Accounting for Wage Inequality in India

PUJA VASUDEVA DUTTA[1]

Poverty Research Unit at Sussex
Department of Economics
University of Sussex

PRUS Working Paper no. 29

Abstract

This paper investigates the evolution and structure of wage inequality among adult male workers engaged in regular and casual wage employment in India during a period of radical economic change. The analysis exploits data from nationally representative employment surveys and uses decomposition techniques to examine the role played by educational achievement and industry affiliation. This paper finds that there are striking differences for the two groups of workers. Wage inequality rose between 1983 and 1999 among regular workers but fell among casual workers. While human capital (as embodied in age and education) is one of the major factors explaining both the level of and change in regular wage inequality, geographic location is the key determinant of casual wage inequality. Industry affiliation plays an equally important role for both sets of workers. These are also consistently the most important contributors to changes in inequality though the directional effects differ among the different sets of workers..

JEL codes: D33, J31, J42

Keywords: India, wage inequality, inequality decomposition, segmented labour market

Work in progress: Comments welcome.
January 2005

Acknowledgements

I am grateful to Dr. Barry Reilly for his suggestions and guidance and to participants at the International Seminar on Wages and Incomes in India held at Mumbai (2004). During the course of this work I have received funding from the British Federation of Women Graduates, the Royal Economic Society, Universities UK and the Wingate Foundation. The survey data were obtained from the National Sample Survey Organisation.

[1] Poverty Research Unit, Department of Economics, School of Social Sciences and Cultural Studies, University of Sussex, Falmer, Brighton, BN1 9SN, UK. Email P.Vasudeva@sussex.ac.uk

1. Introduction

The late 1980s and 1990s were a period of rapid industrial deregulation and trade liberalisation in India. External sector reforms included the market determination of the exchange rate, tariffication of quantitative restrictions and their subsequent reduction, export promotion and the establishment of export processing zones and the removal of restrictions on inflows of foreign capital. Industrial policy reforms were initiated in the mid-1980s with respect to industrial licensing regulations, the role of public sector enterprises, large firms and foreign direct investment. Labour market regulations such as the Industrial Disputes Act that prevent closure of units and the lay-off of workers without prior government approval, however, were not brought under the liberalisation agenda until 2001 (see Nouroz (2001) and Kapila (2001) for a detailed review). The decade of the 1990s was also one when India seemed to fulfil its growth potential, growing at an average of six to seven per cent per annum. There is some evidence that this process of economic change was accompanied by rising wage inequality among workers in the organised manufacturing sector (see Galbraith et al. (2004)). The primary motivation of this paper is to investigate the structure of wage inequality in India during this period of radical economic change, with a particular focus on the role played by education and industry affiliation. This paper exploits three national employment surveys for the years 1983, 1993-94 and 1999-2000 - the first survey can be interpreted as providing insights into the structure of Indian labour markets prior to liberalisation while the latter two provide the basis for delineating a portrait of these structures after the liberalisation process.

Several studies have documented rising wage inequality between skilled and unskilled workers in developed countries, particularly the United States and United Kingdom, and developing countries, particularly Latin American countries, since the 1980s (see Katz and Autor (1999) and Wood (1997) for a review). The explanations put forward to account for this rising wage gap include institutional factors (such as changes in government policy), supply-side factors (such as demographic shifts and immigration rates) as well as demand-side factors (such as changes in the relative demand for skilled labour due to the rise in international trade and skill-biased technological change) (see Machin (2002)). While these explanations apply to changes in wage inequality in general, in practice, empirical studies examine almost exclusively the changes in the skill premia (proxied by education) and/or inter-industry wage differentials. The contribution of these two factors to wage inequality in India is the focus of the current paper.

This paper is structured as follows. The next section describes our characterisation of the Indian labour market in terms of a dual labour market framework, i.e., the wage distributions of workers engaged in regular and casual wage employment are examined separately rather than conflating them into one category. Section 3 examines the trends in wage inequality for both regular and casual workers. There are relatively few estimates of wage inequality for India. A simple accounting approach is used to decompose overall wage inequality into inequality between and within population sub-groups, with the groups defined by education (a proxy for skill) and industry affiliation. Lastly, an alternative decomposition is undertaken that examines the contribution of various explanatory variables to the level of and change in wage inequality within a regression-based framework following Fields (2002). Augmented Mincerian wage equations are estimated for regular and casual workers separately using a set of human capital measures and a variety of worker, industry and state characteristics after correcting for the presence of selection bias. The estimates from these regression models are used in the Fields' decomposition. Section 5 offers some conclusions.

The contribution of this paper is two-fold. First, this study adds to our understanding of the evolution and structure of inequality in India. There are almost no comprehensive empirical studies of trends in wage inequality for India with the work of Galbraith et al. (2004) providing one notable exception. However, Galbraith et al. compute inequality measures using grouped data (grouped by industry and by industry and state) relating to earnings (calculated as the annual wage bill divided by the number of workers) in the organised manufacturing sector. This paper, on the other hand, computes wage inequality measures using data on individual wage rates for all workers engaged in regular or casual wage employment in all economic sectors including agriculture, organised and unorganised manufacturing and services. Second, this study augments the growing literature on the empirical application of a regression-based decomposition technique in order to examine the significance and contribution of various factors to wage inequality. At the same time it should be stressed that this paper is essentially an accounting exercise and there is no attempt to examine the reasons for the change in the underlying factors, such as the returns to education or industry affiliation, that contributed to the change in wage inequality during this period.

However, by examining in detail the structure of wage inequality for different sets of workers this paper sets the stage for a future investigation of these links.

2. The informal economy: regular-casual worker dichotomy

The dual labour market model supposes the existence of two distinct sectors of economic activity usually classified as the organised and unorganised sectors. The organised sector offers more stable jobs with higher pay, better working conditions and promotional opportunities whereas the unorganised sector is associated with unstable jobs and low or even flat returns to schooling, poor pay, bad working conditions and few opportunities for advancement (Dickens and Lang (1985); Taubman and Wachter (1986)). Thus the dual labour market approach argues that there are two distinct types of jobs with separate wage determination processes.

Unni (2001, pp. 2361) argues that the notion of an "informal economy" that characterises workers "depending on the degree of informality of their work status" is more relevant for examining wage structures. Similarly, Tendulkar (2003) and Das (2003) argue that in the Indian context the organised-unorganised dichotomy generally used to analyse labour market outcomes is better represented using a typology reflecting the employment status of the individual. There are two reasons why this is a desirable strategy. First, the National Sample Surveys (NSS) do not report whether the individual is employed in the organised or unorganised sector;[2] they do however report whether the worker has a regular or casual job or is self-employed, unemployed or not participating in the labour market. The 1999 survey reported data on the type of enterprise that can be used to classify it as belonging to the organised or unorganised sector. This classification reveals that about 57 per cent of regular workers were employed in enterprises that were public, semi-public or otherwise in the registered or organised sector but only 10 per cent of casual workers were so employed.

Second, in the dual labour market literature workers in the unorganised sector are engaged in economic activities with low productivity resulting in low incomes, less stable employment

[2] The organised sector covers those enterprises registered under the Factory Act (1948), i.e., all establishments that employ ten or more workers and use power as well as those that employ 20 or more workers but do no use power.

contracts (this includes the self employed) and fewer social security benefits. There is an increasing awareness that the type of work contract is a better indicator of the informality of an individual's employment rather than whether or not the workplace is in the organised sector. For instance, a worker with a temporary contract with no provisions for social security should be considered as belonging to the unorganised sector even though he works in a large factory. In the Indian context this translates directly into the regular-casual worker dichotomy. Regular wage employment is often considered to be the most preferred category of work (Das, 2003). Tendulkar (2003, p. 2) refers to "workers having regular, contractual hired employment" as the "labour aristocracy because of the privileged service conditions this segment enjoys including high wages". Though these high wages reflect at least in part the returns to the higher skill endowments of these workers, redundancy (especially in the public sector) suggests the presence of rents. Regular workers are also covered by labour market regulations that confer some measure of employment security and social security benefits. Casual workers can be considered a subset of the informal labour market - they are generally engaged in economic activity with low wages, unstable employment contracts and little or no social security benefits. This is the approach followed in this paper - regular wage employment is taken as analogous to the ‘primary labour market’ and casual wage employment to the ‘secondary labour market’ in the dual labour market literature.[3]

3. Wage inequality in India

The empirical analysis in this paper uses data drawn from nationally representative large-scale employment surveys undertaken by the National Sample Survey Organisation (NSSO) during January-December 1983, July 1993–June 1994 and July 1999–June 2000 (referred to as 1983, 1993 and 1999 in this paper).[4] The sample is restricted to prime age males aged between 15 and 65 years (see the data appendix for details). The trends in wage dispersion among regular and casual workers are examined using a number of inequality measures including percentile ratios, the Gini coefficient and three Generalised Entropy measures – the mean log deviation (MLD), the Theil index and half the squared coefficient of variation.

[3] There is no attempt in this paper to formally test segmentation between the two types of wage employment (see Heckman and Hotz (1986) for a comprehensive critique of the inadequacy of tests for labour market segmentation).

[4] The employment survey for 1987-88 could not be used as over 76% of observations on rural wages for persons participating in wage employment are missing.

The Gini coefficient varies between zero (indicating no inequality) and one and is defined as follows:

$$Gini = \frac{1}{2n^2\overline{w}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| w_i - w_j \right| \qquad (1)$$

where n is the number of individuals in the sample, $\overline{w}$ is the arithmetic mean wage, w_i is the income of individual *i*, and w_j is the income of individual *j*.

Members of the Generalised Entropy (GE) class of measures have the following general formula:

$$GE(\alpha) = \frac{1}{(\alpha^2 - \alpha)} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{w_i}{\overline{w}} \right)^{\alpha} - 1 \right] \qquad (2)$$

where α is a parameter that represents the weight given to distances between wages at different parts of the wage distribution, and can take any real value. The most commonly used values of α are 0, 1 and 2. A value of $\alpha = 0$ gives more weight to distances between wages in the lower tail, $\alpha = 1$ applies equal weights across the distribution, while a value of $\alpha = 2$ gives proportionately more weight to gaps in the upper tail. The GE measures with parameters 0 and 1 become, with l'Hopital's rule, two of Theil's measures of inequality, the mean log deviation (GE(0) or MLD) and the Theil index (GE(1)) respectively, and with parameter 2 becomes half the squared coefficient of variation (referred to as GE(2) here) (Litchfield, 2003).

The inequality measures reported in Table 1 are computed for real hourly wages.[5] The standard errors are computed by bootstrapping with 1000 replications and 95 per cent confidence intervals constructed around these bootstrapped estimates are also reported. The statistical significance of changes in the point estimates of inequality measures over time can be examined using these confidence intervals. Cowell (1995) provides formulae to calculate the standard errors for the Gini coefficient and GE(2) based on the assumption that the underlying distribution are normal (provided the sample size is large). As the data reveal that the wage distribution approximates a lognormal rather than a normal distribution the bootstrapping approach was used.

[5] Inequality measures computed for real weekly wages are reported in Table A1 of the Appendix and reveal a similar picture of wage inequality.

Table 1: Measures of wage inequality

	Regular workers			Casual workers		
	1983	**1993**	**1999**	**1983**	**1993**	**1999**
Mean real hourly wage (Rs.)	3.1077	4.4056	5.7747	1.0819	1.4381	1.7419
	(2.71)	(3.37)	(5.36)	(0.77)	(0.73)	(0.93)
Median real hourly wage (Rs.)	2.6042	3.6662	4.4708	0.9197	1.2840	1.5577
Population share (%)	48.67	49.99	47.80	51.33	50.01	52.20
Income share (%)	73.14	75.66	75.24	26.86	24.34	24.76
Mean wage by education:						
Completed primary school	2.2813	2.7596	3.4419	1.2547	1.6012	1.8956
Completed middle school	2.6743	3.2265	3.9571	1.3118	1.6726	1.9597
Completed secondary school	3.8915	4.6547	5.9684	1.4042	1.6182	2.0036
Completed graduate school	5.7181	7.1968	9.5482	1.6344	1.5361	2.0918
Ratio graduate to primary wage	2.5065	2.6079	2.7741	1.3027	0.9593	1.1035
Wage inequality measures:						
Inter-quantile range (90-10)	4.9054	7.4567	10.3043	1.2669	1.6470	1.9391
Standard error	(0.0408)	(0.0541)	(0.0823)	(0.0151)	(0.0138)	(0.0131)
Confidence interval	[4.8254, 4.9855]	[7.3505, 7.5629]	[10.1428, 10.4658]	[1.2373, 1.2965]	[1.6199, 1.6741]	[1.9133, 1.9649]
Inter-quantile range (90-50)	3.0917	4.9059	7.1684	0.8757	1.0777	1.2664
Standard error	(0.0387)	(0.0517)	(0.0766)	(0.0142)	(0.0126)	(0.0122)
Confidence interval	[3.0157, 3.1676]	[4.8045, 5.0074]	[7.0181, 7.3187]	[0.8478, 0.9036]	[1.0529, 1.1025]	[1.2425, 1.2904]
Inter-quantile range (50-10)	1.8138	2.5507	3.1359	0.3912	0.5693	0.6726
Standard error	(0.0116)	(0.0225)	(0.0383)	(0.0042)	(0.0056)	(0.0053)
Confidence interval	[1.7910, 1.8365]	[2.5067, 2.5948]	[3.0608, 3.2110]	[0.3830, 0.3994]	[0.5584, 0.5802]	[0.6623, 0.6830]
Gini coefficient	0.3937	0.3923	0.4293	0.2920	0.2644	0.2629
Standard error	(0.0021)	(0.0017)	(0.0020)	(0.0022)	(0.0013)	(0.0013)
Confidence interval	[0.3896, 0.3978]	[0.3890, 0.3955]	[0.4253, 0.4332]	[0.2878, 0.2963]	[0.2619, 0.2670]	[0.2602, 0.2655]
Mean log deviation, GE(0)	0.2855	0.3136	0.3369	0.1431	0.1524	0.117
Standard error	(0.0029)	(0.0032)	(0.0032)	(0.0025)	(0.0026)	(0.0013)
Confidence interval	[0.2797, 0.2913]	[0.3073, 0.3199]	[0.3307, 0.3431]	[0.1382, 0.1479]	[0.1473, 0.1575]	[0.1145, 0.1195]
Theil index, GE(1)	0.2743	0.2561	0.314	0.1585	0.1199	0.1186
Standard error	(0.0041)	(0.0025)	(0.0052)	(0.0055)	(0.0013)	(0.0017)
Confidence interval	[0.2663, 0.2823]	[0.2512, 0.2610]	[0.3038, 0.3243]	[0.1478, 0.1693]	[0.1174, 0.1224]	[0.1153, 0.1220]
Half squared coeff. variation, GE(2)	0.3811	0.2927	0.4303	0.2531	0.1287	0.1424
Standard error	(0.0122)	(0.0054)	(0.0292)	(0.0363)	(0.0016)	(0.0052)
Confidence interval	[0.3573, 0.4050]	[0.2822, 0.3032]	[0.3730, 0.4876]	[0.1818, 0.3243]	[0.1255, 0.1320]	[0.1322, 0.1526]
Sample size	27,356	26,387	27,295	28,855	26,398	29,805

Note: Real hourly wage (Rs.) in constant 1983 prices. The inequality measures are computed without applying household weights (the results are very similar with weights). Figures in parentheses are as labelled - standard deviation for the mean wage, standard errors and 95% confidence intervals for inequality measures obtained by bootstrapping with 1000 replications.

The distribution of real hourly wages is approximately lognormal and positively skewed as suggested from the difference in the mean and median wage. Though regular workers comprise roughly half of all wage workers their wage share is about three-quarters. The wage gap between casual and regular workers is substantial and increased during this period - in 1983 an average casual worker in the labour market earned about 35 per cent of the hourly wage earned by an average regular worker; by 1999 this had fallen to 30 per cent (see also Dubey et al. (2004)). The dispersion of wages among casual workers is much lower than that among regular workers - for instance, the inter-quantile range in 1983 was 1.27 for the former and 4.91 for the latter. It should be noted that these summary statistics and inequality measures are computed after trimming the wage distribution (see the Data Appendix). These estimates of wage inequality, however, are uncorrected for differences in observable individual and job characteristics that determine wages.

Trends in the wage inequality measures over time are examined using confidence intervals (Litchfield, 2003). The Gini coefficient for regular workers is stable between 1983 and 1993 at about 0.393 and rises to 0.429 in 1999. Inequality trends according to the Generalised Entropy measures depend on the measure used because of the different weighting given to different parts of the wage distribution. The mean log deviation reveals that inequality among regular workers rose in both periods while the other two measures reveal a fall during the first period followed by a rise in the second. Overall, the data indicate that wage inequality among regular workers has risen between 1983 and 1999. This pattern is reversed for casual workers though there is some disagreement among the different measures. The inter-quantile range rose in both sub-periods but the Gini coefficient and Theil index fell between 1983 and 1993 before stabilising during the 1990s. The mean log deviation measure was stable in the first sub-period but fell during the 1990s. In contrast, the coefficient of variation rose between 1993 and 1999 after a sharp fall in the first sub-period. Overall the picture is one of declining inequality among casual workers between 1983 and 1999. The differences in the inequality measures suggest that the changes in both the regular and casual wage distribution occurred at the tails of the distribution; measures such as the coefficient of variation and the three inter-quantile ranges that are more sensitive to the tails reveal a rise in inequality during the 1990s. The trends in the wage gap between workers with graduate and primary schooling are similar to those in wage inequality.

These opposing trends in wage inequality for regular and casual workers are confirmed by a comparison of the disaggregated distributions using Lorenz curves (plots of the cumulated wage share against the cumulated population share) (see Figure A1 of the Appendix). These reveal that, for regular workers, the wage distribution for 1983 Lorenz dominates that for 1999 – i.e., the Lorenz curve of this distribution lies nowhere below and at least somewhere above the 1999 distribution indicating that inequality was unambiguously lower in the first year. The pattern for casual workers is exactly the opposite with the 1999 distribution Lorenz dominating the 1983 distribution.

As mentioned earlier, there are few estimates for wage inequality in India though several studies employ casual empiricism to suggest rising inequality during the 1990s. For instance, Acharyya and Marjit (2000) use data on the minimum daily wages for the lowest paid unskilled workers in the organised sector for the periods 1985-86 and 1993-94 to illustrate the widening gap between the minimum and maximum wage from about 0.48 in 1985-86 to 2.56 in 1993-94. To the author's knowledge, only one other study (Galbraith et al., 2004) constructs measures of pay inequality for the organised manufacturing sector in India. The construction of the wage measure is not specified. However, it is safe to assume that it is the product wage (i.e., total emoluments per employee or total wages per worker) as the data are obtained from the Annual Survey of Industries (ASI) datasets. It appears that the wage data are expressed in nominal terms, though this is not explicit in the paper. The authors construct between-component Theil indices using grouped data (grouped by three-digit industry cells at the all-India level or two-digit industry cells at the state level) for the period 1972 to 1998. They document an increase in pay inequality in the organised Indian manufacturing since the early 1980s, with particular increases in the 1990s. This increase is driven primarily by increases in inequality between industry groups (nationally and within states) rather than by regional inequality. It should be noted, however, that these trends are in regard to movements in wage inequality in the organised manufacturing sector only. To the extent that there are inter-linkages between manufacturing, agriculture and services, these trends in wage inequality in organised manufacturing provide some notion of the developments in the rest of the economy. Given the large share of regular workers in manufacturing the trends in wage dispersion among these workers parallels those reported in other studies.

The trend in wage inequality among regular workers also mirrors that for the consumption Gini coefficients computed from the consumption surveys (companion to the employment surveys used in this research) – this Gini coefficient is relatively stable at 0.32 in 1983 and 1993 and rising to 0.38 in 1997 (Datt, 1999; Özler et al., 1996). Various studies have found that this remained relative stable between 1961 and 1993 (Deaton and Dreze, 2002; Panda, 1999). There is some debate about household inequality trends during the 1990s due to the changes in the 1999-2000 consumption survey that render comparison with previous years difficult (notably, the change in the recall period – see Deaton and Dreze (2002) for details). After correcting for these changes, the data suggest that household inequality has risen overall - the variance of log per capita expenditure increased from 0.29 in 1993 to 0.32 in 1999 (the unadjusted data reveals no change at 0.29) (Deaton and Dreze, 2002).

It is also useful to place the findings of this paper in the context of wage inequality estimates for other countries. It should be noted that these comparisons should be treated as indicative due to differences in data sources, samples used, time coverage and the measurement of wages. In particular, the Indian estimates refer to the male wage distribution only while the others refer to all wage workers. Galbraith et al. (2000) demonstrate using UNIDO industrial statistics spanning a period between 1973 and 1997 that inequality in manufacturing pay is lowest among Australia, China, most European countries and Taiwan; next highest in Korea, Japan, North American countries and some middle-income developing countries; and the highest in Africa, Latin America, Russia and South Asia. In addition, they find that inequality in South Asia has risen during this period.

The estimates of wage inequality reported in this paper are comparable with some Asian countries for which data are available. Wage inequality in Vietnam, undergoing rapid economic change from a centrally planned socialist to a market-oriented economy, fell during the 1990s – the Gini coefficient fell from 0.40 to 0.38 between 1993 and 1998 (Gallup, 2002). The Republic of Korea also experienced declining wage inequality between 1976 and 1993 with the Gini falling from approximately 0.40 to 0.30 (Fields and Yoo, 2000). Information on wage inequality in South Asian countries that are at a similar level of development as India is scarce. Mujeri (2002) reports that the ratio of unskilled to skilled wage in Bangladesh fell from 0.88 in 1985 to 0.69 in 1996. Comparable estimates for Indian male workers are: 0.64, 0.88 and 0.77 for 1983, 1993 and 1999 for male casual workers and 0.32, 0.34 and 0.32 in the three years for male regular workers (where unskilled workers are defined as those with no education and skilled workers as graduates). The stability of this ratio for regular workers suggests that the rise in the wage inequality measures occurred at the middle of the wage distribution rather than at the tails. These ratios for India and Bangladesh cannot be compared precisely as the division of workers into skilled and unskilled categories is unclear in the latter. In contrast, wage inequality in India is higher than that prevailing in developed and some formerly socialist European countries but lower than that in the United States and Colombia. Transition economies such as the Czech Republic, Hungary, Poland, Romania and Slovenia experienced rising wage inequality between 1989 and 1997 with the Gini coefficient ranging from 0.20 to 0.33. Others such as Bulgaria and Russia had higher levels of wage inequality (Newell, 2001). Blau and Kahn (1996) report a standard deviation of log hourly wages for ten industrialised European countries, including France, Germany and the United Kingdom, around the mid- to late-1980s of 0.48; the corresponding figure for the United States is 0.77. Attanasio et al. (2004) report standard deviation of log hourly wages for Colombia ranging from 0.77 to 0.89 between 1984 and 1998. The corresponding figures for all male wage workers, regular and casual, in India is 0.50 in 1983, rising to 0.57 and 0.62 in 1993 and 1999.

3.1. Decomposing inequality by education and industry

The contribution of educational achievement and industry affiliation to wage inequality can be examined by decomposing the inequality measures reported above into within- and between-group inequality components (see Cowell (1995) for the methodology and Litchfield (2003) for an application to Brazil). The partitions are those used in the subsequent wage regression models – 5 education groups (no education, those who have completed primary, middle, secondary and graduate school) and 38 industry groups.

Any of the Generalised Entropy class of measures can be decomposed by population sub-group so that the overall inequality (I) can be separated into within-group (I_w) and between-group (I_b) inequality as follows (Cowell and Jenkins, 1995):

$$I = I_w + I_b \tag{3}$$

Within-group inequality is defined as a weighted sum of inequality within each of the sub-groups where the weights are population shares, relative wages or some combination of these two depending on the inequality measure used:

$$I_w = \sum_{j=1}^{k} \vartheta_j GE(\alpha)_j \tag{4}$$

where $\vartheta_j = v_j^{\alpha} f_j^{1-\alpha}$, f_j is the population share and v_j the wage share of each group j (j=1,2,..k). Between-group inequality is computed by assuming each group member receives the mean group wage ($\overline{w}_j$) as follows:

$$I_b = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{k} f_j \left(\frac{\overline{w}_j}{\overline{w}}\right)^{\alpha} - 1\right] \tag{5}$$

The fraction of inequality "explained" by differences between population sub-groups is then given by $R_b = I_b / I$ while the remainder is the unexplained component. The Gini coefficient can only be decomposed under the special circumstance when the groups are non-overlapping (i.e., each individual's income in one group is greater/lower than each individual in the other groups). Due to the differences in the sensitivity to different parts of the wage distribution and in the weights used to summarise within-group inequality the percentage explained by characteristics differ according to the GE measure used.

The decomposition of the mean log deviation, Theil index and coefficient of variation into wage inequality between and within education groups is reported in Table 2.

Table 2: Decomposition of wage inequality by education level

	1983			**1993**			**1999**		
Education level	**GE(0)**	**GE(1)**	**GE(2)**	**GE(0)**	**GE(1)**	**GE(2)**	**GE(0)**	**GE(1)**	**GE(2)**
Regular workers:									
Overall wage inequality	0.2855	0.2743	0.3811	0.3136	0.2561	0.2927	0.3369	0.3140	0.4303
Within-group inequality	0.2025	0.1901	0.2915	0.2360	0.1794	0.2136	0.2525	0.2310	0.3447
Contribution (%)	(70.92)	(69.30)	(76.49)	(75.24)	(70.04)	(72.98)	(74.94)	(73.54)	(80.11)
Between-group inequality	0.0830	0.0842	0.0896	0.0776	0.0767	0.0791	0.0844	0.0831	0.0856
Contribution (%)	(29.08)	(30.70)	(23.51)	(24.75)	(29.96)	(27.02)	(25.06)	(26.46)	(19.89)
Casual workers:									
Overall wage inequality	0.1431	0.1585	0.2531	0.1524	0.1199	0.1287	0.117	0.1186	0.1424
Within-group inequality	0.1376	0.1528	0.2471	0.1486	0.1160	0.1247	0.1128	0.1144	0.1381
Contribution (%)	(96.13)	(96.41)	(97.62)	(97.47)	(96.72)	(96.92)	(96.41)	(96.46)	(96.99)
Between-group inequality	0.0055	0.0057	0.0060	0.0038	0.0039	0.0040	0.0042	0.0042	0.0043
Contribution (%)	(3.84)	(3.62)	(2.37)	(2.50)	(3.25)	(3.11)	(3.57)	(3.58)	(3.03)

Notes: Figures in parentheses are the contribution (%) of each component to overall inequality.

The results reveal that the major portion of wage inequality, particularly for casual workers, is accounted for by inequality among individuals within education groups rather than between individuals with different education levels. This is despite the rise in the returns to graduate education for regular workers (see Table A2 of the Appendix and the discussion below). The between-group component is very small in the case of casual workers possibly because only the majority of these workers have no education or, at most, are educated up to primary school. The mean wages in each sub-group (see Table 1) are steadily increasing in education level but are clustered close to the overall mean. For regular workers, however, mean wages are increasing with wide variations around the overall mean, especially for graduates.

The GE inequality measures are also decomposed into wage inequality between and within 38 industry groups in Table 3. As before, the within-industry group component accounts for over three-quarters of wage inequality for regular workers and more than 80 per cent of wage inequality for casual workers.

Table 3: Decomposition of wage inequality by industry groups

	1983			**1993**			**1999**		
Industry group	**GE(0)**	**GE(0)**	**GE(1)**	**GE(2)**	**GE(1)**	**GE(2)**	**GE(0)**	**GE(1)**	**GE(2)**
Regular workers:									
Overall wage inequality	0.2855	0.2743	0.3811	0.3136	0.2561	0.2927	0.3369	0.3140	0.4303
Within-group inequality	0.2083	0.2099	0.3238	0.2498	0.1997	0.2405	0.2601	0.2425	0.3607
Contribution (%)	(72.98)	(76.53)	(84.95)	(79.66)	(77.98)	(82.18)	(77.19)	(77.21)	(83.82)
Between-group inequality	0.0772	0.0644	0.0574	0.0638	0.0564	0.0522	0.0769	0.0716	0.0696
Contribution (%)	(27.02)	(23.48)	(15.05)	(20.34)	(22.02)	(17.82)	(22.81)	(22.79)	(16.18)
Casual workers:									
Overall wage inequality	0.1431	0.1585	0.2531	0.1524	0.1199	0.1287	0.117	0.1186	0.1424
Within-group inequality	0.1200	0.1345	0.2278	0.1357	0.1028	0.1110	0.1015	0.1029	0.1262
Contribution (%)	(83.83)	(84.88)	(90.00)	(89.06)	(85.72)	(86.28)	(86.71)	(86.72)	(88.64)
Between-group inequality	0.0231	0.0240	0.0253	0.0166	0.0171	0.0177	0.0155	0.0158	0.0162
Contribution (%)	(16.14)	(15.15)	(9.99)	(10.92)	(14.25)	(13.75)	(13.27)	(13.31)	(11.37)

Notes: Figures in parentheses are the contribution (%) of each component to overall inequality.

An examination of the within-industry inequality measures (not reported) suggests that inequality within almost all industries rose during the 1990s.[6] The exceptions are the mining, utilities and services industries, all of which are dominated by the public sector, which could explain the stability in wage inequality in these industries over this period of a decade and a

[6] As before, these trends are examined using the 95% confidence intervals obtained by bootstrapping (not reported).

half.[7] Galbraith et al. (2004) also find that rising manufacturing pay inequality during the 1990s was driven primarily by rising incomes in heavy manufacturing industries.[8]

4. Accounting for inequality in a regression framework

In contrast to examining within- and between-group inequality as described above several methods have been suggested to decompose the wage inequality into its components parts within a regression framework (see Fields (2002) for a review). While the earlier approaches rely on parsimonious specifications of the wage functions many of the more recent approaches are complicated and unwieldy. One exception is the methodology developed by Fields (2002) that decomposes the contribution of various explanatory variables to the level of and change in inequality within a standard semi-logarithmic wage (or income) regression model. The merit of this method is that it is independent of the measure of inequality chosen and is based on robust decomposition rules derived axiomatically.

4.1. Methodology

Assume a semi-logarithmic Mincerian (standard or augmented) wage determining function as follows (suppressing the sub-scripts for time and type of wage employment):

$$\ln(w_i) = \sum_{j=1}^{J} \beta'_j Z_{ij} + \varepsilon_i \tag{6}$$

where ln(.) is the natural log operator, w_i wages, β_j coefficients and Z_{ij} the explanatory variables (j = 1,..J) for individual i (at time t) while ε_i is the random error term. This can be re-written as follows:

$$\ln(w_i) = \sum_{j=1}^{J+1} a'_j Z_{ij} = \mathbf{a'Z} \tag{7}$$

where $\mathbf{a}$ = $[\beta_1.....\beta_J,1]$ and $\mathbf{Z}$ = $[Z_1......Z_J,\varepsilon]$ are vectors of coefficients and explanatory variables respectively. An inequality index I can be defined on the vector of wages (**w**).

[7] Services includes public administration - the largest employer of regular workers.
[8] Unlike this research, Galbraith et al. (2004) also document rising inequality in the electricity sector.

Fields (2002) then applies Shorrocks' theorem to compute the relative factor inequality weights (i.e., the percentage of inequality that is accounted for by the j^{th} factor) as follows:

$$s_j[\ln(w)] = \frac{\operatorname{cov}[a_j Z_j, \ln(w)]}{\sigma^2 \ln(w)} = \frac{a_j \times \sigma(Z_j) \times cor[Z_j, \ln(w)]}{\sigma \ln(w)} \tag{8}$$

where *cov[.]* denotes the covariance, *cor(.)* the correlation coefficient and $\sigma(.)$ the standard deviation. The conditions $\sum_{j=1}^{J+1} s_j[\ln(w)] = 1$ and $\sum_{j=1}^{J} s_j[\ln(w)] = R^2[\ln(w)]$ hold for any inequality index I(w) which is continuous and symmetric (including the Gini coefficient and the Generalised Entropy measures) and for which I(ε) = 0. The fraction of inequality explained by the j^{th} explanatory factor, $p_j[\ln(w)]$, is the following:

$$p_j[\ln(w)] = \frac{s_j[\ln(w)]}{R^2[\ln(w)]} \tag{9}$$

The percentage contributions of factors to the change in inequality can be similarly computed. However this share is no longer independent of the inequality measure used as any change in inequality would depend on the measure used. The difference in the chosen inequality index between two time periods (0,1) can be written as follows:

$$I(w)_1 - I(w)_0 = \sum_j \left[s_{j,1} \times I(w)_1 - s_{j,0} \times I(w)_0 \right] \tag{10}$$

The contribution of the j^{th} factor to the change in equality between the two time periods (0,1) is given by:

$$\Pi_j(I(w)) = \frac{s_{j,1} \times I(w)_1 - s_{j,0} \times I(w)_0}{I(w)_1 - I(w)_0} \tag{11}$$

where $\Pi_j I(w)$ is a function of the inequality measure *I(w)* chosen and $\sum_{j=1}^{J+1} \Pi_j I(w) = 1$.

4.2. The determinants of wages

In the first step of the above regression-based decomposition methodology, separate wage equations are estimated for regular and casual workers with human capital, industry affiliation and various other characteristics as controls. Before the wage regression models are estimated the issue of selection bias is addressed using the generalised framework popularised by Lee (1983).[9] Selection is modelled as a polychotomous outcome between three employment categories – non-wage earners (including non-participants in the labour market, self-employed and unemployed individuals), regular wage workers and casual wage workers. As this bias is mediated through observed wages it is sufficient and computationally more convenient to separate employment status into non-wage earners and two different types of wage earners. As noted in Section 2, the separation of workers based on the nature of employment is based on the notion of a dual economy.

A two-stage model for selection and wage determination is estimated. First, a multinomial logit model is estimated where the probability that individual i is in outcome m can be expressed as:[10]

$$P_1 = \frac{1}{1+\sum_{m=2}^{M}\exp(x_m^{'}\gamma_m)}; \text{ and } P_s = \frac{\exp(x_m^{'}\gamma_m)}{1+\sum_{k=2}^{M}\exp(x_k^{'}\gamma_k)} \quad \gamma_1 = 0\,; m = 2,3;\ k = 2,3 \tag{12}$$

where the vector x comprises exogenous explanatory variables, m and k are categorical variables signifying selection between the different employment categories. Exclusion restrictions are required to identify the parameters of the wage equations so that a set of variables that influence employment status between the alternative outcomes but not wage are included as regressors in the selection equation. The selection bias correction term, λ_k, constructed from the selection equation, is similar to the inverse of the "Mills ratio":

$$\lambda_k = \frac{\phi(\Phi^{-1}(P_k))}{P_k} \tag{13}$$

where $\phi(.)$ and $\Phi(.)$ represent the normal density and distribution functions respectively and P_k is the probability of being in outcome k (k = 2, 3).

An augmented semi-logarithmic Mincerian specification can then be used to estimate consistent coefficients of the wage equations for both sets of workers (suppressing the sub-scripts for time, t, and type of wage employment, k):

---

[9] Selection bias could arise if the selection of individuals into wage employment is systematic. This is because this sample is essentially truncated as data on wages as well as industry affiliation is reported only for those individuals in wage employment.

[10] The MNL model is identified only up to an additive vector. As a result one set of parameters γ s (outcome one, i.e., non-wage earners, in this paper) must be selected as the base category and set to zero.

$$\ln(w_i) = \sum_{j=1}^{J} \beta_j^{'} Z_{ij} - \beta^* \hat{\lambda}_i + \varepsilon_i \qquad (14)$$

where β^* is the coefficient on the selection bias correction term, $\hat{\lambda}_i$, and ε_i the random error terms such that $E(\varepsilon_i \mid z_i; x_i) = 0$.

This two-step procedure controls for the underlying process by which the set of observations actually observed are generated and ensures consistent estimates of the parameters of the wage equations.[11] The sampling distribution for the estimates can be obtained by using a modification to the formula suggested in Trost and Lee (1984) or by bootstrapping. The latter procedure is adopted here and each of the estimated wage regression models has been bootstrapped using 1000 replications. The selection equations are not reported here for brevity; the wage equations are reported in Table A2 of the Appendix. The explanatory variables common to both models are worker characteristics such as age, education, marital status, caste and religious affiliation as well as controls for settlement type, state of residence and seasonality effects. As noted earlier, the parameters of the wage equations are identified by including in the selection equation variables that capture household structure - household size and four dependency variables.[12] The wage regression models also include industry affiliation dummy variables and selection bias correction terms constructed as described above.

The explanatory power of the variables in all three years is quite high – explaining over half and one-third of the variation in log wages for regular and casual workers respectively. The standard error of estimate is the dispersion of the predicted wages after controlling for various individual and other characteristics and has increased by about 10 and 2.5 percentage points for regular and casual workers respectively. Comparing this with the standard deviation of mean regular wages (reported in Table 1) suggests that while the change between 1983 and 1993 is comparable for both statistics, it is far more rapid for the unadjusted standard

[11] The Lee correction was chosen over other methods of selection correction in polychotmous outcome models because of its simplicity, computational convenience and transparent interpretation of the selection effect. It should be noted that parameter estimates of the wage equations obtained in this paper using power series approximations for the selection term following the approach advocated by Newey (1999) were very similar to those obtained using the Lee correction.

[12] As the choice of identifying variables is necessarily *ad hoc* the MNL model was estimated for different specifications of identifying variables. The parameter estimates in the wage equations are not sensitive to the choice of the identifying variables and the coefficient on the correction term itself was not materially different across specifications. On balance, these instruments were also not found to strongly influence wages in most specifications in most years.

deviation. The standard deviation of casual wages was lower in 1993 compared to the first survey, but increased in 1999. The standard error of estimate, on the other hand, suggests rising inequality after adjusting for individual characteristics in both sub-periods, though less so during the 1990s. This rise in the standard error of estimates in the wage equations coupled with the growing importance of selection for regular workers suggests an increase in the returns to unobservable skills that could possibly be related to the liberalisation process.

The explanatory variables have the anticipated signs and the majority yield significant effects at the one per cent level or better. The industry dummy variables are almost all significant at the one per cent level or better indicating the presence of inter-industry wage differentials for both sets of workers. The differences in the returns to human capital embodied in experience and education for different types of workers are consistent with the notion of dual primary and secondary labour markets. Casual workers face at best flat returns to education and experience while the returns to education for regular workers are positive and rising in education level. This pattern of returns increasing in education level has been observed in several country studies for Africa and Asia (see Bennell (1995)), and in national and regional studies within India (Duraisamy, 2002; Kingdon, 1998).

As mentioned earlier, the 1990s were a period of rapid industrial deregulation and trade liberalisation in India. This considerable structural change over the space of a decade is likely to have had some implications for the labour market. This is reflected in the changes in the marginal effects of selected variables during the 1990s. In particular, the returns to graduate education for regular workers rose from 9 per cent to 10 per cent between 1993 and 1999 (t-statistic = 2.33),[13] despite a rise in the supply of graduates. Various explanations for this rise include the existence of patterns of trade that increase the relative demand for skill, i.e., skill-enhancing trade or SET (Robbins, 1996), a structure of trade protection that formerly favoured relatively unskilled-labour intensive sectors (Harrison and Hanson, 1999), and skill-biased technological change (Lawrence and Slaughter, 1993). This rising gap between graduate and primary education among regular workers during the 1990s is reflected in the sharp rise in wage inequality during this period. The contribution of these explanatory variables, including education and industry affiliation, to wage inequality is examined next.

[13] These are computed as the difference in the coefficients (obtained from the wage regression models) between two consecutive education levels divided by the difference in the years of schooling associated with each level. Standard errors (used in the t-tests for change over time) can also be computed.

4.3. Accounting for inequality in India: 1983, 1993 and 1999

The decomposition methodology described above is applied to the wage regression models with one qualification. This approach can be used to examine the contribution of all variables including variables with non-linear effects and categorical variables entered as a string of dummy variables. A problem arises when there are two or more variables enter interactively as the underlying assumption in the decompositions is that the explanatory variables enter the wage function additively. As a result the contribution of seasonality and settlement type included in the wage regression models cannot be decomposed neatly and are presented as a composite effect. One merit of this approach is that it is possible to examine not only the statistical significance of explanatory variables in a regression but also the strength of their explanatory power. For instance, the wage regression models (see Table A2 of the Appendix) indicate that the majority of the explanatory variables are significant at the one per cent level or better. However, as Table 4 reveals, the contribution of these variables to wage inequality differs markedly.

Table 4: Relative factor inequality shares

	Regular workers			**Casual workers**		
Factors	**1983**	**1993**	**1999**	**1983**	**1993**	**1999**
Age	0.1019	0.1159	0.1338	0.0140	0.0193	0.0223
	(18.63)	(24.13)	(25.20)	(4.10)	(6.28)	(6.57)
Marital status	0.0145	0.0167	0.0189	0.0023	0.0038	0.0020
	(2.65)	(3.47)	(3.55)	(0.68)	(1.24)	(0.58)
Educational level	0.2313	0.1735	0.1722	0.0003	-0.0010	0.0048
	(42.29)	(36.12)	(32.43)	(0.09)	-(0.34)	(1.42)
Social and ethnic group	0.0071	0.0040	0.0036	0.0029	-0.0003	0.0019
	(1.30)	(0.82)	(0.67)	(0.85)	-(0.11)	(0.57)
Season and settlement type	0.0393	0.0018	0.0018	0.0428	0.0073	0.0099
	(7.19)	(0.38)	(0.34)	(12.54)	(2.36)	(2.90)
State of residence	0.0182	0.0130	0.0188	0.1798	0.1849	0.2112
	(3.33)	(2.71)	(3.54)	(52.74)	(60.10)	(62.14)
Industry affiliation	0.1222	0.1031	0.1385	0.0837	0.0879	0.0842
	(22.35)	(21.45)	(26.07)	(24.55)	(28.57)	(24.77)
Selection	0.0124	0.0524	0.0436	0.0152	0.0058	0.0035
	(2.27)	(10.92)	(8.20)	(4.46)	(1.89)	(1.04)
Residual	0.4530	0.5196	0.4689	0.6590	0.6923	0.6601
R-squared	0.5470	0.4804	0.5311	0.3410	0.3077	0.3399
Sample size	27,356	26,387	27,295	28,855	26,398	29,805

Note: Relative factor inequality shares refer to $s_j[\ln(w)]$ in equation (8). Figures in parentheses are the share of inequality explained (i.e., $p_j[\ln(w)]$ in equation (9), multiplied by 100).

The differences in the structure of inequality among the two groups of workers are striking. In terms of the explained component for regular workers human capital embodied in experience (proxied by age) and education plays an important role in any given year - age accounted for about a quarter and education for about one-third of the explained level of inequality in 1999. While age does contribute to wage inequality among casual workers the relative factor inequality share is much lower than that for regular workers - age accounts for only about five to seven percent of explained inequality. Geographic factors are the single most important factor in explaining inequality among casual workers – the state of residence variables explain between 50 per cent and 62 per cent of explained wage inequality in the three years. Industry affiliation plays an equally important role for both groups of workers accounting for about a quarter of explained wage inequality.

Surveys of decomposition analyses in developing countries typically reveal that individual attributes (such as education, age, gender, household composition) account for large proportions of income inequality (Fields, 1980; Litchfield, 2003). There is a growing literature on the empirical application of this decomposition methodology. The factor inequality shares for regular workers are somewhat similar to other studies using this accounting methodology though the countries studied are vastly different – Guatemala (Alejos, 2003), the Republic of Korea (Fields and Yoo, 2000), Serbia (Krstic and Reilly, 2003) and the United States (Fields, 2002), among others. Though the values of the factor inequality shares estimated differ somewhat the same factors play an important role in explaining inequality – education, experience and occupation. Alejos (2003) found that education (11%), experience (3%), occupation (24%) and ethnicity (3%) were the major factors explaining inequality in Guatemala among male workers. Fields and Yoo (2000) found that, among others, four factors – job tenure (24-38%), occupation (20-21%), education (13-26%) and potential experience (13-14%) - were the major contributors to inequality among male workers in both the years examined though there were some rank reversals. Industry affiliation explained about 3-6% of explained inequality. Fields (2002) obtains estimates comparable to this research for the United States for experience (26%) but find a greater role for education (46%) and occupation (26%) than for industry affiliation in 1999. Krstic and Reilly (2003) also find that education contributed (32-40%) significantly to wage inequality in Serbia during the 1990s but find a much lower role played by labour force experience (3-4%). The contribution of industry affiliation to inequality is comparable to the estimates in this research though those for Serbia are more volatile – ranging from 9% in 2000 to 33% in 1997. It should be noted that estimates from the Krstic and Reilly study include male and female workers while those from Alejos, Fields and Yoo, Fields and this research refer to male workers only. Other than this research, none of the other studies control for selection. The contribution of industry affiliation to wage inequality in India is also highlighted by Galbraith et al. (2004) - they find that inequality in manufacturing pay is driven by between-industry (and within-state) inequality rather than between-state (and within-industry) inequality.

Alejos (2003) also decomposed wage inequality among different sub-groups of workers. In particular, he found that the contribution of education to inequality among agricultural, livestock and fishery workers was much lower at about three per cent compared to inequality among all workers. As casual workers are predominantly engaged in agricultural and allied activities this supports the findings of a low factor inequality share for education in this paper for casual workers. This suggests that there is equality of educational attainment among casual workers – supported also by the low between-education group share of explained inequality. This equality, however, is a result not of high enrolment rates but of low average level of schooling among these workers. This is unsurprising as, by definition, casual workers compete in a secondary labour market with virtually flat returns to education, indicating either that there is a low demand for skill or that the acquired skills are not useful in the casual labour market.

The relative role of the factors determining wages has changed somewhat over time. Though the wage gap between regular workers with graduate and primary school qualifications has risen between 1983 and 1999, the factor inequality weight of education has fallen during this period. For casual workers, on the other hand, the contribution of education, though small, rose between 1983 and 1999. At the same time the contribution of age has risen for both workers, which is also reflected in the steepening of the age-earnings profile (see Table A2 of the Appendix). The increase in the contribution of selection coupled with the fall in that of education suggests a rising importance for unobservable skill in the probability of employment in regular wage employment that is possibly linked to the process of trade and industrial liberalisation. The contribution of selection into casual wage employment declined during this period. The fall in the importance of season and settlement variables for both workers is driven primarily by the decrease in the negative effect on individual wages associated with residing in rural areas. The relative importance of variables capturing caste and religious affiliation is quite low and has also halved during this period. The contribution of industry affiliation towards wage inequality among regular workers rose between 1983 and 1999, possibly reflecting the change in the structure of protection across industries that was a consequence of the trade liberalisation during this period. In particular, there was substantial reductions in both the level and dispersion of tariff and non-tariff barriers that altered the structure of protection across industries. The data suggest that industries with high initial levels of tariff protection (in 1983) had low wages relative to the economy average and a high share of unskilled regular workers. These industries were also those that experienced the greatest reductions in tariff protection. There is some evidence that the wages of regular workers (predominantly unskilled) in these manufacturing industries fell relative to those in more skill-intensive industries (Vasudeva-Dutta, 2004). This would explain the rise in the contribution of industry affiliation to the level of inequality between 1983 and 1999 for these workers. There is an increase in the factor share of industry affiliation for casual workers in 1993. However, it is less likely that this would have been influenced by the trade and industrial reforms as relative demand shifts are less likely to matter for casual workers who, by definition, comprise a largely homogenous pool of labour.

4.4. Accounting for changes in inequality in India: 1983 to 1999

For regular workers wage inequality trends indicate rising inequality at least during the 1990s. The picture is less clear between 1983 and 1993 and the trends depend on the measure of inequality used. The Gini coefficient was stable between 1983 and 1993 (this implies a very small denominator in terms of equation (11) resulting in improbably large estimates of $\Pi_j(.)$), the Theil index fell during this period while the inter-quantile range rose. For casual workers though there is some disagreement about trends depending on the inequality measure used but the overall picture is one of declining inequality. The contribution of the explanatory factors to the changes in inequality between 1983 and 1999 is reported in Table 5. The inequality measures used are the Gini coefficient, Theil index and the inter-quantile range between the 90th and 10th percentile (IQR). A factor increases (decreases) inequality if the sign of the factor share is the same (opposite) as the direction of change in the inequality measure.

Table 5: Contribution of factors to changes in wage inequality, 1983-1999

	Regular workers			Casual workers		
Factors	**Gini**	**Theil**	**IQR**	**Gini**	**Theil**	**IQR**
Age	0.4881	0.3543	0.1629	-0.0610	-0.0108	0.0380
Marital status	0.0671	0.0489	0.0228	0.0052	0.0033	0.0014
Educational level	-0.4828	-0.2354	0.1185	-0.0403	-0.0131	0.0133
Social and ethnic group	-0.0358	-0.0209	0.0004	0.0116	0.0058	0.0001
Season and settlement type	-0.4141	-0.2570	-0.0323	0.3389	0.1405	-0.0521
State of residence	0.0253	0.0228	0.0193	-0.1027	0.0865	0.2704
Industry affiliation	0.3182	0.2503	0.1532	0.0793	0.0823	0.0851
Selection	0.3889	0.2584	0.0719	0.1202	0.0499	-0.0185
Residual	0.6451	0.5785	0.4834	0.6488	0.6556	0.6623

Note: IQR refers to the inter-quantile range between the 90^{th} and 10^{th} percentiles for real hourly wages. The contribution of factors refers to $\Pi_j(.)$ in equation (11).

As regular wage inequality increased between 1983 and 1999 a positive (negative) sign denotes that a factor was responsible for widening (narrowing) inequality. The majority of factors contribute towards increasing wage inequality during this period. Age, industry affiliation and selection are consistently the most important sources of widening inequality while education serves to narrow inequality among regular workers. The adverse impact of industry affiliation on wage inequality is, to some extent at least, a consequence of the differential impact of the fall in tariff protection across industries. For casual workers there is some disagreement among the different inequality measures - the inter-quantile range rose while the Gini coefficient and Theil index fell between 1983 and 1999. As a result, a positive sign denotes widening inequality for the first measure and narrowing inequality for the latter two measures. While age and education contribute towards widening inequality among casual workers, selection narrows inequality during this period. The role of industry affiliation and geographic location depends on the measure used – the former serves to widen inequality using the inter-quantile range and *vice-versa* using the Gini and the Theil while the latter serves to widen inequality using the IQR and Gini but not the Theil. There is considerable evidence of wide inter-state disparities in economic growth, poverty incidence and income inequality especially in the post-liberalisation period (see, for example, Sachs et al. (2002) and Deaton and Dreze (2002)). As geographic location is the single most important determinant of the level of wage inequality it would be an interesting topic for future research to analyse the possible causes of these inter-state wage disparities.

5. Conclusion

This paper examines the structure of wage inequality for different types of adult male workers – those with regular wage or salaried jobs and those with casual or contractual jobs – using the notion of a dual labour market. The analysis uses microeconomic data for three years - 1983, 1993 and 1999 - spanning a period of rapid economic liberalisation in India.

The major finding of this paper is that there are striking differences in the evolution and structure of wage inequality for the two groups of workers. Wage inequality among regular workers is not only considerably higher than that among casual workers it has also risen between 1983 and 1999, particularly during the 1990s. Though there is some disagreement among the different inequality measures, inequality among casual workers has fallen during this period. This highlights the importance of examining these workers separately rather than as one category. Simple decompositions of wage inequality by education and industry groups reveal that the major portion, particularly for casual workers, is accounted for by inequality among individuals within rather than between these groups.

The contribution of the various factors generally used to explain wages also differs for these workers. Human capital (as embodied in age and education) is one of the major factors explaining both the level of and change in regular wage inequality. For casual workers geographic location plays the most important role in determining wage inequality. Industry affiliation, on the other hand, plays an equally important role explaining about a quarter of wage inequality for both groups of workers. Age, education, industry affiliation and selection are also consistently the most important contributors to changes in inequality among both sets of workers, though their directional effect differ somewhat. The role of education in accounting for wage inequality must be stressed. While educational attainment is an important contributor towards the level of wage inequality its factor inequality share has fallen by about ten percentage points between 1983 and 1999 and education has contributed towards narrowing wage inequality during this period despite the rising skill premium for graduates. At the same time the supply of educated workers increased at all levels of education. For casual workers education serves to widen inequality but only a very small proportion of these workers is educated beyond primary school. This suggests that expanding education through greater access would be a desirable strategy to reduce disparities.

To the author's knowledge this paper is among the first to comprehensively document the evolution and structure of wage inequality in India. This study also augments the growing literature on the empirical application of a regression-based decomposition technique in order to examine the significance and contribution of various factors to wage inequality. The study of inequality and its determinants is important in itself and labour market inequality is a significant determinant of disparities in living standards. In addition, disparities in wages and incomes also have implications for poverty and growth. Ravallion (1997) finds that initial inequality has important implications for poverty reduction by influencing how poverty responds to growth. There is also growing literature on how high levels of inequality hinder growth through political economy arguments imperfect asset markets, and social conflict (Benabou, 1996). Though this study is essentially an accounting exercise, by examining in detail the structure of wage inequality for different sets of workers the empirical analysis reported sets the stage for the future investigation of the reasons for the change in the underlying factors. This agenda for future research could fruitfully focus on the role that returns to education or industry affiliation exert on the change in wage inequality over this period.

REFERENCES

Acharyya, R. and S. Marjit (2000). "Globalisation and Inequality: An Analytical Perspective." *Economic and Political Weekly (September 23)*: 3503-3510.

Alejos, L. A. (2003). Contribution of the Determinants of Income Inequality in Guatemala. Unpublished paper.

Arbache, J. S., A. Dickerson and F. Green (2004). "Trade Liberalisation and Wages in Developing Countries." *The Economic Journal 114(493)*: F73-F96.

Attanasio, O., P. K. Goldberg and N. Pavcnik (2004). "Trade Reforms and Wage Inequality in Colombia." *Journal of Development Economics 74(2)*: 331-366.

Benabou, R. (1996). Inequality and Growth, in. B. S. Bernanke and J. J. Rotemberg (Ed.) *NBER Macroeconomics Annual*, Cambridge, Mass: MIT Press.

Bennell, P. (1995). "Rates of Return to Education in Asia: A Review of the Evidence." *Institute of Development Studies Working Paper No.24, Sussex.*

Blau, F. D. and L. M. Kahn (1996). "International Differences in Male Wage Inequality: Institutions versus Market Forces." *Journal of Political Economy 104(4)*: 791-837.

Cowell, F. A. (1995). *Measuring Inequality*: Harvester Wheatsheaf, Hemel Hempstead.

Cowell, F. A. and S. P. Jenkins (1995). "How Much Inequality can we Explain? A Methodology and an Application to the USA." *Economic Journal 105*: 421-30.

Das, M. B. (2003). Ethnicity and Social Exclusion in Job Outcomes in India: Summary of Research Findings. Unpublished paper, *World Bank Institute*.

Datt, G. (1999). "Has poverty decreased since economic reforms? Statistical data analysis." *Economic & Political Weekly December 11*.

Deaton, A. and J. Dreze (2002). "Poverty and Inequality in India: A Reexamination." *Economic and Political Weekly September 7*: 3729-48.

Dickens, W. T. and K. Lang (1985). "A Test of Dual Labor Market Theory." *American Economic Review 75(4)*: 792-805.

Dubey, A., O. Kharpuri and V. Pala (2004). "Globalisation, Occupational Structure and Wages in India." Paper presented at the International Seminar on Wages and Incomes in India, Dec 12-14, Mumbai.

Duraisamy, P. (2002). "Changes in the Returns to Education in India, 1983-94: By Gender, Age-Cohort and Location." *Economics of Education Review 21(6)*: 609-622.

Fields, G. S. (1980). *Poverty, Inequality and Development*: Cambridge University Press, Cambridge.

Fields, G. S. (2002). Accounting for Income Inequality and Its Change. Unpublished paper, *Cornell University*. Ithaca.

Fields, G. S. and G. Yoo (2000). "Falling Labour Income Inequality in Korea's Economic Growth: Patterns and Underlying Causes." *Review of Income and Wealth 46(2)*: 139-159.

Galbraith, J. K., P. Conceicao and H. Kum (2000). "Inequality and Growth Reconsidered Once Again: Some New Evidence from Old Data." *UTIP Working Paper 17*, University of Texas Inequality Project.

Galbraith, J. K., D. Roy Chowdhury and S. Shrivastava (2004). "Pay Inequality in the Indian Manufacturing Sector, 1979-1998." *University of Texas Inequality Project Working Paper 28*.

Gallup, J. L. (2002). "The Wage Labour Market and Inequality in Vietnam in the 1990s." *World Bank Policy Research Working Paper No. 2896*, Washington D.C.

Harrison, A. and G. Hanson (1999). "Who Gains from Trade Reform? Some Remaining Puzzles." *Journal of Development Economics 59(1)*: 125-154.

Kapila, U., Ed. (2001). *Indian Economy Since Independence*. New Delhi: Academic Foundation.

Katz, L. and D. Autor (1999). Changes in Wage Structure and Earnings Inequality, in. O. Ashenfelter and D. Card (Ed.) *Handbook of Labor Economics*: North Holland.

Kingdon, G. G. (1998). "Does the Labour Market Explain Lower Female Schooling in India?" *Journal of Development Studies 35(1)*: 39-65.

Krstic, G. and B. Reilly (2003). "An Anatomy of Labour Market Earnings Inequality in Serbia - 1995 to 2000." *Discussion Paper No. 97, Department of Economics, University of Sussex.*

Krueger, A. B. and L. H. Summers (1988). "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica 56(2)*: 259-293.

Labour Bureau (various years). *Indian Labour Statistics*. Shimla: Government of India.

Lawrence, R. Z. and M. J. Slaughter (1993). "International Trade and American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?" *Brookings Papers on Economic Activity: Microeconomics 2*: 161-211.

Lee, L.-F. (1983). "Generalised Econometric Models with Selectivity." *Econometrica 51(2)*: 507-512.

Litchfield, J. A. (2003). *Poverty, Inequality and Social Welfare in Brazil, 1981-1995*. Thesis, London: London School of Economics and Political Science.

Machin, S. (2002). "Labour Market Inequality and Changes in the Relative Demand for Skills." Paper prepared for the Royal Statistical Society "Explanations of Rising Economic Inequality" Conference.

Mujeri, M. (2002). "Decomposing Wage Inequality Change in Bangladesh: An Application of Double Calibration Technique." *Paper prepared as part of "Exploring Links between Globalisation and Poverty in South Asia"*, Globalisation and Poverty Programme, DFID.

Murphy, K. M. and F. Welch (1990). "Empirical Age-Earnings Profile." *Journal of Labour Economics 8(2)*: 202-229.

Newell, A. (2001). "The Distribution of Wages in Transition Countries." *Department of Economics Discussion Paper 72*, University of Sussex.

Newey, W. K. (1999). Two Step Series Estimation of Sample Selection Models. Unpublished paper, *Massachusetts Institute of Technology*. Cambridge MA.

Nouroz, H. (2001). *Protection in Indian Manufacturing: An Empirical Study*. New Delhi: MacMillan.

Özler, B., G. Datt and M. Ravallion (1996). *A Database on Poverty and Growth in India*: Poverty and Human Resources Division Policy, Research Department, The World Bank.

Panda, M. K. (1999). Growth with Equity: Policy Lessons from the Experience of India, in. (Ed.) *Growth with Equity: Policy Lessons from the Experiences of Selected Asian Countries*: UNESCAP.

Ravallion, M. (1997). "Can High-Inequality Developing Countries Escape Absolute Poverty?" *Policy Research Working Paper 1775*, World Bank.

Reilly, B. (1990). "Occupational Endogeneity and Gender Wage Differentials for Young Workers: An Empirical Analysis using Irish Data." *The Economic and Social Review 21(3)*: 311-328.

Robbins, D. (1996). "HOS Hits Facts: Facts Win. Evidence on Trade and Wages in the Developing World." *HIID Development Discussion Paper 557*, Harvard University.

Sachs, J. D., N. Bajpai and A. Ramiah (2002). "Why Have Some Indian States Grown Faster than the Others?" *Centre for International Development, Harvard University*.

Taubman, P. and M. L. Wachter (1986). Segmented Labour Markets, in. O. Ashenfelter and R. Layard (Ed.) *Handbook of Labour Economics*: North Holland. *Volume II*.

Tendulkar, S. D. (2003). Organised Labour Market in India: Pre and Post Reform. Unpublished paper, *Paper prepared for the Conference on Anti Poverty and Social Policy on January 2-4 at Alwar, India*.

Trost, R. P. and L.-F. Lee (1984). "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity." *The Review of Economic and Statistics 66(1)*: 151-156.

Unni, J. (2001). "Gender and Informality in Labour Market in South Asia." *Economic and Political Weekly 36(26)*: 2360-2377.

Vasudeva-Dutta, P. (2004). "Trade Protection and Inter-Industry Wages in India." *Poverty Research Unit Working Paper 27*, University of Sussex.

Wood, A. (1997). "Openness and Wage Inequality in Developing Countries: The Latin American Challenge to East Asian Conventional Wisdom." *World Bank Economic Review 11(1)*: 33-57.

DATA APPENDIX

The three large-scale nationally representative employment surveys for January-December 1983, July 1993–June 1994 and July 1999–June 2000 (referred to as 1983, 1993 and 1999 in the paper) are a rich source of data on wages, employment and individual and household characteristics. The final sample generated for use in the empirical analysis comprises data from 17 states – Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Tripura, Uttar Pradesh and West Bengal. Taken together these 17 states accounted for about 96% of total gross state domestic product and 97% of total population in 1999-2000 (see also Özler et al. (1996)). Individuals were divided into three mutually exclusive categories using current weekly status: (i) non-wage earners, i.e., non-participants in the labour market, self-employed and unemployed individuals (ii) regular wage employment and (iii) casual wage employment.

<u>Wages</u>

Nominal weekly wages include payment in cash and kind. Some observations (about 1-2% in the three years) had to be dropped from the sample as there were missing observations on wages, hours worked and industry affiliation. It is assumed that the excluded observations are random as the mean observable characteristics of the workers excluded do not differ significantly from those retained in the sample though this does not take possible differences in unobservables into account. The wage distribution was then trimmed by 0.1% at the top and bottom tails. This is necessarily an *ad hoc* measure: some researchers prefer to trim the wage distribution using specific values (Krueger and Summers, 1988) while others prefer to trim the distribution at the tails (Arbache et al., 2004) as adopted here. These nominal wages were deflated to 1983 prices using official state-level monthly consumer price indices (base year 1960-61) for agricultural labourers (CPIAL) for rural wages and industrial workers (CPIIW) for urban wages (Labour Bureau, various years). Using the survey data on the intensity of work – i.e., no work, part-time or full-time – for each day of the week and assuming a 48 hour week, the number of hours worked and the real hourly wage was constructed.

<u>Variables influencing wages</u>

The standard quadratic form for age is not used as this did not fit the data well and following Murphy and Welch (1990) age splines at ten-year intervals were included instead as a proxy for labour force experience. Marital status is a dummy variable coded one if currently married and zero if never married, widowed, divorced or separated. There is information on the highest level of schooling completed (but not on the number of years of schooling) so dummy variables corresponding to the following education variables were constructed: primary school, middle school, secondary school and graduate and above. The reference category is individuals who are illiterate or have less than two years of formal or informal schooling. Dummy variables for caste and religious affiliation were constructed from household data; the omitted category is all other households. Seasonality effects are captured by dummy variables for the quarter in which the households were interviewed. These quarterly dummies were also interacted with the dummy variable for the rural sector. The variable for industry affiliation was constructed based on the individual's current weekly industrial classification. In order to ensure adequate observations in each industry the three-digit National Industrial Classification codes are aggregated into 38 industries.

APPENDIX

Table A1: Inequality measures, real weekly wages

	Regular workers			Casual workers		
	1983	**1993**	**1999**	**1983**	**1993**	**1999**
Mean real weekly wage (Rs.)	141.479	206.81	267.03	45.02	62.02	74.19
	(112.27)	(153.18)	(220.34)	(31.66)	(34.30)	(42.20)
Median real weekly wage (Rs.)	121.98	173.66	210.29	38.39	56.43	66.34
Wage inequality measures:						
Inter-quantile range (90-10)	219.0349	350.5498	479.0380	58.4178	79.0688	91.1097
Standard error	(1.7247)	(1.7967)	(3.0404)	(0.6075)	(0.6407)	(0.7045)
Confidence interval	[215.6504, 222.4193]	[347.0240, 354.0755]	[473.0716 485.0044]	[57.2257 59.6099]	[77.8116 80.3260]	[89.7272 92.4923]
Inter-quantile range (90-50)	133.6018	229.3618	331.0612	37.8083	48.6347	56.6144
Standard error	(1.6131)	(1.7101)	(3.0191)	(0.5647)	(0.6103)	(0.6648)
Confidence interval	[130.4363, 136.7673]	[226.0059, 232.7176]	[325.1367, 336.9857]	[36.7000, 38.9165]	[47.4371, 49.8322]	[55.3099, 57.9190]
Inter-quantile range (50-10)	85.4331	121.1880	147.9768	20.6096	30.4341	34.4953
Standard error	(0.7536)	(0.8505)	(1.5734)	(0.1853)	(0.2393)	(0.2704)
Confidence interval	[83.9543, 86.9118]	[119.5190, 122.8570]	[144.8893, 151.0643]	[20.2460, 20.9732]	[29.9645, 30.9037]	[33.9648, 35.0258]
Gini coefficient	0.3791	0.3881	0.4210	0.3235	0.2919	0.2896
Standard error	(0.0020)	(0.0016)	(0.0015)	(0.0019)	(0.0013)	(0.0013)
Confidence interval	[0.3752, 0.3830]	[0.3851, 0.3912]	[0.4180, 0.4240]	[0.3198, 0.3272]	[0.2893, 0.2945]	[0.2870, 0.2922]
Mean log deviation, GE(0)	0.2684	0.3103	0.3268	0.1863	0.1829	0.1461
Standard error	(0.0027)	(0.0032)	(0.0024)	(0.0022)	(0.0025)	(0.0014)
Confidence interval	[0.2631, 0.2737]	[0.3040, 0.3166]	[0.3221, 0.3314]	[0.1819, 0.1906]	[0.1779, 0.1879]	[0.1434, 0.1487]
Theil index, GE(1)	0.2477	0.2488	0.2918	0.1852	0.1440	0.1409
Standard error	(0.0034)	(0.0021)	(0.0023)	(0.0032)	(0.0014)	(0.0015)
Confidence interval	[0.2411, 0.2544]	[0.2447, 0.2529]	[0.2873, 0.2963]	[0.1790, 0.1915]	[0.1413, 0.1466]	[0.1380, 0.1437]
Half squared coeff. variation, GE(2)	0.3149	0.2743	0.3404	0.2473	0.1529	0.1618
Standard error	(0.0088)	(0.0031)	(0.0039)	(0.0091)	(0.0018)	(0.0024)
Confidence interval	[0.2976, 0.3321]	[0.2682, 0.2804]	[0.3328, 0.3480]	[0.2294, 0.2652]	[0.1495, 0.1563]	[0.1571, 0.1665]
Sample size	27,356	26,387	27,295	28,855	26,398	29,805

Note: Real weekly wage (Rs.) in constant 1983 prices. The inequality measures are computed without applying household weights. Figures in parentheses are as labelled - standard deviation for the mean wage, standard errors and 95% confidence intervals for inequality measures obtained by bootstrapping with 1000 replications.

Table A2: Wage regression models for regular and casual workers
Dependent variable: Natural log of real hourly wages

	Regular wage workers			Casual wage workers		
	1983	**1993**	**1999**	**1983**	**1993**	**1999**
Individual characteristics:						
Education:						
Completed primary school	0.0658***	0.0426***	0.0485***	0.0088	0.0108*	0.0227***
	(0.0072)	(0.0100)	(0.0114)	(0.0057)	(0.0057)	(0.0056)
Completed middle school	0.1361***	0.0933***	0.1090***	-0.0065	-0.0079	0.0165*
	(0.0085)	(0.0120)	(0.0115)	(0.0092)	(0.0092)	(0.0085)
Completed secondary school	0.3486***	0.2640***	0.2945***	-0.0085	-0.0348**	0.0117
	(0.0110)	(0.0160)	(0.0140)	(0.0174)	(0.0154)	(0.0147)
Completed graduate school	0.6192***	0.5385***	0.6025***	-0.0043	-0.0693*	0.0180
	(0.0150)	(0.0229)	(0.0188)	(0.0540)	(0.0381)	(0.0318)
Experience:[1]						
Age spline: 15-25 years	0.0139***	0.0106***	0.0130***	0.0106***	0.0136***	0.0137***
	(0.0017)	(0.0024)	(0.0022)	(0.0007)	(0.0009)	(0.0010)
Age spline: 25-35 years	0.0174***	0.0165***	0.0203***	-0.0008	-0.0009	0.0019***
	(0.0008)	(0.0012)	(0.0012)	(0.0005)	(0.0006)	(0.0006)
Age spline: 35-45 years	0.0106***	0.0158***	0.0160***	-0.0015**	-0.0003	-0.0007
	(0.0008)	(0.0011)	(0.0011)	(0.0006)	(0.0007)	(0.0006)
Age spline: 45-55 years	0.0055***	0.0129***	0.0159***	-0.0007	-0.0043***	-0.0033***
	(0.0013)	(0.0014)	(0.0014)	(0.0008)	(0.0009)	(0.0009)
Age spline: 55-65 years	-0.0291***	-0.0304***	-0.0252***	-0.0052***	-0.0053***	-0.0034**
	(0.0031)	(0.0049)	(0.0046)	(0.0013)	(0.0015)	(0.0014)
Others:						
Married	0.0647***	0.0781***	0.0878***	0.0271***	0.0327***	0.0196***
	(0.0073)	(0.0095)	(0.0102)	(0.0037)	(0.0045)	(0.0043)
Member of scheduled caste	-0.0472***	-0.0389***	-0.0309***	-0.0033	0.0111*	-0.0060
or tribe [2]	(0.0056)	(0.0074)	(0.0072)	(0.0051)	(0.0060)	(0.0059)
Muslim	-0.0163**	-0.0328***	-0.0436***	0.0249***	0.0245***	0.0130**
	(0.0075)	(0.0093)	(0.0092)	(0.0049)	(0.0055)	(0.0053)
Residence in rural areas	-0.1454***	-0.0468**	-0.0240	-0.0972***	-0.0479***	-0.0329***
	(0.0141)	(0.0207)	(0.0166)	(0.0082)	(0.0079)	(0.0076)
Selection bias correction term	-0.0202	-0.1152***	-0.1149***	0.0610***	0.0682***	0.0391***
	(0.0174)	(0.0270)	(0.0228)	(0.0125)	(0.0129)	(0.0145)
Selection effect [3]	0.0231	0.1281***	0.1336***	-0.0726***	-0.0781***	-0.0448***
	(0.0228)	(0.0334)	(0.0308)	(0.0177)	(0.0169)	(0.0190)
Constant	0.4375***	0.6996***	0.8532***	0.4263***	0.4483***	0.5710***
	(0.0579)	(0.0917)	(0.0846)	(0.0279)	(0.0324)	(0.0352)
Number of observations	27,356	26,387	27,295	28,855	26,398	29,805
Adjusted R^2	0.5458	0.4789	0.5298	0.3393	0.3058	0.3382
Standard error of estimate	0.3483	0.4229	0.4477	0.2127	0.2318	0.2364

Notes: a/ Standard errors in parentheses (obtained after bootstrapping with 1000 replications). b/ * significant at 10%; ** significant at 5%; *** significant at 1%. c/ Also included but not reported - six seasonality dummy variables including interactions of the season and settlement type (the first season (January-March) and the interaction with the rural dummy are the omitted categories); 37 industry dummy variables are included (five agricultural and allied industries, two mining, 21 manufacturing and eight non-tradable industries); food crops is the omitted industry; 16 state dummies (West Bengal is the omitted state).
1/ The estimated coefficients on the age splines are not cumulative. 2/ These terms are derived from the schedules of the Constitution Orders passed in 1950 that listed the names of specific castes and tribes eligible for special treatment from the State in terms of reservations in public sector employment, legislatures and government-funded educational institutions (Das, 2003). 3/ The selection effect is computed as the coefficient on the selection bias correction term times its mean for the nominated outcome - regular or casual wage employment - multiplied by 100. A crude estimate of the standard error of the selection effect is obtained as

follows: the square of the average selection bias correction term times the standard error of its estimated coefficient (Reilly, 1990).

Figure A1 Lorenz curves

Regular workers