Educational Attainment and Wage Inequality in Turkey

Gurleen Popli[1] — Okan Yılmaz[2]

Abstract. This paper analyses the relationship between wage distribution and educational attainment of the workforce. Wage inequality in Turkey decreased over 2002–10; a period over which it also saw an increase in the supply of educated workers. Our findings suggest that decreasing inequality in the bottom half of the distribution was largely due to decreasing returns to education and experience; whereas the moderate decline in inequality in the upper tail of the wage distribution is explained by a fall in returns to the 'routine' occupational tasks. The effect of changes in the composition of workers was found to be moderate.

1. Introduction

Following the dramatic increase in wage inequality observed across several countries, there has been a considerable interest in studying the distribution of wages, over the last three decades. In particular, the steep increase in the wage gap between the college and the high school graduates in the United States has been documented by many authors, starting with Katz and Murphy (1992) and Bound and Johnson (1992). Besides the United States, an increasing educational premium has also been documented for many OECD countries (Berman *et al.*, 1997; Machin and Reenen, 1998). In line with these studies, most of the analysis in the literature has subsequently focused on wage inequality between workers with high and low educational qualifications. In these studies, the increasing wage inequality is mostly explained by the supply–demand approach and attributed to increasing skill demand.[1]

Two most popular theories used to explain increase in the relative demand for skilled workers are the Stolper–Samuelson effects and the skill biased technological change (SBTC). The first theory argues that competition with labour-abundant countries decreases the relative price of labour-intensive goods and accordingly reduces the real wages of less educated workers in both relative and absolute terms (Hanson and Harrison, 1999). The SBTC theory argues that in accordance with the diffusion of higher technologies of information and communication, the labour demand has shifted in favour of skilled workers and increased the skill premiums (Acemoglu, 2002). Although there is no clear agreement,

We are grateful to the Turkish National Statisitics Institute for making the data available. We would like to thank the participants at the 33rd IARIW General Conference, Rotterdam, Netherlands, August 2014, for their useful comments. We would also like to thank Sarah Brown, Ken Clark, Jonathan Perraton, Peter Wright, the editor, and an anonymous referee for their helpful comments. All responsibility for the analysis and interpretation of these data lies with the authors.

[1]Department of Economics, University of Sheffield, 9 Mappin Street, Sheffield S1 4DT, UK

[2]Department of Economics, Bartın University, Ağdacı Campus, Merkez, Bartin, 74110, Turkey Email: oyilmaz@bartin.edu.tr

LABOUR 31 (1) 73–104 (2017) DOI: 10.1111/labr.12083

JEL J31, J82, I24

it is generally believed that SBTC is the main reason for increasing wage inequality. Most of the initial explanations for the increasing demand for skilled workers centred on the argument that the new and efficient technology that is embodied in the new capital equipment complimented the high skilled workers, thus depressing the relative demand for less skilled workers (Berman *et al.*, 1994; Caselli, 1999; Galor and Tsiddon, 1997; Krusell *et al.*, 2000).

More recently a number of papers (using evidence from both the United States and Europe) nuance this argument further and link the earnings inequality to job polarization, routinization and changes in returns to specific 'tasks' all linked with evolving technology and globalization (Acemoglu and Autor, 2011; Firpo *et al.*, 2011; Goos *et al.*, 2009, 2014). The argument here is that the changing nature of technology has resulted in the depression of returns to (and eventual loss of jobs in) 'routine' tasks that can now be performed by technologies. As these routine tasks are performed by the medium skilled workers, this in turn leads to wage polarization where, relative to middle of the wage distribution, the wages at the upper and the lower tail grow faster; this in turn then results in an increases in inequality in the top half of the distribution but a decrease in inequality in the bottom half of the distribution (Naticchioni *et al.*, 2014).

Although a number of studies have looked at the evidence from developed countries, to a lesser extent, evolution of wage distribution has also been analysed for developing countries. Using cross-sectional household data for Argentina, Chile, Costa Rica, Colombia, Malaysia, Mexico, Philippines, Chinese Taipei and Uruguay, Robbins (1996) found that trade liberalization was accompanied by rise in relative wages and the demand for skilled workers. Hanson and Harrison (1999) examined the effects of Mexican trade reforms on wage inequality and they asserted that the rising wage gap was associated with changes internal to industries and could not be explained by the Stolper–Samuelson type effects. Galiani and Sanguinetti (2003) analysed the trade liberalization process which took place in Argentina during 1990s; their results show that trade openness explained only a small proportion of the increase in wage inequality. Berman and Machin (2000) analysed SBTC for developing countries and found evidence that demand for skilled labour increased in middle income countries and this increase was mainly due to skill upgrading within industries rather than a reallocation of employment from low- to high-skill industries.

In the recent years a number of Latin American countries have witnessed a decline in wage inequality (Campos *et al.*, 2012; Lustig *et al.*, 2013). The reasons given for the declining wage inequality echo those for rising wage inequality, in reverse. The observed fall in wage inequality has been largely attributed to fall in skill premiums and more robust government transfers; further much of the fall in the skill premiums is attributed to increases in educational attainments and fall in the demand for the skilled labour (Lustig *et al.*, 2013). Naticchioni *et al.* (2008) in their paper analysed the stable wage inequality in Italy over 1993–2004 and found that the increasing educational attainment of the workforce was countervailed by the stable skill demand.

This paper analyses the relationship between wage distribution and changing educational attainment of the workforce. For our analysis we focus on the Turkish labour market, which experienced a substantial decrease in wage inequality between 2002 and 2010; a period over which it also saw a rapid increase in the educational attainment of the workforce. We explore the link between the changes in the relative supply of skilled labour, fall in the returns to education and decrease in wage inequality. Within this context we also check if there was any evidence of job polarization and its impact on wage inequality for Turkey;

where job polarization is captured at the individual level by using the occupational task measures classification suggested by Goos *et al.* (2009).[2]

To understand the factors affecting wage inequality in Turkey over this period, the change in wage distribution between 2002 and 2010 is, as a first step, decomposed into three components: effect of returns to human capital (changes in coefficients); effect of changes in the composition of human capital (changes in covariates); and changes in residual distribution (price and composition of unmeasured human capital characteristics). To do this we use the decomposition method proposed by Lemieux (2002); furthermore, in order to correct for the possible selection bias related to the participation in wage employment, a two-step procedure, proposed by Dubin and McFadden (1984), is also used within this decomposition. The analysis is done for both men and women separately.

Next we use the Firpo *et al.* (2009) decomposition which allows us to identify, for the entire distribution, the composition effect of the different covariates used in the analysis. This then allows us to look at the role played by education in explaining the decreasing inequality; it also allows us to look at the contribution of occupations to the observed changes, i.e., whether or not job polarization played any role in the changing wage distribution.

Our paper contributes to the existing literature in two main ways. First, it provides the most recent and rigorous detailed decomposition analysis of wage inequality in Turkey; where we look closely at the impact of increase in education of the labour force on wage inequality. Second, it is the first paper to look at occupational task measures and see how they might contribute to changes in the distribution of wages in Turkey.

The key findings of the paper suggest that the increasing supply of educated labour, which was due to an increase in the number of universities and the reform to the education system which increased the years of compulsory education, had a significant effect on the wage distribution in Turkey between 2002 and 2010. Wage inequality in Turkey decreased over this period with much of the decline happening in the lower tail of the wage distribution. Decomposition results reveal that decreasing wage inequality in the bottom half was mainly due to decreasing returns to education and experience. It is also found that the decrease in residual wage dispersion was mainly due to the decreasing price of unmeasured human capital. The moderate decline in inequality in the upper tail of the wage distribution is explained by fall in returns to 'routine' occupational tasks. Finally, the effect of changes in the composition of workers is found to be moderate.

The structure of the paper is as follows; Section 2 reviews the existing literature on the empirical work for Turkey and discusses the education reforms and the macroeconomic environment for the Turkish economy. Section 3 presents the underlying hypothesis and outlines the methodology used in this paper. Section 4 presents the data and descriptive statistics; wage regression estimations, and decomposition results are discussed in Section 5; and finally Section 6 concludes.

2. Turkish economy

2.1 Wage inequality in Turkey

Studies on the wage inequality in Turkey are few. Kizilirmak (2003) analysed the increase in relative demand for skilled labour and wage inequality in the Turkish manufacturing sector for the 1988–2000 period and argued that the change in relative demand for skilled workers was primarily due to the within-industry skill upgrading (which in part she attributed to trade). Elveren and Galbraith (2009) examined the sub-sectors of Turkish manufacturing and found an increase in the sector premiums between 1980 and 2001. Meschi *et al.* (2011) investigate the relationship between trade openness, technology adoption and relative demand for skilled workers in the Turkish manufacturing sector using firm-level data covering the period 1980–2001. They find evidence of skill upgrading within firms as a direct result of increased trade openness; thus lending support to the argument of SBTC in a middle income country like Turkey.

Tansel and Bodur (2012) use the quantile regression technique to analyse the evolution of male wage inequality over the 1994–2002 period, using the Household Budget Survey (HBS). The authors find that there was an overall decline in wage inequality in Turkey; where inequality declined at the lower end of wage distribution while it increased at the top end. Their results indicated that education contributed to higher wage inequality through both within and between dimensions; and finally they find evidence of decreasing returns to education which they attribute to increase in the educational attainment of the labour forces, combined with a relatively stable demand for skilled workers.

Bakis and Polat (2015) study the evolution of wage inequality in Turkey over the period 2002–10, using the Labour Force Survey (LFS) data. The authors do a supply–demand analysis, as in Katz and Murphy (1992). They find a steep increase in the supply of college educated workers relative to both, workers with some college level education and workers with high school education; and the relative wages for the college educated workers over this period decreased. The aggregate decomposition of wage inequality[3] further revels that it is the changing returns to covariates, and not the changing composition, that explain much of the observed change in inequality. The authors provide institutional changes, especially the sharp increase in minimum wage as an explanation for the observed decline in wage inequality.

Our analysis uses the same data set and covers the same period as the Bakis and Polat (2015) paper. However, we differ from their analysis in three key ways: (1) we do the aggregate decomposition, similar to them, however, we decompose the changes in the distribution of residuals as well; (2) we do a detailed decomposition to look at the specific covariates which might contribute to the changing trends in the wage inequality; and (3) given the evidence of Meschi *et al.* (2011) of potential SBTC in Turkey we look at the different occupational task measures to see if they contribute to wage inequality in Turkey.

2.2 *Education system and composition of labour force*

The education system in Turkey has seen numerous changes in the last few decades. First of these was the reform to the compulsory education system in 1997. Before the reform the education system in Turkey was organized as compulsory primary school (5 years), middle (or secondary) school (3 years), high school and vocational high school (3 years) and university education (2–6 years). With the reform in the education system, compulsory primary education was extended to 8 years and middle school was abolished.[4]

The second important development was the rapid increase in the number of universities. The establishment of new universities began with the second Five Year Development Program which was put in practice in 1968 (State Planning Organization, 1967). Over the following years, the number of universities increased gradually with the establishment of 20 universities between 1971 and 1991. In 1991, there were only 29 universities in Turkey.

However, in the following years the number of universities increased dramatically, particularly in 1992 when 24 new universities were established in 1 year. Finally, with the ‘a university in each city’ policy (starting in 2000) and growing private sector participation, in 2011 the number of universities reached 165. As a consequence of these new universities, the number of university students also increased substantially. For instance, from 1994 to 2011, number of students in formal university education increased by more than 300 per cent and rose from 1.2 to 3.7 million (Turkey National Statistics Institute; TurkStat henceforth).

The fast increase in the number of universities in Turkey has been accompanied by a rising difference between the old and the new universities in terms of equipment, funding and resources which can be expected to lead differences in the quality of education. Hatakenaka (2006) argued that the new universities which are not located in metropolitan centres find it difficult to recruit qualified people because they are unwilling to relocate to outer regions even though several measures were tried in the past to address these issues.[5] In addition, due to the central university entrance examination system, more successful students have the chance of studying with other successful students in the universities with better resources and this leads to further skill differences between university graduates.[6]

Considering these two developments, it can be argued that the supply of educated labour in labour force (employed and unemployed) continuously increased starting from the early 1990s. According to TurkStat web database, between 1988 and 2011 the share of individuals with university education in labour force increased from 4 to 21 per cent for women and 5–15 per cent for men. For the same period, the share of individuals with education less than high school in labour force decreased for both men and women. Similar patterns are also observed in the composition of the employed and the unemployed. Based on these observations, a decrease in educational premium can be expected; the increasing supply of workers with higher education is likely to decrease the upward pressure on the wages of university graduates, whereas decreasing the share of workers with low level of education moderates the downward pressure on their wages.

2.3 Macroeconomic environment

Turkey experienced several major economic crises starting from the 1990s. The first crisis occurred with the Gulf war in 1991. The second crisis, which was triggered by the fiscal and external imbalances, occurred in 1994 and a GDP growth rate of negative 6.1 per cent was witnessed. After a short period of recovery, due to the adverse effects of the Asian, Russian and Brazilian crises, the Turkish economy experienced a slowdown in 1998 with a growth rate of 3.1 per cent, and then contracted in 1999 at the rate of negative 3.4 per cent. Even though the economy was in boom in 2000 with a 7.3 per cent growth rate, the heaviest crisis of Turkey’s recent history, mainly due to the major capital outflows, occurred in November 2000 and February 2001 when the GDP declined by 5.7 per cent in 2001 in real terms (Tansel and Bodur, 2012).

In terms of growth rates, the post-2001 period can be defined as the recovery period for Turkey. The real growth rate was 6.2 per cent in 2002 and the economy grew by 6 per cent on average until 2007 when the recent global economic crises first showed its effects on Turkey. On the other hand, in contrast to the fast growth performance across sectors, additional employment could not be generated. The rate of unemployment was 6.5 per cent in 2000 and it increased to 10.4 per cent in 2002. The unemployment rate remained high and never fell below 10 per cent despite the rapid surges in the GDP and exports. This observation is defined as jobless-growth in the literature, characterized by a contraction of formal jobs and increased informalization of economic activities (Telli *et al.*, 2007; Yeldan, 2006).

3. Hypothesis and methodology

3.1 Hypothesis

In his seminal paper, Mincer (1974) proposed an earnings function in which wages are defined as a function of human capital, where individuals who make higher investment in human capital receive higher returns from the labour market. The wage function can accordingly be written as:

$$y_i = X_i\beta + \varepsilon_i, \tag{1}$$

where y_i is log wage of individual i, X_i is the vector of individual characteristics that determine wages; β is the coefficient vector giving the marginal returns to the covariates in X_i; and ε_i is the random error term. The key individual characteristics that Mincer stressed were education (schooling) and experience (years in the labour market, including but not restricted to, on-the-job training). However, in subsequent empirical work vector X_i has included a wide range of wage-determining characteristics.

Although higher levels of education provide higher earnings at the individual level, an increase in the educational attainment of the whole population does not necessarily mean an increase in returns to education. Pritchett (2001) argued that the marginal return to adding an additional year of schooling in whole population can be substantially different from average returns estimated with a Mincerian regression at a single point in time depending on the shifts in skill demand; he asserted that marginal returns to education decrease as the supply of educated labour expands if the demand remains stagnant.

In their seminal paper, Katz and Murphy (1992) analysed the changes in wage inequality in the United States during the 1963–87 period. Their results show that the college wage premium decreased in the 1971–79 period in which there was a large increase in the supply of college graduates. On the other hand, the college wage premium increased in the 1979–87 period in which the growth of the supply of graduates was very small. Accordingly, they argued that, combined with the smooth increase in skill demand, the fluctuations in the growth of the supply of college graduates as a fraction of the labour force played an important role in explaining the large differences in the relative wages of college graduates between these two decades. Based on the findings of Pritchett (2001) and Katz and Murphy (1992), it can be argued that the education premium and wage inequality can increase or decrease depending on the differences between the relative growth rate of the supply and demand for the skills.

Another important implication of estimating the Mincerian regression is that if the demand for skill increases, returns to the unmeasured human capital such as unobservable skills linked to school quality, intrinsic ability and effort, which are the main reasons why workers with the same level of education and experience have different wages, increase as well. In particular, the rate of increase in returns to unmeasured characteristics is greater for individuals who have more education. In econometric terms, residuals in Mincerian-type equations are empirically heteroskedastic.

Lemieux (2002) and Martins and Pereira (2004) also note that school quality differences are more likely to be prevalent at higher schooling levels, because those are the stages that exhibit greater heterogeneity in schooling paths and school quality. Moreover, differences in school quality and the variance of residuals increase even more dramatically if admissions to schools get more selective at higher levels of education (as is in the case of Turkey). Therefore, changes in skill premiums affect wage inequality in two ways simultaneously; first, it affects the wage gap between workers with lower and higher education levels which causes an increase in the between-group inequality component. Second, it affects the wage dispersion within workers who have the same level of education but studied at schools with different qualities, which causes an increase in the within-group inequality component.[7]

The argument of technological changes leading to polarization in wages and jobs requires a distinction between skills and tasks. Based on the models proposed by Acemoglu and Autor (2011) and Firpo *et al.* (2011), the tasks content of individual jobs can be determined by the occupations. These occupation-based task categories can then be incorporated in the standard Mincerian wage regressions to determine the impact of changes in the composition of these tasks categories and returns to them on wage distribution. In general it is argued that routine tasks are more vulnerable to technological changes compared with non-routine tasks; this is due to the fact that routine tasks require repeated physical strength and motions which computer and machines can perform (Goos *et al.*, 2014). Job polarization would require the share of workers in the routine jobs to fall and the share of workers in the non-routine jobs to increase.

3.2 Decomposition method

In our paper we do both an overall decomposition, also called the aggregate decomposition, and a detailed decomposition which looks at the contribution of different covariates to the overall change in the distribution of wages.

The literature on decomposition of wage inequality goes back to the seminal papers of Oaxaca (1973) and Blinder (1973). In their models the change in average wage between two time periods can be decomposed into two components: (1) the 'explained' effect, which is the effect of changes in the distribution of covariates (also referred to as the composition effect); and (2) the 'unexplained' effect, which is the effect of changes in the regression coefficients (changes in the returns to the covariates), this is often also referred to as the 'wage structure' effect.

Following the seminal work of Blinder and Oaxaca, various other decomposition methods have been developed over the last few decades. One of the methods is that proposed by Juhn *et al.* (1993, JMP hereafter) in which they extended the Blinder–Oaxaca decomposition of the mean wages by taking the distribution of residuals into account. In the JMP framework the changes in the distribution over time can be decomposed into three components: (1) changes in the distribution of observable covariates; (2) changes in the regression coefficients (returns to the observable covariates); and (3) changes in the distribution of residuals. However, the JMP decomposition does not account for changes in the distribution of covariates; their method thus ignores the problem of heteroscedasticity.

For the aggregate decomposition we use the decomposition method proposed by Lemieux (2002), which unifies the residual imputation method proposed by JMP and the re-weighting factor method proposed by DiNardo et al. (1996). The Lemieux approach has several advantages compared with the other decomposition methods. First, like JMP it allows for the decomposition of changes in the entire distribution of wages rather than the decomposition of the change in the mean wages only. Second, unlike JMP it accounts for distribution of covariates, therefore, it can account for the problem of heteroskedasticity. Finally, it is also possible to decompose the changes in the residual distribution into the effect of the changes in unobservable characteristics and the effect of changes in returns to those characteristics. This then allows us to test the hypothesis of the human capital approach which asserts that a positive change in returns to observable skills exerts a positive impact on returns to unobservable skills.

Once we have the aggregate decomposition, to look at the contribution of different covariates we do a detailed decomposition, for this we use the recentered influence function (*RIF*) regression method proposed by Firpo *et al.* (2009). This method is an extension of the Blinder–Oaxaca decomposition where in the first step we run a regression where the dependent variable, instead of being log wages, is *RIF* of the statistic of interest, say the τ^{th}-quantile (q_τ) of the log wages, $RIF(y, q_\tau)$. Once the *RIF* regression has been estimated we can do a detailed decomposition, in a way similar to the Blinder–Oaxaca decomposition, for any distributional parameter.

Full econometric specification of different decompositions that we use in our empirical analysis is provided in Appendix A.

3.3 Selection bias

By definition, wage estimations are performed for the individuals who have reported their wages. However, selection into wage-earning sector cannot be assumed as random, resulting in a selection bias. The consequence of ignoring this problem is that the estimators of the wage regressions are biased (Heckman, 1979).

According to TurkStat, nearly 30 per cent of working men in Turkey were self-employed in the year 2000. This share decreased to 23 per cent in 2010. The considerable share of self-employed is an indicator of non-random participation into the wage sector. The selection bias problem becomes even more crucial in wage estimations for women. Over the last 50 years, Turkey's female labour force participation has been decreasing (Goksel, 2012). According to the Global Gender Gap Report 2013 (World Economic Forum, 2013), the female labour force participation rate in Turkey is 30 per cent and this ratio puts Turkey in the 123rd place out of 136 countries. Another problem that may cause selection bias is that there is a substantial share of unpaid family workers who traditionally work in the agricultural sector and are recorded as employed in employment statistics. According to TurkStat, for the year 2002, 49 per cent of total female employment consists of unpaid workers. However, these individuals do not report any form of labour income and so they are automatically omitted from the sample. Given the above issues we take into account the selection bias in our analysis.

In the first step, following Dubin and McFadden (1984), a multinomial logit model is estimated separately for men and women. An individual i is characterized as having three options; these options are different for men and women. The sample of woman is grouped into following three categories: $fw_i = 2$ if working in the wage sector; $fw_i = 1$ if she is an unpaid family worker; and $fw_i = 0$ if she is economically inactive (non-participant). The categories for men are: $fw_i = 2$ if working in the wage sector; $fw_i = 1$ if self-employed; and $fw_i = 0$ if he is economically inactive (non-participant). The predicted probabilities from the multinomial logit model (estimated separately for men and women) are then used to construct a selection correction term, λ_{ij}, by using the formula provided by Hill (1983).

In the second step, augmented wage equations are estimated separately for men and women by including correction term as an additional regressor:

$$y_{ij} = X_{ij}\beta_j + \theta_j\lambda_{ij} + \varepsilon_{ij}, \tag{2}$$

where y_{ij} is the log wage of individual i in sector j; X_{ij} is the vector of sector specific individual characteristics; β_j is the vector of returns to characteristics in sector j; θ_j is the unknown coefficient related with the selection correction term; and ε_{ij} is the independent residual term. The augmented wage equation estimation results are then used in decomposition of the change in wage distributions of men and women over time.

Details of the estimation of the selection-corrected wage regressions, including the econometric specification of the multinomial logit model and how the correction term is estimated, are given in Appendix A.

4. Data

4.1 Sample selection

Data sources are the two waves of Household Labour Force Survey (LFS), which is conducted by TurkStat, for the years 2002 and 2010. The LFS has the largest sample of Turkish labour force and contains information on both the workers' demographic characteristics and characteristics of the main job for each individual within the household.[8]

For our analysis we create two cross-sections of workers. The details of the exclusion criterion used and the observations lost as a result are given in Table 1. The number of individuals surveyed by TurkStat increased between 2002 and 2010 (first row, 'full sample', Table 1). We exclude from our analysis all workers outside the ages of 20 and 64.[9] We also drop from our analysis all those who are coded as students, ill, disabled and retired. As LFS does not provide any information on the wages that are earned from the second job, individuals who have more than one job are also dropped. A very small proportion of the women (men) are self-employed (unpaid family workers) so these are also dropped from the sample. The remaining observations are classified as economically inactive, self-employed (only for men), unpaid family workers (only for women) and wage earners.

Labour Force Survey contains information on net monthly wages. In order to get hourly wages, monthly wages are first divided by 4.3 and then divided by the usual hours of work per week. Finally, hourly wages are deflated by the consumer price index, which is provided by TurkStat, to obtain real hourly wages.

4.2 Descriptive statistics

Tables with the descriptive statistics of the variables for each sample year for women and men are presented in Appendix B (Tables B1 and B2).

Wage-earning women are the most educated, with 40 per cent having a university education in 2010. Unpaid family workers are the least educated, with a third of them having no formal education. Between 2002 and 2010 there has been a shift in the distribution of education for the wage earners, with a fall in the wage earners with primary (high school)

Table 1. Exclusion criterion for the sample

	2002			2010		
	All	**Men**	**Women**	**All**	**Men**	**Women**
Full sample	300,689	146,836	153,853	522,171	255,053	267,118
Exclusion observations						
Age over 64 or <20	135,422	67,780	67,642	227,112	112,765	114,347
Students	3,252	1,969	1,283	6,190	3,401	2,789
Disabled or ill	2,838	1,639	1,199	10,691	4,678	6,013
Retired	10,987	8,249	2,738	18,426	14,159	4,267
More than one job	1,259	1,148	111	4,747	4,344	403
Employers	5,370	5,082	288	7,602	7,032	570
Self-employed (women)	2,552	–	2,552	5,835	–	5,835
Unpaid family workers (men)	2,765	2,765	–	4,024	4,024	–
Sample for analysis[a]	136,244	58,204	78,040	237,544	104,650	132,894
% of the full sample	45%	40%	51%	45%	41%	50%
Economically inactive[b]	72,237	11,426	60,811	115,298	17,704	97,594
	(53%)	(20%)	(78%)	(49%)	(17%)	(73%)
Self-employed (men)[b]	14,007	14,007	–	23,869	23,869	–
	(10%)	(24%)	–	(10%)	(23%)	–
Unpaid family workers (women)[b]	7,787	–	7,787	15,220	–	15,220
	(6%)	–	(10%)	(6%)	–	(11%)
Wage earners[b]	42,213	32,771	9,442	83,157	63,077	20,080
	(31%)	(56%)	(12%)	(35%)	(60%)	(15%)
Sample for wage equations[c]	41,205	32,039	9,166	77,701	59,131	18,570

Notes: [a] This is the sample used in sample selection equation.
[b] The % in the () are % of the sample for analysis.
[c] Any further losses in observations from the row above (wage earners) are for following reasons: reported extreme values of experience (inconsistent with age); reported zero wage; and exclusion of Skilled agricultural and fishery workers occupational category. We lose 212 observations in 2002 and 639 observations in 2010 due to exclusion of this occupation.

education by 6 percentage points (4 percentage points), and an increase in the university educated women by 8 percentage points. Compared with other groups, wage-earning women are more likely to be heads of their household, more likely to come from households where other members of the household are also wage earners; and are less likely to have children present in the household. Women in unpaid work tend to come predominantly from rural households (where other members of the household are more likely to be self-employed). Although a majority of wage-earning women are younger than 35 years, economically inactive and unpaid worker group mainly consist of women who are older than 35.

Wage-earning men are more educated relative to other two groups (economically inactive and self-employed). There has been a shift in the distribution of education between 2002 and 2010, towards university-educated workers in the wage sector; the share of workers with primary education decreased by 8 percentage points, whereas the share of workers with a university degree increased by 5 percentage points. There is not much difference in the mean values of the variables that represent presence of children in the household between self-employed men and wage earning men; for the economically inactive men on the other hand values of these variables are smaller. Self-employed men have a high proportion of other members of the household who do unpaid family work (possibly women in the family). Self-employed men are the oldest group with 70 per cent of them over 35 years old; on the other hand, more than half of the inactive and wage worker groups are younger than 35.

From the descriptive statistics, it can be seen that the change in compulsory education system and the increasing number of universities caused a substantial change in the educational composition of wage workers (and workforce in general). The new education system, which abolished the primary school and made secondary education compulsory, reduced the number of workers with only primary education for both men and women; whereas with the increasing accessibility of university education, the number of workers with a university degree increased for both genders. Based on these observations, it can be argued that the supply of educated labour in Turkey increased between 2002 and 2010.

To capture job polarization, following Goos *et al.* (2009), occupations are classified into three task categories: abstract, routine and service. Details of the occupation classifications used are given in Appendix B, Table B3. Women, relative to men are over represented in the abstract jobs, i.e., among professionals and technicians. This could be due to the selection process as only very high skilled women enter the labour force (40 per cent of the wage earning women have university education). Women are underrepresented in the routine jobs, such as plant and machine operators and assemblers. For both men and women, over time share of wage earners in abstract category has not changed; whereas the share of workers in the routine category has decreased.

4.3 Wage dynamics

Table 2 presents the wage dynamics for men and women over time. The mean log wages for both men and women have increased over time; with a higher increase in the mean wages of women. The mean wage for men is higher than the mean wage for women in 2002 and this relationship reverses in 2010. This observation does not necessarily imply positive discrimination towards women. For instance, using the Structure of Earnings Survey 2006 conducted by the TurkStat, Kaya (2010) shows that on average women earn 2.5 percentage points more than men. She also argues that the higher average wage for women is a consequence of a composition effect as most of the female employees have a university degree, whereas male employees mostly have only primary education. By exploiting a quantile decomposition method, she finds that although the composition effect had a narrowing effect on the gender wage gap (favouring women), the effect of differences in the returns to characteristics had an increasing effect on the gender wage gap (favouring men) showing that the human capital characteristics of women are rewarded less than the characteristics of their male counterparts. Considering Kaya's (2010) findings, the reversal of relative wages between men and women between 2002 and 2010 can be explained by the higher increase in the share of university graduates among female wage workers.

Wage inequality over this period decreased with a fall in inequality reflected in the fall in the 90-10 wage gap which has fallen by 21 per cent for women and almost 18 per cent for men. For both men and women, much of the fall in inequality has been in the lower tail of the distribution (the 50-10 wage gap), with inequality in the upper tail (90-50 wage gap) showing only a moderate decline. The changing pattern of inequality seems to be consistent with the job polarization argument.

Table 2. Wage dynamics

	Mean log wages	Log wage differentials 90–10	90–50	50–10
Women				
2002	5.38 (0.91)	2.169	1.139	1.030
2010	5.65 (0.78)	1.792	1.099	0.693
Total change	0.27	−0.377	−0.041	−0.336
%	5.0	−21.1	−3.7	−48.5
Men				
2002	5.42 (0.80)	1.902	1.091	0.811
2010	5.61 (0.67)	1.617	1.020	0.597
Total change	0.19	−0.285	−0.071	−0.214
%	3.5	−17.6	−6.9	−35.8

5. Empirical results

5.1 Wage regressions

The first step in our empirical investigation is to look at the multinomial logit model for labour market status.[10] Tables 3 and 4 show the relative risk ratios (*RRR*) from the maximum likelihood estimation of the multinomial logit model for women and men, respectively. The omitted category in the model is the group of economically inactive men and women. The *RRR* tells us how the probability of choosing wage employment or unpaid family work (self-employment in the case of men) relative to being economically inactive changes if we increase the independent variable by one unit. If the *RRR* is >1, it means that the individual is more likely to be in wage employment or unpaid family work (self-employment for men) and accordingly, if it is lower than 1, the individual is more likely to be economically inactive.

Table 3 shows that for women education at all levels increases the probability of choosing wage employment. Having other members of the household in the wage sector, being head of the household and (for the year 2002) having a grandmother in the household also increases the probability of being in wage employment. The presence of children in the household aged 5 or less and between 5 and 11, being married, and presence of self-employed members in the household all have a decreasing effect on wage employment.

Table 4 shows that for men education at all levels, the presence of children in the household aged 5 or less, other members of the household in the wage sector, and being the head of the household increase the probability of choosing wage employment. Living in rural area and the presence of unpaid family workers in the household all increase the likelihood of being self-employed. For both men and women (Tables 3 and 4), the RRRs for educational categories are much smaller in 2010 than they were in 2002. This indicates that education does not increase the probability of being a wage worker (as opposed to being economically inactive) in 2010 as much as it did in 2002.

To see the changes in the returns to the human capital characteristics the wage regressions are estimated for each year, for men and women separately. The regression results (Table 5) show that the coefficients for the education variables decreased from 2002 to 2010 for men and women. The percentage decrease is more at lower education levels than

Table 3. Relative risk ratios (women)

Variables	2002 Unpaid family worker	2002 Wage worker	2010 Unpaid family worker	2010 Wage worker
Education				
Primary school	1.10**	1.74***	1.30***	1.56***
	(0.05)	(0.11)	(0.04)	(0.06)
Secondary school	0.68***	3.01***	1.20***	2.42***
	(0.09)	(0.23)	(0.07)	(0.11)
High school	0.66***	5.54***	0.92	4.21***
	(0.08)	(0.38)	(0.08)	(0.18)
Vocational high school	0.69*	8.13***	0.92	5.41***
	(0.13)	(0.60)	(0.09)	(0.24)
University	1.13	33.58***	0.85	21.26***
	(0.26)	(2.38)	(0.10)	(0.90)
Presence of children in the household				
Age ≤4	0.69***	0.52***	0.65***	0.46***
	(0.03)	(0.02)	(0.02)	(0.01)
5 ≤ Age < 11	0.97	0.81***	0.98	0.76***
	(0.04)	(0.03)	(0.03)	(0.02)
11 ≤ Age < 15	0.88***	0.99	0.92**	1.02
	(0.04)	(0.04)	(0.03)	(0.03)
Other members of the household				
Presence of wage workers	0.72***	2.08***	0.82***	1.63***
	(0.04)	(0.07)	(0.03)	(0.04)
Presence of self-employed	14.23***	0.85***	12.21***	0.93**
	(0.80)	(0.05)	(0.42)	(0.03)
Presence of unpaid family workers	4.91***	0.74***	4.88***	0.64***
	(0.23)	(0.08)	(0.18)	(0.05)
Marital status				
Married	1.31***	0.36***	1.70***	0.43***
	(0.10)	(0.02)	(0.10)	(0.01)
Divorced	0.37***	0.99	0.55***	1.39***
	(0.13)	(0.09)	(0.09)	(0.08)
Widowed	0.47***	0.34***	0.85	0.40***
	(0.08)	(0.04)	(0.11)	(0.03)
Age				
25–35	1.57***	1.53***	1.82***	1.55***
	(0.11)	(0.07)	(0.11)	(0.05)
35–45	1.81***	1.44***	2.74***	1.72***
	(0.14)	(0.07)	(0.17)	(0.07)
45–65	1.80***	0.60***	2.51***	0.62***
	(0.14)	(0.04)	(0.16)	(0.03)
Head of household dummy	0.02***	2.02***	0.21***	1.56***
	(0.01)	(0.13)	(0.03)	(0.07)
Grandmother	1.67***	1.20***	1.26***	1.07
	(0.11)	(0.08)	(0.06)	(0.05)
Rural	9.70***	1.00	8.16***	0.88***
	(0.41)	(0.04)	(0.24)	(0.03)
Constant	0.00***	0.07***	0.00***	0.11***
	(0.00)	(0.01)	(0.00)	(0.01)
Observations	7,787	9,442	15,220	20,080

Notes: Robust standard errors are reported in (). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Omitted category: economically inactive.

Table 4. Relative risk ratios (men)

	2002		2010	
Variables	**Self-employed**	**Wage worker**	**Self-employed**	**Wage worker**
Education				
Primary school	1.35***	2.10***	1.19***	1.58***
	(0.09)	(0.14)	(0.06)	(0.07)
Secondary school	1.39***	3.23***	1.29***	2.22***
	(0.11)	(0.23)	(0.08)	(0.10)
High school	1.19**	3.67***	1.09	2.29***
	(0.10)	(0.26)	(0.07)	(0.11)
Vocational high school	0.90	4.27***	1.15**	3.31***
	(0.08)	(0.32)	(0.08)	(0.17)
University	0.63***	4.96***	0.73***	3.96***
	(0.06)	(0.36)	(0.05)	(0.19)
Presence of children in the household				
Age ≤ 4	0.99	0.96	1.00	1.02
	(0.04)	(0.03)	(0.03)	(0.03)
5 ≤ Age < 11	0.90***	0.93**	0.92***	1.00
	(0.03)	(0.03)	(0.03)	(0.02)
11 ≤ Age < 15	0.90**	0.95	0.92***	0.96
	(0.04)	(0.03)	(0.03)	(0.02)
Other members of the household				
Presence of wage workers	0.68***	1.43***	0.73***	1.30***
	(0.03)	(0.04)	(0.02)	(0.03)
Presence of self-employed	0.17***	0.83***	0.35***	0.99
	(0.02)	(0.04)	(0.02)	(0.03)
Presence of unpaid family workers	23.91***	1.06	25.43***	1.35***
	(1.74)	(0.08)	(1.31)	(0.07)
Marital status				
Married	2.19***	2.06***	2.29***	2.02***
	(0.19)	(0.10)	(0.14)	(0.08)
Divorced	1.32	1.13	1.28**	1.01
	(0.24)	(0.15)	(0.14)	(0.08)
Widowed	1.62**	0.96	2.42***	1.05
	(0.30)	(0.18)	(0.41)	(0.18)
Age				
25–35	2.33***	1.62***	2.72***	1.46***
	(0.20)	(0.07)	(0.20)	(0.05)
35–45	2.85***	1.35***	3.69***	1.20***
	(0.27)	(0.07)	(0.29)	(0.05)
45–65	3.03***	0.75***	4.00***	0.62***
	(0.29)	(0.04)	(0.31)	(0.03)
Head of household dummy	2.71***	2.53***	2.39***	2.13***
	(0.19)	(0.11)	(0.11)	(0.07)
Rural	1.66***	0.67***	1.63***	0.66***
	(0.06)	(0.02)	(0.05)	(0.02)
Constant	0.07***	0.29***	0.06***	0.61***
	(0.01)	(0.02)	(0.00)	(0.03)
Observations	14,007	32,771	23,869	63,007

Notes: Robust standard errors are reported in (). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Omitted category: economically inactive.

Table 5. Wage regression estimations

	Women		Men	
Variables	**2002**	**2010**	**2002**	**2010**
Education				
Primary school	0.04	−0.09***	0.15***	0.00
	(0.05)	(0.03)	(0.03)	(0.01)
Secondary school	0.10*	−0.02	0.29***	0.06***
	(0.06)	(0.03)	(0.03)	(0.01)
High school	0.25***	0.06**	0.43***	0.15***
	(0.06)	(0.03)	(0.03)	(0.01)
Vocational high school	0.18***	−0.01	0.45***	0.12***
	(0.06)	(0.03)	(0.03)	(0.01)
University	0.56***	0.27***	0.79***	0.50***
	(0.07)	(0.04)	(0.03)	(0.02)
Tenure	0.02***	0.04***	0.04***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Tenure squared (10^{-3})	−0.51***	−0.76***	−0.77***	−0.29***
	(0.11)	(0.06)	(0.06)	(0.03)
Presence of children in the household				
Age ≤ 4	0.06***	0.08***	−0.02*	−0.01*
	(0.02)	(0.01)	(0.01)	(0.00)
5 ≤ Age < 11	−0.02	0.01	−0.02**	−0.02***
	(0.02)	(0.01)	(0.01)	(0.00)
11 ≤ Age < 15	−0.03*	−0.03***	−0.02**	−0.03***
	(0.02)	(0.01)	(0.01)	(0.01)
Marital status dummies				
Married	0.11***	0.16***	0.07***	0.04***
	(0.02)	(0.01)	(0.01)	(0.01)
Divorced	0.08**	0.03*	−0.01	0.04**
	(0.04)	(0.02)	(0.05)	(0.02)
Widowed	0.13**	0.04	−0.06	0.09
	(0.05)	(0.03)	(0.10)	(0.06)
Age				
25–35	0.12***	0.10***	0.11***	0.10***
	(0.02)	(0.01)	(0.01)	(0.01)
35–45	0.16***	0.16***	0.20***	0.18***
	(0.02)	(0.01)	(0.02)	(0.01)
45–65	0.21***	0.25***	0.25***	0.25***
	(0.04)	(0.02)	(0.02)	(0.01)
Member of social security system	0.55***	0.33***	0.41***	0.27***
	(0.02)	(0.01)	(0.01)	(0.01)
Occupational categories				
Service	0.16***	−0.01	−0.05***	−0.07***
	(0.03)	(0.01)	(0.01)	(0.01)
Abstract	0.43***	0.30***	0.24***	0.25***
	(0.03)	(0.02)	(0.01)	(0.01)
Λ (selection term)	0.15***	0.28***	0.18***	0.28***
	(0.04)	(0.03)	(0.03)	(0.02)
Constant	3.79***	4.81***	4.06***	4.88***
	(0.09)	(0.05)	(0.06)	(0.03)
Industry dummies	YES	YES	YES	YES
Observations	9,166	18,570	32,039	59,131
R-squared	0.61	0.63	0.49	0.56

Notes: Robust standard errors are reported in (). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Details of the industry dummies used are given in Appendix B, Table B4.

at higher education levels. For example, although returns to high school education for women (men) decreased by 76 per cent (65 per cent) over time, considering 2002 as the base year, the returns to university education decreased by 51 per cent (36 per cent). The coefficient for the abstract dummy is positive for both men and women; compared with the base category of routine workers, workers in the occupations classified as abstract earn more, ceteris paribus. For women, however, the premium of being in abstract jobs has fallen over time. The coefficient for the service dummy is negative (except for the 2002 female sample) showing that workers in the service group earn less than the workers in the routine occupations, ceteris paribus. For men, there has been no change in the returns to service sector jobs, whereas for women these returns have fallen over time.

According to the wage regression results, compared with the base category of never married, being married has a positive effect on wages. Also, working in the formal sector (which is captured by the membership of social security system) has a positive effect on wages. The coefficient for the selection term is positive for men and women and statistically significant in each year, which means that the unobservable factors that affect selection into wage employment are positively correlated with the unobservable factors that affect wages.

The descriptive statistics and the regression results show that composition of the wage earners and returns to their characteristics changed significantly over time. However, it is not possible to see to what extent these changes were responsible for the changes in the wage distribution without decomposing the changes in wage inequality.

5.2 Decompositions

Table 6 reports the results of the aggregate decomposition using the method proposed by Lemieux. Looking at the 90-10 wage gap columns, for both men and women, the decomposition results indicate that the systematic part of the wage equation (sum of the covariate and the coefficient effects) explains about 58 and 55 per cent of the fall in

Table 6. Aggregate decomposition of changes in wage distribution

	Women			**Men**		
Inequality measure	**90-10**	**90-50**	**50-10**	**90-10**	**90-50**	**50-10**
2002	2.169	1.139	1.030	1.902	1.091	0.811
2010	1.792	1.099	0.693	1.617	1.020	0.597
Total change %	−0.377	−0.041	−0.336	−0.285	−0.071	−0.214
	(−21.1)	(−3.7)	(−48.5)	(−17.6)	(−6.9)	(−35.8)
Effect of coefficients	−0.174	0.003	−0.178	−0.121	−0.031	−0.091
	(−46.2)	(8.5)	(−52.9)	(−42.7)	(−43.0)	(−42.5)
Pricing function	−0.099	−0.080	−0.019	−0.141	−0.065	−0.075
	(−26.2)	(−196.0)	(−5.6)	(−49.4)	(−92.3)	(−35.2)
Covariates	−0.046	−0.020	−0.026	−0.034	0.016	−0.050
	(−12.3)	(−50.1)	(−7.7)	(−12.1)	(22.2)	(−23.4)
Unexplained	−0.058	0.056	−0.114	0.012	0.009	0.002
	(−15.3)	(137.6)	(−33.9)	(4.1)	(13.1)	(1.1)

Notes: Table presents results from the Lemieux (2002) decomposition.
Percentage shares of each effect in total change are shown in parentheses.

inequality for women and men, respectively. Changes in returns to the observed characteristics (effect of coefficients) had a substantial share in explaining the decreasing inequality both at the upper tail (90-50 gap) and the lower tail (50-10 gap) of the wage distribution for men. For women, however, the changes in the coefficients was inequality increasing at the upper tail, and inequality decreasing in the lower half of the wage distribution.

The effect of covariates, for men, is inequality increasing in the upper tail but inequality decreasing in the lower tail of the wage distribution. For women, the effect of covariates is negative and relatively small in the lower tail of the distribution but larger in the top tail of the wage distribution. Part of the explanation for the differences in these findings for men and women could be due to the fact that the proportion of women with university education (who are more likely to be in the upper tail) working in the wage sector is much higher than the proportion of men in the wage sector with university education. A further increase in the supply of educated women dampens the returns to them, whereas for men it does not.

For women, around 26 per cent of the fall in inequality is explained by returns to the unmeasured characteristics (pricing function), this rises to almost 50 per cent for men. The pricing function has a considerable share in explaining the declining inequality at the top end of the distribution, whereas the size of this effect is found to be relatively small in the lower tail of the distribution, for both men and women.

Table 7 presents the decomposition of changes in wage residuals. The skill pricing function explains almost all of the variation in the change in dispersion of wage residuals in the top half of the distribution; and almost 80–90 per cent of the variation in the bottom part of the distribution. This provides evidence that decreasing returns to human capital were accompanied by a decrease in pricing function of unmeasured characteristics.

Next we do the detailed decomposition using the method proposed by Firpo *et al.* (2009); the results are presented in Table 8; Panel A reports the aggregate decomposition and Panels B and C report the detailed composition and coefficient effects, respectively, for the main factors: education, experience (age dummies, tenure,[11] and tenure square) and the occupational categories (routine, service and abstract).[12]

From Panel A we can see that, for both men and women, there is a bigger fall in inequality in the bottom half (50-10) of the wage distribution and most of this is explained by the coefficient effect. For women, the covariate (composition) effect is negative and

Table 7. Decomposition of changes in wage residuals

	Women			**Men**		
Inequality measure	**90-10**	**90-50**	**50-10**	**90-10**	**90-50**	**50-10**
2002	1.262	0.646	0.616	1.270	0.652	0.618
2010	1.047	0.514	0.534	1.031	0.523	0.509
Total change %	−0.215	−0.132	−0.082	−0.239	−0.130	−0.109
	(−20.5)	(−25.7)	(−15.5)	(−23.2)	(−24.9)	(−21.4)
Effect of covariates	−0.002	0.009	−0.006	−0.015	−0.003	−0.018
	(−1.1)	(6.5)	(−7.4)	(−6.4)	(−2.2)	(−16.7)
Pricing function	−0.217	−0.141	−0.076	−0.224	−0.133	−0.091
	(−98.9)	(−106.5)	(−92.6)	(−93.6)	(−97.8)	(−83.3)

Notes: Table presents results from the Lemieux (2002) decomposition.
Percentage shares of each effect in total change are shown in parentheses.

Table 8. Detailed decomposition of changes in wage distribution

	Women			Men		
Inequality measure	**90-10**	**90-50**	**50-10**	**90-10**	**90-50**	**50-10**
Panel A: Aggregate decomposition						
2002	2.151	1.160	0.991	1.874	1.092	0.782
2010	1.808	1.152	0.656	1.632	1.038	0.594
Total change	−0.344	−0.008	−0.335	−0.241	−0.054	−0.188
% change	−16.0	−0.7	−33.9	−12.9	−4.9	−24.0
Composition effect	−0.042	−0.023	−0.020	−0.012	0.036	−0.048
	(−12.3)	(−269.0)	(−5.9)	(−4.9)	(67.3)	(−25.6)
Specification error	−0.014	−0.001	−0.013	0.034	−0.011	0.045
	(−4.1)	(−8.9)	(−4.0)	(14.0)	(−21.1)	(24.0)
Coefficient effect	−0.244	0.030	−0.274	−0.258	−0.075	−0.182
	(−70.9)	(354.0)	(−81.5)	(−106.8)	(−140.5)	(−97.1)
Reweighting error	−0.044	−0.015	−0.029	−0.005	−0.003	−0.002
	(−12.7)	(−176.2)	(−8.6)	(−2.3)	(−5.6)	(−1.3)
Panel B: Detailed composition effect: Main factors						
Education	0.038	−0.014	0.052	0.052	0.027	0.025
	(10.9)	(−168.2)	(15.4)	(21.7)	(50.2)	(13.6)
Experience	−0.018	0.002	−0.020	−0.017	0.009	−0.026
	(−5.3)	(25.6)	(−6.0)	(−7.2)	(16.1)	(−13.8)
Occupation categories						
Routine	0.003	−0.002	0.005	0.002	0.001	0.001
	(0.9)	(−26.5)	(1.6)	(1.0)	(2.7)	(0.4)
Service	−0.007	−0.004	−0.003	−0.002	−0.001	0.000
	(−2.1)	(−45.4)	(−1.0)	(−0.7)	(−2.6)	(−0.2)
Abstract	−0.001	0.000	−0.001	0.000	0.000	0.000
	(−0.2)	(−1.1)	(−0.2)	(0.2)	(0.6)	(0.1)
Panel C: Detailed coefficient effect: Main factors						
Education	−0.042	0.046	−0.088	0.001	−0.006	0.007
	(−12.2)	(544.7)	(−26.2)	(0.6)	(−10.6)	(3.8)
Experience	−0.055	0.201	−0.256	−0.069	0.123	−0.192
	(−15.9)	(2,509.2)	(−76.3)	(−28.7)	(227.1)	(−102.0)
Occupation categories						
Routine	−0.042	−0.015	−0.027	0.012	−0.021	0.034
	(−12.3)	(−181.2)	(−8.1)	(5.1)	(−39.8)	(18.0)
Service	0.099	0.002	0.097	−0.016	0.000	−0.016
	(28.6)	(23.5)	(28.8)	(−6.7)	(−0.2)	(−8.5)
Abstract	0.009	0.031	−0.022	0.002	0.012	−0.011
	(2.6)	(365.6)	(−6.5)	(0.7)	(22.9)	(−5.6)

Notes: Table presents results from the Firpo *et al.* (2009) decomposition.
Percentage shares of each effect in total change are shown in parentheses.
Education: Primary school, secondary school, high school, vocational high school and university.
Experience: age dummies, tenure, tenure square.

relatively more effective in the top half of the distribution; for men, the total composition effect is positive (inequality increasing) on the top half of the distribution and negative (inequality decreasing) in the bottom of the distribution. This is consistent with the Lemieux decomposition results.[13]

In Panel B of Table 8 we have detailed composition effect. Overall the changes in the composition of educated workers, where there are more of them in the labour market, has been inequality increasing. The only exception is the top tail of the wage distribution for women. So, although it seems that for women, holding the returns to education constant, an increase in the educational qualification lowers difference between the workers who are in the top end of the distribution and those who are around the median of the distribution, the same cannot be said for men. This could be as women have had much bigger gains in university education relative to men. Composition effect of experience has been inequality decreasing for both men and women in the lower tail of the distribution but inequality increasing in the upper tail. Changes in the composition of workers in different occupation-based task categories contribute very little to the overall changes in inequality.

Most of the change in inequality has happened in the bottom half of the distribution; the inequality has fallen in the top half but very little. Focusing on the bottom half (50-10), here most of the decrease in inequality is attributed to the coefficient effect, which explains more than 80 per cent of the fall in inequality for women and almost 97 per cent of the fall in inequality for men. Looking at the detailed coefficient effect, for women much of this is due to fall in returns to education and experience, which together explain almost all of the fall in inequality in the bottom tail; followed by changes in the returns to routine occupations which explain 8 per cent of the fall in inequality. For men, returns to education and returns to routine tasks are inequality increasing in the bottom tail of the distribution. Much of the decrease in inequality for men is explained by the changes in returns to experience.

6. Conclusion

This paper investigates the decrease in wage inequality observed in Turkey between 2002 and 2010, by looking at the relationship between the distribution of wages and the educational attainment of the workforce. To understand the factors behind the changes in the wage distribution we do an aggregate decomposition using the Lemieux (2002) method, which allows us to look at the residuals, and a detailed decpomposition using the Firpo *et al.* (2009) method.

The results of the analysis show that the increasing supply of educated labour, which is attributed to the increase in the number of universities and the reform in the education system which increased the years of compulsory education, had a substantial effect on the wage distribution in Turkey between 2002 and 2010. As a result of the increase in supply, the wage inequality decreased in both the top and the bottom half of the wage distribution. The decrease in wage inequality was relatively low in the top half of the wage distribution, with most of the decline being concentrated in the lower half of the wage distribution.

Using the decomposition methodology proposed by Lemieux, it is found that the decreasing wage inequality observed in Turkey between 2002 and 2010 was mainly due to two factors: (1) the decreasing between group inequality which is related to decreasing coefficients for education, a finding confirmed by the Firpo *et al.* decomposition; and (2) the decreasing within-group inequality due to decreasing skill pricing function of unmeasured skills. The second finding is in accordance with the skill price theory of Lemieux (2002) who argued that negative changes in the coefficient component exerts a negative impact on the residual component along the wage distribution, providing a measure for unmeasured skills pricing. Substantial change in skill pricing function may be linked to high-quality differences between universities.

Using the Firpo *et al.* detailed decomposition we find that for women much of the decrease in inequality in the lower tail is explained by the coefficient effect with falling returns to education and experience explaining almost all of the fall in inequality. For men, the fall in inequality in the bottom tail is mainly explained by the changing returns to experience. We also look at job polarization: there is little role for it in the composition effect. In the coefficient effect, changes in the returns to routine tasks explain the fall in inequality in the upper tail of the wage distribution for both men and women. This is contrary to what we expect from the job polarization hypothesis.

Notes

[1] The impact of institutional factors such as, unionization, minimum wages and collective bargaining practices, on the wage distribution has also been investigated (Lee, 1999; Card *et al.*, 2003).

[2] The task content of work is determined at the occupational level, making occupation a key channel via which technology affects the wage structure (Firpo *et al.*, 2011).

[3] The authors use the DiNardo *et al.* (1996) and Juhn *et al.* (1993) decomposition methods.

[4] Turkish education system was reformed again in 2012 and a more complex system known as the '4+4+4 system' has been adopted.

[5] For instance, there was a requirement for academics to 'serve' in outer areas before being promoted. Today, there is a salary supplement to provide incentives for people to work in universities in the outer areas.

[6] There is also teaching quality differences between high schools in Turkey. Apart from vocational high schools, there are three different types of public high schools in Turkey, namely science high schools, Anatolian high schools and general high schools which have different levels of selectiveness in their admissions.

[7] Card and Krueger (1992) provided evidence that men who are educated at higher quality schools have higher return to an additional year of schooling; and returns are also higher for individuals who studied with better educated teachers.

[8] Previous studies on Turkey have used HBS. The main aim of HBS is to collect consumption and expenditure information, although it has some information on the labour market. LFS has numerous advantages over HBS: in particular, it has a much wider coverage and more observations which allow us to look at women, who are underrepresented in the labour market; all the national statistics are based on LFS, and these are used for policy analysis; and lastly, the main aim of LFS, unlike HBS, is to look at the labour market.

[9] The reason of choosing this age band is that decision to pursue higher education is endogenous as it depends on the returns to higher education. However, individuals who are at the age of 20 or older are likely to have already made their educational decisions. We exclude those above 64, as that is when the retirement decisions are made.

[10] Selection procedure requires the availability of valid instruments, i.e., variables which affect the labour market status but do not affect wages. For identification the following variables are included in the multinomial logit model. Three dummies to capture the labour market status of other members of the household: presence of wage workers, presence of self-employed and presence of unpaid family workers in the household; a dummy for the head of the household; a dummy for whether the worker lives in rural or urban area; and finally, only for women, whether or not there is a grandmother present in the household. Justification for using these variables, along with the full econometric specification, is discussed in Appendix A of the paper.

[11] Tenure is obtained by using the question 'year that you started your latest job/occupation'.

[12] Detailed decomposition for all other variables used in the wage regression is available on request from the authors.

[13]The Lemieux decomposition method decomposes the change in the observed (actual) gap; whereas the Firpo *et al.* decomposition method decomposes the predicted gap.

[14]Unless the number of observations are equal in the two periods, it is not possible to match residuals exactly. Lemieux (2002) suggested a solution to this problem. His idea is to discretize the distribution of residuals in h intervals which contain the same number of observations and replace the actual residuals by the average residual in each interval. In this analysis h is chosen as 500.

[15]We use selection correction in both the wage regressions underlying Lemeiux decomposition and the *RIF* regressions underlying the Firpo *et al.* decomposition. This is done to ensure consistency of analysis and results across the two decomposition methods. As a robustness check analysis was done without selection correction for the male sample, the results were qualitatively similar to those presented in the paper.

References

Acemoglu D. (2002) 'Technical Change, Inequality, and the Labor Market', *Journal of Economic Literature* 40: 7–72.

Acemoglu D. and Autor D. (2011) 'Skills, Tasks and Technologies: Implications for Employment and Earnings' in Card D. and Ashenfelter O. (eds.) *Handbook of Labor Economics*, 4th edn, Amsterdam: Elsevier North Holland: 1043–1171.

Bakis O. and Polat S. (2015) 'Wage Inequality in Turkey, 2002–10', *Economics of Transition* 23(1): 169–212.

Berman E. and Machin S. (2000) 'Skill-Based Technology Transfer Around the World', *Oxford Review of Economic Policy* 16(3): 12–22.

Berman E., Bound J. and Griliches Z. (1994) 'Changes in the Demand for Skilled Labor Within U.S. Manufacturing: Evidence From the Annual Survey of Manufactures', *The Quarterly Journal of Economics* 109(2): 367–397.

Berman E., Bound J. and Machin S. (1997) 'Implications of Skill-Biased Technological Change: International Evidence' CEP Discussion Papers DP0367. Centre for Economic Performance, LSE.

Blinder A. S. (1973) 'Wage Discrimination: Reduced Form and Structural Estimates', *The Journal of Human Resources* 8(4): 436–455.

Bound J. and Johnson G. (1992) 'Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations', *The American Economic Review* 82(3): 371–392.

Calvo-Armengol A. and Jackson M. O. (2004) 'The Effects of Social Networks on Employment and Inequality', *The American Economic Review* 94(3): 426–454.

Campos R., Esquivel G. and Lustig N. (2012) 'The Rise and Fall of Income Inequality in Mexico, 1989–2010', UNU-WIDER, Working Paper No. 2012/10.

Card D. and Krueger A. B. (1992) 'Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States', *Journal of Political Economy* 100(1): 1–40.

Card D., Lemieux T. and Riddell W. C. (2003) 'Unionization and Wage Inequality: A Comparative Study of the U.S, the U.K., and Canada', NBER Working Papers 9473.

Caselli F. (1999) 'Technological Revolutions', *The American Economic Review* 89(1): 78–102.

Cunningham W. (2001) 'Breadwinner Versus Caregiver: Labor Force Participation and Sectoral Choice Over the Mexican Business Cycle' in Katz E. and Correia M. (eds.) *The Economics of Gender in Mexico: Work, Family, State and Market*. Washington, DC: The World Bank: 85–132.

DiNardo J., Fortin N. M. and Lemieux T. (1996) 'Labor Market Institutions and the Distribution of Wages 1973–1992: A Semiparametric Approach', *Econometrica* 64(5): 1001–1044.

Dubin J. and McFadden D. L. (1984) 'An Econometric Analysis of Residential Electric Appliance Holdings and Consumption', *Econometrica* 52(2): 345–362.

Elveren A. Y. and Galbraith J. K. (2009) 'Pay Inequality in Turkey in the Neo-Liberal Era, 1980–2001', *European Journal of Comparative Economics* 6(2): 177–206.

Ettner S. L. (1996) 'The Opportunity Costs of Elder Care', *The Journal of Human Resources* 31(1): 189–205.

Firpo S., Fortin N. M. and Lemieux T. (2009) ‘Unconditional Quantile Regressions’, *Econometrica* 77(3): 953–973.

Firpo S., Fortin N. M. and Lemieux T. (2011) ‘Occupational Tasks and Changes in the Wage Structure’, IZA Discussion Paper No. 5542.

Fortin N., Lemieux T. and Firpo S. (2011) ‘Decomposition Methods in Economics’ in Ashenfelter O. and Card D. (eds.) *Handbook of Labor Economics*, Vol. IV.A. Amsterdam: North-Holland: 1–102.

Galiani S. and Sanguinetti P. (2003) ‘The Impact of Trade Liberalization on Wage Inequality: Evidence From Argentina’, Working Paper 65, Universidad de San Andres, Departamento de Economia.

Galor O. and Tsiddon D. (1997) ‘Technological Progress, Mobility, and Economic Growth’, *The American Economic Review* 87(3): 363–382.

Goksel I. (2012) ‘The Reasons of Decreasing Trend of Female Labour Force Participation In Turkey: The Role of Conservatism’, Working Paper 1205, Izmir University of Economics.

Goos M., Manning A. and Salomons A. (2009) ‘Job Polarization in Europe’, *The American Economic Review* 99: 58–63.

Goos M., Manning A. and Salomons A. (2014) ‘Explaining Job Polarization: Routine-Biased Technological Change and Offshoring’, *The American Economic Review* 104(8): 2509–2526.

Hanson G. H. and Harrison A. (1999) ‘Trade, Technology, and Wage Inequality’, *Industrial and Labor Relations Review* 52: 271–288.

Hatakenaka S. (2006) ‘Higher Education in Turkey for 21st Century: Size and Composition’, World Bank, Last Retrieved May 2016: 444607-1192636551820.

Heckman J. J. (1979) ‘Sample Selection Bias as a Specification Error’, *Econometrica* 47(1): 53–161.

Heitmueller A. (2007) ‘The Chicken or the egg?: Endogeneity in Labour Market Participation of Informal Carers in England’, *Journal of Health Economics* 26(3): 536–559.

Heitmueller A. and Inglis K. (2007) ‘The Earnings of Informal Carers: Wage Differentials and Opportunity Costs’, *Journal of Health Economics* 26(4): 821–841.

Hill M. A. (1983) ‘Female Labor Force Participation in Developing and Developed Countries-Consideration of the Informal Sector’, *The Review of Economics and Statistics* 65(3): 459–468.

Juhn C., Murphy K. M. and Pierce B. (1993) ‘Wage Inequality and the Rise in Returns to Skill’, *Journal of Political Economy* 101(3): 410–442.

Katz L. F. and Murphy K. M. (1992) ‘Changes in Relative Wages, 1963–1987: Supply and Demand Factors’, *The Quarterly Journal of Economics* 107(1): 35–78.

Kaya E. (2010) ‘Quantile Regression and the Gender Wage Gap: Is There a Glass Ceiling in the Turkish Labor Market?’, SSRN:http://ssrn.com/abstract=2320411.

Kizilirmak B. (2003) ‘Explaining Wage Inequality: Evidence From Turkey’, www.politics.ankara.edu.tr/eski/dosyalar/tm/SBF_WP_55.pdf.

Krusell P., Ohanian L. E., Rios-Rull J. and Violante G. L. (2000) ‘Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis’, *Econometrica* 68(5): 1029–1053.

Lee D. S. (1999) ‘Wage Inequality in the U.S. During the 1980s: Rising Dispersion or Falling Minimum Wage?’, *The Quarterly Journal of Economics* 114(3): 977–1023.

Lemieux T. (2002) ‘Decomposing Changes in Wage Distributions: A Unified Approach’, *Canadian Journal of Economics* 35(4): 646–688.

Lustig N., Lopez-Calva L. F. and Ortiz-Juarez E. (2013) ‘‘Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico’, *World Development* 44: 129–141.

Machin S. and Reenen J. V. (1998) ‘Technology And Changes In Skill Structure: Evidence From Seven Oecd Countries’, *The Quarterly Journal of Economics* 113(4): 1215–1244.

Marenzi A. and Pagani L. (2005) ‘The Impact of Elderly Parents on Labour Market Participation of Italian Women’, *Rivista di Politica Economica* 95(2): 155–190.

Martins P. S. and Pereira P. T. (2004) ‘Does Education Reduce Wage Inequality? Quantile Regression Evidence From 16 Countries’, *Labour Economics* 11(3): 355–371.

Maurer-Fazio M., Connelly R., Chen L. and Tang L. (2011) 'Childcare, Eldercare, and Labor Force Participation of Married Women in Urban China, 1982–2000', *The Journal of Human Resources* 46(2): 261–294.

Meschi E., Taymaz E. and Vivarelli E. (2011) 'Trade, Technology and Skills: Evidence From Turkish Microdata', *Labour Economics* 18: S60–S70.

Mincer J. A. (1974) 'Schooling, Experience, and Earnings', National Bureau of Economic Research.

Montgomery J. D. (1991) 'Social Networks and Labor-Market Outcomes: Toward an Economic Analysis', *The American Economic Review* 81(5): 1408–1418.

Naticchioni P., Ricci A. and Rustichelli E. (2008) 'Wage Inequality, Employment Structure and Skill-Biased Change in Italy', *LABOUR: Review of Labour Economics and Industrial Relations* 22 (s1): 27–51.

Naticchioni P., Ragusa G. and Massari R. (2014) 'Unconditional and Conditional Wage Polarization in Europe', IZA Discussion Paper 8465.

Oaxaca R. (1973) 'Male-Female Wage Differentials in Urban Labor Markets', *International Economic Review* 14(3): 693–709.

Pritchett L. (2001) 'Where Has All the Education Gone?', *The World Bank Economic Review* 15(3): 367–391.

Robbins D. J. (1996) 'Evidence on Trade and Wages in the Developing World', OECD Development Centre Working Papers 119.

State Planning Organization. (1967) '2nd Five Year Development Plan', http://www.kalkinma.gov.tr/Lists/Kalknma%20Planlar/Attachments/8/plan2.pdf. (in Turkish)

Tansel A. and Bodur F. (2012) 'Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis', *Review of Economic Development* 16(1): 107–121.

Telli C., Voyvoda E. and Yeldan E. (2007) 'Modeling General Equilibrium for Socially Responsible Macroeconomics: Seeking for the Alternatives to Fight Jobless Growth in Turkey', *METU Studies in Development* 33(2): 255–293.

World Economic Forum. (2013) 'Global Gender Gap Report 2013', Retrieved Sep 19, 2014 from http://www3.weforum.org/docs/WEF_GenderGap_Report_2013.pdf.

Yeldan E. (2006) 'Neoliberal Global Remedies: From Speculative-Led Growth to IMF-Led Crisis in Turkey', *Review of Radical Political Economics* 38: 193–213.

Yun M. (2005) 'A Simple Solution to the Identification Problem in Detailed Wage Decompositions', *Economic Inquiry* 43: 766–772.

Appendix A Econometric specification

1. Decompositions

1.1 Lemieux (2002)

Consider the wage determination equation for period t:

$$y_{it} = X_{it}\beta_t + u_{it}, \qquad \text{[A1]}$$

where y_{it} is the log wage of individual i at time period t; X_{it} is the vector representing skills (human capital), tasks (based on occupations) and other individual characteristics; β_t is the vector of estimated coefficients for returns to all elements in X_{it}; and u_{it} is the regression residual. There is a similar wage determination equation for period s.

The decomposition method developed by Lemieux (2002) unifies the residual imputation method proposed by Juhn *et al.* (1993; JMP hereafter) and the re-weighting factor method proposed by DiNardo *et al.* (1996; DFL hereafter). As a first step, for the Lemieux decomposition, we estimate a separate wage regression for each period (t and s) as given by equation [A1]. To compute the effect of changes in the prices of characteristics, following JMP, we construct a counterfactual wage vector that would prevail in period s (the base year; in our empirical analysis $s = 2002$) if the price of human capital were the same as they were in period t (in our empirical analysis $t = 2010$). To get the counterfactual wage vector, the coefficients from period s wage regression are replaced with coefficients from period t such as:

$$y_{is}^{a} = X_{is}\beta_t + u_{is}. \qquad [A2]$$

Once the counterfactual wages y_{is}^{a} are constructed, we can obtain the share of effect of changes in prices in total change in wage distribution by comparing any inequality measure (such as percentile wage gaps) for y_{is}^{a} and empirical wages for period s (y_{is}).

The second step of the method is to estimate the re-weighting factor which will then be used to modify the original sample weights and calculate the effect of changes in covariates. The concept of modifying sample weights, which are used to calculate sample statistics representative of the population, was originally proposed by DFL. This method attaches a new counterfactual weight to each individual to keep the distribution of characteristics constant and thereby makes it possible to account for changes in covariates. Their idea is to pool the samples of two periods and estimate a logit or probit model for the probability of being in the base year, using the same covariates as in the wage regression. The re-weighting function is defined as:

$$\psi_i = \frac{1 - P_{is}}{P_{is}} \frac{P_s}{1 - P_s}, \qquad [A3]$$

where $P_{is} = \Pr(period = s|x_{is})$ is the predicted probability that an individual in the pooled sample comes from the base year s conditional on covariates; and P_s is the unconditional probability that the observation is in period s. The new sample weights are then computed by multiplying the original sample weights for period s (w_{is}) with the re-weighting factor: $w_{is}^{a} = w_i\psi_i$. Following Lemieux (2002) notation, the counterfactual values of wages that can be generated by using the new sample weights are summarized in Table A1.

The difference between the distributional statistics (e.g. percentile wage gaps) of y_{is}^{a} and y_{is} (both using the original sample weights) gives the coefficient effect. This corresponds to the comparison of the distributions presented in the second and the first rows of the Table A1. On the other hand, the distribution statistics that are calculated by using y_{is}^{a} and the new re-weighted sample weights, w_{is}^{a}, contain the effect of covariates as well as the coefficients. Accordingly, the covariates effect is calculated as the difference between the distributional statistics that are obtained by using y_{is}^{a} with the original sample weights w_{is} and y_{is}^{a} with the new sample weights w_{is}^{a}. This corresponds to the comparison of the distributions presented in the fourth and the second rows of Table A1.

The final stage of the decomposition is to calculate the effect of changes in the residuals on the change in the wage distribution between the two periods; here Lemieux (2002) uses the residual imputation method which is provided by JMP. Consider the following linear form of the model for wage residuals:

$$u_{is} = p_s \eta_{is} + \varepsilon_{is}, \tag{A4}$$

where η_{is} is the unmeasured human capital, p_s is the return to the unmeasured human capital and ε_{is} is a random error term not linked with skills. Then, the variance of residuals can be calculated as:

$$\sigma_s^2 = p_s^2 \sigma_{\eta,s}^2 + \sigma_{\varepsilon,s}^2,$$

where $\sigma_{\eta,s}^2 = Var(\eta_{is})$ and $\sigma_{\varepsilon,s}^2 = Var(\varepsilon_{is})$. In this model, with the assumption that the distribution of unmeasured skills is constant ($\sigma_{\eta,s}^2 = \sigma_{\eta,s}^2 = \sigma_\eta^2$), and $\sigma_{\varepsilon,s}^2$ is zero or stable over time, changes in skill prices are the only source of any change in residual wage inequality:

$$\sigma_t^2 - \sigma_s^2 = (p_t^2 - p_s^2)\sigma_\eta^2.$$

The main disadvantage of this form (equation [A4]) is that the residuals are assumed to be a linear function of unmeasured skills. To understand the effect changes in unmeasured human capital prices have on wage distribution, JMP propose a more general setting in which a non-linear pricing scheme is applied:

$$u_{is} = p_s(\eta_{is}) + \varepsilon_{is}, \tag{A5}$$

where $p_s(\cdot)$ is a monotonic and continuous function; and for simplicity ε_{is} is assumed to be zero.

The JMP model, as given in equation [A5], is more general compared with the form in equation [A4] as it provides more flexibility by making it possible to generate any distribution of u_{is} from an arbitrary distribution of skills η_{is}. For instance, assume without loss of generality, that the η_{is} follows a uniform distribution over the interval of $[0, 1]$: $\eta_{is} = F_s(u_{is})$; where $F_s(\cdot)$ is the cumulative distribution function of u_{is}. Following which equation [A5] can be written as:

$$u_{is} = p_s(\eta_{is}) = F_s^{-1}(u_{is}),$$

where η_{is} can be interpreted as the rank of observation i in the distribution of residuals, whereas the non-linear skill pricing function $p_s(\cdot)$ is the inverse cumulative distribution of u_{is}.

Using the skill pricing function, the counterfactual wages in equation [A2] can be rewritten as:

$$y_{is}^a = X_{is}\beta_t + u_{is} = X_{is}\beta_t + p_s(\eta_{is}). \tag{A6}$$

The decomposition is finalized by replacing the residuals in period s with the residuals that would prevail if the skill pricing function was $p_t(\cdot)$ instead of $p_s(\cdot)$ such that:

$$y_{is}^a = X_{is}\beta_t + p_t(\eta_{is}) = X_{is}\beta_t + u_{is}^b, \tag{A7}$$

where $u_{is}^b = p_t(\eta_{is}) = F_t^{-1}F_s(u_{is})$ is the counterfactual residual for the observation i. To compute the counterfactual residuals, JMP suggest the following procedure: first, the rank $\eta_{is} = F_s(u_{is})$ is computed from the empirical residual distribution in period s and then the residual at the same rank in the residual distribution in period t is picked.[14]

The obtained counterfactual wages now can be used to decompose the changes in wage inequality. Extending JMP, they can also be combined with counterfactual weights to control for the distribution of covariates. In addition, having the counterfactual wage vectors, it is also possible to calculate any measure of inequality to see the effects of different factors on different parts of wage distribution.

Last but not the least an important feature of this decomposition method is that it is also possible to decompose the changes in wage residuals. As sample weights are used to calculate indices that are representing the population, comparison of variances of residual wages using the original sample weights and counterfactual weights provides information about how much of the change in residual distribution is due to the change in covariates and the change in the skill pricing function.

1.2 Firpo, Fortin and Lemieux (2009)

To get the detailed decomposition we estimate the Re-centred Influence Function (RIF) regression, proposed by Firpo *et al.* (2009). The RIF for the τ-quantile (q_τ) of log wages y_t, at time t, is given as (to keep notation simple we suppress the index i for the individual):

$$RIF(y_t, q_\tau) = q_\tau + [\tau - d_{t,\tau}]/f_{Y_t}(q_\tau), \quad \text{[A8]}$$

where $f_{Y_t}(q_\tau)$is the density distribution function of y_t computed at the quantile q_τ and $d_{t,\tau}$ is the dummy variable taking value one if $y_t \leq q_\tau$ and zero otherwise.

Following Fortin *et al.* (2011) we assume *RIF* to be a linear function of covariates in vector x_t, such that:

$$RIF(y_t, q_\tau) = X_t\beta_t^\tau + u_t, \quad \text{[A9]}$$

where β_t^τ is the vector of coefficients for the τ-quantile. Equation [A9] is also referred to as the unconditional quantile regression; it is estimated for each year, t and s.The changes in the τ-quantile can then be decomposed as:

$$q_{\tau,t} - q_{\tau,s} = \bar{X}_t\beta_t^\tau - \bar{X}_s\beta_s^\tau = \underbrace{(\bar{X}_t - \bar{X}_s)\beta_s^\tau}_{\text{composition effect}} + \underbrace{\bar{X}_s(\beta_t^\tau - \beta_s^\tau)}_{\text{coefficient effect}}. \quad \text{[A10]}$$

The first term on the right hand side of equation [A10] reflects the change in the distribution of observed covariates (composition effect), and the second term reflects the changes in the regression coefficients (returns to the observed covariates). Equation [A10]

Table A1. Counterfactual distributions

Variable	Weight	Resulting distribution
y_{is}	w_{is}	Distribution at period s
y_{is}^a	w_{is}	Distribution at period s with β of period t
y_{is}	w_{is}^a	Distribution at period s with covariates of period t
y_{is}^a	w_{is}^a	Distribution at period s with covariates and β of period t

Notes: The base year is $s.w_{is}$ are the original sample weights from the year s.
$w_{is}^a = w_i\psi_i$ are the counterfactual weights, where ψ_i is the re-weighting factor.

can then be used to obtain the detailed decomposition, similar to the Blinder–Oxaca method, such as:

$$q_{\tau,t} - q_{\tau,s} = \sum_{k=1}^{K} (\overline{x}_{k,t} - \overline{x}_{k,s}) \widehat{\beta}^{\tau}_{k,s} + \sum_{k=1}^{K} \overline{x}_{k,t} (\widehat{\beta}^{\tau}_{k,t} - \widehat{\beta}^{\tau}_{k,s}), \qquad [A11]$$

where $\overline{x}_{k,t}$ is the mean of the k^{th} component of the vector X_t, and $\widehat{\beta}^{\tau}_{k,t}$ is the corresponding coefficient.

The *RIF* regressions can be biased as the assumption of linearity (equation [A9]) holds true only locally. To correct for the *specification error* the *RIF* regression is combined with the DFL re-weighting function. This requires estimating the *RIF* regression for period s with the re-weighting function such that the covariates of period s have the same distribution as in period t. Let the re-weighted period s covariates be X^c_s, and the estimated coefficients be $\widehat{\beta}^{c\tau}_s$. The composition effect can then be further decomposed as:

$$\underbrace{(\bar{X}^c_s - \bar{X}_s)\hat{\beta}^{\tau}_s}_{\text{pure composition effect}} + \underbrace{\bar{X}_s(\hat{\beta}^{c\tau}_s - \hat{\beta}^{\tau}_s)}_{\text{specification error}}, \qquad [A12]$$

where the first term of equation [A12] is the pure composition effect and the second term is the *specification error*. Similarly the coefficient effect is decomposed into pure coefficient effect and the re-weighting error component, given as:

$$\underbrace{\bar{X}_t(\hat{\beta}^{\tau}_t - \hat{\beta}^{c\tau}_s)}_{\text{pure coefficient effect}} + \underbrace{(\bar{X}_t - \bar{X}^c_s)\hat{\beta}^{c\tau}_s}_{\text{reweighting error}}, \qquad [A13]$$

where the first term of equation [A13] is the pure coefficient effect and the second term is the *re-weighting error* term.

Further, within the decomposition analysis we use the normalization suggested by Yun (2005) such that the results are not sensitive to the omitted category for the categorical variables.

In our empirical application we use the STATA ado file 'RIFREG' written by Firpo *et al.* (2009), downloaded from: http://faculty.arts.ubc.ca/nfortin/datahead.html.

2. Selection bias

To estimate the correction term a multinomial logit model (Dubin and McFadden, 1984) is estimated separately for men and women. An individual i is characterized as having three options; these options are different for men and women. The sample of women is grouped into the following three categories: $fw_i = 2$ if working in the wage sector; $fw_i = 1$ if she is an unpaid family worker; and $fw_i = 0$ if she is economically inactive (non-participant). The categories for men are: $fw_i = 2$ if working in the wage sector; $fw_i = 1$ if self-employed; and $fw_i = 0$ if he is economically inactive (non-participant). The probability of being in group j is obtained by:

$$P_{ij} = P(fw_i = j|Z_i) = \frac{\exp(Z_i\alpha_j)}{1 + \sum_{j=0}^{2}\exp(Z_i\alpha_j)} \quad \text{for} \quad j = 0, 1, 2, \qquad \text{[A14]}$$

where the numerator is normalized to 1 for $j = 0$; $P_{ij} = P(fw_i = j|Z_i)$ is the conditional probability of individual i belonging to sector j, conditional on Z_i, a vector of covariates; and α_j is a column vector of coefficients corresponding to the j^{th} sector. The predicted probabilities from the multinomial logit model (estimated separately for men and women) are then used to construct a selection correction term, λ_{ij}, by using the formula provided by Hill (1983):

$$\lambda_{ij} = \left(\frac{6}{\pi^2}\right)(-1)^{J+1}\left[\frac{J-1}{J}\ln\widehat{P}_{ij}\sum_{k\neq j}\left(\frac{\widehat{P}_{ik}\ln\widehat{P}_{ik}}{1-\widehat{P}_{ik}}\right)\right] \quad \text{for} \quad j = 0, 1, 2. \qquad \text{[A15]}$$

In the second step, augmented wage equations are estimated separately for men and women by including correction term as an additional regressor:

$$y_{ij} = X_{ij}\beta_j + \theta_j\lambda_{ij} + \varepsilon_{ij}, \qquad \text{[A16]}$$

where y_{ij} is the log wage of individual i in sector j; X_{ij} is the vector of sector specific individual characteristics; β_j is the vector of returns to characteristics in sector j; θ_j is the unknown coefficient related to the selection correction term; and ε_{ij} is the independent residual term. The augmented wage equation estimation results are then used in decomposition of the change in wage distributions of men and women over time.[15]

Selection procedure requires the availability of valid instruments, i.e., variables which affect the labour market status but do not affect the wages. This means there should exist at least one element in vector Z (equation [A14]) which is excluded from vector X (equation [A16]). For identification, the following variables are included in the multinomial logit model. Three dummies to capture the labour market status of other members of the household: the presence of wage workers, the presence of self-employed and the presence of unpaid family workers in the household; a dummy for the head of the household; a dummy for whether the worker lives in rural or urban area; and finally, only for women, whether or not there is a grandmother present in the household.

Justification for using the labour market status of other members of the household in the participation equation comes from the well documented evidence on the importance of social networks in determining the labour market outcomes (Calvo-Armengol and Jackson, 2004; Montgomery, 1991). Presence of other household members in a specific sector is likely to influence the probability of employment within the sector. For instance, Calvo-Armengol and Jackson (2004) argue that, when information about jobs arrives individuals who are unemployed and directly hear of a job use the information to obtain a job; on the other hand, individuals who are already employed, depending on whether the job is more attractive than their current job, might take the job or else might pass information to one (or more) of their direct connections in the network. In accordance with this argument, it is possible that a wage worker (or a self-employed) may share the information about a waged employment (or self-employment) opportunity with his or her household members who are unemployed or self-employed.

Increasingly there is evidence that the labour supply decisions are more dependent on individuals' 'role in the household' and 'composition of the household' rather than individuals' 'sex' (Cunningham, 2001). The head of the household must enter the labour force as she or he is often perceived as the main bread earner; this status within the household is unlikely to have any effect on the wage of the individual.

The presence of other adults in the household can aid in the labour force participation (LFP) of women, by providing child care; however, it can also be detrimental, as other adults may be either considered as labour market substitutes for women or additional care responsibilities for women. In this case a dummy for the presence of a grandmother in the household can aid in LFP of women by providing child care, but at the same time can be a deterrent for LFP as an elderly relative in the household may be an additional care responsibility for the women. For instance, using US data, Ettner (1996) shows that caregiving responsibilities for elders is a significant detriment to LFP for women. Similar results are echoed using the UK data by Heitmueller and Inglis (2007) and Heitmueller (2007). In case of China, on the other hand, Maurer-Fazio *et al.* (2011) found that co-residence with elders had a positive effect on the LFP of women, especially married women. Similarly Marenzi and Pagani (2005) found that in Italy, the presence of elderly parents in the house can be beneficial for the LFP of women, especially those with pre-school children.

Appendix B Additional tables

Table B1. Descriptive statistics for women

	2002			2010		
	Economically inactive	**Unpaid family worker**	**Wage worker**	**Economically inactive**	**Unpaid family worker**	**Wage worker**
Education dummies						
No formal education	0.22	0.33	0.04	0.25	0.32	0.06
Primary school	0.58	0.63	0.28	0.47	0.55	0.22
Secondary school	0.07	0.02	0.07	0.10	0.07	0.08
High school	0.08	0.02	0.17	0.08	0.03	0.13
Vocational high school	0.04	0.01	0.11	0.05	0.02	0.11
University	0.02	0.00	0.32	0.04	0.01	0.40
Tenure	–	19.55	6.42	–	15.27	5.07
	–	(0.15)	(0.07)	–	(0.11)	(0.05)
Presence of children in the household						
Age $\leq$ 4 years	0.33	0.31	0.18	0.33	0.27	0.18
5 $\leq$ Age < 11 years	0.41	0.42	0.29	0.39	0.38	0.27
11 $\leq$ Age < 15 years	0.22	0.28	0.17	0.21	0.25	0.16
Other members of the household						
Presence of wage workers	0.50	0.12	0.64	0.57	0.21	0.64
Presence of self-employed	0.21	0.93	0.10	0.16	0.87	0.10
Presence of unpaid family workers	0.05	0.55	0.02	0.04	0.42	0.02
Marital status dummy						
Never married	0.10	0.13	0.38	0.10	0.09	0.35
Married	0.83	0.85	0.56	0.83	0.89	0.57
Divorced	0.01	0.002	0.04	0.02	0.01	0.06
Widowed	0.05	0.01	0.02	0.05	0.01	0.02
Age						
20–24	0.16	0.16	0.24	0.13	0.10	0.18
25–34	0.32	0.26	0.43	0.30	0.23	0.42
35–44	0.24	0.22	0.24	0.24	0.26	0.28
45 and above	0.27	0.35	0.08	0.32	0.41	0.11
Member of social security system	–	0.01	0.73	–	0.06	0.77
Head of household dummy	0.07	0.00	0.10	0.08	0.00	0.10
Grandmother	0.05	0.13	0.05	0.06	0.13	0.05
Rural	0.30	0.92	0.18	0.23	0.86	0.13
Observations	60,811	7,787	9,442	97,594	15,220	20,080

Note: Standard errors for continuous variables are reported in (). Sample weights are used.

Table B2. Descriptive statistics for men

	2002			2010		
	Economically inactive	**Self-employed**	**Wage worker**	**Economically inactive**	**Self-employed**	**Wage worker**
Education dummies						
No formal education	0.06	0.10	0.02	0.09	0.09	0.03
Primary school	0.52	0.70	0.43	0.39	0.62	0.35
Secondary school	0.12	0.09	0.14	0.18	0.13	0.16
High school	0.14	0.06	0.15	0.13	0.07	0.12
Vocational high school	0.09	0.03	0.12	0.09	0.06	0.13
University	0.08	0.02	0.15	0.11	0.04	0.20
Tenure	–	17.81	8.00	–	15.07	6.20
	–	(0.12)	(0.04)	–	(0.08)	(0.03)
Presence of children in the household						
Age $\leq$ 4 years	0.27	0.32	0.35	0.25	0.28	0.34
5 $\leq$ Age < 11 years	0.35	0.43	0.42	0.33	0.39	0.38
11 $\leq$ Age < 15 years	0.22	0.27	0.21	0.21	0.25	0.19
Other members of the household						
Presence of wage workers	0.30	0.12	0.29	0.34	0.18	0.34
Presence of self-employed	0.18	0.07	0.08	0.13	0.08	0.08
Presence of unpaid family workers	0.06	0.48	0.03	0.04	0.44	0.03
Marital status						
Never married	0.40	0.05	0.18	0.39	0.06	0.22
Married	0.58	0.93	0.81	0.58	0.92	0.77
Divorced	0.01	0.01	0.01	0.02	0.01	0.01
Widowed	0.01	0.01	0.002	0.003	0.01	0.001
Age						
20–24	0.30	0.03	0.12	0.22	0.02	0.11
25–34	0.33	0.24	0.43	0.33	0.19	0.40
35–44	0.20	0.29	0.29	0.21	0.29	0.30
45 and above	0.17	0.43	0.15	0.23	0.49	0.18
Member of social security system	–	0.42	0.73	–	0.39	0.78
Head of household dummy	0.50	0.89	0.76	0.51	0.85	0.71
Rural	0.36	0.66	0.24	0.27	0.56	0.17
Observations	11,426	14,007	32,771	17,704	23,869	63,077

Notes: Standard errors for continuous variables are reported in (). Sample weights are used.

Table B3. Occupation categories: Sample proportions for the wage earners

	Women		Men	
Sample proportions	**2002**	**2010**	**2002**	**2010**
Abstract	0.39	0.39	0.22	0.23
Legislators, senior officials and managers	0.04	0.03	0.04	0.05
Professionals	0.21	0.21	0.10	0.10
Technicians and associate professionals	0.14	0.14	0.08	0.08
Service	0.45	0.50	0.39	0.41
Clerks	0.20	0.20	0.09	0.09
Service workers and shop and market sales workers	0.11	0.14	0.16	0.16
Elementary occupations	0.14	0.16	0.14	0.16
Routine	0.15	0.11	0.37	0.35
Craft and related trades workers	0.11	0.05	0.22	0.19
Plant and machine operators and assemblers	0.04	0.06	0.15	0.16

Note: Sample weights are used.

Table B4. Industry dummies: Sample proportions for the wage earners

	2002 Women	**2010 Women**	**2002 Men**	**2010 Men**
Industry categories				
Agriculture, Hunting, Forestry and Fishing (Base category)	4.7	3.6	1.9	2.2
Mining and Quarrying	0.2	0.1	1.2	1.1
Manufacturing	25.0	20.3	26.3	25.4
Electricity, Gas and Water	0.3	0.2	1.3	1.1
Construction	1.3	1.4	9.8	8.8
Wholesale and Retail Trade and Restaurants and Hotels	13.6	16.0	18.9	19.6
Transport, Storage and Communication	3.0	3.1	7.7	7.2
Financing, Insurance, Real Estate and Business Services	7.8	12.0	4.6	8.7
Community, Social and Personal Services	44.0	43.3	28.3	25.9

Note: Sample weights are used.