For Official Use ECO/CPE/WP1(2011)18/ANN2

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

06-Oct-2011

English - Or. English

ECONOMICS DEPARTMENT
ECONOMIC POLICY COMMITTEE

ECO/CPE/WP1(2011)18/ANN2
For Official Use

Working Party No. 1 on Macroeconomic and Structural Policy Analysis

INCOME DISTRIBUTION

ANNEX 2. THE DRIVERS OF LABOUR EARNINGS INEQUALITY: AN ANALYSIS BASED ON CONDITIONAL AND UNCONDITIONAL QUANTILE REGRESSIONS

This document has been prepared by the Economics Department for the meeting of Working Party No. 1 of the Economic Policy Committee.

Contact:
Isabell Koske, tel: (33-1) 45 24 17 92; isabell.koske@oecd.org
Jean-Marc Fournier, tel: (33-1) 45 24 92 03;jean-marc.fournier@oecd.org

JT03308468

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

TABLE OF CONTENTS

Tables

Figures

Boxes

Related appendix: Further Details on Data Set and Estimation Results (ECO/CPE/WP1(2011)18/ANN2/APP)

ANNEX 2. THE DRIVERS OF LABOUR EARNINGS INEQUALITY: AN ANALYSIS BASED ON CONDITIONAL AND UNCONDITIONAL QUANTILE REGRESSIONS

A2.1. Introduction

1. Despite some common trends, countries differ considerably with respect to both the extent and timing of the rise in earnings inequality over the past decades. In addition, there are notable cross-country differences in the current level of earnings inequality. Since all OECD economies face the same global environment and have essentially benefited from the same technological advances, globalisation and skill-biased technological change should have led to broadly similar shifts in labour demand. Even though countries have differed with respect to supply shifts, a relative-supply-demand shift story is unlikely to fully account for the marked cross-country differences in both the level and the evolution of labour earnings inequality, which hints at a possible role for differences in policy and institutional settings.

2. To shed further light on these issues and in particular on the role of structural policies in shaping the distribution of earnings, this Annex explores the determinants of labour earnings inequality based on household survey data from 31 countries. The empirical analysis makes use of quantile regressions, which allow estimating the effect of the potential determinants on all parts of the earnings distribution, and are thus better suited to answer questions about the drivers of earnings inequality than standard least squares techniques that only allow estimating effects on mean earnings. The regression results are used to decompose cross-country differences in the level of earnings inequality into differences in population characteristics (*e.g.* education) and differences in the returns to these characteristics (*e.g.* returns to education). Similarly, earnings differences between specific subgroups of workers (in particular men/women and natives/immigrants) are broken up into their major components.

3. The following main results emerge from the analysis:

- Bearing in mind that household surveys are not entirely comparable, labour earnings inequality appears to vary considerably across countries. Among full-time employees it is highest in Brazil, the United States, Portugal and Canada, while Belgium, Denmark, Finland and the Czech Republic are the most equal countries.

- The number of hours worked is an important determinant not only of an individual’s earnings but also of earnings inequality among the working population. In almost all countries the reward for working one additional hour is highest for workers at the lower end of the earnings distribution, possibly reflecting the role of overtime pay. This means that a general decrease in the number of hours worked – triggered, for example, by an economic recession – would particularly hurt low-income workers. The numbers of hours worked appears to play a key role in shaping both the within-country distribution of earnings and cross-country differences in earnings inequality.

- In most countries, the returns to an additional year of work experience are higher at lower quantiles, suggesting that work experience plays a larger role in lower-paid jobs and/or that seniority pay is more prevalent in these types of jobs. As a consequence, policies that reduce the likelihood of career breaks or their length (*e.g.* improvements in the availability of formal childcare) may help reduce the dispersion of earnings.

- The link between education and earnings inequality is ambiguous from a theoretical point of view. *First*, *via* a composition effect a rise in the share of highly-educated (high-wage) workers raises earnings inequality up to a certain point, but will then lower it as fewer low-education (low-wage) workers remain. *Second*, a rise in the share of highly-educated workers alters the returns to education, with the direction of the change depending on many factors such as the substitutability between low- and high-education workers. The empirical evidence indicates that policies to increase upper-secondary graduation rates (*e.g.* by providing support to pupils at risk in order to reduce drop outs) should reduce income inequality. Whether similar benefits can be expected from reforms that encourage more students to pursue tertiary studies is unclear and depends on the relative magnitudes of the different offsetting effects.

- For those at the bottom of the earnings distribution, being on a temporary rather than on a permanent contract implies lower labour earnings, even controlling for other individual characteristics. Being self-employed also generally entails an earnings penalty (relative to being employed with a permanent work contract) at lower quantiles. For higher quantiles, the type of contract and work status typically matter less.

- For the majority of countries for which data on union membership are available the results indicate that unions tend to compress the wage distribution.

- Labour earnings vary across different sectors of the economy, but a shift in the sector composition does in general not have a large impact on the overall distribution of earnings. Consequently, the contribution of cross-country differences in the sector composition to cross-country differences in earnings inequality also tends to be fairly small. The only exceptions are agriculture/hunting/forestry/fishing, hotel/restaurants, other community, social and personal service activities/others and financial intermediation, with a rise in the shares of these four sectors being associated with somewhat higher earnings inequality.

- A higher share of the public sector in total employment is found to be associated with lower earnings inequality in a large majority of OECD countries. However, cross-country differences in the size of the public sector or in public/private sector wage structures do not seem to play an important role in explaining cross-country differences in inequality.

- In all countries considered, women earn less than men. Policies to reduce gender differences in working hours (*e.g.* improvements in the availability of, and access to, formal childcare), and in the choice of occupation and sector (*e.g.* gearing curricula and teaching material at school so as to avoid gender stereotyping) could lead to more equal labour market outcomes of men and women. Policies to reduce gender discrimination (*e.g.* empowering specialised bodies to investigate employers and take legal action against those who engage in discriminatory practices) might also be helpful in this regard.

- In about two-thirds of the countries for which data on the country of birth are available, immigrants earn significantly less than natives. Targeted policies such as language courses and transparent systems of recognising foreign qualifications may help to reduce the gap in labour market performance. Decomposing the average earnings gap between immigrants and natives into its major components points to considerable cross-country heterogeneity. This heterogeneity is not explored further here but could reflect cross-country differences in both policy settings and the characteristics of immigrants.

4. This Annex is structured as follows. Section A2.2 first presents the empirical methodology, focusing on the specification of the earnings equation, the estimation and interpretation of conditional and unconditional quantile regressions and the methods used to decompose differences in earnings inequality across countries or across population subgroups. Section A2.3 then briefly discusses the benefits and drawbacks of the dataset that is employed in the empirical analysis before presenting the estimation results. The Annex is accompanied by an Appendix, which provides country-specific quantile regression results and discusses the household survey data that underlie the empirical analysis.

A2.2. The empirical methodology

A2.2.1. The earnings equation

5. The empirical analysis makes use of household survey data for 31 countries.[1] It relates individual labour earnings to personal and employer characteristics, focusing on individuals aged between 15 and 64 who work either part-time or full-time and have positive labour earnings during the reference year.[2] A description of the dataset is provided in the Appendix.

6. The choice of explanatory variables is inspired by the seminal work of Mincer (1958, 1974), who developed a parsimonious model of labour earnings, first using only schooling and later also age and working time as explanatory factors. Numerous supplementary variables have since been added to earnings functions, including gender, ethnicity and union membership, among others (*e.g.* Polachek, 2007). Following this literature, this paper starts by estimating a baseline model which relates the logarithm of an individual's gross labour earnings to the logarithm of working hours, gender, age and age squared, and the highest education level attained. The level of education is captured by two dummy variables, the first one being equal to one for individuals who have *at least* finished upper-secondary education, and the second one being equal to one for individuals who have finished tertiary education. Hence, the coefficient on the first dummy variable gives the impact of an upper-secondary or post-secondary non-tertiary education relative to lower-secondary education or less, and the coefficient on the second dummy variable gives the impact of tertiary education relative to upper-secondary or post-secondary non-tertiary education.

7. Several additional drivers of labour earnings are of interest but are excluded from the baseline because they exist only for a subset of countries and/or cause potential endogeneity bias.[3] These are dummy variables for the sector of employment and the occupation, the number of years of work experience, the number of years of education, and dummy variables for having a temporary as opposed to a permanent work contract, for being self-employed, for being member of a union, for working in the public sector, for having foreign citizenship, for being born in a foreign country, and for having a PhD. These

1. Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, and the United Kingdom (European Union Statistics on Income and Living Conditions); Australia (Household Income and Labour Dynamics in Australia Survey); Canada (Survey of Labour and Income Dynamics); Korea (Korean Labour and Income Panel); Japan (Japan Household Panel Survey); Switzerland (Swiss Household Panel) the United States (Panel Study of Income Dynamics); Brazil and Israel (Luxembourg Income Study)

2. Individuals with zero or negative earnings are excluded from the analysis since the earnings variable is expressed in logarithmic terms (see below). Although rare in the datasets, negative earnings may occur if self-employed individuals make a loss on their business. In total, this concerns less than one percent of all observations.

3. Adding these variables to the baseline specification hardly alters the results obtained for the baseline variables.

variables are added on top of the baseline in a number of alternative specifications, the details of which are set out in Section A2.3.

A2.2.2. Going beyond mean effects with quantile regressions

8. The impact on earnings of the variables listed above is likely to differ across individuals. For example, a tertiary degree may be more valuable for high-income workers as their jobs require such an education, whereas it goes beyond the needs of most jobs of low-income workers (*e.g.* Hartog *et al.*, 2001). Standard OLS techniques ignore this heterogeneity and only provide an estimate of the mean effect of a given variable. As this would severely weaken the analysis (Koenker and Bassett, 1978), this paper makes use of two alternative techniques that allow estimating the impact of explanatory variables on different parts of the income distribution (nonetheless simple OLS results are also shown for the sake of comparison). These are the conditional quantile regression (hereafter CQR) technique proposed by Koencker and Bassett (1978) and the unconditional quantile regression (hereafter UQR) technique proposed by Firpo *et al.* (2007*a*, 2009). While the former has been widely used in the literature, the latter is fairly new and applications are thus still scarce. In general, the estimation of the effects of a given set of variables on the distribution of another variable is still an active area of research and no preferred method has yet emerged from the literature. The choice made here of the methodology by Firpo *et al.* (2007*a*, 2009) over alternative techniques such as the nonparametric approach proposed by Rothe (2010) is mainly motivated by its ease of computation.

9. Conditional quantile regressions focus on the conditional quantile of an individual, which is his/her position in a virtual distribution in which all individuals are assumed to have the same observed characteristics. For example, if individuals would differ only with respect to their education level, the conditional quantile of a low-educated person would be his/her earnings quantile among all low-educated individuals, whereas the conditional quantile of a highly-educated person would be his/her earnings quantile among all high-educated persons. Unconditional quantile regressions, by contrast, focus on the unconditional quantile of an individual, which is his/her earnings quantile in the overall earnings distribution, abstracting from (*i.e.* not controlling for) observed and unobserved characteristics. In the example above, the unconditional quantiles of the two individuals with respectively low and high education would be their earnings quantiles among all individuals in the population.

10. Given the different focus of the two approaches, the types of questions they can answer differ. Conditional quantile regressions – which have often been simply referred to as ‘quantile regressions’ – provide an estimate of the return to a certain characteristic (such as having a tertiary degree), where the return varies across individuals based on the conditional quantile into which they fall (Koenker and Hallock, 2001). The method can thus be used to answer questions such as: what is the impact on an individual’s earnings of increasing the education level by one year, holding everything else constant? The technique assumes in particular, that the conditional quantile of an individual remains the same when his/her characteristics change. Since this assumption may well not hold in practice, the results of CQR must be interpreted with caution (Koencker, 2005).

11. Unconditional quantile regressions, by contrast, allow estimating the effect of a small change in workers’ characteristics on each quantile of the *overall* distribution. They thus provide answers to questions such as: What is the impact on median earnings (or the earning of any particular quantile) of increasing everybody’s education by one year, holding everything else constant?[4] Since the unconditional

4. Firpo *et al.* (2009) extend the interpretation to dummy variables. For example, the coefficient on a dummy variable that takes value zero if the person works in the private sector and value one otherwise can be interpreted as the impact on earnings of raising the probability to work in the public sector by one percentage point.

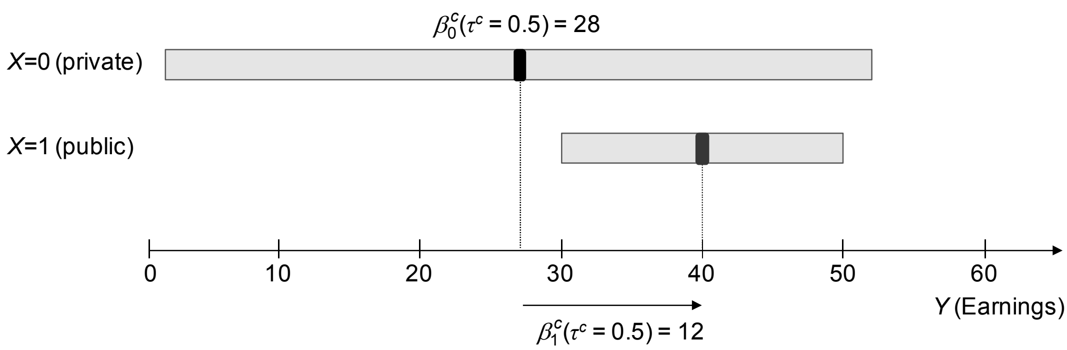
quantile of an individual is simply the share of individuals in the sample population whose earnings are lower than the earnings of the individual of interest, the results of UQRs are easier to interpret than those of CQRs. Since UQRs allow assessing the impact of a particular variable on *overall* earnings inequality, they are also more suitable than CQRs in the context of this project and are thus used as the baseline method. CQRs are computed for two purposes. *First*, this widely used method remains a robustness check *if* the unconditional quantile is likely to remain quite close to the quantile conditional on the variable of interest. Many results are indeed relatively similar with these two methods. *Second*, it provides insights about the comparison of the dispersion of income within different groups, which helps to understand the mechanisms at work. Box A2.1 illustrates the interpretation of CQR and UQR results with the help of a simple example.

Box A2.1. How to interpret the results of conditional and unconditional quantile regressions

To illustrate the interpretation of conditional and unconditional quantile regressions, assume that there are only two explanatory variables, a constant and a dummy variable *X*, which takes value one if an individual is working in the public sector and zero otherwise. Assume further that earnings are higher on average and less dispersed in the public than in the private sector. In Figures A2.1 and A2.2 below, the grey rectangles show the distribution of earnings in the two sectors, with the length of the rectangles indicating the range of earnings and their thickness – assumed here to be the same across the whole distribution for simplicity – indicating the number of persons with a certain earnings level.

The coefficient $\beta_1^c(\tau^c)$ on the public sector dummy obtained from a *conditional quantile regression* gives the change in earnings associated with moving from a private to a public sector job, assuming that the position of the individual among all individuals with the same characteristics does not change. For example, if the individual had the median earnings among all private sector workers before the job change (τ^c = 0.5), he will have the median earnings among all public sector workers after the job change, so that his earnings rise by 12 units in Figure A2.1. The constant β_0^c (τ^c) obtained from a CQR gives the quantile of the individual among the subsample of individuals with a 0-value for the dummy variable, *i.e.* among all private sector workers, which here is 28 for the median.

Figure A2.1. **Interpreting conditional quantile regressions**

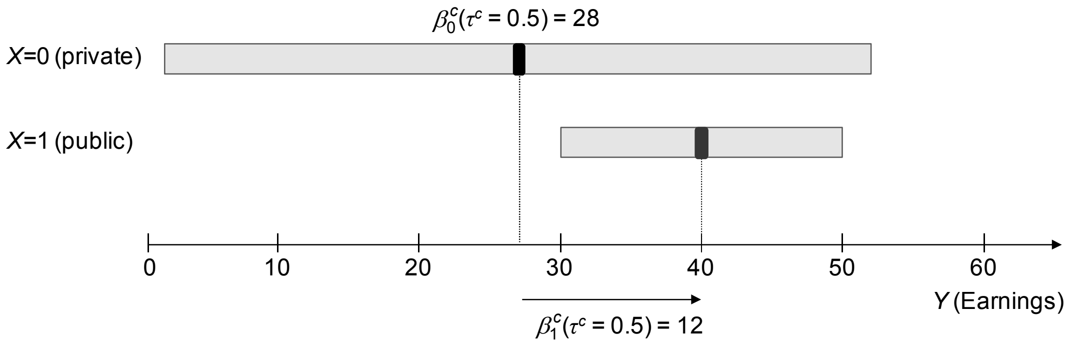


The CQR results can also be used to draw conclusions about the dispersion of earnings within certain subgroups of individuals. In the example above, the coefficient of the dummy variable decreases along the earnings distribution. This reflects that earnings are less dispersed among public sector employees than among private-sector ones.[1] While conclusions about the dispersion of earnings among certain subgroups of workers could also be derived from simple summary statistics such as the variance, CQRs have the advantage that they can control for other determinants of earnings.

The coefficient $\beta_1^{uc}(\tau^{uc})$ on the public sector dummy obtained from an *unconditional quantile regression* gives the change in a certain earnings quantile of the observed distribution – say, the median – associated with an increase in the share of public sector employment by 1 percentage point. As shown in Figure A2.2, the size of the two sectors is affected by such a change with the public sector increasing and the private sector shrinking in size.[2] As there are now more individuals in the economy that earn the higher public sector earnings, the median earnings of the entire population increase. In the example shown in the figure below, median earnings rise by 3, from 33 to 36. The constant cannot easily be interpreted in the case of UQRs and is therefore not shown in Figure A2.2.

Box A2.1. How to interpret the results of conditional and unconditional quantile regressions *(cont.)*

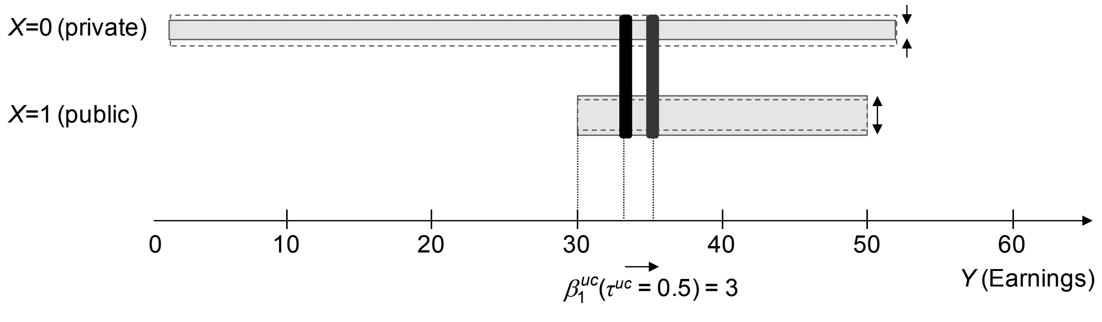
Figure A2.1. **Interpreting conditional quantile regressions**



The CQR results can also be used to draw conclusions about the dispersion of earnings within certain subgroups of individuals. In the example above, the coefficient of the dummy variable decreases along the earnings distribution. This reflects that earnings are less dispersed among public sector employees than among private-sector ones.[1] While conclusions about the dispersion of earnings among certain subgroups of workers could also be derived from simple summary statistics such as the variance, CQRs have the advantage that they can control for other determinants of earnings.

The coefficient $\beta_1^{uc}(\tau^{uc})$ on the public sector dummy obtained from an *unconditional quantile regression* gives the change in a certain earnings quantile of the observed distribution – say, the median – associated with an increase in the share of public sector employment by 1 percentage point. As shown in Figure A2.2, the size of the two sectors is affected by such a change with the public sector increasing and the private sector shrinking in size.[2] As there are now more individuals in the economy that earn the higher public sector earnings, the median earnings of the entire population increase. In the example shown in the figure below, median earnings rise by 3, from 33 to 36. The constant cannot easily be interpreted in the case of UQRs and is therefore not shown in Figure A2.2.

Figure A2.2. **Interpreting unconditional quantile regressions**

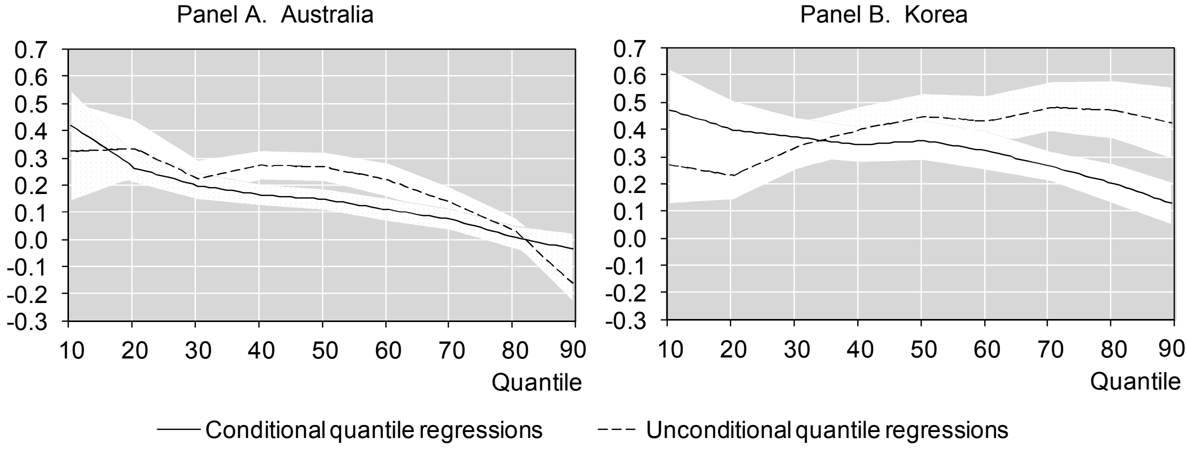


Turning to the analysis carried out in this paper, Figure A2.3 shows the output of the two types of quantile regressions for two countries, Australia and Korea. In addition to the public sector dummy, the specifications include age, age squared, gender, education and hours worked as explanatory variables. When interpreting the results of the CQRs, workers to the left of the figure (*i.e.* workers with a lower conditional quantile) are those, whose earnings are reduced by unobserved characteristics that are not controlled for in the estimation such as ability or work experience. For both countries, these types of workers benefit more from transferring to a public sector job than workers in a high conditional quantile, as indicated by the downward sloping solid lines in the figure.

Box A2.1. How to interpret the results of conditional and unconditional quantile regressions *(cont.)*

Figure A2.3. **Conditional and unconditional quantile regression estimates of the impact on earnings from working in the public sector**

Effect of working for the public sector on log earnings



Note: The shaded areas indicate the 95% confidence interval around the estimates.

The interpretation of the UQRs is very different. Workers to the left (right) of the figure are those with low (high) earnings. The downward sloping dashed line for Australia means that a 1 percentage point increase in the share of the public sector in total employment raises earnings more at the bottom than at the top, whereas the opposite holds in Korea where the dashed line is upward sloping for most quantiles. In other words, a rise in the share of public sector employment reduces earnings inequality in Australia but raises it in Korea.[3] Note that these are partial equilibrium effects however.

While the CQR results are thus very similar for the two countries, the UQR results are exactly the opposite. To better understand the UQR results, it is useful to disentangle the effect of a rise in the public sector employment share into two separate components:

- Assuming for simplicity that earnings in the two sectors are characterised by the same means but differing variances (with earnings being less dispersed in the public sector), starting from an economy where all workers are employed by the private sector, an increase in public sector employment lowers overall earnings inequality in a monotonous way.
- Assuming instead for simplicity that earnings in the two sectors are characterised by zero variances but differing means, starting from an economy where all workers are employed by the private sector, an increase in public sector employment first raises overall earnings inequality (as suddenly not all persons have the same earnings anymore so that the variance of earnings in the total economy becomes strictly positive), but eventually reduces it as fewer private sector employees remain. Once all workers are employed by the public sector, the variance of earnings goes back to zero. The relationship between the public sector employment share and earnings inequality is thus inverted U-shaped.

The effect of a rise in the public sector employment share on earnings inequality thus depends on the initial size of the public sector and the means and variances of earnings in the two sectors. Measuring earnings inequality simply by the log variance of earnings in the total population *Var*, the level of earnings inequality can be expressed as follows (Robinson, 1976):

$$Var = W_0 * Var_0 + W_1 * Var_1 + W_0 (Y_0 - Y)^2 + W_1 (Y_1 - Y)^2$$

where W_0 and W_1 denote the employment shares of the two sectors, Y_0 , Y_1 and Y denote the log mean earnings in the two sectors and the entire economy, and Var_0 and Var_1 denote the log variances of earnings in the two sectors. The variance of earnings in the total economy *Var* peaks when the share of public sector employment W_1 is equal to:

$$\hat{W}_1 = \frac{Var_1 - Var_0}{2(Y_1 - Y_0)^2} + \frac{1}{2}$$

Box A2.1. How to interpret the results of conditional and unconditional quantile regressions *(cont.)*

In the case of Korea, the variances of earnings in the two sectors are very similar, while the earnings of public sector employees are much higher than those of private sector employees. Data from the Korean Labor and Income Panel Study suggest that inequality would peak when the share of public employment reaches around 40% of the total. Since the actual share is much lower (around 10%), an increase in the proportion of public sector employees is associated with a rise in earnings inequality as indicated by the upward-sloping dashed line in Panel B of Figure A2.3. In the case of Australia the dispersion of earnings among public sector employees is much smaller than among private sector employees, so that the formula implies a marginal negative link between inequality and the share of public sector employees, whatever the level of this share. This is consistent with the downward-sloping dashed line in Panel A of Figure A2.3

1. This is strictly true when there are no control variables. As soon as control variables are added (*e.g.* a dummy variable that takes value one if an individual is highly educated and value zero otherwise) the interpretation is slightly different. Let's assume for the sake of demonstration that earnings depend not only on public sector and education dummies, but also on an unobserved determinant such as performance related pay. If the coefficient on the public sector dummy decreases along the earnings distribution, this solely reflects that the dispersion that is due to performance related pay is smaller among public sector employee since the impact of education is picked up by the education dummy.
2. In the CQR example, the size of the two sectors was hardly affected as only one person was assumed to change jobs.
3. A line that peaks in the middle of the distribution would refer to a change that favours the middle class, with an ambiguous effect on overall inequality.

12. Both conditional and unconditional quantile regressions are estimated by breaking up the [0,1] interval of quantiles into 10 intervals of equal length so as to be able to simultaneously estimate 9 quantile regressions for the quantiles 0.1 to 0.9. As a result, for each year and country, the estimation procedures do not yield a single coefficient for each variable of interest, but 9 different coefficients, one for each conditional or unconditional quantile in the range 0.1 to 0.9.[5] In the estimation each observation is weighted by the sampling weight of the individual to correct for imperfections in the representativeness of the sample.[6] The standard errors around the estimated parameter values are obtained using a bootstrap procedure with 200 replications in the case of UQRs, whereas for CQRs an analytical solution exists and is therefore used.[7] The homogeneity hypothesis is rejected in most cases for both CQRs and UQRs, confirming the need to go beyond the mean and the usefulness of quantile regressions. Further technical details of the estimation procedures are provided in Box A2.2.

5. Since the quantile regression estimates vary quite smoothly for small changes in the chosen quantile, the results of the analysis would be fairly similar if other quantiles were chosen. Indeed, estimating quantile regressions for 19 quantiles in the range 0.05 to 0.95 did not change the conclusions.

6. Sampling weights typically compensate for unequal probabilities of selection and non-response and adjust the sample distribution for key variables of interest (for example, age and gender) to make it conform to a known population distribution For Japan, unweighted data are used as no sampling weights are provided in the dataset.

7. Only when testing for the homogeneity of effects across quantiles are bootstrapped standard errors also used for CQRs since the analytical standard errors cannot be used for such tests. The bootstrapped standard errors of the CQR procedure are based on unweighted data. However, while the use of weights is important to obtain correct estimates of key parameters such as the quantiles, they are not crucial for the implementation of the homogeneity test.

Box A2.2 The estimation of conditional and unconditional quantile regressions

Conditional quantile regressions

For the purpose of estimating a CQR, the τ^{th} conditional quantile of a random variable *Y* (*e.g.* earnings) is assumed to be a linear function of randomly distributed exogenous factors *X* (*e.g.* age, gender, hours worked):

$$q_{Y|X}(\tau)[Y] = X\beta(\tau)$$

where τ takes values between 0 and 1, $0 < \tau < 1$. The equation implies that the earnings y_i of an individual *i* with $\tau = \tau_i$ are exactly equal to $x_i\beta(\tau_i)$, where x_i are the characteristics of individual *i*.[1] Any differences in earnings that cannot be explained by differences in personal characteristics are reflected in differences in the constant across quantiles. Similar to standard OLS, the parameter $\beta(\tau)$ is estimated by minimizing the following criteria:

$$\underset{\beta(\tau)}{\arg\min} \sum_i \rho_\tau\left(y_i - x_i\beta(\tau)\right)$$

where the function $\rho_\tau(y_i - x_i\beta(\tau)) = (y_i - x_i\beta(\tau))(\tau - 1)$ if $y_i \le x_i\beta(\tau)$ and $\rho_\tau(y_i - x_i\beta(\tau)) = (y_i - x_i\beta(\tau))\tau$ if $y_i > x_i\beta(\tau)$.

Three special cases of a CQR can help to understand this tool:

- When estimating the impact of *X* on the median of *Y* (*i.e.*, τ = 0.5) the quantile regression estimator becomes equal to the least absolute deviations estimator, which minimizes the sum of absolute deviations, *i.e.* $\operatorname{argmin}_\beta \Sigma_i|y_i - x_i\beta|$. This estimator is more robust to extreme values than the standard OLS estimator which minimises the sum of squared deviations (similar to the higher robustness of the median relative to the mean).

- If the constant is the only explanatory variable, then the estimate of $\beta(\tau)$ is equal to the τ^{th} quantile of *Y*. This can be checked by looking at the first order condition, which can be computed by looking separately at the two cases $y_i < \beta(\tau)$ and $y_i > \beta(\tau)$:

$$\frac{\partial}{\partial\beta}\sum_i \rho_\tau\left(y_i - \beta(\tau)\right) = \sum_{i|y_i<\beta(\tau)}(1-\tau) + \sum_{i|y_i>\beta(\tau)}(-\tau) = 0\,.$$

  This derivative is equal to zero if the share of individuals below $\beta(\tau)$ is equal to τ. In this particular example, the conditional quantile coincides with the unconditional quantile.

- If the only explanatory variables are a constant and a 0/1 dummy variable *X*, then the conditional quantile of an individual with a zero value (value of one) for that dummy variable is the quantile among all individuals with a zero value (value of one) for that dummy variable (the red and black markers in Figure A2.2 in Box A2.2). This can be generalised to multiple explanatory variables as the conditional quantile of an individual *i* is the position of that individual in a virtual population in which all individuals have the same observed characteristics as individual *i*.

Unconditional quantile regressions

The object of interest of a UQR is the marginal effect on the unconditional quantile of a small increase in the characteristic *X*:

$$\gamma(t) = \lim_{t\to 0}\frac{q_Y(\tau)\left[h(X+t,\varepsilon)\right] - q_Y(\tau)\left[h(X,\varepsilon)\right]}{t}$$

where earnings *Y* are a function *h* of observed characteristics *X* and unobserved characteristics ε and $q_Y(\tau)[Y]$ is the τ^{th} quantile of the unconditional distribution of *Y*. The method is rather general as it does not only allow investigating the impact on a particular quantile, but also on other measures of the earnings distribution such as the mean (the usual OLS framework is thus a particular case of the UQR technique) or the Gini index.

The unconditional marginal effect is estimated using the two-step procedure proposed by Firpo *et al.* (2009). The first stage involves the estimation of a so-called recentered influence function (RIF) for each individual:[2]

$$RIF = q_Y(\tau) + \frac{\tau - 1}{f_Y(q_Y(\tau))} \text{ if } Y \le q_Y(\tau) \text{ and } RIF = q_Y(\tau) + \frac{\tau}{f_Y(q_Y(\tau))} \text{ if } Y > q_Y(\tau)$$

Box A2.2 The estimation of conditional and unconditional quantile regressions *(cont.)*

where the density of earnings f_Y is estimated using a Gaussian kernel estimator.[3] In the second stage, the RIF is regressed on the explanatory variables X using OLS, and hence the probability for a worker to have earnings above a certain quantile is assumed to be linear in the observed characteristics. Since the RIF takes on only two different values, a logistic estimator is a natural choice for the second-stage estimation. This estimator is used as an alternative to the OLS estimator to check the robustness of the results. Apart from a few exceptions, the results are not affected by this modification, confirming the conclusion of Firpo *et al.* (2009) who also experiment with alternative methods such as the logistic estimator and non-parametric estimators. The linearisation thus seems to be a reasonable simplification.

Unconditional quantile regressions provide an estimate of the partial equilibrium effect of the variable of interest, assuming that the unobserved heterogeneity is independent from the observed characteristics and that there is no reverse causality. The marginal change in X is assumed to have no impact on the joint distribution of X and Y, meaning that rates of return do not vary in the case of small variations in any of the observed characteristics X. While these assumptions may not hold in practice – for instance, a worker's decision to work extra hours may depend on earnings – a comparison between estimates for low and high quantiles would still be valid in that case as long as the potential bias is the same across the sample population.

The quantile regression techniques are fairly robust to outliers, which is a highly desirable property given the use of household survey data that are prone to measurement error. This is particularly the case for CQRs, where even large measurement errors would have only a small impact on the parameter estimates (provided the number of observations with a large measurement error is reasonably small), whereas estimates from standard OLS techniques would be severely biased in such a case.[4] While the UQR estimator is highly robust to outliers in the dependent variable (Hampel *et al.*, 2005), it is somewhat more sensitive to outliers in the explanatory variables because of the use of OLS in the second stage. In the particular case of tail quantiles, such as the 5th or 95th quantiles, both CQR and UQR estimators hinge crucially on extreme observations and, hence, are more fragile (which is why such quantile estimates are not reported here). In addition, standard errors are generally larger for these extreme quantiles.

---

1. The conditional quantile itself is generally not identifiable for a given individual *i* because the equation $x_i\beta(\tau)= y_i$ may have no or several solutions for the unknown parameter τ.
2. The RIF allows estimating how each explanatory variable affects the probability to have earnings above a given quantile. This is why the dependent variable divides the population into two groups: those below and those above the given quantile. The rescaling factor (inverse of the density of earnings) allows converting the probability of earnings to switch above the given quantile into an earnings gain at that quantile.
3. The results of such a kernel estimator depend on the choice of the bandwidth. This paper employs a bandwidth that minimises the mean integrated squared error under the assumption that the data are Gaussian for all countries with the exception of Brazil where a larger bandwidth is chosen due to the existence of mass points in the data (Firpo *et al.*, 2009 also use a larger bandwidth when applying their methodology to the Current Population Survey of the United States).
4. The coefficients obtained from CQRs are indeed locally estimated. They are barely affected by observations that are not around the percentile of interest. In the neighbourhood of a zero value for the first order condition, the objective function changes only for pivotal individuals for which the conditional quantile is very close to τ, that is when $x_i\ \beta(\tau)$ is very close to y_i. All other individuals have no sizeable weight in this estimation.

A2.2.3. Decomposing labour earnings inequality

13. Earnings differences between two individuals can have two main sources: *i*) differences in personal characteristics such as the level of education and *ii*) differences in the returns to these characteristics. Similarly, cross-country differences in earnings inequality can be decomposed into *i*) differences in the *composition* of the population (for example, inequality should be higher in countries with a more unequal distribution of education endowment) and *ii*) differences in *rates of return* (for example, inequality should be higher in countries with a larger wage gap between high- and low-educated workers).

14. The most commonly used methodology to decompose average earnings differentials between two groups (*e.g.* men and women) is the one proposed by Oaxaca (1973) and Blinder (1973). This method first estimates the rate of return for each observed characteristic (such as the level of education) for the reference group (*e.g.* men) using OLS, and then multiplies this estimated rate of return by the gap in the

average value of the observed characteristic between the two groups (*e.g.* the gap in the average years of schooling between men and women). The resulting value gives the contribution of this characteristic to the earnings gap. The remaining unexplained part of the gap is either due to differences in characteristics that are not controlled for in the regression (*e.g.* the quality of an individual's tertiary diploma) or discrimination.[8] The Oaxaca-Blinder methodology is used in this paper to decompose the average earnings gap between men and women and between natives and immigrants. As regards decompositions of cross-country differences in inequality, several alternative methods have been developed for this purpose (for a recent survey, see Fortin *et al.*, 2011). This paper adopts a methodology that is close to the one proposed by Firpo *et al.* (2007*b*) and builds on the UQRs discussed above.[9] The United States are used as the reference country so that each country's level of earnings inequality is compared to the level of inequality in the United States.[10]

15. An important choice is the measure of earnings inequality to be decomposed. Many studies use the Gini index of the logarithm of earnings because their underlying models consider the logarithm of earnings as the dependent variable. A major drawback of this measure is that the scale independence assumption does not hold, meaning that the value of the measure changes when all earnings are multiplied by a certain scale factor. In addition, by putting less weight on the upper part of the earnings distribution, the Gini index of the logarithm of earnings may yield a country ranking that differs from that of the Gini index. The logarithm of the 90/10 percentile ratio does not have these two weaknesses and is therefore preferred in this paper. The main limitation of this measure is that it only builds on the estimated effect of explanatory variables at the 10th and 90th percentiles, and leaves aside effects on the middle class.

16. For each explanatory variable *k*, the composition effect $C_k^{90/10}$ relies on a comparison of the estimated effects in the two countries (*i.e.* the United States and the country of interest *i*) at the 10th centile and the estimated impact at the 90th centile. If the rise of, say, the proportion of tertiary-educated workers from the US level to the level observed in country *i* is relatively small, the effect can be linearized. To get the effect on the 90/10 centile ratio, the variation of the 90th centile, *i.e.* $(E(X_{k,i}) - E(X_{k,USA}))\beta_{k,i}^{90}$, is compared to that of the 10th centile, *i.e.* $(E(X_{k,i}) - E(X_{k,USA}))\beta_{k,i}^{10}$:

(A2.1) $$C_k^{90/10} = (E(X_{k,i}) - E(X_{k,USA}))\left(\beta_{k,i}^{90} - \beta_{k,i}^{10}\right)$$

where $\beta_{k,i}^{90}$ ($\beta_{k,i}^{10}$) is the coefficient estimate on the variable *k* at the 90th (10th) unconditional quantile for the country of interest *i* (the country to be compared to the United States).

---

8. While the rates of return are assumed to be homogenous within each subgroup, they are allowed to differ between the two subgroups. However, in case they differ, the results depend on the reference group, reflecting the path dependence of this decomposition.

9. The main difference with respect to the methodology proposed by Firpo *et al.* (2007b) is that their method makes use of a regression that is run on the country of interest and assumes that all explanatory variables follow the same distribution in that country as they do in the United States. This regression is omitted in the approach adopted in this paper and the information is instead taken from a regression on the country of interest without changing the distribution of the explanatory variables. This simplified approach assumes that the probability of being above a certain quantile in the distribution is linear in the set of explanatory variables (see Box A2.2). While this assumption facilitates the derivation of the rate of return effects from the estimation results of the unconditional quantile regressions, it may not hold for countries that deviate considerably from the United States in terms of their population characteristics.

10. Although the results depend somewhat on the choice of reference country, the general conclusions are fairly robust to this choice.

17. The rate-of-return effect $R_k^{90/10}$ for variable k is computed by running two separate UQRs – one on the United States and another on the country of interest – and then comparing the coefficients at the 10th and 90th percentiles obtained from the two regressions:

(A2.2) $$R_k^{90/10} = E(X_{k,USA})\left[\left(\beta_{k,i}^{90} - \beta_{k,i}^{10}\right) - \left(\beta_{k,USA}^{90} - \beta_{k,USA}^{10}\right)\right]$$

This rate-of-return effect can also be regarded as the difference between the rate-of-return effect for high-income earners $E(X_{k,USA})\left(\beta_{k,i}^{90} - \beta_{k,USA}^{90}\right)$ and the rate-of-return effect for low-income earners $E(X_{k,USA})\left(\beta_{k,i}^{10} - \beta_{k,USA}^{10}\right)$.

18. This method yields more accurate results for the size of the composition effects than for the size of the rate-of-return effects. In fact, the composition effects strongly rely on differences between the means of the explanatory variables which are known with relatively high precision. By contrast, the rate-of-return effects strongly rely on differences between the estimated rates of return which are intrinsically less accurate. For this reason, solely qualitative conclusions are drawn below regarding the rate-of-return effects.

A2.3. Empirical results: labour earnings inequality and its main determinants

A2.3.1. Benefits and drawbacks of household survey data

19. Household surveys provide a unique source to investigate the determinants of earnings inequality. They allow exploiting information on individual workers, thus involving substantially more variation in the data than aggregate cross-country information. Moreover, they contain specific information on the linkages between earnings and various personal characteristics that cannot be inferred from aggregate data. Compared with administrative data, survey data have the advantage that individuals' answers to survey questions about earnings should pick up all types of labour earnings that are of interest to this study, whereas data from administrative sources often omit some categories such as non-taxable earnings (in the case of fiscal data), or income from additional jobs (in the case of firms' compulsory statements on wages and benefits). Moreover, the household surveys used in this study are designed to cover a representative sample of the whole population, while administrative sources may ignore some sub-samples of the population such as non-taxable workers.

20. However, the use of household survey data also entails a number of drawbacks that need to be kept in mind when interpreting the results. *First*, although the surveys are designed to ensure that the sample population is representative of the entire population in terms of its major characteristics, this is not fully the case for some population characteristics that are of particular importance for the present study (such as the share of temporary workers, for example). *Second*, the number of non-responses can be substantial, not only for the dependent variable but also for some of the explanatory variables (in particular information on the sector of employment or the type of work contract is missing for a larger number of individuals). The analysis assumes that the decision not to respond to a survey question is independent from both the dependent variable and all explanatory variables. To the extent that this assumption is violated, the regression results could be biased. *Third*, it cannot be ruled out that individuals provide wrong answers to some of the questions or that some of the questions are interpreted differently by different individuals. This problem is likely to be negligible in the present study since the quantile regression technique is fairly robust to outliers, as far as these measurement errors are quite rare. *Fourth*, in contrast to some administrative data sources (such as tax data), household surveys do not allow the observation of the most extreme parts of the distribution, such as top income earners.

21. Another issue is the comparability of household survey data across countries. The depth of information varies across surveys, as does the classification of variables and the way in which questions are phrased and thus interpreted by respondents. This paper deals with these problems by starting with the European Union Survey of Income and Labour Dynamics (EU-SILC), which provides a unified framework for 23 OECD countries. For the other eight countries for which household survey data could be collected (Australia, Canada, Israel[11], Japan, Korea, Switzerland, the United States and Brazil), all the variables are then chosen and, if necessary, recoded so as to ensure maximum comparability with the EU-SILC dataset (see the Appendix for details on variable selection and coding).

A2.3.2. A first pass on labour income inequality

22. Keeping in mind comparability issues, countries appear to differ widely with respect to the level of labour earnings inequality. The Gini index for full-time employees is highest in Portugal, Korea, Poland and the United States, while Belgium, Denmark, Finland and the Czech Republic have the most equal distribution (Figure A2.4).[12,13] Inequality is generally higher when looking at all full-time workers (*i.e.* both employed and self-employed), reflecting the greater dispersion of earnings among the self-employed. Extending the analysis to part-time workers, the unemployed and the inactive further raises the Gini index, reflecting the large income differentials between the group of full-time workers and the groups of unemployed and inactive individuals who have zero labour earnings (transfers are not taken into account). The increase in the Gini index is particularly large for countries where part-time workers make up a sizeable share of total employment (*e.g.* Australia, Germany and the Netherlands), and for countries with a high unemployment rate. While the four Gini indices are highly correlated (the correlation coefficients are between 0.8 and 0.9), there are several countries for which the choice of the group matters considerably for the inequality ranking. For example, Korea comes out as the third-most unequal country based on full-time employees, but ranks about average once including part-time and self-employed workers.

---

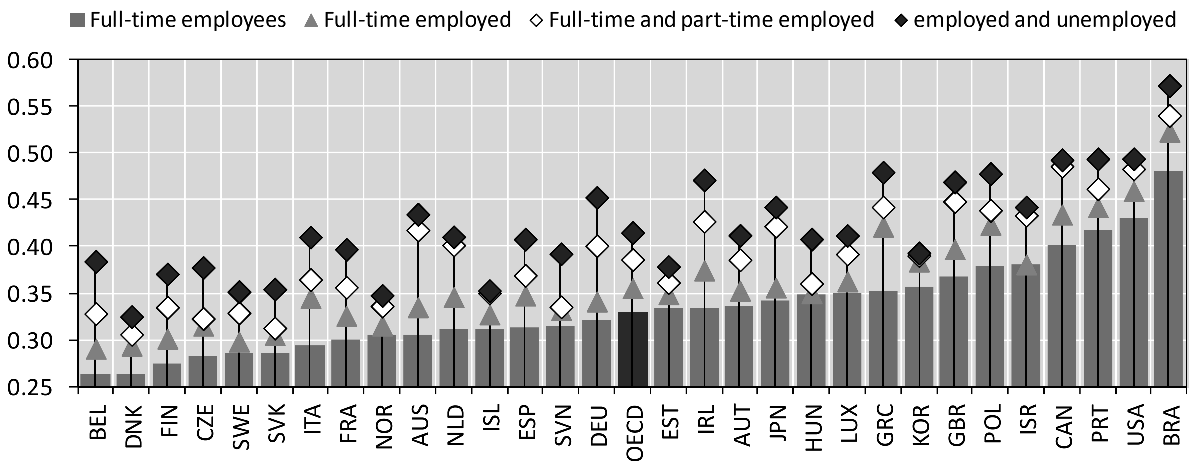
11. The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

12. The difference in Gini indices across most countries is significant because uncertainty around their estimates is small even accounting for the rather small sample sizes for some countries. For instance, in the particular case of a log-normal distribution of earnings, the width of the 95% confidence band around a Gini index of 0.5 is 0.03 when the size of the sample is around 4000 (which is the case for Japan – the smallest sample in the analysis). The size of the confidence band is cut in half when the size of the sample is multiplied by 4.

13. Note that Gini indices put a high weight to inequality at the top of the earnings distribution.

Figure A2.4. **Gini indices for different population subgroups**

15-to-64-year olds, 2008

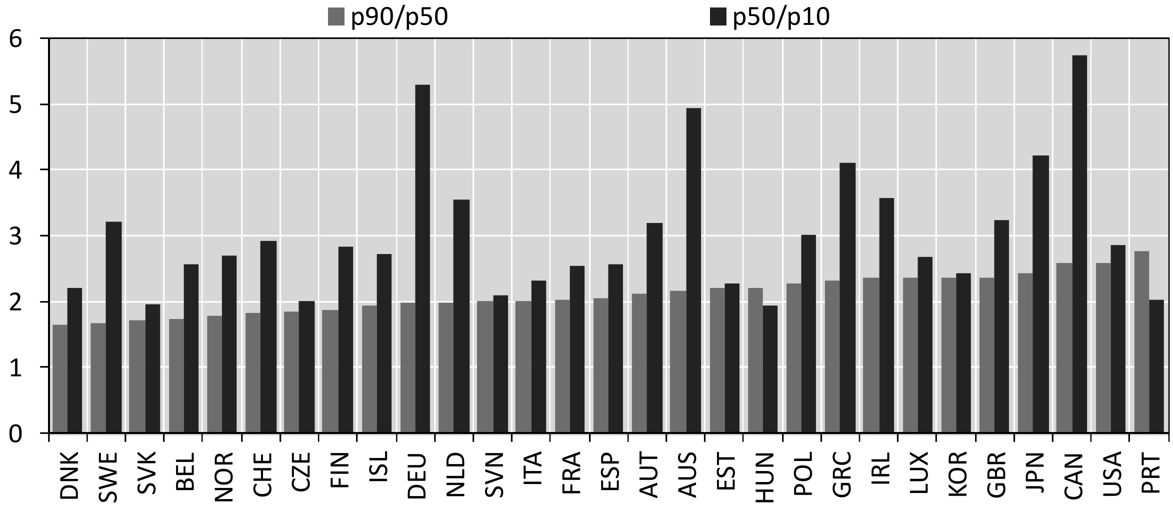


Note: The Gini coefficients take into account labour earnings only; the precise definition of labour earnings differs across countries (see the Appendix for details). 2007 for France, Korea and the United States, 2009 for Australia and Japan.

Source: Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia; Survey of Labour and Income Dynamics (SLID) for Canada; Korean Labour and Income Panel Study (KLIPS) for Korea; Luxembourg Income Study (LIS) for Brazil and Israel; Japan Household Panel Survey (JHPS) for Japan; Panel Study of Income Dynamics (PSID) for the United States; European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

23. While Gini indices provide a simple picture of the overall level of earnings inequality, investigating the relative earnings of different quantiles allows disentangling whether inequality is due to inequality at the top or at the bottom of the earnings distribution. To assess the relative position of low- and high-income workers, Figure A2.5 shows the ratio of the median earnings to the 10th centile and the ratio of the 90th centile to median earnings. Since the sample includes part-time workers and workers who work only part of the year, the earnings at the 10th centile reflect the importance of these two groups among low-income workers. According to the surveys, these working poor are more numerous in Germany, Australia, Japan and the Netherlands – countries where the proportion of part-time workers is relatively high. By contrast, in several central and eastern European countries (the Slovak Republic, the Czech Republic, Slovenia, Estonia, and Hungary) as well as in Portugal, the 10th centile is relatively close to the median. Inequality at the top is the most pronounced in Portugal and the United States, whereas it is rather low in the Slovak Republic, Belgium and several of the Nordic countries.

Figure A2.5. **Percentile ratios for the working population, 2008**



Note: The working population refers to all individuals who work either part-time or full-time and had positive labour earnings during the reference period. 2007 for France, Korea and the United States, 2009 for Australia and Japan.

Source: Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia; Survey of Labour and Income Dynamics (SLID) for Canada; Korean Labour and Income Panel Study (KLIPS) for Korea; Luxembourg Income Study (LIS) for Brazil and Israel; Japan Household Panel Survey (JHPS) for Japan; Swiss Household Panel (SHP) for Switzerland; Panel Study of Income Dynamics (PSID) for the United States; European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

A2.3.3. The determinants of labour earnings – results from quantile regressions

24. This section summarizes the overall conclusions that can be drawn from the quantile regression exercise regarding the linkages between the distribution of labour earnings and hours worked, labour market experience, education, the type of employment, union membership, the sector of employment, gender, and migration. Detailed country results are shown in the Appendix.

Hours worked

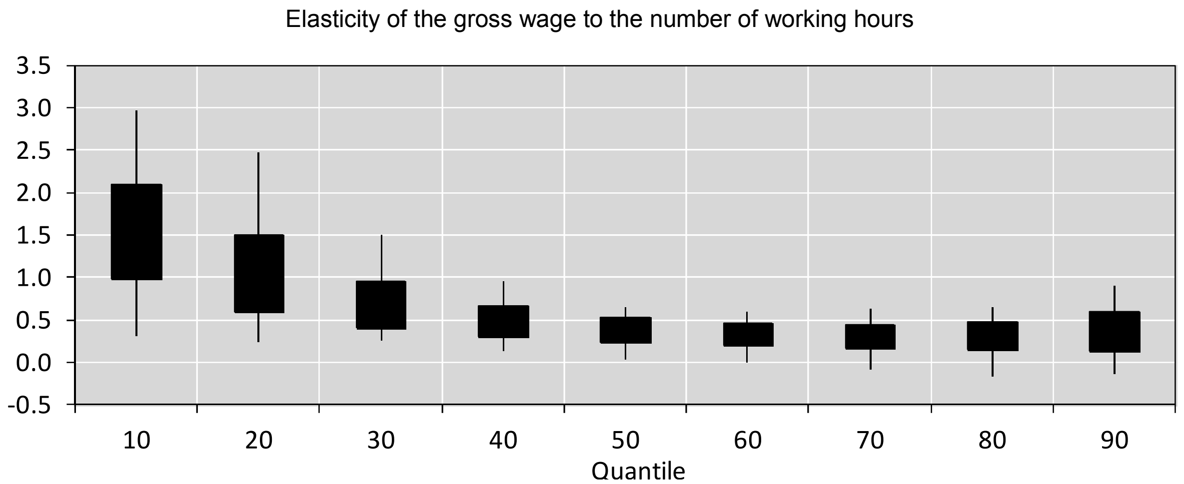
25. An important determinant of earnings inequality among the working population is the number of hours worked (captured by the number of hours worked per month in all jobs).[14] Unconditional quantile regression results indicate that the marginal returns to working one additional hour vary substantially across countries, especially at lower quantiles, which may reflect different labour market policy settings and practices, in particular as regards the role of overtime pay (Figure A2.6). However, one common observation in almost all countries is that the reward for working one additional hour is highest for workers at the lower end of the earnings distribution. This could be due to differences in the extent to which time spent at work is recorded, *i.e.* lower-income workers may be more likely to benefit from overtime pay whereas extra hours by middle and high-income workers may be compensated as part of the basic remuneration package.[15] The results suggest that a general decrease in the number of hours worked,

14. Any cross-country comparison of the regression results for the hours-worked variable needs to be made with great caution as the survey questions used to calculate the number of working hours differ across surveys.

15. While the basic wage rate may also vary with the number of working hours, this is unlikely to adequately explain the observed heterogeneity in the return to hours worked as several studies show that wage offers to part-timers are actually lower than those to full-time workers (*e.g.* Moffitt, 1984; Simpson, 1986; Ermisch and Wright, 1993).

triggered for example by an economic recession, would thus particularly hurt lower-income workers through a fall in overtime pay.

Figure A2.6. Estimated effect across countries of working an additional hour (UQR estimates)



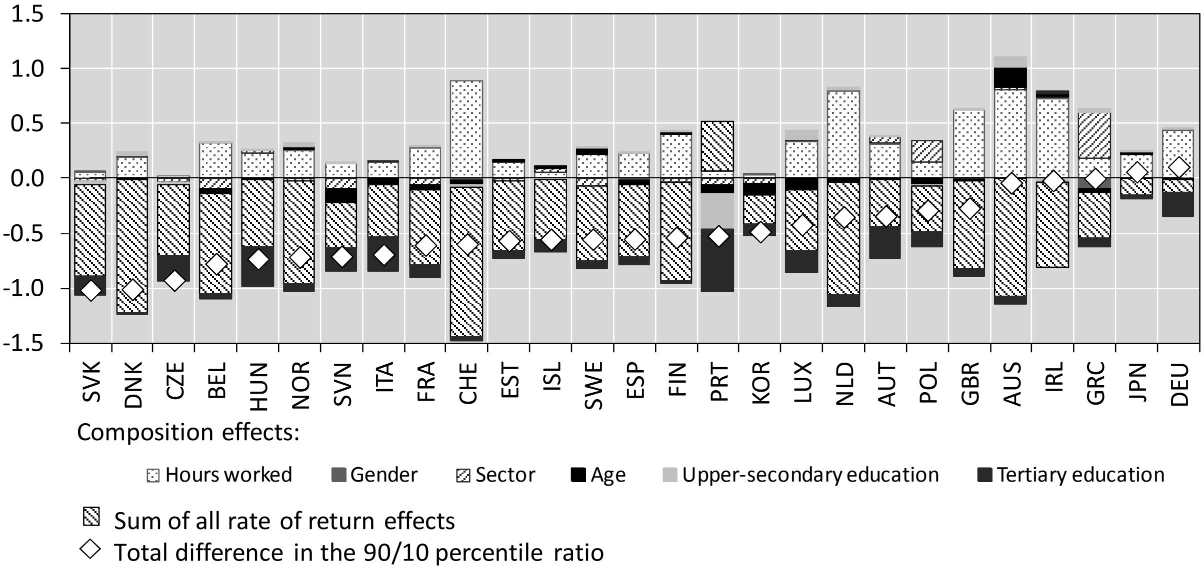
Note: The thick bars depict the cross-country mean of the estimated effect +/- 1 standard deviation across countries, while the thin bars depict the cross-country maximum and minimum of the estimated effect

Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA), the Survey of Labour and Income Dynamics (SLID) for Canada, the Korean Labour and Income Panel Study (KLIPS), the Luxembourg Income Study (LIS) for Brazil and Israel, the Japan Household Panel Survey (JHPS), the Swiss Household Panel (SHP), the Panel Study of Income Dynamics (PSID) for the United States, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

26. The key role played by the numbers of hours worked in shaping the distribution of earnings is confirmed by three other findings. *First*, the inclusion of part-time workers changes the Gini ranking of countries considerably (Figure A2.4). *Second*, the hypothesis of a unit elasticity, and hence a model in which the dependent variable would be the hourly earnings, is rejected for all countries. *Third*, the contribution of cross-country differences in the average number of hours to cross-country differences in earnings inequality is substantial (Figure A2.7). Since hours worked are highest in the United States, the decomposition of countries' inequality gap *vis-à-vis* the United States shows that this factor contributes positively to the gap for all countries. The contribution is highest for countries such as Switzerland, the Netherlands or Canada, where weekly working hours are shorter. To the extent that shorter working hours observed in some countries are not due to personal preferences but reflect labour market rigidities or other impediments to full-time work – *e.g.* a lack of formal care for children and the elderly may drive down working hours among females – there is a role for policy to promote a more equitable distribution of earnings.

Figure A2.7. **Decomposition of cross-country differences in the logarithm of the 90/10 percentile ratio**

The United States is used as reference country, 2007[1]



1. 2008 for Canada; 2009 for Japan.
2. 90/10 percentile ratio of the country shown on the horizontal axis minus 90/10 percentile ratio of the United States.

Note: The decomposition is based on the UQR results. Composition effects can be altered because questionnaires differ across surveys, and hence should be interpreted with care.

Source: Panel Study of Income Dynamics (PSID) for the United States; Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia; Survey of Labour and Income Dynamics (SLID) for Canada; Korean Labour and Income Panel Study (KLIPS) for Korea;, Japan Household Panel Survey (JHPS) for Japan; Swiss Household Panel (SHP) for Switzerland; and European Union Statistics on Income and Living Conditions (EU-SILC) for the other countries.

Labour market experience

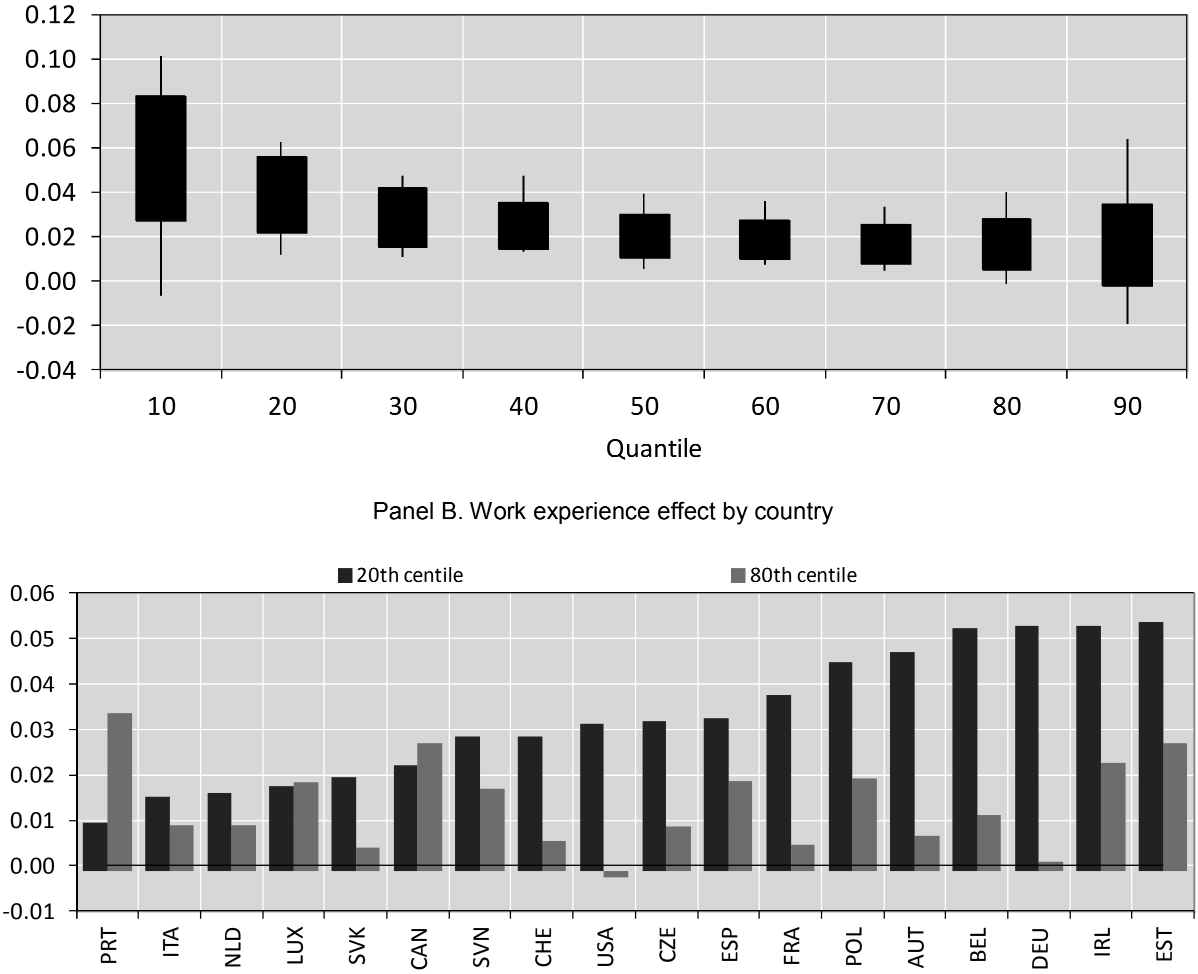
27. The linear and quadratic age terms included in the baseline UQR specification are likely to capture roughly the impact of labour market experience on earnings inequality. The coefficient estimates indicate that the average returns to age are higher for lower quantiles, suggesting that work experience plays a larger role in lower-paid jobs or that seniority pay is more prevalent in these types of jobs (the main exceptions are Brazil, Israel, Japan, Korea and Portugal, where the returns to age are larger at higher quantiles). As a consequence, policies that reduce the likelihood of career breaks or their length (*e.g.* improvements in the availability of formal childcare and less duality in labour markets) may help to reduce the dispersion of earnings. The results also reveal a sizeable variation in the returns to age across countries, possibly reflecting cross-country heterogeneity in the prevalence of seniority pay. The returns to age are particularly high in Belgium, Germany and Poland.

28. To explore the issue further, the baseline specification is augmented with a variable measuring individuals' work experience (typically the number of years worked since the first job) for all countries for which this information is available. The results from this augmented specification confirm the two conclusions drawn from the baseline specification, *i.e.* the returns to experience are larger at the lower end of the earnings distribution and they vary considerably across countries (Figure A2.8). The coefficients on the age terms become smaller once controlling for work experience, as could be expected, but remain significant, particularly for the lower quantiles. There thus seems to be an age-specific reward that goes beyond the pure work experience effect.

Figure A2.8. The effect on earnings of having one additional year of work experience (UQR estimates)

Effect of raising the work experience by one year on log earnings.

Panel A. Work experience effect by quantile



Note: In Panel A, the thick bars depict the cross-country mean of the estimated effect +/- 1 standard deviation across countries, while the thin bars depict the cross-country maximum and minimum of the estimated effect. The specification includes the number of years of work experience and its square. The chart shows the effect for a worker with 20 years of work experience.

Source: UQR estimates for employed individuals using data from the Survey of Labour and Income Dynamics (SLID) for Canada, the Swiss Household Panel (SHP) for Switzerland, the Panel Study of Income Dynamics (PSID) for the United States, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 15 EU member countries.

Education

29. Theory suggests that the link between education and labour earnings inequality is far from straightforward. The impact of a change in the educational composition of the workforce can be thought of as the combination of two separate effects (Knight and Sabot, 1983): *i*) a composition effect, whereby a rise in the share of highly-educated (high-wage) workers raises earnings inequality up to a certain point, but will then lower it as fewer less-educated (low-wage) workers remain; and *ii*) a rate-of-return effect, whereby a rise in the share of highly-educated workers alters the returns to education.[16] The direction of

16. Other theoretical models decompose the total effect instead into the contributions of educational inequality and average education: While higher educational inequality should unambiguously raise the dispersion of

the change in (relative) returns depends on many factors such as the substitutability or complementarity between less-educated workers, highly-educated workers and capital, the interplay between innate ability and schooling (Dur and Teulings, 2004), and the signalling role of education (*e.g.* Hendel *et al.*, 2005).

30. The composition effect of a change in the educational level of the workforce depends on *i*) the variance of wages among high-educated workers relative to the variance of wages among low-educated workers; *ii*) the average wage gap between high- and low-educated workers; and *iii)* the initial share of high and low-educated individuals in the total workforce. Specifically, if earnings are more dispersed among highly-educated individuals, then an increase in the share of highly-educated individuals raises earnings inequality, *ceteris paribus*. A second inverted-U-shaped effect is then superimposed on this monotonic first effect, whereby (starting from zero) a rise in the share of highly-educated individuals initially raises earnings inequality as the earnings of some workers now differs from that associated with a low education level, but eventually inequality declines as more and more individuals have higher education and heterogeneity in education attainment is reduced.

31. The unconditional quantile regressions provide an estimate of the returns to education for 9 different earnings quantiles. The variation in the rates of return across quantiles can be interpreted as the composition effect of a change in the educational composition of the workforce. For upper-secondary or post-secondary non-tertiary education, the UQR estimates show that the returns fall along the earnings distribution for most countries, meaning that the dispersion of earnings would fall as more individuals get upper-secondary or post-secondary non-tertiary degrees − a result that is to be expected as the majority of individuals in the countries considered already have upper-secondary education (Figure A2.9, Panel A).[17,18] By contrast, a rise in the number of tertiary graduates changes the composition of the workforce in such a way that earnings become more dispersed (Figure A2.9, Panel B): the rate of returns to a tertiary degree rises along the earnings distribution. For the three countries, for which more detailed information on education are available (Australia, Switzerland and the United States), splitting up the tertiary education dummy into a dummy for bachelor and master degrees and a dummy for PhD degrees shows that a rise in the share of workers with a PhD is associated with a rise in earnings inequality and that this effect is concentrated on the top part of the earnings distribution.[19]

---

earnings, the impact of increased average schooling may be positive or negative, depending on how rates of return to education evolve (*e.g.* Schultz, 1961; Becker, 1964). Empirically, a higher average level of schooling appears to wage inequality (*e.g.* Gregorio and Lee, 2002; Checchi and García-Peñalosa, 2008).

17. Two notable exceptions are Portugal and Brazil, where upper secondary and post-secondary non-tertiary education is found to be more profitable for those at the top of the earnings distribution. This could be due to the lower average education level compared with the other countries in the sample. The results for Portugal are in line with existing empirical evidence (*e.g.* Machado and Mata, 2001; Hartog *et al.*, 2001).

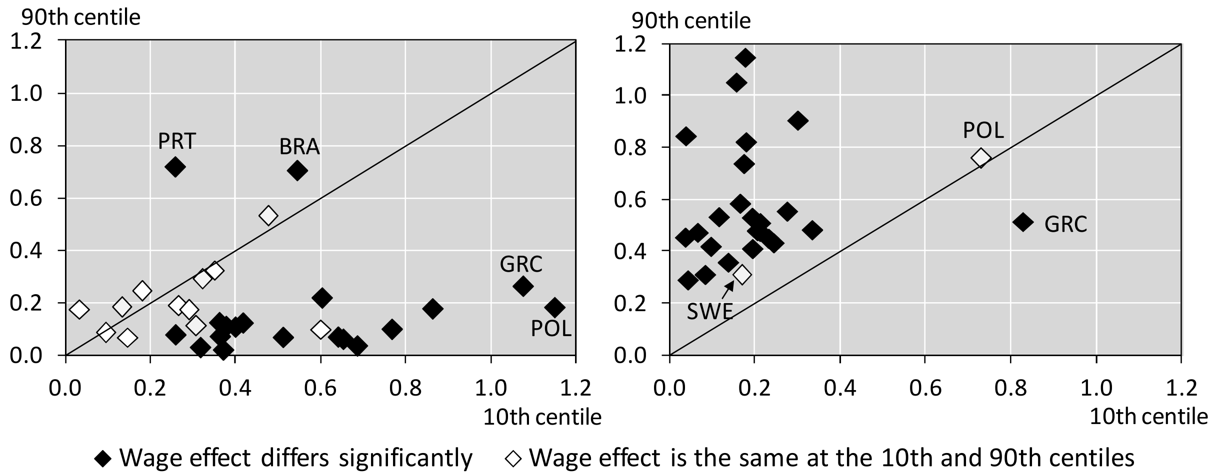
18. The results depend somewhat on the choice of the estimator in the second step of the UQRs. When using the logistic estimator instead of the OLS estimator, the finding still holds for 12 countries, while for roughly one-third of the countries the effect at the 90th quantile is then above that at the 10th quantile. For seven countries, the hypothesis of equal coefficients across the entire range of quantiles cannot be rejected when using the logistic estimator, meaning that, a rise in the share of workers with upper-secondary or post-secondary non-tertiary degrees does not alter the distribution of earnings.

19. The regressions that make use of the logistic estimator in the second step of the UQRs can, however, not confirm this finding, potentially related to the small share of individuals with a PhD in the working population.

Figure A2.9. **The impact of education on the distribution of earnings (UQR estimates)**

Panel A. Effect of raising the number of upper-secondary or post-secondary non-tertiary graduates on log earnings

Panel B. Effect of raising the number of tertiary graduates on log earnings



Notes: The horizontal axis shows the impact of a 1 percentage point increase in the proportion of workers with respectively upper-secondary (Panel A) and tertiary (Panel B) education on the log earnings of the 10th quantile. The vertical axis shows the impact of the same change on the log earnings of the 90th quantile. A data point below (above) the 45 degree line indicates that the change in the educational composition of the workforce is associated with a fall (rise) in earnings inequality. The equality test is performed at the 5% level.

Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia, the Survey of Labour and Income Dynamics (SLID) for Canada, the Korean Labour and Income Panel Study (KLIPS) for Korea, the Luxembourg Income Study (LIS) for Brazil and Israel, the Japan Household Panel Survey (JHPS) for Japan, the Swiss Household Panel (SHP) for Switzerland, the Panel Study of Income Dynamics (PSID) for the United States, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

32. Using the UQR estimates to decompose cross-country differences in the 90/10 percentile ratio suggests that differences in the educational composition of the workforce play an important role (Figure A2.7). *Ceteris paribus* (*i.e.* assuming in particular that the relative rates of return to education remain unchanged), the high shares of workers with tertiary education in countries such as Ireland and the United States imply a high 90/10 percentile ratio relative to other counties, while low tertiary education attainment in Portugal and Hungary implies the opposite. The share of workers with an upper-secondary or post-secondary non-tertiary degree does in general not play a major role in explaining cross-country differences in earnings inequality, reflecting *first* that most countries do not differ much in the share of workers holding such a degree and *second* the smaller impact of this factor in shaping the distribution of earnings in most countries.[20]

33. The impact of changing the educational composition of the workforce on earnings inequality, as inferred from the UQRs, reflects only a composition effect that assumes unchanged returns to education. However, as discussed above, a change in the educational composition of the workforce may alter the relative returns to education. The resulting repercussions on earnings inequality may strengthen or weaken the composition effect as estimated by the UQRs. A simple cross-country time series regression of the average returns to a certain education degree (obtained from an OLS estimation of the baseline

20. The only exception is Portugal where this factor contributes to reduce the inequality gap vis-à-vis the United States. This reflects the very low share of upper-secondary or post-secondary non-tertiary educated workers in Portugal combined with a strong positive link between the share of such workers and the level of earnings inequality.

specification) on the share of individuals holding such a degree and country fixed-effects tentatively indicates that a rise in the number of tertiary graduates significantly lowers the relative returns to tertiary degrees,[21] while the returns to upper-secondary and post-secondary non-tertiary degrees are not influenced by the share of workers with such degrees. This means that the impact of a rise in the share of tertiary educated workers on earnings inequality is likely to be smaller than estimated with UQRs (and may even be negative), while for upper-secondary and post-secondary non-tertiary education, the UQR results can be regarded as the total effect. Overall, the results of the analysis thus suggest that reducing the number of pupils that leave school with a lower-secondary degree or less (*e.g.* by providing support to pupils at risk in order to reduce drop outs) should reduce income inequality by raising the share of workers with upper-secondary education. The impact of a rise of tertiary education on inequality remains ambiguous by contrast and depends on the relative sizes of the (inequality-increasing) composition effect and the (inequality-reducing) rate of return effect.

34. An individual's earnings may not only be influenced by his own education but, through spillover effects, also by the education level of individuals with whom he interacts. To investigate this issue, the baseline UQR specification for the United States is augmented with three additional variables which measure the proportion of workers in an individual's state of residence who hold an upper-secondary or post-secondary non-tertiary, a tertiary or a PhD degree. For the majority of quantiles, the three additional variables are not significant, suggesting that the spillover effects are at best small. This is in line with the findings of Acemoglu and Angrist (1999) who conclude that spillovers are significant but small.

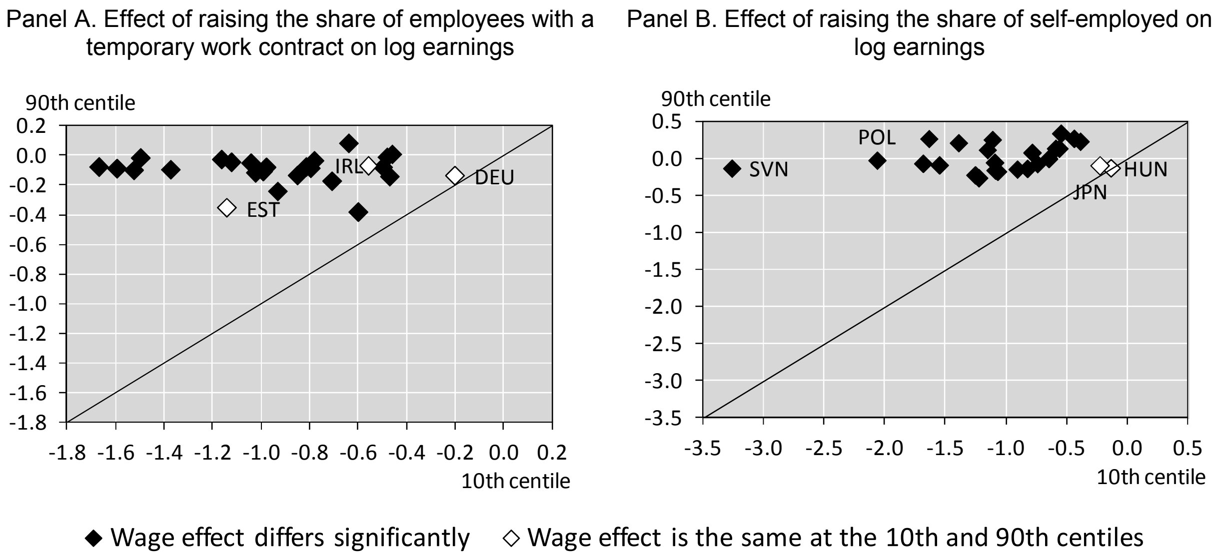
Type of employment

35. The impact of the type of employment on labour earnings inequality is assessed by augmenting the baseline UQR specification with a dummy variable for being self-employed and a dummy variable for holding a temporary work contract (therefore dependent employed individuals with a permanent work contract serve as the reference group). UQR results provide robust evidence that employees on temporary contracts who are at the bottom of the earnings distribution earn less than those on permanent contracts (Panel A of Figure A2.10) – a loss that comes on top of the intrinsic lack of job stability. The earnings penalty at the 10th quantile is particularly large in Austria, Belgium, Finland, Luxembourg and Sweden. The earnings of high-income employees are less dependent on the type of the work contract: in almost all countries the coefficient on the contract dummy is smaller in absolute magnitude at the 90th quantile than at the 10th quantile and in about half them it is not significantly different from zero.[22]

21. The estimated coefficient is -0.23 with a standard error of 0.11.

22. According to preliminary results, a significantly negative average earnings effect of having a temporary work contract can hold even when controlling for person-specific fixed effects.

Figure A2.10. The impact of the type of employment on the distribution of earnings (UQR estimates)



Note: The horizontal axis shows the impact of a 1 percentage point increase in the proportion of workers who are respectively dependent-employed with a temporary work contract (Panel A) or self-employed (Panel B) on the log earnings of the 10th quantile. The vertical axis shows the impact of the same change on the log earnings of the 90th quantile. A data point below (above) the 45 degree line indicates that the change in the composition of the workforce as regards the type of employment is associated with a fall (rise) in earnings inequality. The equality test is performed at the 5% level.

Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia, the Survey of Labour and Income Dynamics (SLID) for Canada, the Korean Labour and Income Panel Study (KLIPS) for Korea, the Japan Household Panel Survey (JHPS) for Japan, the Swiss Household Panel (SHP) for Switzerland, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 19 EU member countries as well as for Iceland and Norway.

36. Being self-employed also generally entails an earnings penalty (relative to being employed with a permanent work contract) at lower quantiles according to unconditional quantile regression estimates (Panel A of Figure A2.10). The effect is particularly sizeable in Poland, Slovenia and, to a lesser extent, Finland, the Netherlands and Sweden, whereas it is relatively small (though still statistically significant) in Hungary and Japan. For higher quantiles, the regressions yield more diverse results. In about one-third of the countries earnings at the 90th quantile do not depend on whether the worker is self-employed or dependent employed with a permanent contract. In another third of the countries, self-employed workers earn significantly more than their counterparts on permanent work contracts (potentially driven by self-employed in the professional services sectors) and in the remaining third of the countries they earn less. The magnitude of this earnings gap at the 90th quantile is rather small in all countries considered, implying that the earnings of self-employed individuals are more dispersed that those of dependent employed individuals who have a permanent work contract.

Unionisation

37. For reasons already discussed above, the relationship between the number of workers covered by collective agreements and overall earnings inequality is inverted U-shaped, with the shape of the curve depending on the relative means and variances of earnings in the two groups of workers. The influence of unions on wage inequality thus depends on both the number of workers who are covered by collective agreements – be it through union membership or through administrative extensions of collective agreements – and the influence of unions on bargained earnings (in terms of the wage gap between union

and non-union members and the dispersion of wages of union members relative to those of non-union members).

38. For all six countries for which data on union membership are available (Australia, Canada, Japan, Korea, Switzerland and the United States) simple descriptive statistics indicate that the earnings of union-members are higher and less dispersed than those of non-union members (Table A2.1). To explore this issue further and properly control for the influence of other personal characteristics, the baseline specification is augmented with a dummy variable that takes value one if a worker is a member of a union and this augmented specification is then estimated with the CQR technique. The coefficient on the union membership dummy is positive for most quantiles and smaller for higher than for lower quantiles, implying that the earnings of union members are higher and less dispersed than those of other workers, even if one controls for the influence of other factors such as age, education and gender. These findings are in line with earlier evidence (*e.g.* Gosling and Machin, 1995; Machin, 1997). The lower dispersion of earnings among union members may reflect that unions push for greater wage equality among their members or that individuals with higher earnings potential have lower incentives to join a union.

Table A2.1. **A comparison of wages among union and non-union members**

	Union members		Non-union members	
	Average earnings	Standard deviation	Average earnings	Standard deviation
Australia	3794	2153	3190	3059
Canada	3734	2304	3033	3854
Japan	4680	2150	3276	2880
Korea	3078	1457	1979	1495
Switzerland	6821	5225	5906	6861
United States	4502	2591	5078	7077

1. 2009 for Australia and Japan; 2008 for Canada; 2007 for Korea and the United States.

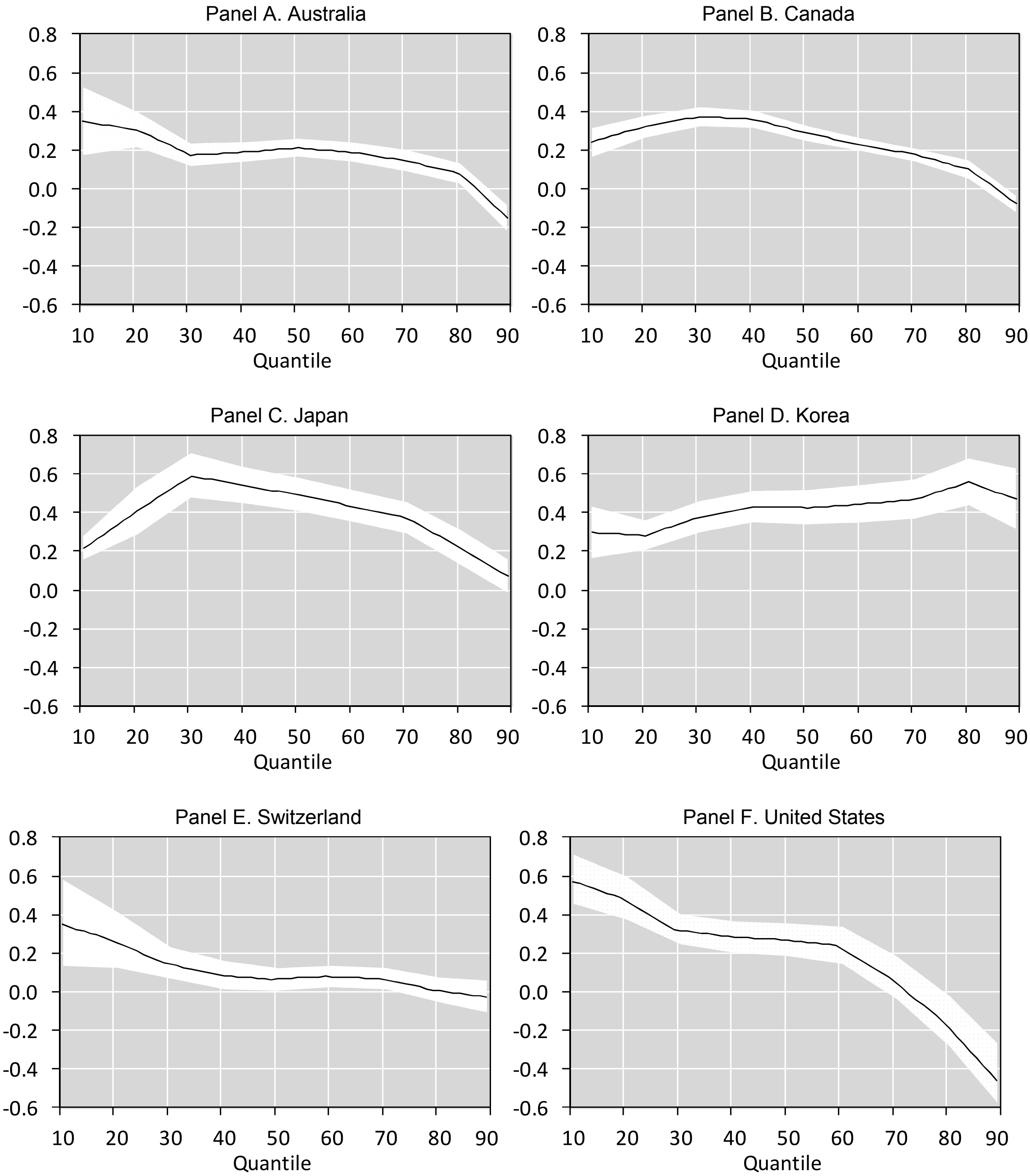
Source: Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia; Survey of Labour and Income Dynamics (SLID) for Canada; Japan Household Panel Survey (JHPS) for Japan; Korean Labour and Income Panel Study (KLIPS) for Korea; Panel Study on Income Dynamics (PSID) for the United States.

39. Based on the unconditional quantile regression results, it appears that higher union density would help to reduce earnings inequality among workers in Australia, Canada, Japan, Switzerland and the United States (Figure A2.11). In all five countries a broad-based (marginal) increase in union membership (spread evenly across the population) has a strong positive effect on the lower quantiles of the earnings distribution, while having no (or even a negative) effect on higher quantiles. For the United States, this finding confirms the results obtained by Firpo *et al.* (2009) using data from the Current Population Survey (CPS).[23] However, insofar as the same changes turned out to have adverse employment effects (see the discussion in the main paper), the inequality-reducing impact associated with the more compressed wage distribution would be reduced. The results obtained for Korea differ: there, union membership benefits all workers to a similar extent, so that a rise in unionisation does not alter the distribution of earnings.

23. Card *et al.* (2004), who, in their analysis of Canada, the United Kingdom and the United States, explicitly distinguish between men and women, show that unions reduce wage inequality among men, but not among women.

Figure A2.11. **The impact of union membership on the distribution of earnings (UQR estimates)**

Effect of raising the share of workers affiliated to a union by 1 percentage point on log earnings



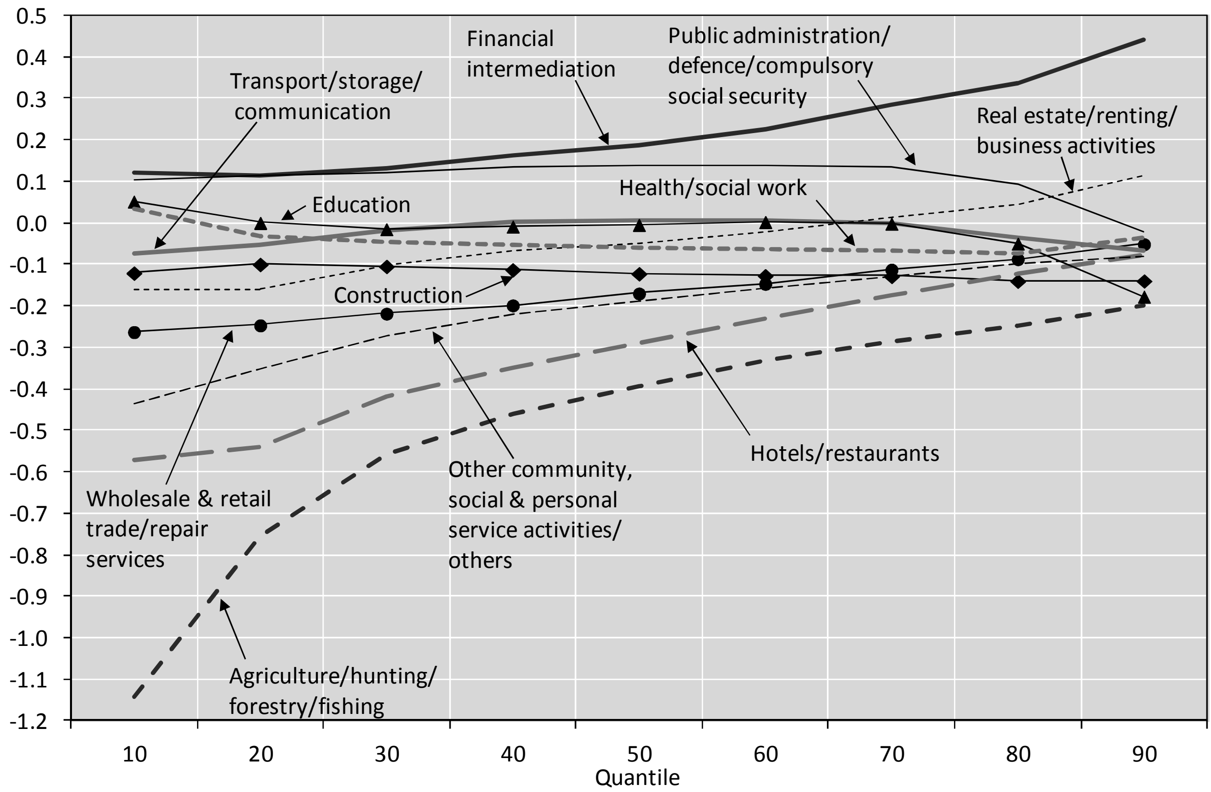
Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA), the Survey of Labour and Income Dynamics (SLID) for Canada, the Japan Household Panel Survey (JHPS), the Korean Labour and Income Panel Study (KLIPS), the Swiss Household Panel (SHP), and the Panel Study on Income Dynamics (PSID) for the United States.

Sector of employment

40. To explore whether the sector composition of the economy has an influence on earnings inequality, the baseline specification is augmented with eleven dummy variables, one for each of the following sectors: Agriculture/hunting/forestry/fishing; construction; wholesale and retail trade/repair services; hotels/restaurants; transport/storage/communication; financial intermediation; real estate/ renting/business activities; public administration/defense/compulsory social security; education; health/social work; other community, social and personal service activities/others. The omitted sector mining/quarrying/manufacturing/electricity, gas and water supply serves as the reference sector. The UQR results suggest that a shift in the sector composition would not in general have a large impact on the distribution of earnings. As shown in Figure A2.12, for most sectors the earnings effect is roughly constant along the earnings distribution. Four sectors, that show some variation along the earnings distribution, are agriculture/hunting/forestry/fishing, hotel/restaurants, other community, social and personal service activities/others and financial intermediation. A rise in the share of the first three sectors is associated with a decrease of earnings at the lower end of the earnings distribution. A rise of the share of financial intermediation implies higher inequality for a different reason: the earnings gain is concentrated at the higher end of the earnings distribution. In line with the rather small role played by the sector of employment in driving the distribution of earnings, the contribution of cross-country differences in the sector composition to cross-country differences in earnings inequality is in general fairly limited (Figure A2.7).

Figure A2.12. **The impact of the sector composition on the distribution of earnings (UQR estimates)**

Effect of increasing the share of a certain sector by 1 percentage point (relative to mining & quarrying/manufacturing/electricity gas & water supply) on log earnings, unweighted cross-country average



Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA), the Korean Labour and Income Panel Study (KLIPS), the Luxembourg Income Study (LIS) for Brazil and Israel, the Japan Household Panel Survey (JHPS), the Swiss Household Panel (SHP), the Panel Study of Income Dynamics (PSID) for the United States, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

41. A related question is the role played by the public sector. For the six countries for which data on the type of employer are available (Australia, Brazil, Japan, Korea, Switzerland and the United States), descriptive evidence indicates that average wages are generally higher and less dispersed in the public than in the private sector (Table A2.2). This is consistent with existing empirical evidence (*e.g.* García *et al.*, 2001). To go beyond simple descriptive statistics and control for other observable characteristics when comparing the dispersion of earnings of public versus private sector workers (for example, the higher dispersion of private sector wages could be because private sector workers are more diverse in terms of their education level), conditional quantile regressions are run on a specification that includes all baseline variables plus a dummy variable capturing whether a person is employed by the government or a government-related entity. For countries where no data on the type of employer are available, this dummy variable is created using information on an individual's sector of employment and information on the share of public employment by sector. The CQRs reveal that the lower dispersion of public sector earnings cannot be explained by other observed characteristics. It could be due to greater prevalence of centralised wage bargaining in that sector (Grimshaw, 2000), less reliance on performance-related pay (especially in Europe), or the purposeful use of public sector employment to achieve redistribution (Alesina *et al.*, 2000).[24]

Table A2.2. **A comparison of earnings among public and private sector workers**

USD, latest available year[1]

	Public sector		Private sector	
	Average earnings	Standard deviation	Average earnings	Standard deviation
Australia	3835	2267	3180	3043
Brazil	730	926	361	506
Japan	4877	2560	3416	3054
Korea	3028	1539	1912	1448
Switzerland	5896	3237	6137	3790
United States	4470	4045	4991	6796

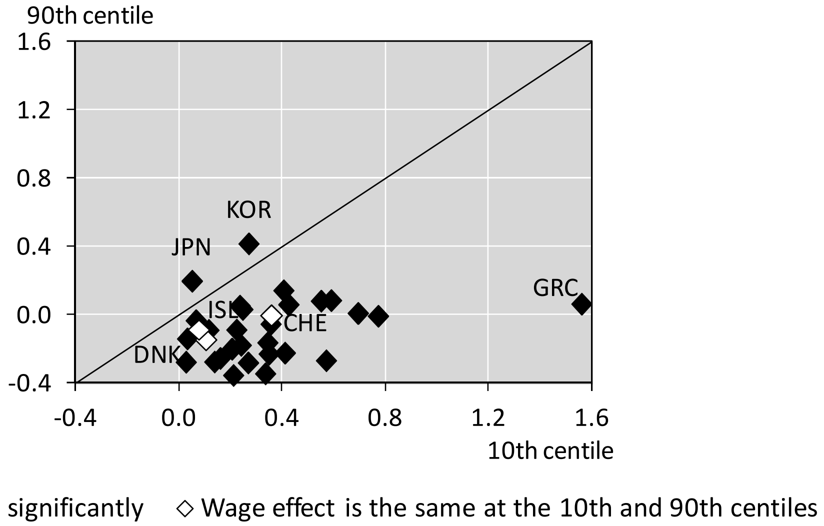
1. 2009 for Australia and Japan; 2008 for Switzerland; 2007 for Korea and the United States; 2006 for Brazil.
Source: Household Income and Labour Dynamics in Australia Survey (HILDA); Luxembourg Income Study (LIS) for Brazil; the Korean Labour and Income Panel Study (KLIPS); Japan Household Panel Survey (JHPS); Swiss Household Panel (SHP); Panel Study of Income Dynamics (PSID) for the United States.

42. The implications of the higher average level and the lower dispersion of public sector earnings for overall earnings inequality are explored using unconditional quantile regressions. They are a priori ambiguous, as already discussed in Box A2.1. In practice, the results indicate that in the majority of OECD countries a (marginal) rise in the public-sector employment share would tend to lower earnings inequality by raising earnings at the lower end of the earnings distribution while leaving those at the upper tail broadly unchanged or even reducing them (Figure A2.13). As with the results above regarding union membership or education, this finding should be interpreted with care as it ignores possible changes in the relative earnings of public and private sector workers that would result from such a shift. However, cross-country differences in the size of the public sector or in the public/private sector wage structures do not seem to play an important role in explaining cross-country differences in inequality (results not shown).

24. Alesina *et al.* (2000) propose a model of public sector employment, in which the latter is not chosen solely based on efficiency considerations, but based on its redistributive impact.

Figure A2.13. **The impact of public sector employment (UQR estimates)**

Effect of a 1 percentage point increase in the share of public sector employment on log earnings



Notes: A data point below (above) the 45 degree line indicates that a rise in the public sector employment share is associated with a fall (rise) in the q90/q10 ratio.

Source: UQR estimates using data from the Panel Study of Income Dynamics (PSID) for the United States, the Household Income and Labour Dynamics in Australia Survey (HILDA), the Survey of Labour and Income Dynamics (SLID) for Canada, the Korean Labour and Income Panel Study (KLIPS), the Luxembourg Income Study (LIS) for Brazil and Israel, the Japan Household Panel Survey (JHPS), the Swiss Household Panel (SHP), and the European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

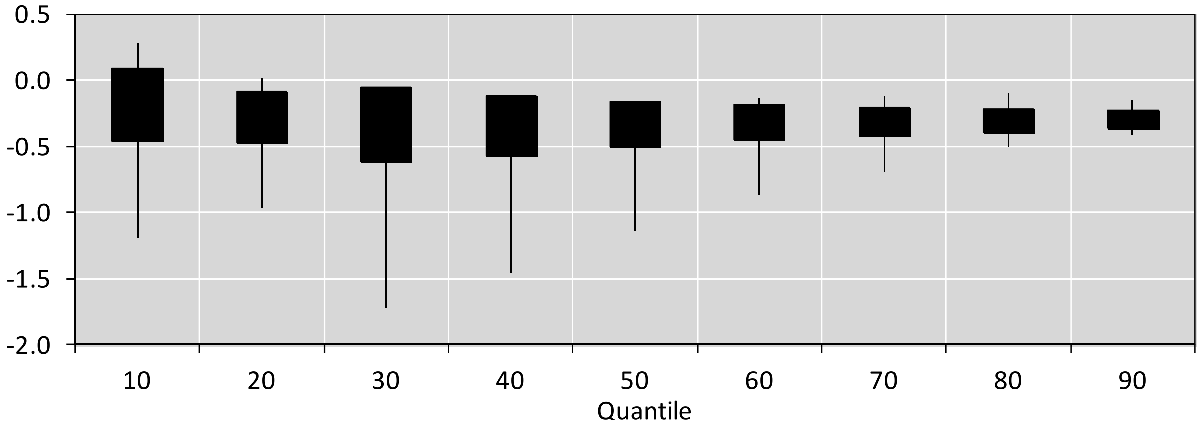
Gender inequality

43. Despite some decline over the past decades, gender differences in labour market performance are still striking in most OECD countries. Women are less likely to be employed than men and those who work typically earn less than their male counterparts (OECD, 2010). For all countries considered, the unconditional quantile regression estimates confirm that women earn less than men – the coefficient on the gender dummy in the baseline specification is significantly negative at almost all quantiles (Figure A2.14) – even after controlling for factors such as education and the number of working hours.[25] Whether this earnings gap is bigger at the lower or upper tail of the earnings distribution varies by country, with no clear overall pattern.

25. The same result is obtained using conditional quantile regressions, suggesting that the conclusion is robust to the choice of the estimation technique.

Figure A2.14. **The gender earnings gap (UQR estimates)**

Effect of being female relative to being male on the log earnings



Note: The thick bars depict the cross-country mean of the estimated effect +/- 1 standard deviation across countries, while the thin bars depict the cross-country maximum and minimum of the estimated effect

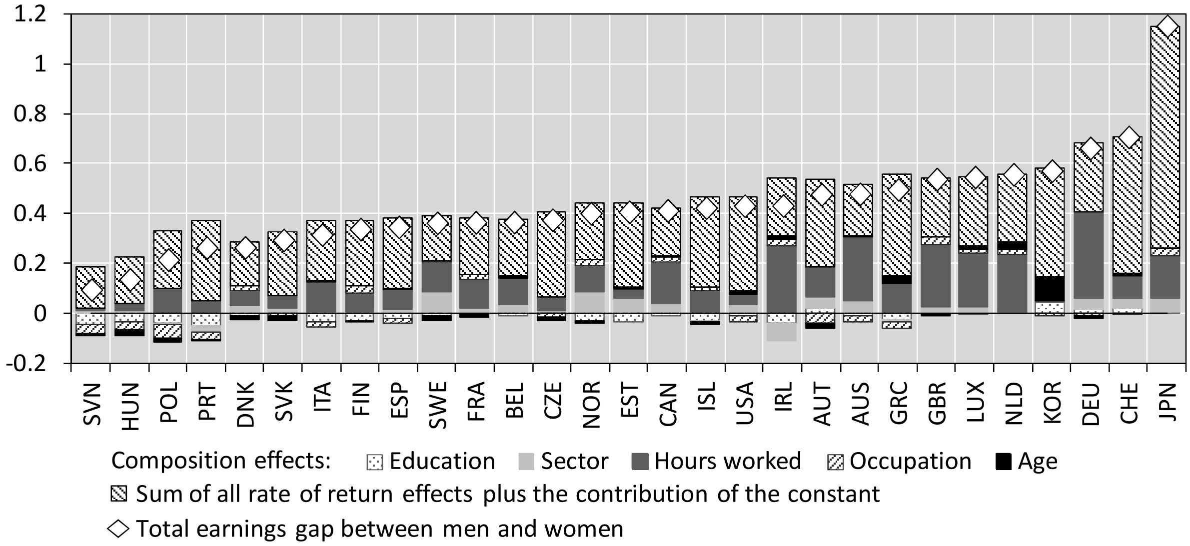
Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA), the Survey of Labour and Income Dynamics (SLID) for Canada, the Korean Labour and Income Panel Study (KLIPS), the Japan Household Panel Survey (JHPS), the Swiss Household Panel (SHP), and the European Union Statistics on Income and Living Conditions (EU-SILC) for 19 EU member countries as well as for Iceland and Norway.

44. Using an Oaxaca-Blinder-type decomposition (Oaxaca, 1973; Blinder, 1973) allows breaking down the gender earnings gap into differences in personal characteristics and differences in the returns to these characteristics. This decomposition is based on the estimation of two separate ordinary least square regressions: one for men and one for women. The following main results emerge from analysis (Figure A2.15):

- A large part of the gender earnings gap is explained by women's shorter working hours, most likely reflecting the fact that women take on more caring obligations for children and elderly relatives than men.
- Secondary and tertiary attainment rates of women are equal to or exceed those of men in most countries, resulting in a zero or negative contribution of education to the gender earnings gap.
- In most countries, gender differences in the sector of employment and the occupation raise the gender wage gap – notable exceptions being Ireland, Poland and Portugal. To the extent that differences in occupational choice do not result from personal preferences but from gender stereotyping they need to be addressed, for example by gearing curricula and teaching material to avoid gender stereotyping and by encouraging teachers to motivate girls to study science, technology, (financial) engineering and mathematics, where they are currently underrepresented (OECD, 2011).
- Rates of return to various characteristics also tend to be lower for women. While this effect picks up factors that are not properly controlled for in the regression (*e.g.* differences in fields of study that might be associated with different returns to tertiary education), it is also likely to reflect at least partially gender discrimination. To combat discrimination, it can help to have effective legal rules, for example by empowering well-resourced specialised bodies to investigate employers even in the absence of individual complaints and to take legal action against those who engage in discriminatory practices (OECD, 2011).

Figure A2.15. **Decomposition of the gender earnings gap**

2008 or latest available year[1]



1. 2009 for Australia and Japan; 2007 for France, Korea and the United States.

Source: Panel Study of Income Dynamics (PSID) for the United States; Household Income and Labour Dynamics in Australia Survey (HILDA); Survey of Labour and Income Dynamics (SLID) for Canada; Korean Labour and Income Panel Study (KLIPS); Japan Household Panel Survey (JHPS); Swiss Household Panel (SHP),; and European Union Statistics on Income and Living Conditions (EU-SILC) for the other countries.

Migration

45. Migration may influence labour earnings inequality both because immigrants alter the labour market outcomes of natives and because immigrants may fare differently in the labour market.[26] The latter issue is investigated here by augmenting the baseline specification with two alternative dummy variables, capturing whether a person has foreign citizenship and whether he/she was born in a foreign country, respectively.[27] The unconditional quantile regression results point to a substantial cross-country variation in the earnings gap between foreigners and natives (Figure A2.16).[28] Focusing on the country-of-birth dummy variable, there is no significant earnings gap between natives and foreigners for most parts of the earnings distribution in 6 out of the 21 countries considered (Australia, Germany, Norway, Portugal, Slovenia and Switzerland). In the remaining countries foreigners typically earn less than natives, with the exception of Brazil where they tend to earn more – even controlling for other factors such as education. Whether the size of this earnings gap differs across quantiles depends again widely on the country considered. The large cross-country differences in the earnings gap between natives and foreigners have been found to reflect in part differences in the structure of the immigrant population (in terms of country of

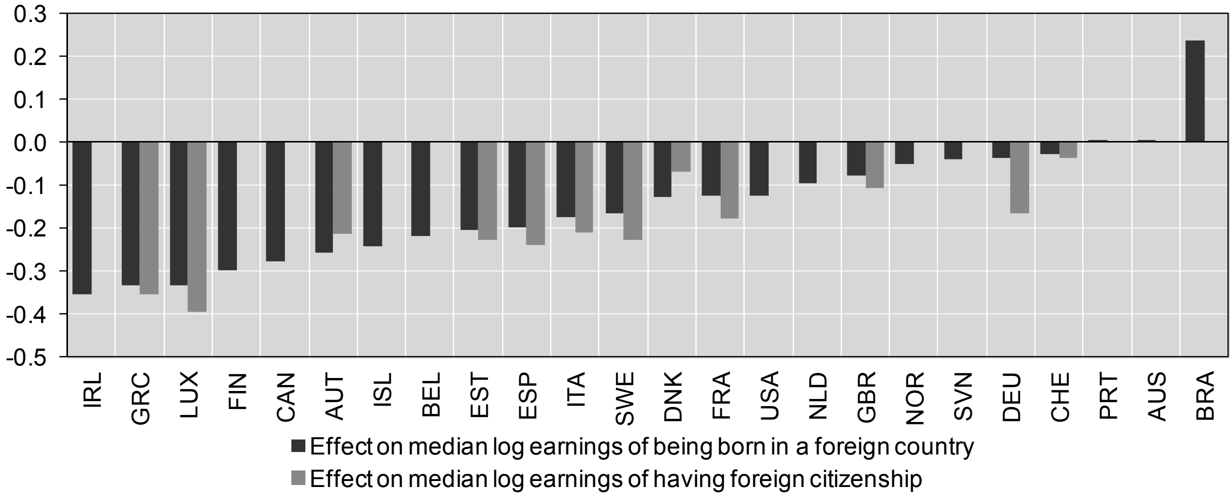
26. At the same time, the level of earnings inequality in the destination country (relative to that in the source country) may influence migration flows (see Liebig and Sousa-Poza, 2004, for a brief overview of the theoretical underpinnings as well as empirical evidence).

27. Data on the two dummy variables are only available for a subset of countries. For all countries for which the analysis is based on the SILC dataset, foreign citizenship means non-EU citizenship and being born in a foreign country means being born outside of the EU. The differences (in terms of education system, culture, etc.) are assumed to be bigger between those born outside the EU and those born inside the EU, as compared to the differences between different EU countries

28. Conditional quantile regressions yield the same conclusion.

origin, timing of immigration or motivation) and differences in countries' policy settings (Jean *et al.*, 2010).[29]

Figure A2.16. **The earnings gap between natives and immigrants (UQR estimates)**

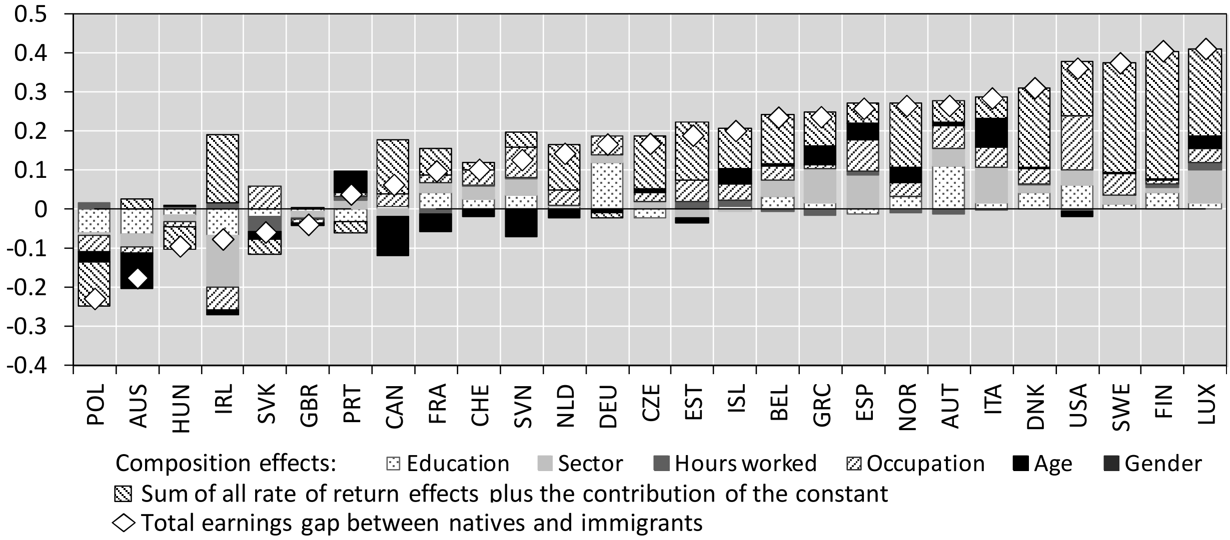


Source: UQR estimates for employed individuals using data from the Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia, the Survey of Labour and Income Dynamics (SLID) for Canada, the Luxembourg Income Study (LIS) for Brazil, the Swiss Household Panel (SHP) for Switzerland, the Panel Study of Income Dynamics (PSID) for the United States, and the European Union Statistics on Income and Living Conditions (EU-SILC) for 21 EU member countries as well as for Iceland and Norway.

46. An Oaxaca-Blinder decomposition of the average earnings gap between natives and foreign-born individuals (*i.e.* the average across all quantiles) into differences in personal characteristics and differences in the returns to these characteristics also points to considerable cross-country heterogeneity (Figure A2.17). Two factors that contribute positively to the gap in many (but not all) countries are a lower average level of education of immigrants and an unfavourable choice of the sector of employment and the occupation, with immigrants working in sectors and occupations that pay less. In addition, rate-of-return effects contribute positively to the gap, meaning that migrants earn less than natives even when they have exactly the same characteristics (in terms of age, gender, education level and so on). This may, for example, reflect firms' difficulties in properly assessing qualifications obtained in a foreign country, immigrants' lack of work experience in the host country (which is an important channel of integration, but not controlled for in the regression), and also discrimination. Targeted policies such as (occupation-specific) language courses and transparent systems of recognising foreign qualifications may therefore help to reduce the gap in labour market performance, as well as employment-friendly labour market policies as these tend to benefit immigrants disproportionately (Causa and Jean, 2007).

29. In the case of the United States, being black has a negative impact on earnings that is more pronounced for higher quantiles. No evidence of an earning gap is found for other colours. In the case of Brazil, a similar pattern penalizes black workers. In addition, indigenous individuals and individuals of mixed origin also suffer from lower earnings, with the impact stronger for higher quantiles.

Figure A2.17. **Decomposition of the earnings gap between immigrants and natives**
2003 to latest available year



Note: Individuals are defined as immigrants if they are born outside the country or, in case of EU member countries, outside the EU. 2009 for Australia and Japan; 2008 for Canada.

Source: Panel Study of Income Dynamics (PSID) for the United States, Household Income and Labour Dynamics in Australia Survey (HILDA) for Australia, Survey of Labour and Income Dynamics (SLID) for Canada, Swiss Household Panel (SHP) for Switzerland, and European Union Statistics on Income and Living Conditions (EU-SILC) for the other countries.

BIBLIOGRAPHY

Acemoglu, D. and J. Angrist (1999), "How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws", *NBER Working Paper*, No. 7444, National Bureau of Economic Research.

Alesina, A., R. Baqir and W. Easterly (2000), "Redistributive Public Employment", *Journal of Urban Economics*, Vol. 48, pp. 219-241.

Blinder, A. S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates", *The Journal of Human Resources*, Vol. 8, pp. 436-455.

Burkhauser, R.V., B.A. Butrica, M.C. Daly and D.R. Lillard (2000), "The Cross-National Equivalent File: A Product of Cross-National Research", in: *Soziale Sicherung in einer dynamischen Gesellschaft* (Social Insurance in a Dynamic Society), Irene Becker, I., N. Ott and G. Rolf (eds.)

Causa, O. and S. Jean (2007), "Integration of Immigrants in OECD Countries: Do Policies Matter?", *OECD Economics Department Working Paper*, No. 564.

Dur, R. A. J. and C. N. Teulings (2004), "Are Education Subsidies an Efficient Redistributive Device?", in *Labor market, institutions and public regulation*, Agell, J., M. Keen ans A. Weichenrieder (eds.), pp. 123-161, Cambridge (MA), MIT Press.

Firpo, S., N. Fortin and T. Lemieux (2007a) "Unconditional Quantile Regressions", *NBER Technical Working Paper*, No. 339, National Bureau of Economic Research.

Firpo, S., N. Fortin and T. Lemieux (2007*b*), "Decomposing Wage Distributions using Re-centered Influence Function Regressions", unpublished manuscript, University of British Columbia, June.

Firpo, S., N. Fortin and T. Lemieux (2009), "Unconditional Quantile Regressions", *Econometrica*, Vol. 77, pp. 953–973, May.Fortin, N., T. Lemieux, T. and S. Firpo (2011), "Decomposition Methods in Economics", in: *Handbook of Labor Economics*, Elsevier.

García, J., P.J. Hernandez and A. Lopez-Nicolas (2001), "How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression", *Empirical Economics*, Vol. 26, No. 1, pp. 149-167.

Gosling, A. and S. Machin (1995), "Trade Unions and the Dispersion of Earnings in British Establishments, 1980-90", *Oxford Bulletin of Economics and Statistics*, Vol. 57, No. 2, pp. 167-184.

Grimshaw, D. (2000), "Public Sector Employment, Wage Inequality and the Gender Pay Ratio in the UK", *International Review of Applied Economics*, Vol. 14, No. 4, pp. 427-448.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," Journal of the American Statistical Association, 60, pp. 383–393.

Hampel F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel (2005), *Robust Statistics: The Approach Based on Influence Functions,* Wiley – Interscience.

Hartog, J., P. Pereira and J.A. Vieira (2001), “Changing Returns to Education in Portugal during the 1980s and Early 1990s: OLS and Quantile Regression Estimators”, *Applied Economics*, Vol. 33, pp. 1021-2037.

Hendel, I., J. Shapiro and P. Willen (2005), "Educational Opportunity and Income Inequality," *Journal of Public Economics*, Vol. 89(5-6), pp. 841-870, June.

Jean, S., O. Causa, M. Jiménez and I. Wanner (2010), “Migration and Labour Market Outcomes in OECD Countries”, *OECD Economic Studies*, Vol. 2010, OECD Publishing.

Knight, J.B. and R.H. Sabot (1983), “Educational Expansion and the Kuznets Effect”, *The American Economic Review*, Vol. 73, No. 5, pp. 1132-1136.

Koencker R. (2005), *Quantile Regression*, Cambridge University Press.

Koencker R. and G. Basset (1978), “Regression Quantiles”, *Econometrica*, Vol. 46, pp. 33-50.

Koenker, R. and K. F. Hallock, (2001) "Quantile Regression", *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143–156.

Machin, S. (1997), “The Decline of Labour Market Institutions and the Rise in Wage Inequality in Britain”, *European Economic Review*, Vol. 41, No. 3-5, pp. 647-657.

Mincer J. (1958), “Investment in Human Capital and the Personal Income Distribution”, *Journal of Political Economy*, Vol. 66, No. 4, pp 281-302.

Mincer, J. (1974, *Schooling, Experience, and Earnings*, NBER Press.

Oaxaca, R. (1973), “Male-Female Wage Differentials in Urban Labor Markets”, *International Economic Review*, Vol. 14, pp. 693–709.

Polachek, S. W. (2007), "Earnings Over the Lifecycle: The Mincer Earnings Function and Its Applications," *IZA Discussion Papers*, No. 3181, Institute for the Study of Labor (IZA).

OECD (2010), Education at a Glance 2010, OECD Publishing.

OECD (2011), Report on the Gender Initiative: Gender Equality in Education, Employment and Entrepreneurship, meeting of the OECD Council at Ministerial Level, 25-26 May 2011.

Robinson, S. (1976), “Note on the U Hypothesis Relating Income Inequality and Economic Development”, *The American Economic Review*, Vol. 66, June, pp. 437-440.

Rothe, C. (2010), “Non-parametric Estimation of Distributional Policy Effects”, *Journal of Econometrics*, Vol. 55, No. 1, pp. 56-70.