**Methodological Appendix**

In what follows, we present some technical details of the approach we adopt. A.1. summarizes and provides some examples of the non-parametric estimate approach for the distribution of wages; A.2. summarizes the inverse probability weighted approach, which allows us to answer some “what” questions related to the economies; A.3. discusses the choice of the set of regressors to model the probability of being a middle manager for international comparison and presents some empirical exercises; A.4. presents some empirical exercises; and A.5. discusses estimations of “what if” questions related to the economies analysed and a reference economy.

**A.1. Non-parametric approach to estimating the distribution of wages**

The appeal of estimating and comparing the distribution (of the *ln* of) wages relies especially on the possibility of theorizing directly about the entire distribution of wages, which, in turn, negates the need for a representative agent.

Distributions can be estimated using parametric and nonparametric frameworks. However, the former typically make use of only a given number of moments (usually the first two - the conditional mean and the variance), while the latter allows for analysis of the entire distribution of the variables of interest. The use of a nonparametric approach when estimating distributions allows us to go beyond the idea of a representative agent. This explains its appeal in analyses of behaviour in large and heterogeneous samples of individuals, as in our case.

Following Silverman (1986), let us define a wide class of nonparametric density estimators (the Rosenblatt-Parzen kernel density estimator) as:

1. ,

where *N* is the number of observations, *w* is the ln of wage and *h* is the bandwidth. *K(.)* is the kernel density function which satisfies:

1. .

Many alternative kernel functions can be used, each of which offers different advantages and disadvantages, especially in terms of efficiency and smoothing power. In our analysis, we make use of the Gaussian kernel, which is the height of the standard normal distribution evaluated at (*w* – *wi*) given the bandwidth *h*. We use the Gaussian kernel because of its property of monotonicity of peaks and valleys with respect to changes in the smoothing parameter (Sheather, 2004).

*h* is the crucial decision in the estimation of distributions using a nonparametric approach. A number of bandwidth selectors is available. In what follows, we report the results based on the average of the optimal rule of thumb proposed by Silverman (1986) since it is generally acknowledged to be the best performing, especially when the kernel density function applied to the data is Gaussian, as in our case. It should be noted that, given the large numbers in our sample, the choice of kernel function and bandwidth selector are unlikely to affect the results significantly. Marron and Schmitz (1992) suggest that comparisons across distributions should be made under the condition that the same kernel *K()*, and the same smoothing parameter *h* are adopted in Eq. (1).

In Figure A.1., Panels 1 and 2 report the results for the sample of employees and the subsample of middle managers and productive employees in the UK. Panels 3 and 4 report the results for Germany.

**A.2. Answering “what if” questions: the semi-parametric approach to estimate distributions**

The differences across the distributions in Panel 2 (and Panel 4 for Germany) cannot be taken as being the premiums for middle management in the UK and Germany, because they do not take account of the difference in the personal characteristic endowments of middle managers with respect to the control group. In other words, use of the fully nonparametric density estimation approach makes it difficult to perform hypothesis testing exercises.

Therefore, as suggested by DiNardo, Fortin and Lemieux (1996), we compare the estimate of the nonparametric distribution of wages with the distribution estimated under some “what if” questions, the so-called counterfactual distributions, that is, the distributions that would have prevailed in the economy if the impact of “being a middle manager” were removed.



Since Oaxaca (1973) and Blinder (1973), a number of methods have been proposed to decompose differences in the means of an outcome variable that are attributable to a number of independent variables. We are interested in the distribution of wages and rely on methods for decomposing distributional parameters other than the mean. In our work, we adopt the Inverse Probability Weighting (IPW) originally proposed by DiNardo, Fortin and Lemieux (1996) in the context of the gender wage gap literature. This method is preferred if the aim, as in our case, is decomposition of the overall difference in the distribution of the outcome variable into its explained and unexplained components, which is described as aggregate decomposition. Its main advantage lies in its simplicity. Let:

1. 

be the actual distribution of wages, where *w* is the wage,  is the conditional density of wages and  is the density of the vector of *x* characteristics, among which the dummy variable *m* is equal to 1 if the individual is a middle manager and zero otherwise. The approach consists of comparing the distribution in Eq. (1) with the distribution of wages that would prevail if, other things being equal, none of the individuals was a middle manager:

1. ,

where *nm* denotes ‘not managers’ and  is the conditional density of wages associated to this group of observations. The wage premium to management is taken as the horizontal difference between the distributions in Eq. (1), the actual distribution of wages, and Eq. (2), the counterfactual distribution of wages, that is, the distribution of wages that would prevail if none of the individuals was a middle manager, *other things being equal*. This condition is obtained by using in Eq. (2) the density of the vector of *x* characteristics  associated to the entire sample instead of , and the density of the *x* characteristics associated to the subsample of observations for which *m=0*.

The distribution in Eq. (2) is obtained following DiNardo, Fortin and Lemieux (1996), who suggest exploiting Bayes law to obtain its estimate:

1. ,

where  and  are the unconditional and conditional probabilities respectivelyof not being a middle manager. Use Eq. (3) in Eq. (2) to obtain:

1. ,

where:

1. .

The authors note that Eq. (4) is the distribution of wages associated to the *nm* subsample up to the unknown re-weighting function, . This reweighting function is the crucial element in the decomposition of the distribution of wages and is built by estimating a probit model to obtain the conditional probability of supervising other employees given *x*:

1. ,

and using the predicted values:

1. ,

where  is the fitted probability.

Eq. (7) gives the estimate of  needed to reweight the distribution of wages for the subsample of employees who are not middle managers in Eq. (2). These weights are used to compute the otherwise fully non-parametric Rosenblatt-Parzen kernel density estimator:

1. .

The decomposition is performed for the UK and Germany separately. Comparison of these results provides evidence in favour of or against our hypothesis (1).

**A.3. Adapting the method to international comparisons: which control variables?**

Apart from using the same kernel and the same smoothing parameter for the two samples, the decomposition exercise requires the same specification of the probit model in Eq. (8) for the UK and Germany. Therefore, in estimating the models, we do not apply a general-to-specific approach since this is likely to make the reduced form economy-specific, and we test the joint significance of the following groups of variables:

1. variables for individual characteristics: education (4 categories: at most lower secondary school, upper secondary, post-secondary education, at least tertiary); work experience (and its squared, cubic and quartic values); gender; marital status; and two dummies for citizenship type (national/non national, European/non-European);
2. variables associated with job characteristics (part time, full time, temporary, permanent);
3. variables associated with firm characteristics (size measured by 3 dummy variables, sector of economic activity based on 13 dummy variables);
4. variables for individual skills (4 dummy variables measuring the skills required for the task).

The variables described in (i), (ii) and (iii) are drawn from DiNardo, Fortin and Lemieux (1996). Combined with the regressors described in (iv), they help to deal with the likely heterogeneity in the data and potential self-selection mechanisms. For instance, if the individual skills of middle managers and other employees differ systematically because the former self select into more rewarding tasks (Card, 1999), then the measure we estimate is an unknown combination of the premium to management and the reward for individual skills endowments. The decision to supervise other employees or not depends upon the abilities required for the particular responsibility, which will result in skilled individuals systematically self-selecting into that job (Cameron and Heckman, 1998).

The literature on the gender wage gap suggests that this problem should be handled by adopting a panel data approach (Fortin, Lemieux and Firpo, 2010). However, in our context, this is not feasible since the EU-SILC survey data are cross sectional. In our context, using panel data would be unattractive also because: 1) the estimating probability model is non-linear so the assumption of separability is unlikely to hold; and 2) disentangling individual effects is problematic since individual fixed effects can be computed only if the individual experiences both statuses (being a supervisor and not being a supervisor) over the (eventually available) time span. Therefore, we follow Picchio and Mussida (2011) and use the International Standard Classification of Occupations (ISCO-88). These variables are associated with the type of job chosen by the individual. The categories range from jobs that require a relatively low level of skill, such as plant and machinery operators and assemblers, to jobs that require professional skills such as legislators and senior officials, and CEO/non-CEO positions. The variables refer explicitly to the skill level required for the job:

the basis for the classification in the ISCO-88 scheme is the nature of the job itself and the level of skills required. A job is defined as the set of tasks and duties to be performed. Skills are the abilities to carry out the tasks and duties of a job. Skills consist of two dimensions: skills level and domain specialization (EU-SILC 2009).

We take the number of accurate predictions for the different models as the criteria for choosing among the alternative reduced form models. We note that, because the counterfactual wage premium is computed as the difference between premiums, the potential residual bias due to self-selection mechanisms is further dissipated unless there are reasons to expect that skilled individuals working in the UK self-select into jobs differently from German individuals.

**A4. Wage as a proxy for ability, and endogeneity**

The four sets of variables described in (i), (ii), (iii) and (iv) help to deal with the likely heterogeneity in the data, and self-selection mechanism. We note that the decision to supervise other employees depends also upon the wage paid for this responsibility. In turn, this may result in skilled individuals systematically self-selecting into that job (Cameron and Heckman, 1998). Hence, another variable that could be used to control for self-selection into middle management jobs is the wage attached to the job. It is worth studying the impact of wage on the forecasting ability of the reweighting functions when added to the set of regressors. Results reported in Table A1 are for the auxiliary probit regression when (the log of) wage is added to the set of regressors.



Table A1 Columns (a) and (b) report the results for Germany. As above, we test for the joint significance of groups of variables using the *F*-statistic (*p*-values are in brackets). Column (a) reports the results for the model with wage included in the sets of variables for individual, job and firm characteristics. As expected, wage is statistically significant. If both the set of ISCO variables and the log of wage are included in the estimated models (Column b), the *F*-statistics associated to the statistical significance of (the log of) wage and the joint statistical significance of the coefficients of the ISCO dummies, reduce relative to their counterparts in Column (b) Table A1 in the main part of the paper. This is likely due to correlation between the ISCO variables and the log of wages, suggesting that the former, as expected, help to control for the self-selection mechanism. However, with respect to the results for our preferred model reported in Column (b) in Table A1, both these models have a significantly lower percentage of correct predictions; recall, that this is what these auxiliary regressions are meant to help with. In Table A1, Columns (c) and (d) report the results applied to the UK observations. The pattern is similar to the pattern for Germany. As before, adding the (log of) wage to the set of regressors reduces the percentage of correct predictions and, in the case of the UK, also the pseudo-R² of the model. The results in Column (d) show that the set of ISCO variables is correlated to wage and they have higher *F*-statistics.

The evidence that the models that include in the set of regressors the log of wage, have a lower percentage of correct predictions, suggesting the use of the models where this variable is excluded from the set of regressors. This adds to the problem of finding a credible instrumenting strategy for wage. Wage is likely to be jointly determined with the probability of being a supervisor and, therefore, its use is conditional on the availability of the set of instruments to be used in the instrumental variables approach (Currie and Madrian, 1999). The problem lies in identifying one or more credible instruments. This is difficult in our context because the instrument must be valid (and the same) for the two economies under investigation.

The literature suggests adopting variables measuring individual health. It is suggested that poor health has substantive effects on compensation and labour market participation. It has been suggested also that the relationship between health status and the task the individual performs is not strong, the idea being that health is likely to affect all the tasks the individual chooses, in a similar manner (Currie and Madrian, 1999). In the EU-SILC survey, physical well-being is measured by limitation on activities due to health problems, and general health (including health status and chronic illness or condition) and is summarized by a variable that takes values from 1 to 5, with higher values indicating poorer health. The variable is not statistically significant in the probit model, suggesting that the exclusion restriction is likely to hold. Moreover, correlation analysis supports this conclusion since the variable is found to be highly correlated to the wage and uncorrelated to the probability of being a supervisor. Table A2 reports the results with individual health status used as an instrument for (the log of) wage.



Table A2 Column (a) reports the results for Germany. Wage is statistically significant at the 1% s.l. The Hausman and Wu test for exogeneity rejects the null hypothesis that the variable can be treated as exogenous. The percentage of correct predictions is lower than that reported for our preferred model. The results in Column (b) show that, if the set of dummies controlling for unobserved skills is added to the set of regressors, the percentage of correct predictions reduces further, and the coefficient associated to (the log of) wage also reduces. The test for exogeneity does not reject the null that wage can be treated as exogenous. Columns (c) and (d) report results for the UK observations. If we instrument the (log of) wage we find that the null hypothesis of the Hausman and Wu test is not rejected. This suggests that, despite the proposal in Currie and Madrian (1999), the necessary exclusion restriction is unlikely to hold for both the UK and Germany. The log of wage is not statistically significant.

**A.5. Answering “what if” using questions related to the economies with the UK as the reference economy**

To test hypothesis 2, we need a measure of how much a German middle manager would earn if, all other things being equal, s/he was performing the same task in the UK. This amount, when compared to the management premium actually earned, will reveal the impact of the German compared to the UK institutional background.

Fortin, Lemieux and Firpo (2010) conduct this exercise in two steps. The first step consists of estimating the counterfactual distribution that would prevail if German employees were working in the UK, all else being equal:

1. .

As above, the key element in comparing across samples (in this case, countries) is the reweighting function, which keeps all the conditioning variables as in the UK. Then, the reweighting factor for the observations in Germany is:

1. .

They suggest implementing the decomposition by pooling the data for Germany and the UK, and running a probit model for the probability of being employed in the Germany given the set of characteristics *x*:

1. .

and, using the sample proportions  and , they construct the distribution of wages that would have prevailed were all the German employees working in the UK:

1. .

where two unconditional probabilities of the first ratio,  and , are equal to the percentage of observations for the German economy over the percentage of observations for the UK economy. This vector of weights, when applied to the sample of German employees, gives:

1. ,

which is an estimate of the wage distribution that would prevail in Germany under UK institutions. The first step in the exercise provides an evaluation of the role of institutional background, that is, how much any German worker would earn in the UK. A similar exercise can be found in DiNardo, Fortin and Lemieux (1996) for the case of the gender wage gap. It focuses on a single economy and constructs counterfactual distributions at different points in time. The impact of a particular factor on changes to the wage distribution over time is constructed by considering the counterfactual state of the world where the distribution of this factor remains fixed across time. Blau and Kahn (1996), again in the context of the gender wage gap literature, propose a comparison across economies. As in our case, they adopt the UK economy as the benchmark and investigate what would be the average wage premium for a German woman working in the British institutional context.

The second step consists of estimating how much a German middle manager would earn in the UK, other things being constant. This vector of weights comes from the interaction of the vector of weights in Eq. (12) and Eq. (7), from which we deduce the reweighting function (Blau and Kahn, 1996; Gottschalk and Joyce, 1998; Katz and Autor, 1999):

1. .

which, when applied to the sample of German employees, gives:

1. ,

This provides the premium that would be paid to German middle managers working in the UK. The counterfactual wage premium to management is taken as the horizontal difference between the distributions in Eq. (1), the actual distribution of wages in Germany, and Eq. (15). In turn, the difference between the wage premium German middle managers are paid and the counterfactual wage premium they would receive if UK institutions prevailed, provides a measure of the role of the institutional context in the German economy.

**References**

Blau FD, Kahn LM. 1996. Wage structure and gender earnings differentials(1996) - An international comparisson.pdf. *Economica, New Series* **63**(250): S29–S62.

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

Cameron V, Heckman JJ. 1998. Life cycle schooling and dynamics selection bias: Models and evidence for five cohorts of American males. *The Journal of Political Economy* **106**(2): 262–333.

Card D. 1999. The causal effect of education on earnings. In *Handbook of Labor Economics*, 3: 1801–1863.

Fortin N, Lemieux T, Firpo S. 2010. Decomposition methods in ecnomics. *NBER Working Paper No. 16045* : 1–118. Available at: http://www.nber.org/papers/w16045.

Gottschalk P, Joyce M. 1998. Cross-national differences in the rise in earnings inequality: Market and institutional factors. *Review of Economics and Statistics* **80**(March): 489–502.

Katz LF, Autor DH. 1999. Changes in the wage structure and earnings inequality. *Handbook of Labor Economics* **3**: 1463–1555.

Marron JS, Schmitz H-P. 1992. Simultaneous Density Estimation of Several Income Distributions. *Econometric Theory*. Cambridge University Press **8**(4): 476–488. Available at: http://www.journals.cambridge.org/abstract\_S0266466600013177.

Oaxaca Ronald. 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*.

Picchio M, Mussida C. 2011. Gender wage gap: A semi-parametric approach with sample selection correction. *Labour Economics* **18**(5): 564–578.