.27	47.05	47.50	48.54	50.20	44.37	45.37	45.01	45.29	44.92
Female	0.08	0.07	0.08	0.09	0.12	0.13	0.10	0.10	0.11	0.13
Sc	0.25	0.19	0.16	0.13	0.10	0.23	0.15	0.11	0.08	0.06
St	0.22	0.15	0.15	0.14	0.10	0.04	0.06	0.07	0.09	0.07
Primary	0.12	0.15	0.17	0.18	0.14	0.18	0.16	0.13	0.10	0.05
Middle	0.10	0.13	0.15	0.19	0.20	0.18	0.22	0.21	0.17	0.10
Secondary	0.04	0.05	0.07	0.09	0.15	0.07	0.16	0.20	0.20	0.16
Higher	0.01	0.02	0.03	0.04	0.07	0.03	0.08	0.11	0.13	0.13
College	0.01	0.01	0.02	0.04	0.11	0.03	0.11	0.18	0.28	0.49
Self employed	0.52	0.64	0.70	0.73	0.69	0.44	0.44	0.38	0.33	0.28
Agriculture	0.77	0.77	0.76	0.71	0.62	0.13	0.09	0.06	0.05	0.04
Manufactures	0.05	0.06	0.05	0.05	0.05	0.18	0.18	0.18	0.15	0.14
Services	0.15	0.15	0.17	0.22	0.29	0.62	0.66	0.66	0.68	0.66

Author's calculation, 1993–1994 and Consumer Expenditure Survey (NSSO) the Zamindari system was being removed (Jha 2004).[4] Both urban and rural poverty and inequality decreased during this period, though rural India experienced a much larger fall in poverty and inequality compared to the urban sector.

The second phase of development occurred between 1964 and 1990, during which the average growth rate of GDP was 4.3%. The Green Revolution (1967–1978) was implemented during this period, which aimed at improving agricultural technology through the use of high-yielding variety seeds. The success of the Green Revolution not only generated economic growth in the agricultural sector but also in related industries, such as fertilizers, pesticides, and other chemicals. This period witnessed a large reduction in poverty, with the urban and rural head count ratio falling 11 and 14% points, respectively. Inequality within the urban and rural sectors, however, fell negligibly during this period (Jha 2004).

The third phase of development occurred from 1991 until the late 1990s, with a higher mean GDP growth rate of 5.1%. This period is characterized by the country's widespread liberalization reforms in international trade and investment as well as a continuation of industrial delicensing implemented during the 1980s. Even though not specifically urban biased, India's 1991 trade reforms targeted the manufacturing sector by lowering tariffs, eliminating quotas and import license requirements, liberalizing technology imports, and removing export controls for manufacturing industries. Foreign direct investment was also liberalized to a limited extent, resulting in an increase of FDI from $233 million to $3.3 billion during the 1990s. While poverty in urban and rural India declined slightly between 1991 and 1997, inequality (as measured by the Gini index) increased.

In general, the literature on poverty and inequality in India has focused on the period immediately following the country's economic reforms, i.e. 1993–1994 to 1999–2000. Moreover, this literature has examined inequality across states or *within* urban and rural India rather than the urban–rural gap in welfare. The exception is Deaton and Dreze (2002), who measure urban–rural inequality within Indian states.

Sen and Himanshu (2003) show that the Gini index for rural India increased by three points between 1993–1994 and 1999–2000. Similar conclusions are drawn by Deaton and Dreze (2002) and Sundaram and Tendulkar (2003a). For the 1990s, Sen and Himanshu (2003) show that consumption expenditure among the top two quintiles of the rural sector increased substantially whereas that of the bottom three quintiles decreased. Further, the gap between the urban and rural rich widened considerably during the 1990s.

Deaton and Dreze (2002) find three changing patterns of inequality between 1993–1994 and 1999–2000. Using per capita consumption expenditure as a measure of welfare, they find that inter-state inequality increased during this period and that urban–rural inequality increased not only throughout India but also within states. Jha (2004) also finds higher inequality in both urban and rural sectors during the post-reform period compared to the early 1990s.

Banerjee and Piketty (2001) investigate the share of income by India's top 1% of the income distribution. The authors find a U-shape pattern for this share from the 1960s

[4] The Zamindari system is an agriculture tenure arrangement between landlords and peasants whereby most of the land belongs to a landlord and peasants pay the landlord for the right to cultivate his land.

until the late 1990s, with the lowest levels of inequality reached during the early 1980s. This pattern appears to be consistent with the country's economic policies—while the socialist policies during the 1960s to 1980s reduced the share of income earned by the richest Indians considerably, the post-1991 economic reforms allowed the richest sections of the population to get richer.

Unlike the majority of studies on poverty and inequality in India, this paper examines the urban–rural *gap* in welfare for the entire welfare distribution during an important period of economic liberalization—1993–1994 to 2004. The change in the urban–rural welfare gap from a convex to a more concave pattern over the welfare distribution is shown in Fig. 1 and reflects important changes in the distributions of labor market characteristics and the returns to these characteristics. Our decomposition analysis provides evidence that a lower covariate effect and higher returns effect may be responsible for changes in the welfare gap at the lower half of the distribution whereas a higher covariate effect and lower returns effect may be driving changes in the top half of the distribution.

3 The data and variables

We use data from the Consumer Expenditure Survey of the National Sample Survey Organization (NSSO). This nationally representative survey has been conducted by the NSSO in 1983 (Round 38), 1987–1988 (Round 53), 1993–1994 (Round 50), 1990–2000 (Round 55), and 2004 (Round 60). We use Rounds 50 and 60 of this survey, the first of which was conducted between July 1993 and June 1994 and the second of which was conducted between January 2004 and June 2004.[5] A stratified multi-stage design was implemented by the NSSO in choosing the surveyed households. Thus, we account for the effects of stratification and clustering in our estimations when we compute standard errors. There is a wide range of information about households and household heads available in the data. Our sample consists of 69,184 rural households and 46,123 urban households in 1993–1994. In 2004, we have 18,975 rural households and 10,656 urban households.

Table 1 provides summary statistics for the variables that we use by various percentiles of log RMPCE. Our dependent variable is log RMPCE, which is the log of real monthly per capita consumption expenditure of a household, measured in 1982 Rupees.[6]

[5] Our measure of welfare (log RMPCE) is only comparable among four rounds—Rounds 38, 43, 50, and 60. These four rounds used a 30-day recall period for consumption expenditure of food items, fuel, and some miscellaneous items and a 365-day recall period for other goods, compared to Round 55 which used a 7-day *and* 30-day recall period for consumption expenditure of food items, fuel, and some miscellaneous items and a 365-day recall period for other goods. It has been argued that the inclusion of a 7-day recall period inflated the 30-day recall period expenditures during this round. Deaton and Dreze (2002) explains how different reference periods across surveys can lead to biased poverty and inequality estimates. In comparing poverty and inequality in 1993–1994 and 1999–2000, most of the studies described in Sect. 2 use *adjusted* monthly per capita expenditure estimates for 1999–2000.

[6] Each household provides their expenditure on food items, fuel, and some miscellaneous items (entertainment, non-institutional medical services, rent, etc.) during the thirty days prior to the survey. In addition, each household is asked about their expenditure on education, institutional medical services, clothing,

We include several household demographic characteristics as explanatory variables of log RMPCE. These include the number of individuals living in the household ($hhsize$), the age of the household head and his/her age squared divided by 100 (age and $agesq$), an indicator for the gender of the household head ($female = 1$ for female headed households), and indicators for the household's caste ($sc = 1$ for Scheduled Castes (SC), $st = 1$ for Scheduled Tribes (ST), the omitted group being non-SC and non-ST households).

The human capital level of the household head is also included in all regressions. Rather than include a single variable that captures the total years of education an individual has completed, we include indicators for whether or not the household head has completed primary ($primary$), middle ($middle$), secondary ($secondary$), higher ($higher$), and college ($college$) education, the omitted group being those with no or less than complete primary schooling.[7] These indicators capture any non-linearities that may exist for returns to different levels of education.

In addition to human capital indicators, we include an indicator for whether or not the household head is self employed ($self$). The data we use include detailed information on an individual's industry of employment (1987 and 1998 National Industrial Classifications for 1993–1994 and 2004, respectively). We categorize industries into agriculture, manufacturing, and services, and using agriculture as the omitted group, include indicators for whether or not the household head works in the manufacturing ($manufactures$) or service ($service$) sector.[8]

Even though several explanatory variables may be endogenous, the focus of this paper is not to estimate a *causal* effect of the covariates on log RMPCE. Rather, the goal of this paper is to estimate urban–rural inequality across the entire welfare distribution and to see how this varies by certain household and labor market characteristics. We therefore focus on the correlation between our explanatory variables and log RMPCE rather than on any causal interpretation of our regression coefficients.

4 Changes in urban–rural inequality between 1993–1994 and 2004

As a starting point, we calculate the Theil index of inequality for 1993–1994 and 2004 and the change in inequality over this time period, using the methodology described in Athanasopoulos and Vahid (2003). We do this for India (both urban and rural sectors together) and separately for urban and rural sectors. We also calculate between-sector and within-sector inequality for India. The Theil inequality measure has two major advantages over the Gini index. First, it satisfies the Pigou-Dalton strong principle of

Footnote 6 continued
footwear, and durable goods (furniture, electronics, appliances, etc.) during the past 365 days prior to the survey. Each household's average monthly per capital consumption expenditure is calculated using this information as well as the number of individuals in the household. We use the Consumer Price Index for Industrial Workers (CPIIW) and Consumer Price Index for Agricultural Laborers (CPIAL) to deflate urban and rural expenditures, respectively.

7 Primary school consists of grades 1–5, middle school of grades 6–8, secondary school of grades 9–10, high school of grades 11–12, and college education consists of any college degree (undergraduate or graduate).

8 Table 2 describes the industries included in each industry group.

Table 2 Industry groups

Industry categories	
Agriculture	Agriculture, hunting, and forestry; fishing; mining and quarrying
manufactures	manufacturing
services	Electricity, gas, and water supply; construction; wholesale and retail trade, and repairs; hotels and restaurants; transport, storage, and communication; financial intermediation, real estate, renting, and business activities; public administration and defense; education; health and social work; other community, social, and personal service activities

transfers (Cowell 1995; Athanasopoulos and Vahid 2003), such that it is more sensitive to income transfers in the lower tail of the income distribution than it is to income transfers in the upper tail. Second, it is additively decomposable and is therefore ideal for measuring income inequality based on population characteristics, such as sector of residence.

The Theil index, T, is given by Eq. 1

$$T = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \ln \left(\frac{y_i}{\bar{y}} \right), \tag{1}$$

where n is the population size, y_i is individual i's income, and $\bar{y}$ is the mean income in the population. This index can be decomposed into between- and within-sector inequality as

$$T = \sum_{j=1}^{J} s_j^* \ln \left(\frac{n}{n_j} s_j^* \right) + \sum_{j=1}^{J} s_j^* \sum_{i=1}^{n_j} s_i^j \ln \left(n_j s_i^j \right), \tag{2}$$

where s_j^* is the share of the total income earned by individuals in sector j where $j = \{urban, rural\}$, s_i^j is the share of total income in group j earned by individual i, and n_j is the population in group j. The first term on the right-hand side of the equality represents inequality between urban and rural sectors whereas the second term is the weighted average of inequality within each sector.

Panel A of Table 3 reports the Theil index for India (both urban and rural sectors together) and between- and within-sector inequality. In Panel B, we report the Theil index separately for the urban and rural sectors. In calculating these measures we account for the stratified sampling design of the data.[9] We report standard errors of the Theil index in parentheses in the first two columns and 95% confidence intervals for the change in inequality in parentheses in the third column. Standard errors are calculated via 1,000 bootstrapped replications and the reported confidence intervals are Hall's percentile confidence intervals, which are calculated as described in Athanasopoulos and Vahid (2003) and Hall (1994).

[9] See Athanasopoulos and Vahid (2003) for details.

Table 3 Theil inequality measure

	1993–1994	2004	Change
Panel A: India			
Theil index	0.2510 (0.0011)	0.1959 (0.0003)	−0.0550*(−0.071, −0.034)
Between sector inequality	0.0177 (0.0001)	0.0164 (0.0000)	−0.0012 (−0.003, 0.000)
Within sector inequality	0.2332 (0.0011)	0.1795 (0.0002)	−0.0537* (−0.070,-0.034)
Panel B: urban and rural India			
Theil index—urban sector	0.2820 (0.0013)	0.2095 (0.0004)	−0.0725* (−0.095,−0.054)
Theil index—rural sector	0.1915 (0.0007)	0.1606 (0.0002)	−0.0309*(−0.043, −0.018)

Standard errors and 95% Hall's percentile confidence intervals reported in parentheses. * Change in inequality measure significantly different from zero at 5% level of significance. Source: Author's calculation, 1993–1994 and 2004 Consumer Expenditure Survey (NSSO)

For India (Panel A), inequality as measured by the Theil index fell by roughly 5% points and this decline was statistically significantly different from zero at the 5% level of significance. Most of the decline in inequality was due to a reduction in within sector inequality, and as Panel B shows, this was greater in urban than in rural areas, where inequality fell by 7 and 3% points, respectively. While the inequality estimates presented in Table 3 provide useful information about changes in aggregate inequality between 1993–1994 and 2004 in India, these estimates do not allow us to make any inferences about inequality changes at different points of the welfare distribution. We thus turn to quantile regression to examine urban–rural inequality across the entire welfare distribution between 1993–1994 and 2004 in India.

5 Urban–rural inequality across the welfare distribution

5.1 Quantile regression

We use quantile regressions to examine how welfare is correlated with labor market characteristics of households and how this varies by sector of residence. Quantile regression has two particularly attractive features in comparison to ordinary least squares (OLS) regression (Cameron and Trivedi 2005). First, OLS regression is sensitive to the presence of outliers and can therefore be inefficient when the dependent variable has a highly nonnormal distribution whereas quantile regression is more robust. Second, quantile regression allows us to study the correlation of welfare with each covariate along the entire welfare distribution and therefore provides a richer analysis of the data.

Let $Q_\theta[y|x]$ for $\theta \in (0, 1)$ denote the θth quantile of the distribution of log RMPCE, y, given a vector of covariates, x, which consists of a constant and household and labor market characteristics. These conditional quantiles are given by

$$Q_\theta[y|x] = x'\beta(\theta), \tag{3}$$

Table 4 Estimates of India's urban–rural welfare gap: OLS and quantile regressions

Year	Variable	OLS	Percentiles				
			5th	25th	50th	75th	95th
1993–1994	Rural	4.7708	4.0014	4.4199	4.7272	5.0745	5.6861
		(0.0040)***	(0.0049)***	(0.0040)***	(0.0039)***	(0.0050)***	(0.0077)***
	Urban	0.3976	0.2264	0.2845	0.3723	0.4902	0.6633
		(0.0077)***	(0.0091)***	(0.0076)***	(0.0078)***	(0.0096)***	(0.1315)***
2004	Rural	4.9858	4.2435	4.6317	4.932	5.2788	5.9138
		(0.0053)***	(0.0068)***	(0.0051)***	(0.0056)***	(0.0071)***	(0.0151)***
	Urban	0.3985	0.1633	0.3319	0.4357	0.4972	0.5069
		(0.0092)***	(0.1333)***	(0.0111)***	(0.0098)***	(0.1172)***	(0.0245)***

Standard errors are reported in parentheses where ***, **, and * represent significance at the 1, 5, and 10% levels. Robust standard errors are reported for OLS estimates. For quantile regressions, bootstrapped standard errors are computed on 1,000 replications and account for the effects of stratification and clustering. Source: Author's calculation, 1993–1994 and 2004 Consumer Expenditure (NSSO)

where $\beta(\theta)$ is a vector of the quantile regression coefficients. The θth quantile regression estimator, $\hat{\beta}\,(\theta)$ minimizes over $\beta(\theta)$ the objective function

$$Q\left(\beta\left(\theta\right)\right) = \sum_{i:y_i \geq x_i'\beta}^{N} \theta\left|y_i - x_i'\beta\left(\theta\right)\right| + \sum_{i:y_i < x_i'\beta}^{N} \left(1-\theta\right)\left|y_i - x_i'\beta\left(\theta\right)\right| \tag{4}$$

where $0 < \theta < 1$ and $\beta(\theta)$ represents the value of β for the θth quantile.

5.2 Urban–rural differences

We start by estimating the mean differences in urban–rural welfare across various percentiles of the welfare distribution. We do this by estimating a series of quantile regressions (Eq. 3) for 1993–1994 and 2004. The vector of covariates, x, includes an intercept (which measures the average rural log RMPCE) and an urban dummy, u. Table 4 presents quantile regression results for the 5th, 25th, 50th, 75th, and 95th percentiles in comparison to OLS results and shows vast differences in the rural (intercept) and urban coefficients between the two specifications. The coefficient of the urban dummy represents the difference in log RMPCE between the θth percentile of the urban and rural distributions. These coefficients, multiplied by 100, represent the approximate percentage by which urban real monthly per capita expenditure exceeds those of rural households. All the coefficients on the urban dummy are highly significant and increase across the percentiles. Compared to 1993–1994, the urban coefficient in 2004 is lower at the 5th and 95th percentiles, more or less identical at the 75th percentile, and higher at the 25th and 50th percentiles, showing the same pattern observed in Fig. 1.

Table 5 Coefficient estimates for 1993–1994: quantile regressions

Variable	5th Percentile		25th Percentile		50th Percentile		75th Percentile		95th Percentile	
	Coef	Z	Coef	Z	Coef	Z	Coef	Z	Coef	Z
Urban	0.13	1.57	0.33	7.94	0.41	10.09	0.48	10.73	0.61	7.48
hhsize	−0.05	−22.96	−0.05	−34.43	−0.05	−41.05	−0.05	−34.14	−0.05	−16.31
hhsize*u	−0.04	−14.73	−0.06	−25.49	−0.06	−31.66	−0.06	−28.38	−0.06	−13.78
Age	0.01	8.27	0.01	9.66	0.00	5.33	0.00	1.22	0.00	0.65
Age*u	0.01	1.99	0.00	1.01	0.00	0.31	0.00	−0.52	−0.01	−2.13
agesq	−0.01	−5.56	0.00	−3.81	0.00	1.81	0.01	4.02	0.01	2.30
Agesq*u	−0.01	−1.87	0.00	−0.75	0.00	0.23	0.00	1.33	0.01	2.70
Female	−0.05	−4.00	0.00	0.50	0.04	6.03	0.08	7.98	0.12	6.56
Female*u	−0.05	−2.16	−0.04	−2.81	−0.04	−3.27	−0.03	−1.82	0.01	0.32
Sc	−0.11	−9.57	−0.10	−14.64	−0.11	−18.34	−0.13	−18.00	−0.21	−13.23
Sc*u	−0.03	−2.01	−0.02	−1.90	−0.01	−0.52	0.01	1.05	0.06	2.15
St	−0.12	−7.46	−0.08	−7.26	−0.08	−8.35	−0.09	−7.28	−0.15	−7.74
St*u	0.16	5.28	0.18	9.46	0.17	9.49	0.14	6.76	0.11	2.87
Primary	0.16	14.37	0.17	22.28	0.19	27.80	0.19	24.86	0.19	13.77
Primary*u	0.00	−0.12	0.01	0.82	−0.02	−1.57	−0.02	−1.47	−0.02	−0.81
Middle	0.21	18.42	0.23	24.44	0.25	29.06	0.27	28.22	0.31	15.41
Middle*u	0.05	2.79	0.04	2.97	0.02	1.69	0.01	1.00	0.01	0.19
Secondary	0.33	26.16	0.37	35.14	0.42	44.57	0.48	39.77	0.57	23.25
Secondary*u	0.09	4.46	0.09	7.43	0.05	4.32	0.03	2.00	0.01	0.42
Higher	0.44	30.15	0.45	32.26	0.50	30.94	0.55	31.79	0.62	19.80
Higher*u	0.04	2.09	0.09	5.19	0.07	3.31	0.08	3.21	0.07	1.49
College	0.51	30.10	0.56	32.07	0.65	39.06	0.73	40.22	0.84	19.89
College*u	0.22	9.35	0.26	12.21	0.24	11.34	0.23	10.34	0.16	3.22
Self	0.15	17.61	0.17	30.04	0.18	39.84	0.19	27.45	0.20	13.81
Self*u	−0.13	−10.14	−0.15	−16.33	−0.14	−17.36	−0.14	−12.83	−0.10	−4.22
Manufactures	0.05	3.98	0.04	3.26	0.04	3.42	0.06	4.05	0.07	2.21
Manufactures*u	0.11	4.40	0.11	5.88	0.08	4.47	0.05	2.34	0.00	−0.05
Services	0.11	13.52	0.13	17.37	0.14	21.02	0.15	19.88	0.16	9.03
Services*u	0.04	2.09	−0.01	−0.95	−0.04	−2.94	−0.07	−5.51	−0.12	−4.37
Constant	3.79	106.91	4.21	191.62	4.55	237.84	4.89	165.02	5.38	91.89

Z statistics are based on bootstrapped standard errors, which are computed on 1,000 replications and account for the effects of stratification and clustering Source: Author's calculation, 1993–1994 Consumer Expenditure (NSSO)

Next, we estimate the full model for percentiles 1–99, which includes household demographic, human capital, and employment controls. The vector of covariates now includes a constant, an urban dummy, u, other relevant household and labor market characteristics, z, as well as interactions of the urban dummy with all other covariates, $u'z$. Tables 5 and 6 present the quantile regression coefficients, standard errors, and

Table 6 Coefficient estimates for 2004: quantile regressions

Variable	5th Percentile		25th Percentile		50th Percentile		75th Percentile		95th Percentile	
	Coef	Z	Coef	Z	Coef	Z	Coef	Z	Coef	Z
Urban	0.07	0.48	0.41	5.40	0.58	8.83	0.67	8.75	0.91	5.51
hhsize	−0.05	−18.29	−0.05	−34.75	−0.05	−33.72	−0.05	−36.12	−0.06	−23.24
hhsize*u	−0.04	−7.64	−0.06	−17.32	−0.06	−16.59	−0.05	−14.03	−0.04	−6.87
Age	0.02	6.79	0.01	7.40	0.01	5.12	0.01	2.85	0.00	0.97
Age*u	0.00	0.47	0.00	−0.47	−0.01	−2.14	−0.01	−2.62	−0.02	−2.51
Agesq	−0.01	−5.07	−0.01	−3.64	0.00	−0.83	0.00	1.36	0.01	2.11
Agesq*u	0.00	−0.37	0.00	0.60	0.01	2.00	0.01	2.26	0.02	2.32
Female	−0.03	−1.45	0.04	2.76	0.09	7.05	0.16	8.44	0.24	7.38
Female*u	0.01	0.39	−0.05	−1.76	−0.02	−1.03	−0.05	−1.69	−0.09	−1.73
Sc	−0.11	−6.48	−0.11	−11.59	−0.12	−11.12	−0.14	−8.66	−0.18	−8.90
Sc*u	−0.03	−0.92	−0.01	−0.49	0.00	−0.02	0.00	−0.09	0.03	0.64
St	−0.17	−8.10	−0.12	−8.87	−0.10	−6.22	−0.09	−5.72	−0.10	−3.07
St*u	0.27	6.62	0.27	8.79	0.23	9.06	0.18	5.86	0.10	1.68
Primary	0.12	6.56	0.14	12.79	0.14	11.86	0.15	12.34	0.17	7.57
Primary*u	0.06	1.81	0.03	1.40	0.01	0.61	0.00	−0.17	−0.11	−2.37
Middle	0.19	9.42	0.22	19.89	0.24	18.46	0.27	23.84	0.32	13.09
Middle*u	0.10	3.23	0.08	3.83	0.05	2.28	0.00	0.05	−0.10	−2.20
Secondary	0.23	10.56	0.29	24.77	0.36	18.26	0.43	21.11	0.52	20.24
Secondary*u	0.29	7.67	0.19	8.59	0.09	3.51	0.02	0.80	−0.12	−2.97
Higher	0.28	7.07	0.34	15.35	0.41	17.07	0.47	16.14	0.58	13.54
Higher*u	0.31	5.62	0.25	8.72	0.14	5.01	0.11	2.84	0.03	0.33
College	0.45	11.86	0.52	24.87	0.62	24.25	0.69	21.86	0.94	12.13
College*u	0.29	5.88	0.24	9.67	0.14	4.95	0.12	3.12	−0.04	−0.41
Self	0.15	11.89	0.15	18.75	0.15	14.98	0.15	13.40	0.18	8.50
Self*u	−0.14	−5.79	−0.15	−9.24	−0.14	−8.90	−0.13	−8.23	−0.13	−3.91
Manufactures	−0.01	−0.55	0.00	−0.07	0.01	0.60	0.05	1.63	0.07	2.15
Manufactures*u	0.15	3.47	0.10	3.36	0.06	2.37	0.00	0.02	−0.03	−0.50
Services	0.04	2.68	0.12	10.59	0.13	12.67	0.15	11.47	0.16	5.81
Services*u	0.10	2.77	−0.05	−2.23	−0.08	−3.99	−0.11	−4.49	−0.11	−2.25
Constant	3.85	47.71	4.32	101.73	4.63	109.11	4.93	98.88	5.42	68.18

Z statistics are based on bootstrapped standard errors, which are computed on 1,000 replications and account for the effects of stratification and clustering Source: Author's calculation, 2004 Consumer Expenditure (NSSO)

z-ratios for the 5th, 25th, 50th, 75th, and 95th percentiles for the two survey years, respectively. In order to compare quantile regression estimates to OLS estimates, Table 7 presents the OLS results of these regressions for both survey years. Not only are quantile regression coefficient estimates vastly different across various quantiles, but also consistently differ from OLS estimates.

Table 7 OLS estimates for 1993–1994 and 2004

Variable	1993–1994		2004	
	Coef	SE	Coef	SE
Urban	0.3867	(0.0322)***	0.5109	(0.0648)***
hhsize	−0.0488	(0.0010)***	−0.0542	(0.0013)***
hhsize*u	−0.0586	(0.0017)***	−0.0498	(0.0031)***
Age	0.0064	(0.0008)***	0.0106	(0.0016)***
Age*u	0.0007	(0.0014)	−0.0056	(0.0027)**
Agesq	−0.0004	(0.0008)	−0.0028	(0.0015)*
Agesq*u	0.0003	(0.0014)	0.0050	(0.0027)*
Female	0.0454	(0.0068)***	0.1018	(0.0128)***
Female*u	−0.0362	(0.0112)***	−0.0376	(0.0211)*
Sc	−0.1302	(0.0045)***	−0.1324	(0.0087)***
Sc*u	−0.0038	(0.0086)	−0.0041	(0.0166)
St	−0.0929	(0.0050)***	−0.1147	(0.0093)***
St*u	0.1540	(0.0103)***	0.2257	(0.0194)***
Primary	0.1847	(0.0054)***	0.1494	(0.0092)***
Primary*u	−0.0138	(0.0094)	0.0074	(0.0186)
Middle	0.2597	(0.0060)***	0.2476	(0.0094)***
Middle*u	0.0184	(0.0093)**	0.0335	(0.0175)*
Secondary	0.4333	(0.0081)***	0.3755	(0.0134)***
Secondary*u	0.0624	(0.0110)***	0.0985	(0.0200)***
Higher	0.5140	(0.0115)***	0.4201	(0.0195)***
Higher*u	0.0753	(0.0148)***	0.1785	(0.0266)***
College	0.6582	(0.0126)***	0.6310	(0.0194)***
College*u	0.2350	(0.0148)***	0.1642	(0.0245)***
Self	0.1785	(0.0037)***	0.1482	(0.0074)***
Self*u	−0.1350	(0.0063)***	−0.1273	(0.0121)***
Manufactures	0.0483	(0.0078)***	0.0242	(0.0156)
Manufactures*u	0.0729	(0.0114)***	0.0559	(0.0230)**
Services	0.1378	(0.0047)***	0.1299	(0.0088)***
Services*u	−0.0404	(0.0084)***	−0.0749	(0.0163)***
Constant	4.5370	(0.0184)***	4.6171	(0.0390)***

Robust standard errors are reported in parentheses where ***, **, and * represent significance at the 1, 5, and 10% levels. Source: Author's calculation, 1993–1994 and 2004 Consumer Expenditure (NSSO)

Once we control for household and labor market characteristics, the coefficient on the urban dummy measures the urban–rural welfare gap that is unexplained by these characteristics. The unexplained gaps are large, positive, statistically significant for the 25th, 50th, 75th, and 95th percentiles, and increasing across the distribution. Moreover, the unexplained gaps are larger in 2004 than in 1993–1994. Since we cannot identify these unobserved factors, we focus instead on changes in the distributions of observed covariates and their returns during this period. Table 1 presents changes in

Table 8 Differences between coefficients estimated at 75th and 25th percentiles, standard errors of the differences, and Z-ratios of the differences: 1993–1994 and 2004

Variable	1993–1994			2004		
	Diff	SE	Z	Diff	SE	Z
Urban	0.1413	0.0531	2.66	0.2602	0.0922	2.82
hhsize	−0.0028	0.0013	−2.12	−0.0017	0.0019	−0.89
hhsize*u	−0.0057	0.0021	−2.75	0.0094	0.0043	2.18
Age	−0.0074	0.0011	−6.77	−0.0070	0.0024	−2.86
Age*u	−0.0028	0.0022	−1.27	−0.0069	0.0038	−1.83
Agesq	0.0089	0.0011	7.83	0.0093	0.0024	3.88
Agesq*u	0.0041	0.0023	1.83	0.0054	0.0038	1.41
Female	0.0787	0.0107	7.38	0.1107	0.0203	5.45
Female*u	0.0069	0.0182	0.38	0.0008	0.0335	0.02
Sc	−0.0304	0.0083	−3.65	−0.0266	0.0136	−1.95
Sc*u	0.0386	0.0152	2.54	0.0081	0.0253	0.32
St	−0.0106	0.0114	−0.93	0.0276	0.0175	1.58
St*u	−0.0412	0.0212	−1.94	−0.0895	0.0317	−2.83
Primary	0.0248	0.0086	2.9	0.0126	0.0133	0.94
Primary*u	−0.0307	0.0136	−2.26	−0.0361	0.0296	−1.22
Middle	0.0332	0.0098	3.38	0.0491	0.0156	3.14
Middle*u	−0.0234	0.0161	−1.45	−0.0794	0.0296	−2.68
Secondary	0.1138	0.0134	8.49	0.1480	0.0203	7.28
Secondary*u	−0.0557	0.0197	−2.83	−0.1704	0.0333	−5.12
Higher	0.1014	0.0192	5.29	0.1320	0.0258	5.12
Higher*u	−0.0137	0.0253	−0.54	−0.1399	0.0351	−3.99
College	0.1743	0.0226	7.71	0.1669	0.0343	4.86
College*u	−0.0350	0.0271	−1.29	−0.1209	0.0440	−2.75
Self	0.0249	0.0060	4.13	0.0044	0.0105	0.42
Self*u	0.0109	0.0119	0.91	0.0174	0.0204	0.85
Manufactures	0.0221	0.0117	1.89	0.0524	0.0263	1.99
Manufactures*u	−0.0595	0.0159	−3.73	−0.0972	0.0392	−2.48
Services	0.0192	0.0080	2.39	0.0317	0.0148	2.14
Services*u	−0.0626	0.0151	−4.15	−0.0591	0.0268	−2.21
Constant	0.6737	0.0256	26.27	0.6040	0.0577	10.46

Bootstrapped standard errors are computed on 1,000 replications and account for the effects of stratification and clustering. Source: Author's calculation, 1993–1994 and 2004 Consumer Expenditure (NSSO)

the distributions of covariates whereas Tables 5 and 6 document differences in returns across the distributions. Table 8 shows the differences in coefficients estimated at the 75th and 25th percentiles and the statistical significance of these differences.[10]

[10] Even though many of the inter-quantile differences in coefficients are insignificant, the majority (83% in 1993–1994 and 76% in 2004) are statistically significant at the 10% level.

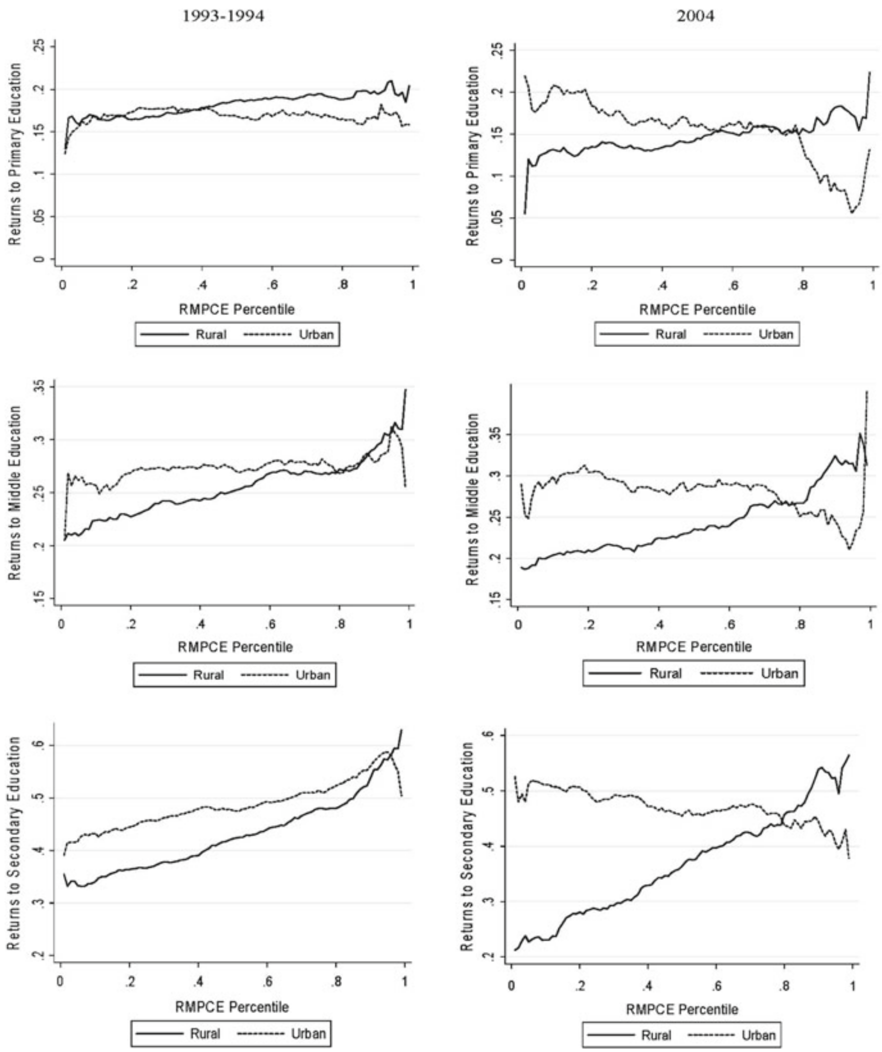

Fig. 2 Returns to education: primary, middle, and secondary

5.3 Returns to education and sector of employment

In this section, we examine changes in the returns to two key labor market characteristics—namely, education and sector of employment—between 1993–1994 and 2004 in urban and rural India. Returns to these characteristics across the welfare distribution are presented in Figs. 2, 3, 4.[11]

[11] The return to each covariate in the rural sector is simply the coefficient on the covariate. In the urban sector, the return to each covariate is the sum of the coefficient on the covariate and the coefficient on the interaction of the covariate with the urban dummy.

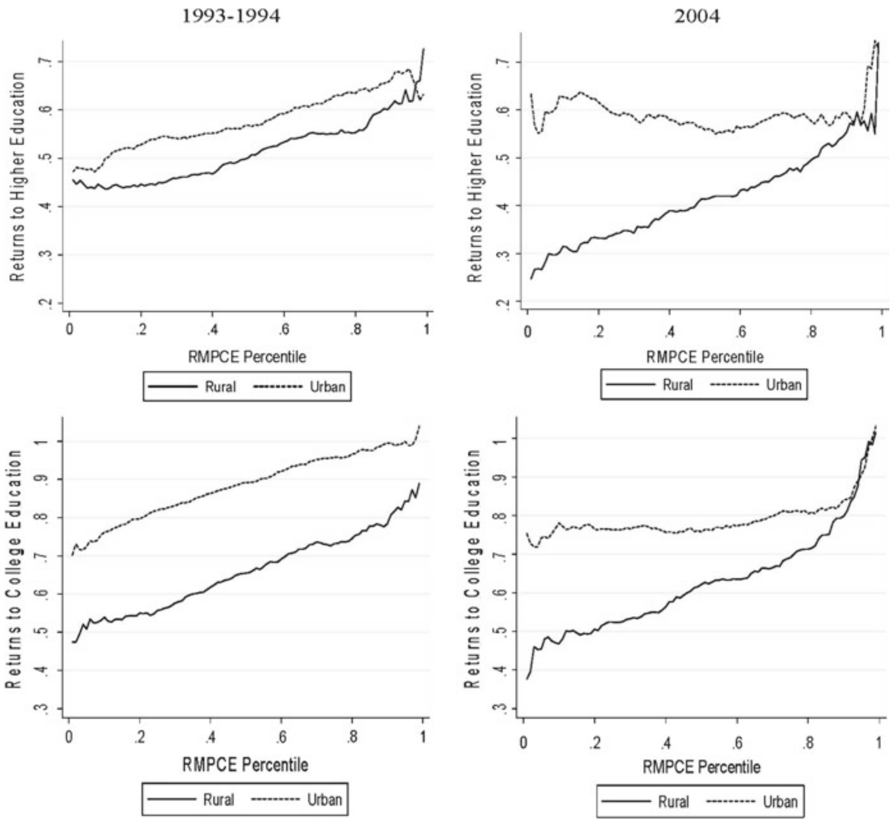

Fig. 3 Returns to education: higher and college

5.3.1 Returns to education

Contrary to our expectations, the returns to education are not generally higher in urban than in rural India throughout the welfare distribution. At lower percentiles of the distribution, returns to all levels of education are higher in urban than in rural India in 1993–1994. The exception is primary schooling, where rural returns are sometimes higher than urban returns at the lowest percentiles. At the highest percentiles, returns to all levels of education other than college are higher in rural than in urban areas in 1993–1994.

This pattern intensifies in 2004, with the positive and negative urban–rural gaps in returns to education widening at lower and higher percentiles, respectively. The exceptions are higher education, where urban returns remain greater than rural returns even at the highest percentiles and college education, where urban returns are approximately equal to rural returns at the highest percentiles. The widening of the positive urban–rural gap at lower percentiles indicates that for the lower part of the welfare distribution, all levels of education pay off more for urban than rural labor in 2004 compared to 1993–1994. This suggests that India's economic reforms during this ten

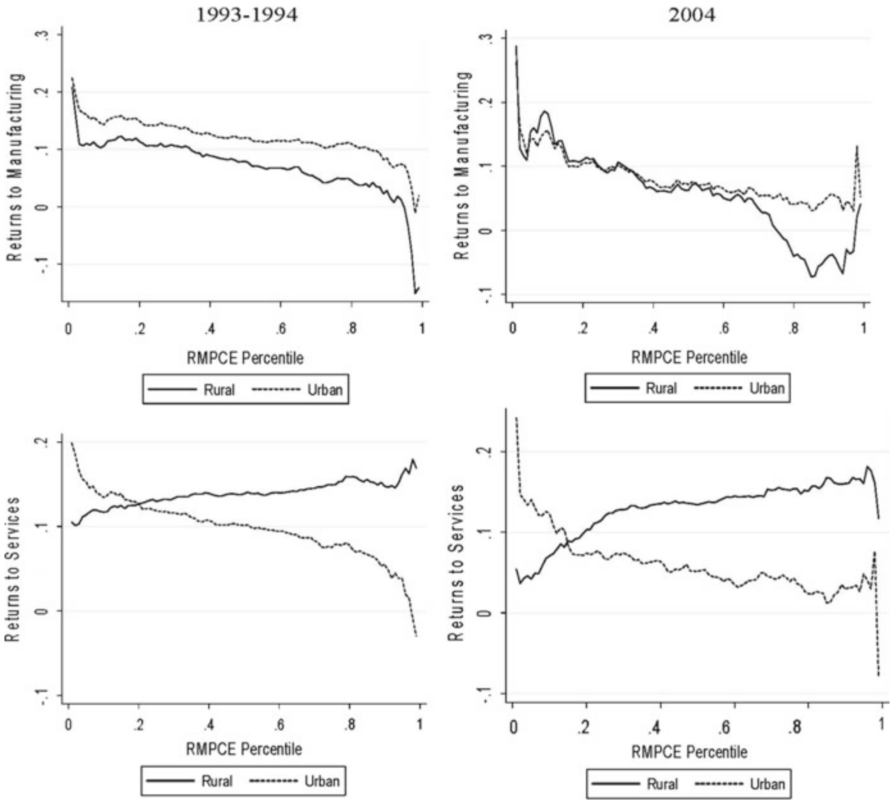

Fig. 4 Returns to manufacturing and services: 1993–1994 and 2004

year period may have been urban-biased for households in the bottom two-thirds to three-quarters of the distribution, at least in terms of generating demand for educated labor. Contributing to this, perhaps, was the acquisition of more education by rural labor, which according to Table 1 was especially high for primary, middle, and secondary education from 1993–1994 to 2004 in rural India. Thus, for urban households at lower percentiles of the distribution, an increase in the demand for educated labor most likely dominated any increase in the supply of educated labor. On the other hand, demand for educated labor was probably insufficient to absorb a more educated work force in rural India among households at the lower portion of the distribution. The opposite pattern appears to have generated a widening negative urban–rural gap in returns to education at higher percentiles of the distribution: demand for educated labor may have been insufficient to absorb a more educated work force in the urban sector whereas in rural India the supply of educated labor may have been scarce compared to its demand.

5.3.2 Returns to manufacturing and services

Returns to employment in the manufacturing sector decreased for most of the welfare distribution in both the urban and rural sectors between 1993–1994 and 2004. Despite this overall decrease, there were interesting changes in the urban–rural gap in returns to employment in the manufacturing sector. In 2004, the urban–rural gap in returns to manufacturing employment narrowed substantially up to around the 70th percentile, after which a sharp decrease in rural returns generated a large urban–rural gap for the top 30% of the distribution. Thus, in terms of returns, rural households engaged in manufacturing employment were closer to their urban counterparts in 2004 compared to 1993–1994 for the bottom 70% of the distribution.

Like manufacturing, returns decreased for rural and especially urban households employed in the service sector between 1993–1994 and 2004. In the service sector, however, changes in the urban–rural gap were more dramatic than in manufacturing. In 1993–1994, urban households engaged in service sector employment enjoyed higher returns than their rural counterparts up to around the 20th percentile of the distribution. For the top 80% of households, however, rural returns to services were higher than urban returns. In 2004, the reversal of the urban–rural gap in returns to services from positive to negative also occurred at around the 20th percentile of the distribution. However, the positive urban–rural gap upto the 20th percentile and the negative urban–rural gap after the 20th percentile widened substantially.

Together, these results suggest that between 1993–1994 and 2004 urban growth may have had spillover effects in rural India between 1993–1994 and 2004. There is evidence that the country's expanding manufacturing and service sectors have been shifting into rural India, mostly as a result of high costs associated with increasing congestion in Indian cities. Our results suggest that India's rural population appears to have benefited from non-farm jobs moving from urban to rural sectors.

6 Decomposing urban–rural inequality

The previous section documented a systematic variation in returns to education and sector of employment across three dimensions—the welfare distribution, urban and rural India, and between 1993–1994 and 2004. On the other hand, Table 1 shows differences in covariates across the welfare distribution in urban and rural India in both years. We now turn to the Machado and Mata (2005) technique to decompose the urban–rural welfare gap across each percentile into two components—one due to urban–rural differences in the distributions of covariates and the other due to urban–rural differences in the distributions of returns to those covariates.

The decomposition involves creating three distributions of log RMPCE—urban, rural, and counterfactual. The urban (rural) distribution of log RMPCE is calculated using urban (rural) covariates and urban (rural) returns to those covariates. The urban and rural distributions are denoted as $F(y^U|x^U, \beta^U)$ and $F(y^R|x^R, \beta^R)$, respectively, where y is log RMPCE, x is the distribution of covariates, and β is the vector of returns to covariates at the various percentiles of the welfare distribution. The counterfactual distribution, $F(y^*|x^U, \beta^R)$, is the distribution of log RMPCE that would exist if rural households were endowed with urban characteristics but received rural returns to labor market characteristics. Using the Machado and Mata (2005) algorithm, $F(y^*|x^U, \beta^R)$ is constructed via the following procedure. First, for each percentile, $\theta = 0.01, 0.02, \ldots, 0.99$, quantile regression coefficients, $\beta^R(\theta)$, are estimated using the rural data. Next, urban data is used to generate predicted values $y^*(\theta) = x^U \beta^R(\theta)$. For each θ this generates a sample of size N^U predicted values, where N^U is the size of the urban sub-sample. Finally, 100 elements are randomly selected from $y^*(\theta) = x^U \beta^R(\theta)$ and stacked into a 99×100 element matrix, y^*. The empirical CDF of y^* is the estimated counterfactual distribution of log RMPCE.[12]

The decomposition consists of comparing the counterfactual distribution (y^*) with the urban (y^U) and rural (y^R) distributions. Defining the θth percentile of these distributions as $y^*(\theta)$, $y^U(\theta)$, and $y^R(\theta)$, the urban–rural welfare gap at the θth percentile of the welfare distribution is given by

$$y^U(\theta) - y^R(\theta) = [y^U(\theta) - y^*(\theta)] + [y^*(\theta) - y^R(\theta)]. \tag{5}$$

The overall urban–rural gap consists of the *returns effect*, which is the first term on the right-hand side, and the *covariate effect*, the second term on the right-hand side of Eq. 5. The returns effect measures the contribution of differences in returns to covariates to the urban–rural gap at the θth percentile whereas the covariate effect measures the contribution of differences in covariates to the overall urban–rural gap at each percentile.

Bootstrapped standard errors are calculated for the estimated returns and covariates effects, from which 95% confidence intervals are constructed. Bootstrapped standard errors are computed by randomly resampling the data. We use 1,000 replications and account for the effects of stratified random sampling. One advantage of calculating standard errors by nonparametric bootstrap is that it allows us to incorporate the stratified sampling design of the data, which reduces the overall variance estimate. If the variance within each sampling strata is smaller than the variance in the entire sample, then variance estimates need to be adjusted for this variance-reducing property of stratified samples. When resampling the original data, we resample independently within each stratum to create each of the 1,000 resampled datasets.

We plot the returns and covariates effects with 95% confidence intervals for 1993–1994 and 2004 in Fig. 5, which shows two striking features. First, the covariate effect is larger at higher percentiles in both years but is steeper in 2004 compared to 1993–1994. The returns effect, on the other hand, is more or less constant in 1993–1994 but decreasing in 2004, especially at higher percentiles of the distribution.

The second feature that stands out in Fig. 5 is the difference in the contribution of the covariate and returns effects at the bottom and top halves of the welfare distribution in 1993–1994 compared to 2004. Figure 6 illustrates this point more clearly by showing the relative contribution of the covariate and returns effects (as a percentage of the overall predicted urban–rural gap) in two separate graphs. For roughly the bottom 40% of the welfare distribution, the covariate effect became smaller and the returns effect became larger in 2004 compared to 1993–1994. The opposite is true for the top

[12] Alternatively, the last step could be dropped. Rather than randomly selecting 100 elements from $y^*(\theta) = x^U \beta^R(\theta)$, all the observations in $y^*(\theta)$ can be treated as a sample from the counterfactual distribution. This approach is followed by Autor et al. (2006) and Blaise (2007). We conduct our analysis using both methods but present results from only the first method since they both generate very similar results.

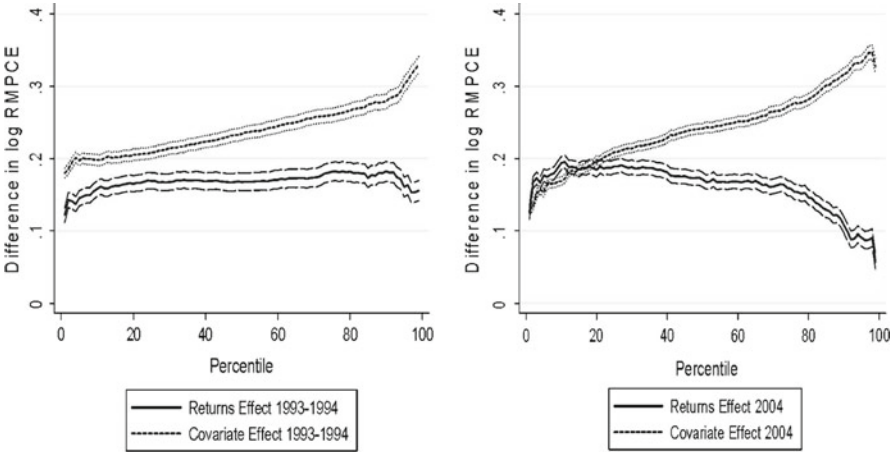

Fig. 5 Covariate and returns effects: 1993–1994 and 2004

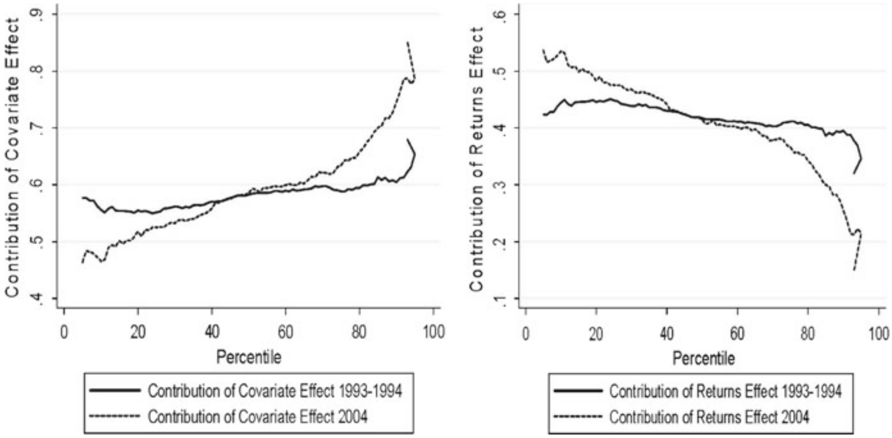

Fig. 6 Contribution of covariate and returns effects to urban–rural gap: 1993–1994 and 2004

40% of the distribution—there is a larger covariate effect and smaller returns effect in 2004 than in 1993–1994.

Our decomposition reveals very different patterns in the urban–rural gap depending on where households are in the welfare distribution. As Fig. 1 shows, the urban–rural welfare gap narrowed below the 20th and above the 80th percentiles of the distribution. Our decomposition shows that greater urban–rural equality for the bottom 20% of the distribution between 1993–1994 and 2004 can be attributed mostly to rural labor aquiring labor market characteristics that were more similar to their urban counterparts. On the other hand, higher returns to rural compared to urban labor market characteristics were primarily responsible for the narrowing urban–rural welfare gap above the 80th percentile of the distribution. Figure 1 also shows an increase in urban–rural inequality between the 20th and 80th percentiles between 1993–1994 and 2004. The widening urban–rural gap between the 20th and approximately 50th percentiles appears to have been driven by lower returns to rural compared to urban labor market characteristics whereas the urban–rural gap in labor market characteristics was responsible for greater urban–rural inequality between roughly the 50th and 80th percentiles.

7 Conclusion

In this paper, we document several interesting features of urban–rural welfare inequality during 1993–1994 and 2004 in India, a period of widespread economic liberalization and growth. In our analysis, we estimate quantile regressions and decompose urban–rural inequality into two parts: one due to differences in household characteristics and the other due to differences in the returns to these characteristics. Our decomposition uses the Machado–Mata technique, which has been applied in numerous studies, in order to examine urban–rural inequality in India during a period of massive economic reforms.

While the covariate effect is increasing across the welfare distribution in both years, the returns effect is relatively constant over the distribution in 1993–1994 but decreasing in 2004. Consistent with changes in the returns to education and non-farm employment, our decomposition also reveals a larger returns effect for the bottom two quintiles and a larger covariate effect for the top two quintiles of the distribution between 1993–1994 and 2004. Thus, the country's economic reforms appear to have aggravated urban–rural differences in returns to superior labor market characteristics for poorer households but diminished these differences for richer households, indicating some spillover of urban growth to rural areas for households at the top of the welfare distribution.

During this ten year period, poorer rural households also acquired labor market characteristics more similar to their urban counterparts whereas the opposite was true for richer rural households. This may reflect education policies targeted towards the rural poor, such as the school meal program and district primary education program. Rural-to-urban migration is another factor that may have contributed to similar labor market characteristics in urban and rural sectors among poorer households. As poorer rural migrants move to and remain in urban areas in search of better employment opportunities, poorer urban households begin to look more similar to their rural counterparts. On average, rural-to-urban migration between the 1991 and 2001 Population Censuses was 7% in India (Mitra and Murayama 2008), the majority of which was driven by unemployed and less educated males searching for better employment prospects in Indian mega cities.

Given that over 70% of India's population continues to reside in the rural sector, greater urban–rural equality remains a critical component of the country's future development. In terms of policy, our results suggest that future efforts to generate urban–rural equality may require very different policies for poorer and richer sections of the Indian economy.

Since differences in returns to covariates became more important at the bottom than at the top of the welfare distribution, policies that seek to equalize urban–rural returns to labor market characteristics may be necessary for the lower half of the welfare distribution. Providing firms with financial incentives to relocate to poorer rural areas may generate greater demand for superior labor market characteristics, such as education. Improving infrastructure in rural India may be critical to the success of firm relocation to rural areas and therefore to generating greater labor market flexibility and less urban–rural inequality. Even though the contribution of differences in urban–rural returns to the welfare gap increased between 1993–1994 and 2004, covariate differences still contributed approximately 50% of the gap even at the lowest percentiles. Thus, policies that provide education and other labor market skills to the rural poor are also essential to reduce rural poverty and urban–rural inequality.

Because urban–rural differences in the endowment of labor market characteristics were primarily responsible for urban–rural inequality among richer households, efforts to reduce urban–rural inequality at the top half of the welfare distribution, where it remains the highest, may require programs that improve rural labor market characteristics, such as education and skill among richer rural households. Given that rich rural households can afford investments in education and skills, what may be needed is making these households aware of the labor market benefits of these investments as well as better access to higher levels of education in rural India.

References

Ahluwalia M (2002) State level performance under economic reforms in india. In: Krueger A (ed) Economic policy reforms and the indian economy. University of Chicago Press, Chicago

Albrecht J, Bjorklund A, Vroman S (2003) Is there a glass ceiling in Sweden. J Labor Econ 21:145–177

Albrecht J, Nguyen B, Vroman S, Westbrook M (2006) A quantile regression decomposition of urban–rural inequality in vietnam. J Develop Econ 83:466–490

Athanasopoulos G, Vahid F (2003) Statistical inference and changes in income inequality in Australia. Econ Rec 79:412–424

Autor D, Katz L, Kearney M (2006) Rising wage inequality: the role of composition and prices, nBER Working Paper No. W11628

Banerjee A, Piketty T (2001) Are the rich growing richer: Evidence from Indian tax data, available at http://www.worldbank.org/indiapovertyworkshop

Blaise M (2007) Estimation of counterfactual distributions using quantile regression, Working Paper

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, London

Cowell F (1995) Measuring inequality, 2nd edn. lSE handbook in economics series. Prentice Hall, London

Deaton A (1997) The analysis of household surveys: a microeconometric approach to development policy. Johns Hopkins University Press, Baltimore

Deaton A (2005) Adjusted Indian Poverty Estimates in 1999/2000. In: Deaton A, Kozel V (eds) Data and dogma: the great Indian poverty debate. Macmillan, New Delhi

Deaton A, Dreze J (2002) Poverty and inequality in India: a re-examination. Econ Polit Wkly 3729–3748

Ganguli I, Terrell K (2005) The gender gap in Ukraine's transition, iZA Discussion Paper No. 1776

Ghosh J, Chandrasekhar C (2003) Per capita income growth in the states of India, available at http://www.macroscan.com/fet/aug03/fet100803SDP1.htm

Hall P (1994) Methodology and theory for the bootstrap. In: Engle R, McFadden D (eds) Handbook of applied econometrics, vol 4. North Holland, Amsterdam

Jha R (2004) Reducing poverty and inequality in India: has liberalization helped? In: Cornia G (ed) Inequality, growth and poverty in an era of liberalization and globalization. Oxford University Press, New York

Machado J, Mata J (2005) Counterfactual decomposition of changes in wage distributions using quantile regression. J Appl Econom 20:445–465

Machado J, Portugal P, Guimaraes J (2006) U.S. unemployment duration: has long become longer or short become shorter? iZA Discussion Paper No. 2174

Mitra A, Murayama M (2008) Rural to urban migration: a district level analysis for India, iDE Discussion Paper No. 137, Institute of Developing Economies

Sen A, Himanshu (2003) Poverty and inequality in India: getting closer to the truth. In: Deaton A, Kozel V (eds) Data and dogma: the great indian poverty debate. Macmillan, New Delhi

Singh N, Bhandari L, Chen A, Khare A (2003) Regional inequality in India: a fresh look. Econ Polit Wkly 1069–1073

Sundaram K, Tendulkar S (2003a) Poverty in India in the 1990s: an analysis of changes in 15 major states. Econ Polit Wkly 1385–1393

Sundaram K, Tendulkar S (2003b) Poverty in India in the 1990s: revised results for all-India and 15 major states for 1993–94. Econ Polit Wkly 4865–4872