

## Further results on the regression-based approach to inequality decomposition with evidence from India

Maurizio Bigotta · Jaya Krishnakumar ·  
Uma Rani

Received: 27 June 2012 / Accepted: 14 March 2014 / Published online: 7 June 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** The paper revisits regression-based inequality decomposition, derives further theoretical results on the factor shares and applies them in an empirical setting. Noting that the approach based on Shorrocks and Fields is not directly applicable to an important welfare-based inequality index, namely Atkinson's inequality index, we generalise it to derive shares for this index. We also derive the asymptotic distribution of all share estimators for obtaining their standard errors necessary for drawing inference. Finally, we use our theoretical results to examine the major factors that contribute to income inequality in India. Our results show that education and household size are the two most dominant factors contributing to income inequality in both rural and urban areas, followed by employment status and regional differences.

**Keywords** Income inequality · Regression-based decompositions · Education · Labour status · India

**JEL Classification** D63 · C21 · J24

---

M. Bigotta  
National Centre for Competence in Research, project LIVES,  
University of Geneva, Geneva, Switzerland

J. Krishnakumar (✉)  
Institute of Economics and Econometrics, University of Geneva,  
40 Bd. du Pont d'Arve, 1211 Geneva 4, Switzerland  
e-mail: jaya.krishnakumar@unige.ch

U. Rani  
Research Department, ILO, Geneva, Switzerland

## 1 Introduction

According to [ILO \(2008\)](#), income inequality has been rising since the early 1990s, and the gap between rich and poor is widening rapidly. Between the early 1990s and the mid-2000s, which was a period of relatively rapid economic growth, the total income of high-income households expanded faster than that of low-income households in about two-thirds of the countries. Similarly, [OECD \(2008\)](#) shows that income inequality has been rising in two-thirds of the OECD countries in this decade. The income of the richest 10 % of people is, on average across OECD countries, nearly nine times that of the poorest 10 %.

One may ask why is it important to analyse inequality as such and not simply study economic growth? This is mainly because the literature does not provide any clear relationship between economic growth, poverty and inequality. Depending on how the benefits of growth are shared among different sections of the population, inequality may increase or decrease as a result of economic growth and hence one needs to directly examine inequality in order to gain a clear picture on this front. From a philosophical angle, one can also argue that the existence of a large inequality in a society may reflect an 'unfair' state of affairs where access to opportunities and resources is denied for certain sections of the population. Examining the factors contributing to inequality in outcomes will provide the necessary insight and lead to possible solutions to correct this underlying inequity.

The methodology to examine income inequality has generated a lot of debate in recent years. Depending upon the question that is raised, different methods can be utilised to address income inequality. We identify four different methods of decomposing inequality in the literature: decomposing differences in means; decomposing inequality in a population into within-group and between-group components; decomposing inequality according to factor components; and decomposing inequality according to the linear income-generating functions.

The first method decomposes inequality in the human capital tradition explaining variation in incomes by means of years of schooling and experience (and other controls) and pioneering work using this method has been undertaken by [Mincer \(1958\)](#), [\(1970\)](#), [Becker \(1964\)](#) and [Oaxaca \(1973\)](#). The second method examines inequality by decomposing it into mutually exclusive population sub-groups. In this tradition, accounting for inequality largely involved decomposing by population groups to quantify how education, age structure, sex, race, etc. affect inequality ([Cowell and Jenkins 1995](#); [Ameida dos Reis and de Barros 1991](#); [Lam and Levison 1991](#); [Shorrocks 1984](#)). The third method relates to the decomposition of inequality into components attributable to different sources—labour income, capital income and land income ([Fei et al. 1978](#); [Pyatt et al. 1980](#)). [Fei et al. \(1978\)](#) and [Shorrocks \(1982\)](#) provide an additive decomposition of total inequality into the contribution of each income source. Another method used in this context is the [Shapley \(1953\)](#) decomposition method which extends the concept of Shapley value in a cooperative game to the decomposition setting.

A final method of decomposing inequality is a regression-based approach allowing for the possibility to control for endogeneity of certain factors explaining income. This approach can be used in a context where income flows are associated with and explained by certain characteristics of the household. As income is regressed on a

number of explanatory variables, the effects of the different variables can be quantified in a *ceteris paribus* context.

A general approach to regression-based decomposition was proposed by Morduch and Sicular (2002) who use the formula given by Theorem 1 of Shorrocks (1982) for decomposing an inequality measure say  $I(Y)$  where  $Y$  denotes the distribution of incomes. This formula involves the different components of income, say  $Y_i$ , as well as functions  $a_i(Y)$  that multiply the different components  $Y_i$  and these  $a_i(Y)$ s differ from one inequality measure to another. Thus, the factor shares obtained by Morduch and Sicular (2002) vary with the inequality measure chosen.

Fields' (2003) regression-based method uses another theorem of Shorrocks, Theorem 3, which gives the formula for factor shares for any inequality measure that satisfies Shorrocks' (1982) Assumptions 1–6 and this class includes Gini, variance, coefficient of variation, Theil's measure, etc. This method goes through a decomposition of income  $Y$ , rather than inequality of income, into its components  $Y_k$  and replaces  $Y_k$  by their estimates  $\hat{\beta}_{k,x_k}$  obtained from a regression. Thus, Fields' method is different from that of Murdoch and Sicular though they are both regression based.

In Sect. 2, we revisit Fields' (2003) method and derive further theoretical results on the shares obtained from the regression model. We calculate the asymptotic distribution of the inequality shares derived by Fields' (2003) which give us their standard errors necessary for drawing inference. In Sect. 3, we show that the procedure cannot be directly applied to Atkinson's inequality index as one of Shorrocks' (1982) six assumptions (the consistent decomposition assumption) needed for applying the procedure is not satisfied by Atkinson's index. However, Atkinson's index is often preferred in the empirical context due to its nice welfare interpretation. So we propose a solution for deriving the shares for the Atkinson's index based on Shorrocks' (1982) weak consistency assumption. We also derive the asymptotic distribution of Atkinson's shares for inference purposes. Finally, in Sect. 4, we provide an empirical example applying our methodology, and analyse the key contributions to income inequality in India using three large nationally representative surveys (three rounds of the National Sample Survey, see Sect. 4). We end the paper with some concluding remarks in Sect. 5.

## 2 Statistical properties of share estimators for many inequality indices

In this section, we first briefly present Fields' (2003) method and then derive the asymptotic distribution of the vector of shares. As is well known, standard errors (or confidence intervals) provide us with fundamental information on the precision with which model parameters are estimated, and are crucial for carrying out inference. In our setting, they will therefore be extremely important for a correct assessment of the magnitude of factor contributions to inequality.

### 2.1 Inequality shares

Our starting point is Shorrocks' (1982) theorem which provides a convenient decomposition of the inequality of total income into different components. Let us write the total income of an individual  $i$  ( $i = 1, \dots, N$ ) as a sum of different components (sources) as

$$Y_i = \sum_{k=1}^K Y_{ik} \quad (1)$$

and let  $I(Y)$  denotes an inequality measure of the distribution of incomes of the total population<sup>1</sup>, denoted as  $Y = [Y_1, \dots, Y_N]'$ . Shorrocks' theorem decomposes inequality  $I(Y)$  into different components  $S_k(Y_k, Y)$  (i.e.  $\sum_{k=1}^K S_k(Y_k, Y) = I(Y)$ ) and gives the share of each component  $s_k$  to total inequality as follows<sup>2</sup>:

### Theorem 1

$$\frac{S_k(Y_k, Y)}{I(Y)} \equiv s_k = \frac{\text{cov}(Y_k, Y)}{V(Y)} \quad \text{with} \quad \sum_{k=1}^K s_k = 1 \quad (2)$$

provided that six assumptions, detailed in Appendix A, are satisfied by  $I(Y)$ .

*Proof* See Shorrocks' (1982), Theorem 3, page 204.  $\square$

Fields (2003) extended this result to the regression framework in which income generation is explained by a certain number of exogenous factors, which can be considered as contributing factors:

$$Y_i = \sum_{k=1}^K \beta_k X_{i,k} + \epsilon_i = \beta' X_i + \epsilon_i \quad (3)$$

with

$$\beta' = [\beta_1, \dots, \beta_K] \quad \text{and} \quad X_i = \begin{bmatrix} X_{i,1} \\ \vdots \\ X_{i,K} \end{bmatrix}$$

This formulation has the same additive form as the sum of incomes from each source seen before (cf. Eq. (1)), so, letting  $Y = [Y_1, \dots, Y_N]'$  and  $X_k = [X_{1,k}, \dots, X_{N,k}]'$ , we can directly apply Shorrocks' theorem and obtain the  $K + 1$  shares as

$$s_k = \frac{\text{cov}(\beta_k X_k, Y)}{V(Y)} \quad k = 1, \dots, K \quad (4)$$

$$s_\epsilon = \frac{\text{cov}(\epsilon, Y)}{V(Y)} \quad (5)$$

One can thus obtain the estimates of the  $K$  shares  $s_k$  from an estimation of  $\beta$ , the vector of the parameters, on a sample of  $n$  individuals as follows:

$$\hat{s}_k = \hat{\beta}_k \frac{\hat{Cov}(X_k, Y)}{\hat{V}(Y)} \quad k = 1, \dots, K \quad (6)$$

<sup>1</sup> Hereafter capital letters refer to concepts relating to the whole population and small letters to the sample counterparts.

<sup>2</sup> Note that  $Y_k$  is the vector containing  $Y_{ik}$ s, constructed in the same way as  $Y$ .

with the share of the residual given by

$$\hat{s}_\epsilon = 1 - \sum_{k=1}^K \hat{s}_k$$

These shares are independent of the particular inequality measure chosen provided the measure satisfies Shorrocks' six assumptions. Note that the shares can be negative if  $\hat{\beta}$  and  $\text{corr}(x_k, y)$  have different signs, that is when the partial correlation turns out to be the opposite of the simple correlation.

Following the reasoning of Fields (2003), the change in inequality between two periods, say 0 and 1, of any given inequality measure  $I$ , can be written as

$$I_1 - I_0 = \sum_k [s_{k,1}I_1 - s_{k,0}I_0]$$

and thus the contribution of factor  $k$  to the change in inequality can be defined as

$$\hat{\Pi}_k = \frac{\hat{s}_{k,1}I_1 - \hat{s}_{k,0}I_0}{I_1 - I_0} \quad (7)$$

Because of the presence of  $I$  in the equation of  $\Pi_k(I)$ , it is obvious that the contribution of the  $k$ -th factor to the *change* in inequality will not only depend on the change in  $x_k$ ,  $y$  and  $\hat{\beta}_k$ , but also on the chosen measure of inequality.

## 2.2 Asymptotic distribution of share estimators

In this sub-section, we compute the asymptotic distribution (and variance) of the shares, which is necessary for doing inference on them, in particular for testing the significance of the contribution of each factor to inequality. The asymptotic distribution allows us to avoid using re-sampling methods, such as bootstrap, which are computer and time intensive. Although the bootstrap has a higher convergence speed and gives better values for small sample inference compared to the asymptotic approximation, it takes an extremely long time for our large data set (approximately 5–10 h per model specification) thus becoming impractical.

**Theorem 2** *The vector of estimated shares  $\hat{s}$  given in Eq. (6) and of the share of residuals<sup>3</sup>*

$$\hat{s} = \begin{bmatrix} \hat{s}_2 \\ \vdots \\ \hat{s}_K \\ \hat{s}_\epsilon \end{bmatrix}$$

<sup>3</sup> The share of the constant term,  $s_1$ , is zero and including it will lead to a singular variance–covariance matrix.

has the following limiting distribution<sup>4</sup>:

$$\sqrt{n} (\hat{s} - s) \xrightarrow{d} N\left(0; \begin{bmatrix} \Sigma & -\Sigma\iota \\ -\iota'\Sigma & \iota'\Sigma\iota \end{bmatrix}\right)$$

where  $\Sigma$  represents the variance–covariance matrix of the limiting distribution of the shares for the factors excluding the residual term,  $\iota'\Sigma\iota$  is the asymptotic variance of the share for the residual term and  $\iota$  is a vector of ones of size  $K - 1$ .  $\Sigma$  is given by

$$\Sigma = \frac{1}{n} \sigma_\epsilon^2 \frac{(I_K \otimes \beta' Q_x) L Q_x^{-1} L' (I_K \otimes Q_x \beta)}{(\beta' Q_x \beta + \sigma_\epsilon^2)^2}$$

$\sigma_\epsilon^2$  is the variance of the error term,  $I_K$  is the identity matrix of order  $K$  and  $L$  is a selection matrix that will be introduced in Appendix B along with  $Q_x$ .

*Proof* See Appendix B. □

The asymptotic variance–covariance matrix is therefore estimated as

$$Asy.\hat{Var}(\hat{s}) = \begin{bmatrix} \hat{\Sigma} & -\hat{\Sigma}\iota \\ -\iota'\hat{\Sigma} & \iota'\hat{\Sigma}\iota \end{bmatrix}$$

where

$$\hat{\Sigma} = \frac{1}{n} \hat{\sigma}_\epsilon^2 \frac{(I_K \otimes \frac{1}{n} \hat{\beta}' x' M x) L (\frac{1}{n} x' M x)^{-1} L' (I_K \otimes \frac{1}{n} x' M x \hat{\beta})}{(\frac{1}{n} \hat{\beta}' x' M x \hat{\beta} + \hat{\sigma}_\epsilon^2)^2}$$

where  $M = I_n - \frac{1}{n} \iota_n \iota_n'$ . The limiting distribution of the shares being normal, one can calculate their 95 % confidence intervals as  $\hat{s}_k \pm 1.96\sqrt{\hat{\Sigma}_{kk}}$  where  $\hat{\Sigma}_{kk}$  denotes the  $k$ -th diagonal element of  $\hat{\Sigma}$ .

In order to investigate the validity of the asymptotic approximation in finite samples, we conducted a simulation experiment with a simple regression model explaining  $y$  with a constant term, a continuous explanatory variable  $x_2$  and a dummy explanatory variable  $d_3$ . The details of the experiment are presented in Appendix C. From Table 3 in this appendix, we can see that for a sample size of 10,000 (relevant for our study), the asymptotic variance of the shares is close to the finite sample variance up to 5 significant digits for both  $x_2$  and  $d_3$ , and the skewness–kurtosis tests for normality cannot be rejected. If we plot the two variances of the shares against the sample size, we find that they in fact become close from a sample size of about 2,000 onwards. Normality is also confirmed by the q–q plots of the sample distribution of shares versus the standard normal one.

Finally, the asymptotic distribution of the change in inequality can be obtained as follows.

<sup>4</sup> The reader should pay attention to the fact that the notation ‘I’ used here for the Identity matrix should not be confused with the inequality measure introduced in the previous section.

**Theorem 3**

$$\sqrt{n} \left( \hat{\Pi}_k - \Pi_k \right) \xrightarrow{d} N \left( 0; \frac{I_1^2 \Sigma_{kk,1} + I_0^2 \Sigma_{kk,0}}{(I_1 - I_0)^2} \right)$$

where  $\Sigma_{kk,0}$  and  $\Sigma_{kk,1}$  are the variances of the limiting distributions of the  $k$ -th share at time 0 and 1, respectively, and where we have kept the same notations  $I_0$  and  $I_1$  to denote the inequality indices corresponding to the whole population at time 0 and 1, respectively.

*Proof* Starting from the definition of  $\hat{\Pi}_k$  in Eq. (7) and assuming independence over time, the above result is easily obtained by combining the asymptotic distribution of shares in Theorem 2 with the fact that when the size of the sample  $n$  goes to infinity, the inequality index will go to its true value given by the inequality in the whole population, i.e.  $I(y) \xrightarrow{p} I(Y)$ . □

**3 Atkinson’s index: a different case**

Although the regression-based approach presented in the previous section applies to the most inequality indices, there are notable exceptions and Atkinson’s index is one of them. We see several reasons for giving a special attention to Atkinson’s index, specifically: (a) it is explicitly derived from an underlying welfare function that ensures a complete ordering of distributions; (b) it is acceptable to social values as it attaches more weight to transfers to the lower end of the distribution whereas Gini attaches more weight to transfers at the middle-income levels; and (c) it is flexible enough to allow specification of increasing degrees of (relative) inequality aversion.

Fields’ formula for deriving the shares cannot be applied to Atkinson’s index as the latter does not respect Assumption 4 of Shorrocks’ theorem, introduced in Theorem 1, which imposes that the inequality index should be decomposable as a sum of terms (i.e. shares). In this section, we propose a suitable way of calculating the shares for Atkinson’s index by adapting the general proposal made by Shorrocks (1982), and using appropriate transformation and aggregation functions.

Let us first write down the expression of Atkinson’s index:

$$I_A(Y) = 1 - \frac{1}{\mu} \left[ \frac{1}{n} \sum_i (y_i)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}, \quad \epsilon \geq 0; \quad \epsilon \neq 1$$

As already identified by Shorrocks, this index is not decomposable as a sum of factor shares, so the author also suggests a ‘generalisation’ to Theorem 1, where the ‘adding up’ rule of Assumption 4 need not be the operator sum, but can be any general aggregation function  $F(\cdot)$ . All but one of the earlier six assumptions still hold and Assumption 4 is now modified as Assumption 4’.

**Assumption 4’** (*Weak consistency assumption*): For all  $Y_1, Y_2, Y$ ,

$$\begin{aligned} S(Y_1, Y_2, Y) &= F(S(Y_1, Y), S(Y_2, Y)) \\ S(0, Y) &= 0, \quad S(Y, Y) = I(Y) \end{aligned}$$

where  $F(\cdot)$  is some aggregator function.

The above modification of the consistency assumption enables Shorrocks (1982) to give the solution to the additive problem in the form of the following theorem:

**Theorem 4** *The weak consistency assumption implies the existence of a continuous and strictly monotonic function:  $f : \mathfrak{R} \rightarrow \mathfrak{R}$  such that  $f(0) = 0$  and:*

$$\tilde{I}(y) = \sum_k \tilde{S}(y_k, y)$$

where  $\tilde{I}(y) = f(I(y))$  and  $\tilde{S}(y_k, y) = f(S(y_k, y))$ .

*Proof* See Shorrocks' (1982), Theorem 4, page 206. □

Now the *transformed* inequality measure and its shares satisfy Assumption 4 of Theorem 1. So, by substituting  $I(\cdot)$  with  $\tilde{I}(\cdot)$  and  $S(\cdot)$  with  $\tilde{S}(\cdot)$  and applying Theorem 1, we obtain the shares for the *transformed* inequality measure.

We can then retrieve the shares  $s_k$  for the 'original' inequality index using the inverse function of  $f(\cdot)$ <sup>5</sup>:

$$s_k = \frac{S_k}{I_A} = \frac{f^{-1}(\tilde{S}_k)}{I_A} = \frac{f^{-1}(\tilde{s}_k \tilde{I}_A)}{I_A} = \frac{f^{-1}(\tilde{s}_k f(I_A))}{I_A} \quad (8)$$

It is important to note that the above formula gives the shares of the 'original' inequality measure (and *not* of the transformed measure) and that the values of these shares do depend on the index.

In the case of Atkinson's index, Shorrocks suggested the following function  $f(y) = (1 - y)^{1-\epsilon}$ . However, this function does not satisfy the condition  $f(0) = 0$ , which is required for Theorem 4<sup>6</sup>. So we correct Shorrocks' function by simply subtracting 1 from the function he proposed i.e.  $f(y) = (1 - y)^{1-\epsilon} - 1$  which then satisfies the conditions of Theorem 4 in particular  $f(0) = 0$ . Then, we decompose Atkinson's inequality index using this transformation. A detailed derivation of the factor contributions including the transformed measures  $\tilde{I}$  and  $\tilde{S}$  is presented in Appendix D.

One can easily see that the inverse of the function  $f(y) = (1 - y)^{1-\epsilon} - 1$  is given by  $f^{-1}(y) = 1 - (y + 1)^{\frac{1}{1-\epsilon}}$  and using it in Eq. (8) we obtain the shares which are presented in the empirical section. It is important to stress that the adjustment regarding the consistency assumption should not be neglected and the calculations have to be modified accordingly to obtain the correct values of shares. Further, the asymptotic variances also have to be corrected taking the transformation into account which we have done in our calculations (See Appendix E).

<sup>5</sup> Note that these shares  $s_k$  do not satisfy the original assumptions of Shorrocks, only the transformed shares do.

<sup>6</sup>  $f(0) = 1$  for the above  $f(\cdot)$ .



The following section examines the major factors that contribute to income inequality in India using the methodology described in this section. The empirical analysis is not intended to be an exhaustive analysis of inequality in India, but should rather be viewed as a pertinent application of the methodology discussed in the previous sections.

#### 4 Empirical application: decomposing income inequality in India

The Indian economy has grown rapidly with an annual average growth of about 6 % over the past two decades. There was an attitudinal shift towards ‘pro-business’ policies in the 1980’s, which resulted in the expansion of the industrial sector—through the abolition of investment licencing, simplification of rules and reduction of the exclusive monopolies of the public sector—and helped to achieve high growth. The boom in information technology and other related services along with other measures has been central to the growth process since the 1990’s (Kohli 2006; Rodrik and Subramanian 2005; Das 2003).

One might have expected that the high growth over the past decades would have translated into better incomes for all workers both at the higher and lower rungs and would have reduced inequality. However, inequality as measured by Gini coefficient has not declined and the most recent data available show a rise in inequality. The evidence with regard to inequality and poverty in the Indian economy is quite mixed (Sundaram and Tendulkar 2003; Sengupta et al. 2008; Bhaduri 2008; Dev and Ravi 2007).

In this paper, we do not intend to explore the effects of growth on inequality but rather to understand at the micro level the determinants of inequality and find out the relative contributions of various independent factors using the theoretical results derived in the earlier sections.

##### 4.1 Regression results: determinants of income

We begin with the estimation of the income equation for rural and urban areas separately. The data are based on multiple rounds of the Employment–Unemployment survey along with the consumption expenditure survey undertaken by the National Sample Survey Organisation (NSSO) every five years, covering all major Indian States. We use three rounds corresponding to the years 1983 (38th round), 1993–1994 (50th round) and 2004–2005 (61st round)<sup>7</sup>. The dependent variable is income for which we use monthly per capita household consumer expenditure as a proxy. Explanatory variables comprise household head’s age, sex, social group, religion, employment status, education level, industry group, region, household size and per capita land owned. Appendix F presents data sources along with the list of variable names and definitions. It also states the reference group for each of the categorical variables. The analysis is undertaken for all persons aged 15–64 years in the sample.

<sup>7</sup> There is a debate on methodological comparability across the different NSSO rounds on the choice of the reference period. However, we have tried to overcome this problem in this paper as we use the 30-day uniform reference period, which is comparable across all rounds and we do not take into consideration the 55th round (1999–2000) for the analysis.

We have excluded the top 1 % of the sample for our estimations because of the huge jumps in income in this category leading to unstable estimation results thus making the interpretation and comparison difficult over different specifications and time periods. If we look at what is excluded in terms of the share of total income, the top 1 % possesses an average of about 10 % of the total income in our sample (average across the different time periods and rural/urban, ranging from a minimum of 3.9 % for 1993 urban to a maximum of 17.8 % for 1983 rural). This means the inequality that we study corresponds to about 90 % of the total income and is thus smaller than what it would be if we took the entire sample. However, the main reason for this exclusion is the fact that we are using a regression-based approach and ‘stable’ regression coefficients are necessary for a sound interpretation of the regression results. Removing the top 1 % stabilises the coefficients (for instance excluding another 1 %—i.e. excluding the top 2 % of the sample—does not change the results significantly). Hence we decided to use the lower 99 % of the sample.

The regression results for 1983, 1993–1994 and 2004–2005 rural and urban areas are presented in Table 1. We have applied OLS but other procedures could be used, for instance instrumental variables method if endogeneity is suspected. The variables included explain about 78 % of the variance in incomes in the urban areas in 1983 and 80 % in 2004–2005, whereas in rural areas the variables included explain about 80 % of the variance in incomes in 1983 and 86 % by 2004–2005. The explanatory variables have the expected signs and most of the variables included are statistically significant at the one percent level using the standard  $z$ -tests. As our sample size is large, it is expected that the coefficients turn out to be significant. Deaton (1997) recommends a correction of the critical value for large samples based on Schwarz (1978), Leamer (1978) and Chow (1983). In particular, Leamer (1978) suggests using the Bayes factor to conclude in favour of or against a set of linear restrictions  $H_0$  as it takes into account both the asymptotic distribution (square of the standard normal i.e.  $\chi^2$ ) as well as the sample size  $T$ . More specifically this leads to a critical value of  $\ln T$  for a  $\chi^2$  statistic which translates as  $\sqrt{\ln T}$  for the asymptotic standard normal  $z$ -statistic and we have applied this correction in our tests<sup>8</sup>.

The dummies for the education variable are statistically significant, which confirms the differences in returns to human capital: as education level increases the difference with respect to the reference group (highest education level) decreases. The only exception is in the rural areas in 1983, having a tertiary level of education brings lower returns than having a middle or secondary level of education. The size of the coefficient of the education dummies has increased between 1983 to 2004–2005 in both rural and urban areas, which brings out the importance of education in reducing inequality.

The industry dummies are also statistically significant for all sectors, in both rural and urban areas, which indicate the presence of industry effects on income. The specific industry dummies show that when compared to agriculture (reference group), all other sectors have a positive differential in both rural and urban areas. These differentials are much higher in rural compared to urban areas for the non-service sectors,

<sup>8</sup> For our case with only one restriction, Leamer’s critical value is given by  $(T - K)(T^{1/T} - 1)$  (cf. Leamer (1978), p. 114) which goes to  $\ln T$  as  $T$  goes to  $\infty$ .

Table 1 Regression results

	Urban				Rural							
	1993–1994		2004–2005		1983		1993–1994		2004–2005			
	1983											
Age	89.44 (0.907)	***	295.22 (2.452)	***	545.44 (4.157)	***	42.75 (0.710)	***	124.85 (1.532)	***	298.98 (2.166)	***
Gender	29.26 (2.060)	***	49.24 (5.803)	***	78.10 (10.072)	***	5.97 (0.865)	***	0.71 (2.311)	***	-19.86 (4.017)	***
Household size	-17.82 (0.262)	**	-62.45 (0.825)	***	-109.13 (1.526)	***	-6.30 (0.130)	***	-16.37 (0.368)	***	-41.19 (0.533)	***
Land ownership							4.17 (0.127)	***	14.56 (0.418)	***	20.06 (0.520)	***
Scheduled tribe	-15.79 (2.397)	***	-23.04 (7.295)	***	-31.11 (12.395)	**	-21.18 (0.689)	***	-43.65 (1.805)	***	-68.06 (3.265)	***
Scheduled caste	-13.72 (1.501)	***	-31.39 (3.990)	***	-103.51 (6.881)	***	-10.87 (0.582)	***	-22.37 (1.479)	***	-51.96 (2.589)	***
Hindu	-1.33 (1.448)		13.32 (4.249)	***	23.17 (7.703)	***	0.76 (0.746)	***	5.09 (2.152)	**	20.74 (3.422)	***
Muslim	19.08 (2.875)	***	52.37 (8.507)	***	164.33 (15.287)	***	9.17 (1.230)	***	34.87 (3.231)	***	123.50 (5.455)	***
Christian	22.32 (3.101)	***	42.41 (9.000)	***	108.97 (17.705)	***	8.95 (1.573)	***	9.15 (3.833)	**	103.57 (6.685)	***
Manufacturing	17.62 (1.816)	***	61.71 (5.642)	***	70.94 (10.651)	***	4.73 (1.010)	***	21.76 (2.661)	***	4.76 (3.856)	***
Electricity, gas and water	32.82 (4.059)	***	99.18 (10.585)	***	225.52 (23.605)	***	24.57 (2.815)	***	32.97 (6.024)	***	93.89 (10.988)	***
Construction	19.76	***	84.97	***	82.88	***	11.27	***	33.84	***	37.33	***

Table 1 continued

	Urban				Rural				
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
Low-skilled services sector	(2.689) 9.33 (1.667)	(6.925) *** 34.39 (4.934)	(11.498) *** 29.92 (9.115)	(1.560) 2.58 (0.884)	(3.652) *** 31.02 (2.255)	(3.886) *** 20.99 (3.085)	(1.560) 2.58 (0.884)	(3.652) *** 31.02 (2.255)	(3.886) *** 20.99 (3.085)
High-skilled services sector	(2.303) 17.19 (2.303)	(6.828) *** 63.20 (6.828)	(12.787) *** 183.14 (12.787)	(1.920) 24.09 (1.920)	(4.665) *** 48.19 (4.665)	(6.634) *** 79.03 (6.634)	(1.920) 24.09 (1.920)	(4.665) *** 48.19 (4.665)	(6.634) *** 79.03 (6.634)
Illiterate	(2.660) -124.32 (2.660)	(6.557) *** -460.91 (6.557)	(11.459) *** -883.22 (11.459)	(2.842) -14.26 (2.842)	(5.314) *** -134.66 (5.314)	(6.951) *** -336.98 (6.951)	(2.842) -14.26 (2.842)	(5.314) *** -134.66 (5.314)	(6.951) *** -336.98 (6.951)
Primary education	(2.547) -105.42 (2.547)	(6.237) *** -402.23 (6.237)	(10.978) *** -765.15 (10.978)	(2.832) 0.76 (2.832)	(5.287) *** -92.51 (5.287)	(6.860) *** -260.59 (6.860)	(2.832) 0.76 (2.832)	(5.287) *** -92.51 (5.287)	(6.860) *** -260.59 (6.860)
Middle-level education	(2.639) -82.03 (2.639)	(6.444) *** -345.63 (6.444)	(11.422) *** -639.47 (11.422)	(2.870) 16.39 (2.870)	(5.401) *** -57.72 (5.401)	(7.015) *** -189.57 (7.015)	(2.870) 16.39 (2.870)	(5.401) *** -57.72 (5.401)	(7.015) *** -189.57 (7.015)
Secondary education	(2.637) -41.01 (2.637)	(6.308) *** -220.85 (6.308)	(11.541) *** -406.85 (11.541)	(3.071) 36.33 (3.071)	(5.466) *** -10.65 (5.466)	(7.140) *** -111.94 (7.140)	(3.071) 36.33 (3.071)	(5.466) *** -10.65 (5.466)	(7.140) *** -111.94 (7.140)
Self-employed	(2.315) 0.61 (2.315)	(6.791) 1.24 (6.791)	(12.264) *** -45.53 (12.264)	(0.986) 18.88 (0.986)	(2.836) *** 36.99 (2.836)	(5.120) *** 54.03 (5.120)	(0.986) 18.88 (0.986)	(2.836) *** 36.99 (2.836)	(5.120) *** 54.03 (5.120)
Casual worker	(2.428) -27.35 (2.428)	(6.645) *** -100.51 (6.645)	(12.538) *** -230.55 (12.538)	(0.996) 3.00 (0.996)	(2.742) *** 4.95 (2.742)	(5.029) *** -13.55 (5.029)	(0.996) 3.00 (0.996)	(2.742) *** 4.95 (2.742)	(5.029) *** -13.55 (5.029)
Salaried worker	(2.566) 7.91 (2.566)	(7.526) *** -10.67 (7.526)	(13.699) *** -55.31 (13.699)	(1.468) 23.21 (1.468)	(4.311) *** 69.12 (4.311)	(7.028) *** 96.61 (7.028)	(1.468) 23.21 (1.468)	(4.311) *** 69.12 (4.311)	(7.028) *** 96.61 (7.028)
Andhra Pradesh	(2.309) -23.60 (2.309)	(7.122) *** -76.30 (7.122)	(12.009) *** -47.73 (12.009)	(1.146) -9.85 (1.146)	(3.000) *** -24.84 (3.000)	(4.785) *** 12.98 (4.785)	(1.146) -9.85 (1.146)	(3.000) *** -24.84 (3.000)	(4.785) *** 12.98 (4.785)
Assam	(3.363) -26.00 (3.363)	(10.609) *** -100.20 (10.609)	(18.876) *** -96.01 (18.876)	(1.074) -15.02 (1.074)	(2.600) *** -74.14 (2.600)	(4.414) *** -67.63 (4.414)	(1.074) -15.02 (1.074)	(2.600) *** -74.14 (2.600)	(4.414) *** -67.63 (4.414)

Table 1 continued

	Urban			Rural		
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
Bihar	-22.81 (2.623)	*** (7.518)	*** (13.302)	-27.46 (0.931)	*** (2.382)	*** (3.805)
Gujarat	-10.58 (2.506)	*** (8.006)	*** (13.982)	-4.39 (1.329)	*** (3.555)	*** (6.456)
Haryana	7.41 (4.769)	-48.03 (12.056)	*** (17.930)	26.39 (2.374)	*** (6.175)	*** (8.160)
Karnataka	3.31 (2.946)	-59.26 (7.976)	*** (14.185)	-9.35 (1.354)	*** (3.347)	*** (5.493)
Kerala	-2.09 (3.545)	-23.47 (9.255)	** (16.957)	3.25 (1.544)	*** (4.373)	*** (7.883)
Madhya Pradesh	-23.95 (2.371)	*** (7.225)	*** (12.372)	-21.78 (1.073)	*** (2.760)	*** (4.551)
Maharashtra	6.48 (2.288)	19.07 (6.956)	*** (11.870)	-20.91 (1.133)	*** (3.078)	*** (4.912)
Orissa	-18.09 (3.579)	*** (10.147)	*** (14.493)	-23.69 (1.191)	*** (2.908)	*** (4.225)
Punjab	3.59 (3.240)	16.82 (9.239)	* (15.287)	36.32 (1.904)	*** (5.125)	*** (8.898)
Rajasthan	1.17 (3.096)	-48.43 (8.655)	*** (14.260)	-1.16 (1.513)	2.87 (3.578)	*** (4.976)
Tamil Nadu	-21.73 (2.256)	*** (7.179)	*** (11.828)	-21.30 (1.228)	*** (3.297)	*** (5.160)
Uttar Pradesh	-30.22 (3.545)	*** (8.283)	*** (11.229)	-22.82 (1.544)	*** (3.347)	*** (5.493)

Table 1 continued

	Urban			Rural		
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
West Bengal	(2.157) -11.37	(6.664) -76.98	(10.849) -34.61	(0.947) -18.50	(2.602) -35.57	(3.873) -17.36
Observations	(2.506) 40,260	(7.263) 43,927	(12.332) 43,423	(1.056) 74,959	(2.794) 66,249	(4.444) 76,194
Adjusted R-squared	0.783	0.784	0.804	0.803	0.831	0.857

Dependent variable: *monthly per capita expenditure (mpce)*. Source NSSO (1987), NSSO (1997), NSSO (2006a), NSSO (2006b). For the definition of the variables, see Appendix F

\*\*\*, \*\*, \* Significant at the 1, 5, 10 percent respectively

while for low and especially high skilled, the returns are higher in urban areas. When compared across the three periods, the size of the coefficients of electricity, gas and water and high-skilled services increased in both rural and urban areas while that of other industry groups has reduced or remained stable between 1993–1994 and 2004–2005.

Social groups play a very important role in Indian society and one can observe their significant effects on income. In both areas, with respect to the reference group (other castes), scheduled tribes and scheduled castes have negative differentials. In rural areas, the differential is higher in absolute value for the former, and in urban areas, the differentials are higher for scheduled castes. In both rural and urban areas, their magnitude of the impact on income increased over the whole period. Similarly, across religious groups, the income differential (compared to ‘Others’) is much higher for Muslim and Christian households than for Hindus and the effects increase over time in both rural and urban areas.

#### 4.2 Factors accounting for income inequality

The decomposition methodology described in the earlier section is applied to our regression model to compute the inequality shares of explanatory variables. Table 2 reports the contributions of various household characteristics to inequality in income, measured using Atkinson’s index with an inequality aversion coefficient of 2 ( $\epsilon = 2$ )<sup>9</sup>, for both rural and urban areas from 1983 to 2004–2005. The shares are computed using Eqs. (6) and (8). In order to facilitate the interpretation of the results, the factor contributions have been normalised to add upto 100 % and aggregated for categorical variables by summing the contributions of constituent dummy variables<sup>10</sup>. The complete set of results without normalising or aggregating the shares is presented in Table 4 in Appendix G<sup>11</sup>. Both the tables also report the standard errors for the share estimators obtained using the asymptotic distribution derived in Theorem 2 combined with Eq. (13) for Atkinson’s shares given in Appendix D.

Before commenting on the different share values, we would like to make two remarks: (a) the low inequality shares of certain factors should not be interpreted in terms of the relevance of these factors in the fit of the data; indeed not only do these explanatory factors globally explain well the observed *variance* in income (the  $R^2$  value shows a good fitting) but also their individual contributions to the overall explanation (i.e. the partial correlations of these factors with the endogenous variable) are important<sup>12</sup>; (b) we also computed the shares for the group of indices that satisfy

<sup>9</sup> The procedure was also applied for  $\epsilon = 0.5$ ; the results are consistent with the ones presented in the paper showing the same pattern across the studied period.

<sup>10</sup> For example, the total contribution of industry groups is made up of five terms (dummy variables): mining and manufacturing; electricity, gas and water; construction; low-skilled service sector; and high-skilled service sector. This aggregation is also done for social groups, employment status, education levels, industry groups and regions.

<sup>11</sup> Note that if a coefficient is not significant then its contribution to income inequality will be equal to zero. So we set the coefficient equal to zero and do not compute the shares for the insignificant parameters.

<sup>12</sup> We do not report the partial correlations here but they are available upon request.

Table 2 Atkinson's index: aggregate factor shares, their asymptotic standard errors and contributions to the change of inequality

	Rural											
	Urban					Rural						
	Shares (%)		Change in inequality			Shares (%)		Change in inequality				
1983	1993–1994	2004–2005	1983–1993	1993–2004	1983	1993–1994	2004–2005	1983–1993	1993–2004	1983–1993	1993–2004	
Age	-5.3 (0.022)	*** (0.017)	*** (0.024)	*** -0.08	0.8 0.44	0.8 (0.023)	*** (0.020)	*** (0.015)	*** -0.01	3.3 (0.015)	*** (0.015)	-0.10
Gender	0.5 (0.037)	*** (0.008)	*** (0.008)	*** 0.01	-0.01	-0.02 (0.002)	*** (0.002)	0.1 (0.002)	*** -	0.1 (0.002)	*** (0.002)	-
Household size	21.3 (0.182)	*** (0.060)	*** (0.062)	*** 0.24	-0.42	7.0 (0.031)	*** (0.017)	11.7 (0.019)	*** 0.06	7.4 (0.017)	*** (0.019)	-0.20
Land Ownership												
Social Group	1.1 (0.002)	*** (0.025)	*** (0.027)	*** 0.02	0.08	3.1 (0.028)	*** (0.019)	2.3 (0.011)	*** 0.04	2.5 (0.031)	*** (0.020)	0.03
Religion	0.8 (0.051)	*** (0.028)	*** (0.030)	*** 0.01	0.03	0.5 (0.026)	*** (0.018)	1.4 (0.017)	*** 0.01	0.3 (0.018)	*** (0.017)	-0.06
State dummies	3.2 (0.093)	*** (0.034)	*** (0.036)	*** 0.03	-0.06	6.9 (0.051)	*** (0.031)	10.8 (0.026)	*** 0.02	9.3 (0.031)	*** (0.026)	-0.01
<i>Household head's</i>												
Industry	1.3 (0.210)	*** (0.082)	*** (0.104)	*** 0.02	0.18	2.4 (0.043)	*** (0.038)	1.6 (0.029)	*** 0.02	2.6 (0.038)	*** (0.029)	0.09
Education	21.1 (0.302)	*** (0.088)	*** (0.092)	*** 0.16	0.48	8.1 (0.194)	*** (0.108)	11.2 (0.053)	*** 0.04	10.3 (0.108)	*** (0.053)	0.04
Employment status	3.9 (0.352)	*** (0.060)	*** (0.122)	*** 0.05	0.02	4.2 (0.100)	*** (0.085)	4.2 (0.051)	*** 0.02	5.1 (0.085)	*** (0.051)	0.11



Table 2 continued

	Urban				Rural					
	Shares (%)		Change in inequality		Shares (%)		Change in inequality			
	1983	1993–1994	2004–20–05	1983–1993	1993–2004	1983	1993–1994	2004–2005	1983–1993	1993–2004
Residuals	52.1 (0.199)	*** 49.1 (0.075)	*** 47.6 (0.079)	*** 0.55	0.26	62.5 (0.059)	*** 56.8 (0.044)	*** 48.5 (0.036)	*** 0.74	1.09
Sum	100.0	100.0	100.0			100.0	100.0	100.0		
Atkinson ( $e=2$ )	0.623	0.302	0.326			0.330	0.219	0.189		

Source Same as Table 1. – indicates a non-significant beta. Figures in brackets are standard errors. The shares above have been normalised to add upto 100 % to facilitate the interpretation of the results; we also aggregated the shares for the variables coded as dummy to see the total effect of the categories

\*\*\*, \*\*, \* Significant at the 1, 5, 10 percent, respectively

Shorrocks' original assumptions (such as Gini, coefficient of variation, the generalised entropy indices, the centile measures) and the interpretations are qualitatively similar<sup>13</sup>.

Education emerges as the most dominant factor contributing to inequality in urban and rural areas, although this factor's share in inequality is much higher in urban (28 %) than in rural (11.2 %) areas in 2004–2005 (Table 2). The relative contribution of education has increased both in rural and urban areas over the three periods. Education explains most of the inequality in urban areas (53 % of the inequality explained by our model), while both education and household size tend to be equally important in rural areas. The dominance of education in explaining income inequality has also been confirmed by others (Londono 1996; Deininger and Squire 1998; Barro 2000).

Household size is another major factor contributing to the inequality and this factor contribution has declined by almost 7 percentage points between 1983 and 2004–2005 in urban areas attesting itself at 14.3 %. In rural areas, the contribution of the household size increased from 7.0 to 11.7 % during the same period. This may reflect the fact that bigger households may mean either more nuclear families within the same household (a specific feature of the Indian society) and therefore possibly more wage-earners, or more dependent members thus accounting for disparity.

The contribution of employment status to inequality is comparatively higher in rural (4.2 %) compared to urban (2.7 %) areas in 2004–2005. The relative inequality weight of employment status has declined by one percentage point between 1993–1994 and 2004–2005 in rural areas. At the disaggregate level, this decline is in the self-employed and salaried category, as casual workers see their contribution to inequality rise by 1.2 percentage points. In urban areas, the factor inequality shares declined by half percentage point in the overall period; the salaried category is the sole reason for the decline, with a decline of almost 1 percentage points partially compensated by the other two categories. These results imply that variations in casual labour status tend to be influential for inequality in both areas.

Regional factors, as captured by State dummies, turn out to be important in explaining income inequality especially in rural areas. In India, income varies across States and also within States, and this inter-State variation is to a large extent due to the differences in the level of development. The relative factor contribution of this regional variable to inequality has increased over the different periods in rural areas from 6.9 to 10.8 % and in urban areas, it diminished attesting itself at 2.8 %, less than three times that of rural areas. At the disaggregate level, we find that the factor inequality shares of households residing in developed States are lower than that of poorer States especially in rural areas, reiterating that infrastructure and industrial development are important factors in reducing inequality.

Land ownership is another important factor explaining inequality in rural areas. This factor's inequality share is about 5 % in 2004–2005 and the relative factor inequality only marginally increased between 1983 and 2004–2005. Though caste and religion do have a significant influence in India, the inequality share results do not seem to reflect

<sup>13</sup> We do not discuss the differences between Atkinson's index and Gini to avoid a lengthy paper but the Gini shares are available with the authors upon request.

it<sup>14</sup>. The relative importance of variables representing caste and religious affiliation is quite low. The contribution of social group in rural areas declined from 3.1 to 2.3 % between 1983 and 2004–2005, in the same period in urban areas, the share remained stable at 1.1/1.2 %.

The above results are confirmed if we look at the *change* in inequality (cf. Table 2). Between 1993–1994 and 2004–2005, education accounts for a huge increase in the Atkinson's coefficient in urban India. The other variables which substantially contribute to change in inequality are age, industry and social groups in urban areas. In rural India, education and employment status account for the increase in Atkinson coefficients, while household size and religion account for the decline in the Atkinson coefficient. The analysis clearly brings out the emergence of education as the most important factor contributing to inequality (increasing or declining).

## 5 Concluding remarks

This paper completes and extends the methodology of [Shorrocks \(1982\)](#) and [Fields \(2003\)](#) for decomposing inequality into its factor shares. We first derive the asymptotic distribution of the share estimators given by Fields. Then we show that Fields' formula is not applicable to Atkinson's index, and provide an extension to obtain the shares for Atkinson's index. We also derive the asymptotic variances of Atkinson's shares emphasising the point that all economic interpretations have to be based not only on the estimates of parameters but also on their standard errors calculated in a correct manner.

We apply our methodology to examine the factors contributing to income inequality in rural and urban India. The empirical analysis shows that education and household size emerge as the most dominant factors contributing to inequality in both rural and urban areas. The relative importance of education has increased much more in rural areas. The employment status of the household head is another major factor contributing to inequality, and its importance is much bigger in rural than in urban areas. In rural areas, land ownership, regional factors and caste to a lesser extent do play some role whereas other factors like religion or industry group are only marginal in explaining inequality. In urban areas, regional differences are important and industry's share, although slightly less than region, increased between 1993–1994 and 2004–2005.

**Acknowledgments** The authors are grateful to Gary Fields for useful comments on an earlier version of this paper. We also thank Ajit Ghose for helpful discussions. The authors would like to thank the referees for their valuable comments which helped to improve the paper substantially.

<sup>14</sup> This could unfortunately be due to the nature of available data. In the analysis, we compare SCs and STs with 'Others', the latter being an all encompassing category that includes everyone else. This is a large heterogeneous category that includes castes that are very low in the hierarchy, not necessarily very different from the SCs and STs in status and in economic conditions. It is possible that this classification actually underestimates the relative disadvantage of scheduled castes and scheduled tribes with respect to the 'higher' castes.

## Appendices

Appendix A: Shorrocks' theorem and its six assumptions:

Shorrocks' theorem can be stated as follows:

Let the total income of an individual  $i$  ( $i = 1, \dots, N$ ) be written as a sum of different components (sources)

$$Y_i = \sum_{k=1}^K Y_{ik} \quad k = 1, \dots, K$$

and let  $I(Y)$  be an inequality measure of the distribution of incomes denoted as  $Y = [Y_1, \dots, Y_N]'$ . Then the share of each component  $s_k$  to the total inequality is given by:

$$s_k = \frac{\text{cov}(Y_k, Y)}{\sigma^2(Y)}$$

such that:

$$\sum_k s_k = 1$$

provided the following six important assumptions<sup>15</sup> are satisfied by  $I(Y)$ .

1.  $I(Y)$  is continuous and symmetric, and it is equal to 0 if and only if all individuals have the same income,  $Y' = [\mu \dots \mu]$ .
2.  $S_k(Y_1, \dots, Y_K, K) = S_k(Y_k, Y)$ , the contribution of factor  $k$ , is continuous in  $Y_k$  (*continuity*), and if  $\pi_1, \dots, \pi_K$  is a permutation of  $1, \dots, K$ , then  $S_k(Y_{\pi_1}, \dots, Y_{\pi_K}, K) = S_{\pi_k}(Y_1, \dots, Y_K, K)$  (*symmetric treatment of factors*).
3. The contribution of a factor does not depend on how the others are grouped (*independence of the level of disaggregation*).
4. The sum of the contributions is equal to the inequality measure (*consistent decomposition*):  $\sum_k S_k(Y_1, \dots, Y_K, K) = I(Y)$ .
5. If  $P$  is any permutation matrix, then  $S_k(Y_k, Y) = S_k(Y_k P, Y P)$  (*population symmetry*) and if all individuals have the same income for factor  $k$  its contribution is  $S_k(Y_k, Y) = 0$  (*normalisation for equal factor distribution*).
6. (*Two factor symmetry*) Suppose that the incomes of factor 2 are a permutation of those of factor 1 ( $Y_2 = Y_1 P$ ). Then if there are only these two sources of income, they should receive the same value in the decomposition:  $S_1(Y_1; Y = Y_1 + Y_1 P) = S_2(Y_2 = Y_1 P; Y = Y_1 + Y_1 P)$ .

As one can see from the substance of these assumptions inserted in parentheses, they are reasonably basic and self-explanatory. They can also be considered realistic as the majority of inequality measures satisfy them. The assumption of consistent decomposition (i.e. Assumption 4) is the only one that can cause some problems as an

<sup>15</sup> A combination of the formulations by Shorrocks (1982) and Fields (2003) has been used here.

additive form cannot be obtained for a few measures and a workaround is needed. An alternative approach is introduced here for Atkinson’s measure which does not satisfy Assumption 4 and it can be replicated for other such measures.

Appendix B: Proof of Theorem 2

Let us recall some important limit results in the form of a lemma which will be useful for future reference<sup>16</sup>.

- Lemma 1**
1.  $x_n \xrightarrow{d} x, \quad y_n \xrightarrow{p} \alpha \Rightarrow x_n + y_n \xrightarrow{d} x + \alpha$
  2.  $x_n \xrightarrow{d} x, \quad y_n \xrightarrow{p} 0 \Rightarrow y_n x_n \xrightarrow{d} 0$
  3.  $x_n \xrightarrow{d} x, \quad A_n \xrightarrow{p} A \Rightarrow A_n x_n \xrightarrow{d} Ax$

where  $x_n, x$  and  $y_n$  are random variables,  $A_n$  is a deterministic sequence, and  $\alpha$  and  $A$  are constants.

We make two basic assumptions on  $x$  and  $\epsilon$ :

$$\frac{1}{n}x'Mx \xrightarrow{p} Q_x, \quad \text{and} \quad \frac{1}{n}x'M\epsilon \xrightarrow{d} 0$$

where  $M = I_n - \frac{1}{n}t_n t_n'$ .

To prove the theorem we need to derive the limiting distribution of the vector of the shares for the factors (excluding the constant term), say  $\hat{s}^*$ . From Eq. (6), we have:

$$\hat{s}^* = \begin{bmatrix} \hat{s}_2 \\ \vdots \\ \hat{s}_K \end{bmatrix} = \begin{bmatrix} \frac{\hat{\sigma}_{x_2 y}}{\hat{V}(y)} \hat{\beta}_2 \\ \vdots \\ \frac{\hat{\sigma}_{x_K y}}{\hat{V}(y)} \hat{\beta}_K \end{bmatrix} = \frac{1}{\hat{V}(y)} \begin{bmatrix} \hat{\sigma}_{x_2 y} & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_{x_K y} \end{bmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{pmatrix} \equiv \frac{1}{\hat{V}(y)} D \hat{\beta}$$

using the notation  $D$  for the diagonal matrix in the middle of the above expression. We denote as  $\tilde{\beta}$  is the sub-vector of  $\beta$  when the constant term is excluded and  $\hat{\tilde{\beta}}$  is its estimator. Standard estimators of variances and covariances are given by:

$$\hat{V}(y) = \frac{1}{n} \sum_i (y_i - \bar{y})^2 = \frac{1}{n} y' M y$$

$$\hat{\sigma}_{x_k y} = \frac{1}{n} \sum_i (x_{k,i} - \bar{x}_k) (y_i - \bar{y}) = \frac{1}{n} y' M x_k$$

<sup>16</sup> The notation  $\xrightarrow{p}$  denotes convergence in probability and  $\xrightarrow{d}$  convergence in distribution. The results of Lemma 1 will not be proved here as they are available in many econometrics textbooks (e.g. see Greene (2007), page 1049).

We can now write  $D$  as:

$$D = \begin{bmatrix} \frac{1}{n}y'Mx_2 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n}y'Mx_K \end{bmatrix} \\ = \frac{1}{n} \begin{pmatrix} y'M & & 0 \\ & \ddots & \\ 0 & & y'M \end{pmatrix} \begin{pmatrix} x_2 & & 0 \\ & \ddots & \\ 0 & & x_K \end{pmatrix} = \frac{1}{n} (I_{K-1} \otimes y'M) x^*$$

where the notation  $x^*$  is used for the last diagonal matrix in the above expression. The OLS estimator  $\hat{\beta}$  is<sup>17</sup>:

$$\hat{\beta} = (x'Mx)^{-1} x'My$$

Replacing  $\hat{V}(y)$  and  $D$  in the expression of  $\hat{s}^*$ , we get:

$$\hat{s}^* = \frac{1}{\hat{V}(y)} D \hat{\beta} = \frac{\frac{1}{n} (I_{K-1} \otimes y'M) x^* \hat{\beta}}{\frac{1}{n} y'My}$$

We can now write  $\sqrt{n}(\hat{s}^* - s)$  as:

$$\sqrt{n}(\hat{s}^* - s) = \sqrt{n} \left( \frac{\frac{1}{n} (I_{K-1} \otimes y'M) x^* \hat{\beta}}{\frac{1}{n} y'My} - \frac{\frac{1}{n} (I_{K-1} \otimes y'M) x^* \beta}{\frac{1}{n} y'My} \right) \\ = \frac{\frac{1}{n} (I_{K-1} \otimes y'M) x^*}{\frac{1}{n} y'My} \sqrt{n} (\hat{\beta} - \beta)$$

Let us now study the different parts of the equation, starting with  $\frac{1}{n}y'My$ :

$$\frac{1}{n}y'My = \frac{1}{n} (\beta_1 l_n + \tilde{x}\tilde{\beta} + \epsilon)' M (\beta_1 l_n + \tilde{x}\tilde{\beta} + \epsilon) \\ = 0 + \frac{1}{n} \tilde{\beta}' \tilde{x}' M \tilde{x} \tilde{\beta} + \frac{1}{n} \epsilon' M \tilde{x} \tilde{\beta} + \frac{1}{n} \tilde{\beta}' \tilde{x}' M \epsilon + \frac{1}{n} \epsilon' M \epsilon$$

The asymptotic limit of the first term is given by

$$\frac{1}{n} \tilde{\beta}' \tilde{x}' M \tilde{x} \tilde{\beta} \xrightarrow{p} \tilde{\beta}' Q_x \tilde{\beta},$$

<sup>17</sup> Here we use the standard OLS estimator but other estimators could be used in which case the computation of the asymptotic distribution will have to be modified accordingly. Note that the constant term disappears with the  $M$  transformation.

The second and third terms go to 0 and the last term goes to the variance of  $\epsilon$ . So at the end we have:

$$\frac{1}{n}y'My \xrightarrow{p} \tilde{\beta}'Q_x\tilde{\beta} + \sigma_\epsilon^2 \tag{9}$$

We know the asymptotic distribution<sup>18</sup> of  $\sqrt{n}(\hat{\beta} - \tilde{\beta})$ :

$$\sqrt{n}(\hat{\beta} - \tilde{\beta}) \xrightarrow{d} N(0; \sigma_\epsilon^2 Q_x^{-1}) \tag{10}$$

The last part to be examined is  $\frac{1}{n}(I_{K-1} \otimes y'M)x^*$ . To analyse its asymptotic behaviour, let us go back to the matrix notation:

$$\frac{1}{n}(I_{K-1} \otimes y'M)x^* = \begin{bmatrix} \frac{1}{n}y'Mx_2 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n}y'Mx_K \end{bmatrix}$$

and look at the  $k$ -th term for example,  $\frac{1}{n}y'Mx_k$ . Using a selection vector  $l_k$ , that only selects the  $k^{th}$  element<sup>19</sup> and substituting for  $y'$  we can write:

$$\frac{1}{n}y'Mx_k = \tilde{\beta}'\tilde{x}'M\tilde{x}\frac{1}{n}l_k + \frac{1}{n}\epsilon'M\tilde{x}l_k$$

It can be easily verified that:

$$\frac{1}{n}y'Mx_k \xrightarrow{p} \tilde{\beta}'Q_x l_k + 0 = \tilde{\beta}'Q_x l_k$$

Going back to the matrix notation we obtain<sup>20</sup>:

$$\frac{1}{n}(I_{K-1} \otimes y'M)x^* \xrightarrow{p} (I_{K-1} \otimes \tilde{\beta}'Q_x)L \tag{11}$$

<sup>18</sup> See for example [Greene \(2007\)](#), page 67.

<sup>19</sup>  $l_k$  is a vector of dimension  $K - 1$  with 0 everywhere and 1 on the  $(k - 1)^{th}$  position:  $l_k = [0 \dots 010 \dots 0]'$ .

Using this vector we have  $x_k = \tilde{x} \cdot l_k = [x_2 \dots x_K] \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ .

<sup>20</sup> Let  $L = \begin{bmatrix} l_2 & & 0 \\ & \ddots & \\ 0 & & l_K \end{bmatrix}$ .

Putting together the results in Eqs. (10) and (11) and using Lemma 1.3, we get :

$$\frac{1}{n} (I_{K-1} \otimes y' M) x^* \sqrt{n} (\hat{\beta} - \tilde{\beta}) \xrightarrow{d} N\left(0; \sigma_\epsilon^2 \left( I_{K-1} \otimes \tilde{\beta}' Q_x \right) L Q_x^{-1} L' \left( I_{K-1} \otimes Q_x \tilde{\beta} \right)\right)$$

Now using Lemma 1.1 we find the result:

$$\sqrt{n} (\hat{s}^* - s) \xrightarrow{d} N\left(0; \sigma_\epsilon^2 \frac{(I_K \otimes \beta' Q_x) L Q_x^{-1} L' (I_K \otimes Q_x \beta)}{(\beta' Q_x \beta + \sigma_\epsilon^2)^2}\right)$$

The variance–covariance matrix obtained through the asymptotic results can then be written as:

$$Asy.Var(\hat{s}^*) = \Sigma = \frac{1}{n} \sigma_\epsilon^2 \frac{(I_K \otimes \beta' Q_x) L Q_x^{-1} L' (I_K \otimes Q_x \beta)}{(\beta' Q_x \beta + \sigma_\epsilon^2)^2}$$

We can now move on to compute the variance of the all vector of shares  $\hat{s}$  (i.e. including the share for the residuals). We have  $\hat{s}_\epsilon = 1 - \sum_{k=1}^K \hat{s}_k = 1 - l' \hat{s}^*$  and therefore:

$$\hat{s} = \begin{bmatrix} \hat{s}_2 \\ \vdots \\ \hat{s}_K \\ \hat{s}_\epsilon \end{bmatrix} = \begin{bmatrix} \hat{s}^* \\ 1 - l' \hat{s}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} I_k \\ -l' \end{bmatrix} \hat{s}^*$$

We can than easily compute the variance of the vector of all shares as:

$$Asy.Var(\hat{s}) = 0 + \begin{bmatrix} I_k \\ -l' \end{bmatrix} Asy.Var(\hat{s}^*) \begin{bmatrix} I_k & -l' \end{bmatrix} = \begin{bmatrix} \Sigma & -\Sigma l \\ -l' \Sigma & -l' \Sigma l \end{bmatrix}$$

Thus we have:

$$\sqrt{n} (\hat{s} - s) \xrightarrow{d} N\left(0; \begin{bmatrix} \Sigma & -\Sigma l \\ -l' \Sigma & l' \Sigma l \end{bmatrix}\right)$$

□

### Appendix C: Simulation experiment

We simulate a simple model specified as follows:

$$y = \beta_1 + x_2 \cdot \beta_2 + d_3 \cdot \beta_3 + \epsilon \tag{12}$$

where  $y$  could be the income variable,  $x_2$  a continuous explanatory variable such as age in our case, and  $d_3$  a dummy explanatory variable such as gender. We generate the two explanatory variables and the error term as follows:



**Table 3** Summary of results for the simulation

$n$	$\beta$	$\hat{\beta}$	$V(\hat{\beta})$	$\hat{s}$	$V_s(\hat{s})$	$Asy.Var(\hat{s})$	Skewness–Kurtosis test			
							Skewness	Kurtosis	Joint	
$x_2$	100	3	2.999	0.731	0.117	3.3E-03	9.6E-04	0.000	0.937	0.000
	1'000	3	3.001	0.067	0.112	3.2E-04	9.5E-05	0.468	0.631	0.684
	10'000	3	3.002	0.007	0.111	3.4E-05	9.5E-06	0.693	0.312	0.554
$D_3$	100	40	39.960	427.8	0.041	1.2E-03	3.4E-04	0.000	0.000	0.000
	1'000	40	40.109	38.43	0.034	1.0E-04	2.9E-05	0.000	0.002	0.000
	10'000	40	39.913	4.230	0.033	1.1E-05	2.8E-06	0.139	0.822	0.327
$\epsilon$	100				0.842	4.0E-03	1.3E-03	0.000	0.220	0.000
	1'000				0.854	3.9E-04	1.2E-04	0.388	0.338	0.436
	10'000				0.856	4.2E-05	1.2E-05	0.533	0.970	0.822

$$\begin{aligned}
 x_2 &\sim N(45, 12) \\
 d_3 &= 1 \text{ if } u \geq 0.6 \text{ and } 0 \text{ otherwise, where } u \sim U[0, 1] \\
 \epsilon &\sim N(0, 100)
 \end{aligned}$$

Inspiring from our empirical results, we fix  $\beta_1 = 500$ ,  $\beta_2 = 3$  and  $\beta_3 = 40$  and generate  $y$  according to Eq. (12). Then  $\hat{\beta}$ , the shares  $\hat{s}$  and their asymptotic variances are computed. We repeat the procedure 1'000 times and for different sample sizes  $n = 100, 1,000$  and  $10,000$ .

Table 3 presents a summary of the results, including the small sample variance  $V_s(\hat{s})$ , the asymptotic variance  $Asy.Var(\hat{s})$  and the p-values of the Skewness–Kurtosis test for normality<sup>21</sup>.

We also explored different values for the coefficients and the results are similar.

#### Appendix D: Solving the problem of consistent decomposition for Atkinson’s Index

In this appendix we derive the decomposition function for Atkinson’s Index. We start by writing the measure as:

$$I_A(Y) = 1 - \frac{1}{\mu} \left[ \frac{1}{n} \sum_i (y_i)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}, \quad \epsilon \geq 0; \quad \epsilon \neq 1$$

Simply rearranging the expression of the inequality measure we obtain:

$$(1 - I_A(Y))^{1-\epsilon} = \frac{1}{\mu^{1-\epsilon}} \frac{1}{n} \sum_i (y_i)^{1-\epsilon}$$

<sup>21</sup> We present the variance instead of MSE because the estimation of  $\beta$  is unbiased even with a small sample size; note that  $V(\hat{\beta})$  denotes the sample variance of  $\hat{\beta}$ .

The above expression gives us the transformation function proposed by [Shorrocks \(1982\)](#) (i.e.:  $f(y) = (1 - y)^{1-\epsilon}$ ). As observed in Sect. 3, this function does not satisfy the condition  $f(0) = 0$ , so we modify it, by subtracting one, to make it respect this condition. We then obtain the following expression for the transformed index:

$$\tilde{I}_A = f(I_A) = (1 - I_A(Y))^{1-\epsilon} - 1 = \frac{1}{\mu^{1-\epsilon}} \frac{1}{n} \sum_i (y_i)^{1-\epsilon} - 1$$

The above expression can be rewritten as:

$$\tilde{I}_A = \frac{1}{\mu^{1-\epsilon}} \frac{1}{n} \sum_i (y_i)^{-\epsilon} \left[ \sum_k y_{i,k} - \frac{1}{n} \right]$$

Thus we can compute the contribution to inequality of the k-th factor as follows:

$$\tilde{S}_k = \frac{1}{\mu^{1-\epsilon}} \frac{1}{n} \sum_i (y_i)^{-\epsilon} \left[ y_{i,k} - \frac{1}{nK} \right]$$

It is easy to show that the two transformed measures ( $\tilde{S}$  and  $\tilde{I}$ ) respect the six assumptions of Shorrocks' theorem and hence the general formula of Theorem 1 can be applied to  $\tilde{I}_A$  to get  $\tilde{s}_k, k = 1, \dots, K$  which can in turn be reconverted into  $s_k$  using  $f^{-1}(\cdot)$  as in (8).

### Appendix E: Variances and covariances of Atkinson shares

Recall from (8) that

$$s_k = \frac{f^{-1}(\tilde{s}_k \tilde{I}_A)}{I_A}$$

with

$$f^{-1}(y) = 1 - (y + 1)^{\frac{1}{1-\epsilon}}$$

Thus

$$\begin{aligned} \frac{\partial s_k}{\partial \tilde{s}_k} &= \frac{1}{I_A} \frac{\partial f^{-1}(\cdot)}{\partial (\cdot)} \tilde{I}_A \\ &= \frac{1}{I_A} \frac{-1}{(1-\epsilon)} (\tilde{s}_k \tilde{I}_k + 1)^{\frac{\epsilon}{(1-\epsilon)}} \tilde{I}_A \end{aligned}$$

and

$$\frac{\partial s_k}{\partial \tilde{s}_j} = 0 \text{ for } k \neq j$$

Hence,

$$V(\hat{s}_k) \cong \left(\frac{\partial s_k}{\partial \hat{s}_k}\right)^2 V(\hat{s}_k)$$

where  $V(\hat{s}_k)$  is given by our Theorem 2.

Denoting the vectors of shares as

$$\hat{s} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ \hat{s}_K \end{bmatrix}; \quad \hat{s} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ \hat{s}_K \end{bmatrix}$$

and writing

$$\frac{\partial s}{\partial \hat{s}'} = \begin{bmatrix} \frac{\partial s_1}{\partial \hat{s}_1} & 0 & \dots & 0 \\ 0 & \frac{\partial s_2}{\partial \hat{s}_2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \frac{\partial s_K}{\partial \hat{s}_K} \end{bmatrix} \equiv \text{say } J$$

we have:

$$Asy.V(\hat{s}) \cong \hat{J}[Asy.V(\hat{s})]\hat{J}' \equiv \text{say } Q \tag{13}$$

Therefore, the variance of the sum of shares, for instance  $l'\hat{s}$ , can be derived as follows:

$$V(l'\hat{s}) \cong l'Ql$$

### Appendix F: Data sources, variable names and definitions

The data are based on multiple rounds of the employment–unemployment survey along with the consumption expenditure survey undertaken by the NSSO every five years, covering major Indian States. We use three rounds corresponding to the years 1983 (38th round), 1993–1994 (50th round) and 2004–2005 (61st round). The detailed characteristics of all household members including sex, age, caste/religion, marital status, relation to the household head, education level, employment status, occupation, industry and the region are provided in the survey. The monthly per capita consumption expenditure that is used as a proxy variable for income is obtained for the same set of households from the consumer expenditure survey. The sample is restricted to the age group 15–64 years and the variables are defined below:

- **Income:** The income variable is proxied by monthly per capita consumer expenditure.
- **Age:** Age of the individual in logarithm.
- **Household size:** Number of persons in the household.

- **Gender:** Dummy variable, indicating female=1, 0 otherwise.
- **Land ownership:** Per capita land possession obtained as land possession owned by the household divided by the number of persons in the household.
- **Social group:** Social group consists of scheduled tribes, scheduled castes and others (other backward caste and forward caste). Two dummy variables for scheduled tribes and scheduled castes are constructed with “Others” as the reference category.
- **Religion:** Religion comprises of Hindus, Muslims, Christians and Others. We have constructed three dummy variables for Hindus, Muslims and Christians separately, with “Others” as the reference category.
- **Education:** We classify education into five categories: illiterate, primary, middle, secondary and above secondary. We generate four dummy variables for illiterate, primary, middle and secondary and the reference category is “above secondary”.
- **Employment status:** The employment status categories that we consider are self-employment, casual worker, salaried and unemployed. The self-employment comprises own account workers, employers and unpaid family workers; salaried workers comprises regular salaried and waged employee and the casual workers comprises casual labour in public works or other type of works. We create three dummy variables with “unemployed” as the reference category.
- **Industry:** We aggregate the industries classified under National Industrial Classification to six industry groups with similar qualitative characteristics: agriculture (comprises agriculture, forestry and fishing); manufacturing (comprises mining and manufacturing); electricity, gas and water; construction; low-skilled services sector (comprises trade, hotels and restaurant, transport and personal services) and high-skilled services sector (comprises banking and insurance, communication, real estate, business services and public administration). The categorization of the service sector into two groups is justified on the basis of skill and capital requirements. “agriculture” is used as reference category and we constructed five dummy variables for each of the other industry groups.
- **State dummies:** We have generated state dummies for 15 major States in India and the remaining states are used as the reference category. The 15 major states for which we have generated dummies are Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal.

Appendix G: Shares of inequality and their asymptotic variances for Atkinson’s index

Table 4 Atkinson's index: disaggregate factor shares, their asymptotic standard errors and contribution to the change of inequality

	Urban				Rural				
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
Age	-0.105 (4.4E-04)	*** (5.3E-04)	*** (7.4E-04)	*** (6.0E-04)	*** (7.6E-04)	*** (6.8E-04)	0.021 (6.0E-04)	0.057 (7.6E-04)	0.152 (6.8E-04)
Gender	0.010 (7.5E-04)	*** (2.6E-04)	*** (2.5E-04)	*** (6.2E-05)	*** (6.5E-04)	*** (9.9E-05)	0.000 (6.2E-05)	–	0.007 (9.9E-05)
Household size	0.428 (3.7E-03)	*** (1.9E-03)	*** (1.9E-03)	*** (8.2E-04)	*** (1.2E-03)	*** (9.0E-04)	0.182 (8.2E-04)	0.282 (6.5E-04)	0.544 (9.0E-04)
Land ownership							0.117 (1.1E-03)	0.163 (1.2E-03)	0.223 (9.4E-04)
Scheduled tribe	0.006 (1.8E-04)	*** (3.8E-05)	*** (9.2E-05)	*** (4.6E-04)	*** (3.7E-04)	*** (9.5E-05)	0.048 (4.6E-04)	0.053 (3.7E-04)	0.053 (9.5E-05)
Scheduled caste	0.017 (1.2E-03)	*** (7.8E-04)	*** (8.2E-04)	*** (5.8E-04)	*** (6.1E-04)	*** (5.0E-04)	0.034 (5.8E-04)	0.042 (6.1E-04)	0.055 (5.0E-04)
Hindus	–	–0.008 (7.1E-04)	*** (6.5E-04)	*** (7.1E-04)	*** (1.6E-04)	*** (1.4E-04)	–	-0.002 (1.6E-04)	-0.004 (1.4E-04)
Muslims	0.005 (7.1E-04)	*** (4.1E-04)	*** (5.2E-04)	*** (3.4E-04)	*** (4.4E-04)	*** (5.9E-04)	0.005 (3.4E-04)	0.009 (4.4E-04)	0.037 (5.9E-04)
Christians	0.009 (7.2E-04)	*** (3.5E-04)	*** (3.9E-04)	*** (5.9E-04)	*** (5.2E-04)	*** (5.1E-04)	0.009 (5.9E-04)	0.004 (5.2E-04)	0.031 (5.1E-04)
Manufacturing	-0.002 (2.1E-04)	*** (1.0E-04)	*** (2.8E-04)	*** (6.4E-05)	*** (3.2E-05)	–	0.002 (6.4E-05)	0.007 (3.2E-05)	–
Electricity, gas and water	0.006 (3.3E-04)	*** (1.2E-04)	*** (2.8E-04)	*** (1.5E-04)	*** (6.3E-05)	*** (1.2E-04)	0.005 (1.5E-04)	0.002 (6.3E-05)	0.005 (1.2E-04)
Construction	-0.006	*** (-0.020)	*** (-0.015)	*** (-0.001)	*** (-0.002)	*** (-0.007)	-0.001 (-0.001)	-0.002 (-0.002)	-0.007 (-0.007)

Table 4 continued

	Urban			Rural		
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
Low-skilled services sector	(4.4E-04) -0.012 (1.4E-03)	(3.6E-04) -0.024 (1.0E-03)	(8.8E-04) -0.011 (1.3E-03)	(2.5E-05) 0.001 (1.2E-04)	(1.5E-05) 0.019 (2.5E-04)	(2.4E-04) 0.006 (1.6E-05)
High-skilled services sector	0.040 (4.0E-03)	0.065 (2.3E-03)	0.095 (2.7E-03)	0.055 (1.1E-03)	0.072 (1.4E-03)	0.069 (1.3E-03)
Illiterates	0.348 (4.3E-03)	0.581 (2.0E-03)	0.489 (2.1E-03)	0.097 (4.5E-03)	0.522 (3.5E-03)	0.737 (2.1E-03)
Primary education	0.204 (2.5E-03)	0.374 (1.4E-03)	0.357 (1.5E-03)	- 0.030	-0.049 (2.8E-04)	0.017 (4.0E-04)
Middle-level education	-0.007 (4.7E-05)	0.086 (4.6E-04)	0.112 (4.6E-04)	0.030 (1.3E-03)	-0.062 (8.8E-04)	-0.101 (4.1E-04)
Secondary education	-0.123 (3.5E-03)	-0.220 (1.1E-03)	-0.106 (9.9E-04)	0.085 (1.8E-03)	-0.021 (1.9E-03)	-0.133 (1.2E-03)
Self-employment	-	-	0.013 (1.1E-03)	0.066 (1.1E-03)	0.072 (1.2E-03)	0.042 (2.1E-04)
Casual worker	0.051 (2.9E-03)	0.101 (1.9E-03)	0.096 (2.1E-03)	-0.019 (2.0E-03)	-0.021 (2.4E-03)	0.031 (1.6E-03)
Salaried worker	0.028 (6.0E-03)	-	-0.026 (2.9E-03)	0.063 (1.2E-03)	0.143 (1.8E-03)	0.120 (1.7E-03)
Andhra Pradesh	0.010 (5.3E-04)	0.017 (2.5E-04)	0.002 (1.3E-04)	-0.004 (3.3E-05)	-0.002 (9.3E-06)	0.003 (1.8E-05)
Assam	0.000 (9.3E-06)	0.000 (4.8E-05)	0.000 (2.4E-05)	-0.002 (4.5E-05)	0.011 (2.1E-04)	0.000 (7.9E-05)

Table 4 continued

	Urban				Rural				
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
Bihar	0.006 (4.0E-04)	*** 0.022 (4.2E-04)	*** 0.011 (4.9E-04)	*** 0.053 (6.6E-04)	*** 0.121 (7.8E-04)	*** 0.067 (4.2E-04)	*** 0.026 (1.3E-04)	*** 0.062 (2.7E-04)	*** 0.007 (1.3E-04)
Gujarat	0.001 (1.1E-04)	*** 0.001 (6.2E-05)	*** 0.004 (1.6E-04)	*** -0.001 (7.4E-05)	*** -	*** 0.026 (2.6E-04)	*** 0.003 (6.2E-05)	*** 0.139 (6.6E-04)	*** 0.032 (3.1E-04)
Haryana	-	*** -0.001 (1.5E-05)	*** 0.001 (6.0E-05)	*** 0.022 (3.1E-04)	*** 0.003 (6.2E-05)	*** 0.007 (1.3E-04)	*** 0.139 (6.6E-04)	*** 0.006 (2.7E-04)	*** 0.007 (1.3E-04)
Karnataka	-	*** 0.002 (6.9E-05)	*** -	*** -0.002 (3.8E-05)	*** 0.003 (4.1E-04)	*** 0.139 (6.6E-04)	*** 0.023 (5.6E-04)	*** 0.031 (3.5E-04)	*** 0.008 (1.0E-04)
Kerala	-	*** -0.001 (6.9E-05)	*** 0.002 (1.7E-04)	*** 0.003 (4.1E-04)	*** 0.018 (3.5E-04)	*** 0.032 (3.1E-04)	*** 0.139 (6.6E-04)	*** 0.032 (3.1E-04)	*** 0.032 (3.1E-04)
Madhya Pradesh	0.012 (5.8E-04)	*** 0.015 (2.8E-04)	*** 0.011 (3.6E-04)	*** 0.018 (3.5E-04)	*** 0.031 (3.5E-04)	*** 0.032 (3.1E-04)	*** 0.139 (6.6E-04)	*** 0.032 (3.1E-04)	*** 0.032 (3.1E-04)
Maharashtra	0.006 (1.0E-03)	*** 0.009 (6.2E-04)	*** 0.018 (5.2E-04)	*** 0.001 (8.7E-05)	*** 0.008 (1.0E-04)	*** -0.004 (2.2E-05)	*** 0.007 (1.3E-04)	*** 0.007 (1.3E-04)	*** 0.007 (1.3E-04)
Orissa	0.001 (1.2E-04)	*** 0.003 (1.1E-04)	*** 0.011 (3.8E-04)	*** 0.016 (2.6E-04)	*** 0.054 (4.2E-04)	*** 0.087 (2.2E-05)	*** 0.071 (5.5E-04)	*** 0.071 (5.5E-04)	*** 0.071 (5.5E-04)
Punjab	-	*** 0.001 (3.1E-04)	*** 0.003 (1.9E-04)	*** 0.039 (8.2E-04)	*** 0.066 (6.7E-04)	*** 0.071 (5.5E-04)	*** 0.071 (5.5E-04)	*** 0.071 (5.5E-04)	*** 0.071 (5.5E-04)
Rajasthan	-	*** 0.002 (6.2E-05)	*** -	*** -	*** -	*** 0.006 (4.8E-06)	*** 0.006 (4.8E-06)	*** 0.006 (4.8E-06)	*** 0.006 (4.8E-06)
Tamil Nadu	0.007 (5.0E-04)	*** 0.015 (2.4E-04)	*** 0.000 (1.1E-04)	*** 0.001 (1.1E-04)	*** -0.003 (1.7E-05)	*** -0.002 (1.0E-05)	*** -0.002 (1.0E-05)	*** -0.002 (1.0E-05)	*** -0.002 (1.0E-05)
Uttar Pradesh	0.024	*** 0.027	*** 0.023	*** 0.024	*** 0.007	*** 0.006	*** 0.006	*** 0.006	*** 0.006

Table 4 continued

	Urban			Rural		
	1983	1993–1994	2004–2005	1983	1993–1994	2004–2005
West Bengal	(1.1E-03) -0.003	(4.9E-04) -0.005	(5.6E-04) 0.000	(3.8E-04) 0.014	(1.7E-04) 0.007	(3.1E-04) 0.002
	(4.3E-04)	(4.3E-05)	(3.0E-05)	(2.4E-04)	(1.0E-04)	(1.2E-04)
Residuals	1.045	1.533	1.446	1.628	2.155	2.248
	(4.0E-03)	(2.4E-03)	(2.4E-03)	(1.5E-03)	(1.7E-03)	(1.7E-03)

Source Same as Table 1. - indicates a non-significant beta. Figures in brackets are standard errors. The above shares do not add upto 1 as mentioned in the text on page 7; they have been normalised to add upto 100 % in Table 2 to facilitate the interpretation

\*\*\*, \*\*, \* Significant at the 1, 5, 10 percent, respectively



## References

- Ameida dos Reis JG, de Barros RE (1991) Wage inequality and the distribution of education. *J Dev Econ* 36(1):117–143
- Barro RJ (2000) Inequality and growth in a panel of countries. *J Econ Growth* 5(1):5–32
- Becker GS (1964) Human capital. National Bureau of Economic Research, New York
- Bhaduri A (2008) Predatory growth: commentary. *Econ Political Wkly* 43(16):10–14
- Chow G (1983) *Econometrics*. McGraw-Hill, New York
- Cowell FA, Jenkins SP (1995) How much inequality can we explain?: a methodology and an application to the United States. *Econ J* 105(429):421–430
- Das DK (2003) Manufacturing productivity under varying trade regimes: India in the 1980s and 1990s. Working Paper 107, Indian Council for Research on International Economics Relations, New Delhi
- Deaton A (1997) Analysis of household surveys: a microeconomic approach to development policy. World Bank Publications, Washington DC
- Deininger K, Squire L (1998) New ways of looking at old issues: inequality and growth. *J Dev Econ* 57(2):259–287
- Dev MS, Ravi C (2007) Poverty and inequality: all India and states, 1983–2005. *Econ Political Wkly* 42(6):509–521
- Fei JC, Ranis G, Kuo SWY (1978) Growth and the family distribution of income by factor components. *Q J Econ* 92:17–53
- Fields GS (2003) Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the United States. *Res Labour Econ* 22:1–38
- Greene WH (2007) *Econometric analysis*, 6th edn. Prentice Hall, New York
- International Labour Organisation (LO) (2008) World of Work Report: Inequalities in the age of Financial Globalisation. International Institute of Labour Studies, Geneva
- Kohli A (2006) Politics of economic growth in India, 1980–2005, Part II: the 1990s and beyond. *Econ Political Wkly* 41(14):1361–1370
- Lam D, Levison D (1991) Declining inequality in schooling in Brazil and its effects on inequality in earnings. *J Dev Econ* 37:199–225
- Leamer E (1978) *Ad Hoc inference with nonexperimental data*. Wiley, New York
- Londono JL (1996) Poverty, inequality and human capital development in Latin America 1950–2025. World Bank Latin American and Caribbean Studies, Washington DC
- Mincer J (1958) Investment in human capital and personal income distribution. *J Political Econ* 66:281–302
- Mincer J (1970) The distribution of labour incomes: a survey with the special reference to the human capital approach. *J Econ Lit* 8(1):1–26
- Morduch J, Sicular T (2002) Rethinking inequality decomposition, with evidence from rural China. *Econ J* 112:93–106
- National Sample Survey Organisation (NSSO) (1987). Employment and unemployment situation in India 1983. NSS 38 Round (January 1983–December 1983); No. 341. Department of Statistics, Government of India, New Delhi (Raw data from this survey was used)
- National Sample Survey Organisation (NSSO) (1997). Employment and Unemployment Situation in India 1993–94. NSS 50 Round (July 1993–June 1994); No. 409 Department of Statistics, Government of India, New Delhi (Raw data from this survey was used)
- National Sample Survey Organisation (NSSO) (2006a). Employment and unemployment situation in India 2004–5 (Part-I). NSS 61 Round (July 2004–June 2005); No. 515 (61/10/1). Department of Statistics, Government of India, New Delhi (Raw data from this survey was used)
- National Sample Survey Organisation (NSSO) (2006b). Employment and unemployment situation in India 2004–5 (Part-II). NSS 61 Round (July 2004–June 2005); No. 515 (61/10/1). Department of Statistics, Government of India, New Delhi (Raw data from this survey was used)
- Oaxaca RL (1973) Male–female wage differentials in urban labour markets. *Int Econ Rev* 14(3):693–709
- Organisation for economic co-operation and development (OECD) (2008) *Growing Unequal? Income Distribution and Poverty in OECD Countries*, OECD, Paris
- Pyatt G, Chen C, Fei J (1980) The distribution of income by factor components. *Q J Econ* 95(3):451–473
- Rodrik D, Subramanian A (2005) From Hindu growth to productivity surge: the mystery of the Indian growth transition. *IMF Staff Papers* 52(2):193–228
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464

- Sengupta A, Kannan KP, Raveendran G (2008) India's common people: who are they, how many are they and how do they live? *Econ Political Wkly* 43(11):49–63
- Shapley L (1953) A value for  $n$ -person games. In: Kuhn HW, Tucker AW (eds) *Contributions to the theory of games*, vol 2. Princeton University Press, Princeton
- Shorrocks AF (1982) Inequality decomposition by factor components. *Econometrica* 50(1):193–211
- Shorrocks AF (1984) Inequality decomposition by population sub-groups. *Econometrica* 52:1369–1385
- Sundaram K, Tendulkar SD (2003) Poverty in India in the 1990s: revised estimates. *Econ Political Wkly* 38(46):4865–4872

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.