Unobserved Worker Ability, Firm Heterogeneity, and the Returns to Schooling and Training[†]

Ana Sofia Lopes* and Paulino Teixeira**

* Departamento de Gestão e Economia, ESTG/Instituto Politécnico de Leiria, Portugal
** GEMF/Faculdade de Economia, Universidade de Coimbra, Portugal

January 2009

Keywords: Human Capital, Unobserved Heterogeneity, Earnings, LEED.
JEL Codes: J24, J31, C23, C81

Corresponding author:
Ana Sofia Lopes
Instituto Politécnico de Leiria
Campus 2 – Morro do Lena – Alto do Vieiro
Apartado 4163 | 2411-901 Leiria – Portugal
Email: analopes@estg.ipleiria.pt

[†] We are very grateful for the comments and suggestions made by António Alberto Santos, José Murteira, and Óscar Lourenço.

Abstract

It is well known that unobserved heterogeneity across workers and firms seriously impacts the computation of the determinants of individual earnings in standard human capital earnings functions. Following the tradition of AKM (Abowd, Kramarz, and Margolis, 1999), this paper offers an alternative way of controlling unknown worker and firm heterogeneity by taking full advantage of a matched employee-employer dataset based on two key Portuguese micro databases. Our modelling strategy assumes that the gap between individual and firm average wages, unexplained by differences in observable characteristics, gives the extent to which the unobserved ability of a given individual deviates from the unobserved worker average ability in the firm. This methodology has, in particular, the advantage of not relying exclusively on information on job switchers to identify worker and firm effects, thus avoiding any bias arising from endogenous worker mobility. Another important aspect of our treatment is that it allows the estimation of worker effects without risk of contamination from firm effects. To test our modelling we use an original 2-year longitudinal LEED dataset, comprising of more than 400 thousand workers and 1,500 firms in each year. We focus on two separate sets of individuals (i.e. stayers and switchers) and provide a variety of robustness tests, including replication of the original AKM methodology. After controlling worker and firm effects, our results show that the acquisition of schooling, labor market experience, and training, inter al., pays off. Moreover, we do find evidence of a large bias in standard OLS return rates to typical covariates. Evidence from Monte Carlo simulation and bootstrapping also shows that our estimated rates of return to human capital do not seem to be sensitive to changes in various assumptions. Our study does provide therefore further evidence that a wide set of individual and firm characteristics is crucial to understanding the true role of human capital variables in labor markets.

The accident of birth plays a powerful role in explaining variability in lifetime income.
James J. Heckman (2008)

1. Introduction

Measuring human capital is essential to understanding the determinants of individual earnings in the labor market. However, and despite substantial improvement in quality of available micro datasets, there are still sizeable differences in productivity and wages across firms and individuals requiring further explanation. The most common indeed is to observe workers with apparently identical attributes employed in apparently similar firms earning different wages. This discrepancy must of course be due to unmeasured differences both at firm and worker level.

The omission of relevant variables put human capital earnings functions under considerable stress as observable and unobservable attributes are likely to be correlated. For instance, workers with higher intrinsic abilities are expected to select themselves into higher levels of schooling. In this case, schooling is deemed to be endogenous, which means that the corresponding OLS estimate will reflect the direct effect of schooling on wages as well as a self-selection effect. In the limit, this contamination invalidates any meaningful interpretation of regression coefficients.

In this study, we follow Abowd, Kramarz, and Margolis (1999) – AKM hereafter – tradition as we deliberately try to control for firm and worker (unobserved) effects. But by using a longitudinal LEED dataset, obtained from matching *Quadros de Pessoal* and *Balanço Social*, our approach follows a different route. Indeed, we develop an original modelling in which we try to take full advantage of the fact that in our dataset we can observe not only a rather comprehensive set of individual- and firm-level characteristics – including firm-provided training – but also follow individuals longitudinally.

Firstly, we start by considering a *Mincerian* model to analyse the relation between the individual hourly wage and a set of observable worker and firm attributes. Secondly, we use a similar model, run at firm level, to predict firm average wages to next assume that the difference between the expected wage and the expected firm average wage is explained by the gap in observable characteristics between the individual and the firm average. Then, once taken the observed characteristics into account, any difference left is attributed to differences in unobserved heterogeneity. We also assume that the firm unobserved effect contains workers' average ability, plus a firm-specific effect.

Our methodology has some interesting aspects which should be mentioned up front. In the first place, it has the advantage of not relying exclusively on switchers to capture unobserved effects, which means that we avoid any bias arising from endogenous mobility. Secondly, our treatment allows the estimation of worker effects without running the risk of contamination from firm unobserved effects. Finally, our proposed route is easy to implement in standard packages such as STATA.

To test our modelling, we use an original 2-period (year) longitudinal LEED dataset, comprising of more than 400 thousands workers and 1,500 firms in each year, and focus on two main sets of separate individuals – stayers and switchers. According to our estimates, the correlation between observed and unobserved attributes implies an upward bias in the standard (OLS) 'return to education' of roughly 80 percent. Returns to training and labor market experience in standard OLS seem also highly contaminated by the omission of unobserved heterogeneity. In turn, our results point to a substantial reduction in the gender gap once worker and firm effects are taken into account.

The sensitivity of the results is examined using Monte-Carlo simulation and bootstrapping. We also apply the original AKM methodology to our data to obtain some useful benchmarking.

This article is organised as follows. In the next section we present the modelling strategy and in the third section we describe the construction of our longitudinal LEED dataset and the corresponding subsamples of stayers and switchers. Section 4 presents the results and a whole set of robustness tests, including Monte-Carlo and bootstrap. The main conclusions are drawn in section 5.

2. Modelling

2.1 Measuring worker and firm unobserved heterogeneity

Let us start with a standard Mincer (1974) formulation in which the individual (log) wage is a function of a set of (observable) employee and employer attributes, that is:

$$Ln\ w_{it} = X_{it}\beta + Z_{j(i)t}\gamma + u_{it}, \tag{1.1}$$

where $Ln\ w_{it}$ denotes the logarithm of the wage of individual (worker) *i*. X_{it} is the vector of his/her observable characteristics, $Z_{j(i)t}$ contains the observable characteristics of firm *j* – the firm in which worker *i* is employed in period *t* – and u_{it} denotes the error term.[1] Clearly, in model (1.1) u_{it} is not necessarily independent and identically distributed as it includes unobservable characteristics of workers and firms that may well be correlated with the observed variables *X* and *Z*.

Based on equation (1.1), the expected value of the wage of worker *i*, conditional on *X* and *Z*, is given by $E\left(Ln\ w_{it}\middle|X_{it}, Z_{j(i)t}\right)$, while the corresponding error is given by:

$$u_{it} = Ln\ w_{it} - E\left(Ln\ w_{it}\middle|X_{it}, Z_{j(i)t}\right). \tag{1.2}$$

In turn, the log average wage in firm *j*, $Ln\,\overline{w}_{jt}$, can be formulated as a function of observable firm characteristics, Z_{jt}, and average characteristics of workers, $\overline{X}_{jt}$, giving:

$$Ln\ \overline{w}_{jt} = \overline{X}_{jt}\beta + Z_{jt}\gamma + \upsilon_{jt}, \tag{1.3}$$

with $Ln\ \overline{w}_{jt} = Ln\sum_{i=1}^{N_{jt}}\frac{w_{it}}{N_{jt}}$ and $\overline{X}_{jt} = \sum_{i=1}^{N_{jt}}\frac{X_{it}}{N_{jt}}$.[2]

Using model (1.3), the mathematical expectation of the average earnings in a given firm, $E\left(Ln\ \overline{w}_{jt}\middle|\overline{X}_{jt}, Z_{jt}\right)$, will of course depend on $\overline{X}$ and *Z*, while the corresponding errors are given by

$$\upsilon_{jt} = Ln\ \overline{w}_{jt} - E\left(Ln\ \overline{w}_{jt}\middle|\overline{X}_{jt}, Z_{jt}\right). \tag{1.4}$$

In this context, it is fair to assume that the difference between what a worker is entitled to receive, given *X* and *Z*, and the expected firm average wage, conditional on $\overline{X}$ and

[1] The log wage empirical distribution is, in general, very close to a normal distribution (Card, 1999, Ch. 30). This specification also offers a ready-to-use interpretation, especially with respect to the 'return to education'.
[2] N_{jt} is the number of workers in firm *j* in period *t*. Equation (1.3) follows from (A1.1) in Appendix A1.

Z, or $E\left(Ln\ w_{it}\left|X_{it}, Z_{j(i)t}\right.\right) - E\left(Ln\ \overline{w}_{jt}\left|\overline{X}_{jt}, Z_{jt}\right.\right)$, depends on the gap between worker's observed attributes and the mean attributes of his/her counterparts in the same firm. Under this assumption, we expect, on average, that a worker with a higher schooling level than his/her average co-worker, for example, will have a higher wage.

Let us then assume that, for individual *i*, we have

$$Ln\ w_{it} - Ln\ \overline{w}_{jt} > E\left(Ln\ w_{it}\left|X_{it}, Z_{j(i)t}\right.\right) - E\left(Ln\ \overline{w}_{jt}\left|\overline{X}_{jt}, Z_{jt}\right.\right). \tag{1.5}$$

Under the assumption that the set of observed variables is sufficiently representative of both individual and firm characteristics, one may hypothesize that the inequality (1.5) holds if there is any gap between unobserved ability of worker *i* and the average unobserved ability in firm *j*.[3]

Let us now assume that α_i is the (time-invariant) innate ability of worker *i*, while $\phi_{j(i)}$ is the (time-invariant) unobserved effect specific to firm *j*; $\overline{\alpha}_{j(i)}$ is the unobserved worker average ability in firm *j* in period *t*, with $\overline{\alpha}_{j(i)} = \sum_{i=1}^{N_{jt}} \frac{\alpha_i}{N_{jt}}$. Then, using (1.2), we set:

$$\begin{aligned} u_{it} &= \left[Ln\ w_{it} - E(Ln\ w_{it} \mid X_{it}, Z_{j(i)\,\mathrm{t}})\right] \\ &= \alpha_i + \phi_{j(i)}. \end{aligned} \tag{1.6}$$

On the other hand, the error from equation (1.4) can be explained by unobserved firm heterogeneity, ψ_j, in which case we have:

$$\begin{aligned} \upsilon_{jt} &= \left[Ln\ \overline{w}_{jt} - E(Ln\ \overline{w}_{jt} \mid \overline{X}_{jt}, Z_{jt})\right] \\ &= \psi_j, \end{aligned} \tag{1.7}$$

with $\psi_j = \overline{\alpha}_j + \phi_j$. (A normal error term can be easily added to formulations (1.6) and (1.7).)

Under these assumptions, we can now give a clear interpretation to inequality (1.5) as we have

[3] While innate ability cannot be directly measured, we know how it is rewarded. 'Value' and 'volume' in this framework are equivalent.

$$
\begin{aligned}
&(Ln\ w_{it} - Ln\ \overline{w}_{jt}) - \left[E(Ln\ w_{it} \mid X_{it}, Z_{j(i)t}) - E(Ln\ \overline{w}_{jt} \mid \overline{X}_{jt}, Z_{jt})\right] \\
&\quad = \left[Ln\ w_{it} - E(Ln\ w_{it} \mid X_{it}, Z_{j(i)t})\right] - \left[Ln\ \overline{w}_{jt} - E(Ln\ \overline{w}_{jt} \mid \overline{X}_{jt}, Z_{jt})\right] \\
&\quad = (\alpha_i + \phi_{j(i)}) - (\psi_{j(i)}) \\
&\quad = (\alpha_i + \phi_{j(i)}) - (\overline{\alpha}_{j(i)} + \phi_{j(i)}) = (\alpha_i - \overline{\alpha}_{j(i)}).
\end{aligned} \tag{1.8}
$$

Clearly, (1.5) holds if and only if $\alpha_i > \overline{\alpha}_{j(i)}$, which means that any worker with a higher unobserved human capital than the average co-worker in firm *j* will be expected to have a higher wage. To simplify the notation we will make $\theta_i \equiv \alpha_i - \overline{\alpha}_{j(i)}$.

In what follows, we will assume that $\overline{\alpha}_{j(i)}$ is also time-invariant, which of course makes ψ_j time-invariant as well. Meanwhile, we note that relaxing this assumption produces no material changes in the results.[4]

2.2 Estimation

Let us now recap by considering the data generator process implicit in equation (1.6), that is:

$$Ln\ w_{it} = X_{it}\beta + Z_{j(i)t}\gamma + \alpha_i + \phi_{j(i)} + \varepsilon_{it}. \tag{2.1}$$

This model follows directly from equation (1.1), under the assumption that the error term, u_{it}, is given by $u_{it} = \alpha_i + \phi_{j(i)} + \varepsilon_{it}$. This data generating process for individual earnings is similar to the one used in AKM.

Considering the logarithmic of the average wage in firm *j*, we have:

$$Ln\ \overline{w}_{jt} = \overline{X}_{jt}\beta + Z_{jt}\gamma + \overline{\alpha}_j + \phi_j + \omega_{jt}. \tag{2.2}$$

which, in turn, is equivalent to model (1.3) under $\upsilon_{jt} = \overline{\alpha}_j + \phi_j + \omega_{jt}$.

The error terms in models (2.1) and (2.2) are assumed to have the following properties:

$\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$; $\omega_{jt} \sim IID(0, \sigma_\omega^2)$; $E\left(\varepsilon \middle| X, Z, D, F\right) = 0$ and $E\left(\omega \middle| X, Z, D, F\right) = 0$,

[4] Indeed, an alternative modelling with a non-constant $\overline{\alpha}_{j(i)}$ term generates similar results. In particular, the correlation between the unobservable effects obtained from the two alternatives is very high, at 0.90 and 0.86, in the case of $\hat{\theta}_i$ and $\hat{\psi}_j$, respectively.

where D denotes a $NT \times N$ matrix of dummies that identify the worker over T periods and F is a $JT \times J$ matrix of dummies representative of firms. N denotes the number of workers in the dataset, J the number of firms, and T the length of the time series.

Estimation of model (2.1) by *OLS* faces two major obstacles. The first one has to do with the possible correlation between observable characteristics, X and Z, and unobserved heterogeneity, α_i and $\phi_{j(i)}$. Indeed, both the standard Hausman test and the F-statistic test reject the null of no correlation between the unobservable effects and the regressors X and Z.[5] The second major difficulty arising from applying *OLS* to (2.1) is the non-orthogonality of α_i and $\phi_{j(i)}$.

Our empirical approach is as follows. Firstly, we use equation (2.1) and note that

$$Ln\ w_{it} - (X_{it}\beta + Z_{j(i)t}\gamma + \phi_{j(i)} + \overline{\alpha}_{j(i)} + \varepsilon_{it}) = \alpha_i - \overline{\alpha}_{j(i)}, \qquad (2.3)$$

which, under the assumption that θ_i is given by $\alpha_i - \overline{\alpha}_j$, is equivalent to have

$$\theta_i = Ln\ w_{it} - (X_{it}\beta + Z_{j(i)t}\gamma + \phi_{j(i)} + \overline{\alpha}_{j(i)} + \varepsilon_{it}). \qquad (2.4)$$

In turn, by manipulating (2.2), we get

$$\phi_j + \overline{\alpha}_j = Ln\ \overline{w}_{jt} - \left(\overline{X}_{jt}\beta + Z_{jt}\gamma\right) - \omega_{jt}. \qquad (2.5)$$

and, substituting (2.5) into (2.4), we finally have

$$\theta_i = Ln\ w_{it} - \left(X_{it}\beta + Z_{j(i)t}\gamma + Ln\ \overline{w}_{j(i)t} - \left(\overline{X}_{j(i)}\beta + Z_{j(i)t}\gamma\right) - \omega_{j(i)t}\right) - \varepsilon_{it}$$

$$(=)\ \theta_i = \left(Ln\ w_{it} - Ln\ \overline{w}_{j(i)t}\right) - \left(X_{it} - \overline{X}_{j(i)t}\right)\beta + \omega_{j(i)t} - \varepsilon_{it}, \qquad (2.6)$$

which is equivalent to (1.8).[6]

In particular, we note that equation (2.6) may also take the form

$$\left(Ln\ w_{it} - Ln\ \overline{w}_{j(i)t}\right) = \left(X_{it} - \overline{X}_{j(i)t}\right)\beta + \theta_i + \left(\varepsilon_{it} - \omega_{j(i)t}\right), \qquad (2.7)$$

which means that the wage gap $\left(Ln\ w_{it} - Ln\ \overline{w}_{j(i)t}\right)$ can be explained by the difference in observed characteristics $\left(X_{it} - \overline{X}_{j(i)t}\right)$ and by θ_i (or $\alpha_i - \overline{\alpha}_j$), plus a stochastic term, $\varepsilon - \omega$.

[5] The F-statistic test is used to find the statistical significance of γ in the regression of an auxiliary model given by $y_{it} - \hat{\lambda}\overline{y}_i = (1 - \hat{\lambda})\mu + (x_{it} - \hat{\lambda}\overline{x}_i)\beta + (x_{it} - \overline{x}_i)\gamma + e$ (see Johnston and Dinardo, 1997, p. 404).

[6] An alternative route to obtain (2.6) is of course to subtract (2.2) from (2.1).

In matrix notation, the equation (2.7) is equivalent to

$$\left(LW - L\overline{W}^{j}\right) = \left(X - \overline{X}^{j}\right)\beta + D\theta + (\varepsilon - \omega), \tag{2.8}$$

with $E\left(\varepsilon_{it}\omega_{j(i)t}\right) = 0$ and $\left(\varepsilon_{it} - \omega_{j(i)t}\right) \sim IID(0, \sigma_{\varepsilon}^{2} + \sigma_{\omega}^{2})$.[7]

Multiplying equation (2.8) by $M_D = I - P_D$, where P_D denotes the matrix that provides an orthogonal projection in D, we obtain

$$M_D\left(LW - L\overline{W}^{j}\right) = M_D\left(X - \overline{X}^{j}\right)\beta + M_D D\theta + M_D(\varepsilon - \omega). \tag{2.9}$$

By definition, we have $M_D D\theta = 0$, and therefore

$$M_D\left(LW - L\overline{W}^{j}\right) = M_D\left(X - \overline{X}^{j}\right)\beta + M_D(\varepsilon - \omega), \tag{2.10}$$

which yields the same estimates and residuals as model (2.8).[8] We also note that the first element of matrix $M_D(LW - L\overline{W}^{j})$, for example, is given by

$$\left(Ln\ w_{1,1} - Ln\ \overline{w}_{j(1),1}\right) - \frac{\left(Ln\ w_{1,1} - Ln\ \overline{w}_{j(1),1}\right) + \left(Ln\ w_{1,2} - Ln\ \overline{w}_{j(1),2}\right)}{2}, \tag{2.10$'$}$$

and that for $M_D\left(X - \overline{X}^{j}\right)$ we have:[9]

$$\left(x_{1,1}^{1} - \overline{x}_{j(1),1}^{1}\right) - \frac{\left(x_{1,1}^{1} - \overline{x}_{j(1),1}^{1}\right) + \left(x_{1,2}^{1} - \overline{x}_{j(1),2}^{1}\right)}{2}. \tag{2.10$''$}$$

The corresponding estimator of β can be then written as

$$\widehat{\beta} = \left(\left(X - \overline{X}^{j}\right)^{T} M_D\left(X - \overline{X}^{j}\right)\right)^{-1}\left(X - \overline{X}^{j}\right)^{T} M_D\left(LW - L\overline{W}^{j}\right). \tag{2.11}$$

From equation (2.8) we also have

$$\left(LW - L\overline{W}^{j}\right) - \left(X - \overline{X}^{j}\right)\beta = D\theta + (\varepsilon - \omega), \tag{2.12}$$

which means that the associated $\hat{\theta}$ can be written as

$$\hat{\theta} = \left(D^{T}D\right)^{-1} D^{T}\left(LW - L\overline{W}^{j} - \left(X - \overline{X}^{j}\right)\widehat{\beta}\right). \tag{2.13}$$

---

[7] These assumptions indicate that both the variance of the error term in (2.8) and the variance-covariance of $\widehat{\beta}$ and $\hat{\theta}$ depend on the variance of the error terms in models (2.1) and (2.2).
[8] According to the Frisch and Waugh theorem.
[9] $x_{1,1}^{1}$ denotes the first observable variable from worker 1 in period 1.

Finally, we tackle the unobserved firm heterogeneity issue. Thus, using equation (2.5) and making $\psi_j = \overline{\alpha}_j + \phi_j$, we have

$$\psi_j = Ln\ \overline{w}_{jt} - \left(\overline{X}_{jt}\beta + Z_{jt}\gamma\right) - \omega_{jt}. \tag{2.14}$$

In matrix notation, equation (2.14) becomes

$$L\overline{W}^j = \overline{X}^j\beta + Z^j\gamma + F\psi + \omega. \tag{2.15}$$

where, we recall, F is a ($JT \times J$) matrix of dummies flagging the J firms.

Consider now the matrix of orthogonal projection in F, $P_F = F(F^T F)^{-1} F^T$, and the matrix M_F, given by $M_F = I - P_F$. Multiplying equation (2.15) by M_F, we have:

$$M_F L\overline{W}^j = M_F \overline{X}^j\beta + M_F Z^j\gamma + M_F F\psi + M_F\omega, \tag{2.15)'}$$

where the first element of the matrix $M_F Z$, for example, is given by $z_{1,1}^1 - \dfrac{z_{1,1}^1 + z_{1,2}^1}{2}$.[10]

By definition, we have $M_F F\psi = 0$, which means that the estimator of γ can be written as:

$$\hat{\gamma} = \left(Z^T M_F Z\right)^{-1} Z^T M_F \left(L\overline{W}^j - \overline{X}^j\hat{\beta}\right). \tag{2.16}$$

Substituting $\hat{\beta}$ and $\hat{\gamma}$ into equation (2.15) we finally have

$$\hat{\psi}' = \left(F^T F\right)^{-1} F^T \left(L\overline{W}^j - \overline{X}^j\hat{\beta} - Z^j\hat{\gamma}\right). \tag{2.17}$$

Now we elaborate further on the case where Z contains some time-invariant characteristics. We note first that by pre-multiplying (2.15) by the matrix M_F, we are in practice getting rid of all time-invariant (observed) firm-specific characteristics (e.g. sector, legal status, and location). This implies that $\hat{\psi}_j'$ in (2.17) will not only capture the unobserved firm effect but also the effect of time-invariant (observed) firm characteristics. Since we only want to capture ψ_j, that is, the contribution of (time-invariant) unobserved firm characteristics, model (2.17) is not appropriate.

Let us then denote the subset of time-invariant firm characteristics by Z_j'. Then, using $\hat{\psi}_j'$ (from (2.17)), we run the model

---

[10] $z_{1,1}^1$ ($z_{1,2}^1$) denotes the first characteristic of firm 1 in period 1 (2).

$$\widehat{\psi}_j{}' = \psi_j + \kappa Z_j{}', \tag{2.18}$$

where ψ_j denotes the unobserved effect of firms excluding time-invariant (observable) characteristics.[11] Model (2.18) is estimated by feasible *GLS* using the estimated variance of $\widehat{\psi}_j{}'$ (denoted by $\widehat{\sigma}^2_{\psi_j}$), obtained from (2.17) and noting that $\widehat{\sigma}^2_{\psi} = \left(F^T F\right)^{-1} \widehat{\sigma}^2_{w}$.[12] In particular, for firm *j*, we have $\widehat{\sigma}^2_{\psi_j} = \dfrac{\widehat{\sigma}^2_w}{\sum_t N_{jt}}$ (see Appendix A2). Using $\hat{\kappa}$ and $\widehat{\psi}_j{}'$ we are therefore in a position to have an estimate of ψ_j by simply solving (2.18) in order to ψ_j.

We finally note that one key aspect of our methodology is that neither $\widehat{\psi}_j$ nor $\widehat{\psi}_j{}'$ depend on $\hat{\theta}_i$, which means that worker and firm unobservable effects are estimated separately.

2.3 Unobservable heterogeneity among job switchers

The methodology described in the previous section cannot be directly applied to a panel of job switchers. In this case, as it will be shown below, the fixed effects approach is not powerful enough to capture worker and firm unobserved effects.

Let us consider that, in period 1, worker *i* is in firm *j*. Then, using equation (2.7), we have

$$\left(Ln\ w_{i1} - Ln\ \overline{w}_{j(i)1}\right) = \left(X_{i1} - \overline{X}_{j(i)1}\right)\beta + \left(\alpha_i - \overline{\alpha}_{j(i)1}\right) + \left(\varepsilon_{i1} - \omega_{j(i)1}\right). \tag{3.1}$$

In period 2, assuming worker *i* moves to firm *s*, we have

$$\left(Ln\ w_{i2} - Ln\ \overline{w}_{s(i)2}\right) = \left(X_{i2} - \overline{X}_{s(i)2}\right)\beta + \left(\alpha_i - \overline{\alpha}_{s(i)2}\right) + \left(\varepsilon_{i2} - \omega_{s(i)2}\right). \tag{3.2}$$

Clearly, applying the fixed effects approach – or taking first differences – the derived model will contain an unknown element, that is, $\overline{\alpha}_{s(i)2} - \overline{\alpha}_{j(i)1}$. A direct extension of the

[11] This procedure is equivalent to the approach followed by AKM to distinguish the effect of (time-invariant) schooling from unobservable (time-invariant) worker attributes.
[12] We use feasible GLS to deal with the eventual heteroscedasticity caused by firm-level aggregation. Andrews, Schank and Upward (2006) argue that robust OLS would be sufficient.

methodology described in section 2.1 will yield therefore biased results as the term $\overline{\alpha}_{s(i)2} - \overline{\alpha}_{j(i)1}$ will be simply assumed away.

Let us start again with model (2.1) and consider an individual who is in firm *j* in period 1 and in firm *s* in period 2. Taking first differences, we have

$$Ln\ w_{i2} - Ln\ w_{i1} = (X_{i2} - X_{i1})\beta + \left(Z_{s(i)2} - Z_{j(i)1}\right)\gamma + \left(\phi_{s(i)2} - \phi_{j(i)1}\right) + \left(\varepsilon_{i2} - \varepsilon_{i1}\right). \qquad (3.3)$$

This equation explains the difference in wages received by worker *i* in periods 1 and 2 as a function of changes in his/her observable characteristics and observed and unobserved characteristics of firms *s* and *j*. (Note that the wage gap does not depend on α_i since this component is time-invariant.)

In matrix notation, we have

$$HLW = HX\beta + HZ\gamma + P\phi + \varepsilon, \qquad (3.4)$$

where *H* is the first difference operator and *P* is a $M \times J$ matrix, with $p_{it,l}$ given by[13]

$$p_{it,l} = \begin{cases} 1 \text{ if } l \text{ denotes the enterprise where worker } i \text{ is employed in period } t \\ -1 \text{ if } l \text{ denotes the enterprise where worker } i \text{ was employed in period } t-1 \\ 0 \text{ otherwise} \end{cases}$$

From (3.4) we get an unbiased estimate of β, γ, and ϕ.[14] Finally, considering two periods and using (2.1), we proxy α_i by the over time average of the two estimated individual effects, that is:

$$\alpha_i = \tfrac{1}{2}\left[\left(Ln\ w_{i1} - X_{i1}\hat{\beta} - Z_{j(i)1}\hat{\gamma} - \hat{\phi}_{j(i)1}\right) + \left(Ln\ w_{i2} - X_{i2}\hat{\beta} - Z_{s(i)2}\hat{\gamma} - \hat{\phi}_{s(i)2}\right)\right]. \qquad (3.5)$$

Thus, the parameter α_i will be given by the over time average wage unexplained by the observable characteristics of workers nor by the characteristics (observable and not observable) of firms at which workers have been employed.

[13] M is the number of switchers.
[14] See Wooldridge (2002, Ch.10).

2.4 Computing the size of the bias

The standard *Mincerian* approach looks at the relationship between individual earnings and individual attributes, controlling for firm characteristics. Based on the developments in sections 2.1 and 2.2, we want, in particular, to assess the impact on the rate of return of typical covariates (e.g. schooling and training) after controlling directly for unobserved worker and firm heterogeneity. To this end, we will use the model

$$Ln\ w_{it} = X_{it}\beta + Z_{j(i)t}\gamma + \theta_i + \psi_{j(i)} + \varepsilon_{it}, \tag{4.1}$$

which is identical to model (2.1) as, by definition, $\theta_i + \psi_j = \alpha_i + \phi_j$, and then, assuming away all unobserved effects, we run ordinary least squares on the model

$$Ln\ w_{it} = X_{it}\beta' + Z_{j(i)t}\gamma' + \varepsilon_{it}'. \tag{4.2}$$

To obtain the unbiased estimates of β and γ, we then add $\hat{\theta}_i$ and $\hat{\psi}_{j(i)}$ and re-run the model

$$Ln\ w_{it} - \hat{\theta}_i - \widehat{\psi}_{j(i)} = X_{it}\beta + Z_{j(i)t}\gamma + \varepsilon_{it}. \tag{4.3}$$

By comparing the results from these two models – that is, $\hat{\beta}$ with $\hat{\beta}'$ (and $\hat{\gamma}$ with $\hat{\gamma}'$) – we will be in a position to measure the bias resulting from the omission of unobservable (worker and firm) heterogeneity. Clearly, in this framework, $\hat{\beta}$ and $\hat{\gamma}$ will be conditional on *X* and *Z*, but also on $\hat{\theta}$ and $\hat{\psi}$. Given models (2.1) and (2.2), and the corresponding assumptions on ε and ω, both $\hat{\theta}$ and $\hat{\psi}$ are unbiased, but even if $\hat{\theta}$ and $\hat{\psi}$ are biased, problems will arise only if $(\theta - \hat{\theta})$ and $(\psi - \hat{\psi})$ are correlated with *X* or *Z*, a possibility that seems unlikely.

3. Data

Our linked employer-employee dataset (LEED) was obtained by matching the information from *Quadros de Pessoal* (worker-level information) and *Balanço Social* (firm-level information), both from *Gabinete de Estudos e Planeamento* (GEP) of the Ministry of Labor, Portugal. The matching was made using firm's (and worker's) unique identification number which allowed us to match individuals and firms not only in a given year but also longitudinally. Our raw LEED data, in particular, contains two data points (1998 and 1999), covering approximately 900,000 workers, employed in some 2,200 firms with at least 100 employees.[15]

Information on firm characteristics is mainly extracted from *Balanço Social*, and it includes value added, the wage bill, number of employees, location (five regions of continental Portugal), sectoral activity (twenty seven sectors), and legal status (three categories). *Balanço Social* also contains information on average characteristics of workers such as age, gender, schooling, tenure, and skill. A key feature of *Balanço Social* is that it contains unique information on firm-provided training, namely the number of training sessions, the number/share of training participants by occupation level, the number of training hours and the corresponding training costs (direct and indirect). Each of these items are subdivided in on-the-job and off-the job training categories.

In turn, the information on individual worker attributes is extracted from *Quadros de Pessoal*. It includes monthly earnings, hours of work, age, gender, schooling level, skill, tenure, job occupation, and whether the individual is a full or part-time worker, inter al.[16] Based on the detailed information on training at firm level (from *Balanço Social*), we also used a model to impute training participation at worker level. (This procedure is available upon request from the authors.)

Our estimation sample was obtained by applying several filters to the raw data. In particular, we dropped part-time workers and all individuals who were younger than 16 years old or older than 65. (There are some 100,000 part-time workers in the raw sample.) Apprentices and individuals with earnings less than the statutory minimum wage were also eliminated, as well as those who were employed in firms located in Madeira and Açores.

[15] A total of 535,254 individuals were observed in both sample years, while 178,435 were only observed in 1998 and 182,996 in 1999.

[16] *Quadros de Pessoal* contains information on basic and total earnings, with the latter being obtained by adding to basic earnings other elements such as compensation for night shifts and productivity bonus. Typically, total earnings show greater cross-section and over time variability. *Quadros de Pessoal* also contains information on firm characteristics which were used for double-checking proposes. As described in Appendix Table 1, our selected earnings measure is *total earnings*.

After applying these filters, and eliminating all observations in which at least one selected variable is missing, we ended up with a balanced panel of 401,258 individuals who were observed consecutively.

The summary statistics at individual and firm level are presented in Table 1. In the first place, we note that although our dataset should present, in principle, a comparatively lower degree of (observed) heterogeneity – no firm in the sample has less than 100 workers, we recall – there is a considerable dispersion in earnings. In fact, using column (1) of the table, we obtain a coefficient of variation of approximately 0.4, while, for example, in Germany this indicator is only 0.1.[17] Worker-level means of the selected characteristics are also typically different from firm-level means, while the standard deviation of the variables earnings, age, schooling, and tenure in column (2) are roughly ½ of the corresponding value in column (1), an indication that there is a sizeable sorting of individuals across firms.

Table 2 gives the summary statistics of our estimation samples: in column (1), we have the subsample of individuals who are in the same firm in 1998 and 1999, and, in column (2), the subsample of switchers, that is, those individuals who were employed in different firms in two consecutive years. We note that from the initial set of 401 thousand individuals in Table 1, some 382 thousand are stayers, while 19 thousand are switchers. Then, due to missing observations on the selected variables, we lost an additional total of 20 thousand stayers to end up with 357,081 useable observations in Sample 1. The number of switchers also turned up to be much lower than the initial 19 thousand. Indeed, a closer inspection of the data in 1998 and 1999 reveals that a sizeable fraction of these individuals do stay in the same firm, being the firm identification code different exclusively due to changes in firm activity classification and ownership. (Mergers and acquisitions are generally at the root of the problem.) These cases were detected by looking carefully at the worker tenure variable, on the one hand, and at firm labor turnover rates, on the other, and by observing massive worker flows across two (artificially) distinct firms. (Other time-invariant firm characteristics were also checked to be absolutely sure about our procedures.) All individuals associated to these artificial worker flows were dropped from our sample of switchers.[18]

According to Abowd, Creecy and Kramarz (2002), the identification of firm (and worker) effects should be computed within a 'connected' group of workers and firms. Thus,

---

[17] See, for example, Addison, Teixeira and Zwick (2009), Table 1A.

[18] The final set of switchers is therefore substantially smaller than the original sample. This seemingly drop shows quite emphatically how sensitive is worker mobility data in practice.

applying their procedure to our sample of switchers, we created a first *workers/firms* group that includes *all* the switchers who were employed in any firm in the group at some point over the sample period and *all* the firms at which at least one of the individuals in the group was ever employed. In a second step, we selected *all* the workers employed in those firms. Similarly, for the second, third, …, groups, the condition being that the intersection between any pair of groups ought to be empty either in terms of individuals or firms. In our case, only 25 workers of the initial sample of switchers did not belong to the first group (say Group 1) as they were not connected to any firm in the group. By the same token, a total of 17 firms were excluded from Group 1 as they never employed any worker in the group.[19] Based on this procedure, we ended up with a final sample of 4,069 switchers (i.e. 99% of a total of 4,094 individuals) and 802 firms. The corresponding summary statistics are presented in Table 2, column (2). Clearly, worker mobility is non-random, as switchers do present a set of attributes quite distinct from the characteristics of stayers. In particular, switchers are younger and have higher levels of schooling than stayers. They also seem to have participated more often in training. Worker mobility seems also to be more concentrated in Lisboa and Vale do Tejo, in large firms, and in the service sector. (This information is not reported in the Table.)

4. Results

4.1 Unobservable characteristics of workers

Table 3 shows the summary statistics of $\hat{\theta}_i$, obtained from model (2.13). We note that, by definition, the mean of $\hat{\theta}_i$ is equal to zero. The median, in turn, is slightly negative which is due to the fact that the median of the log wage distribution is slightly below the mean. (The kurtosis and skewness are equal to 7.58 and 1.61, respectively; see Figure 1.) The standard deviation of $\hat{\theta}_i$, at 0.33, also confirms the presumption that unobserved heterogeneity across workers is quite substantial. Finally, we note that the computed value for the standard deviation is very close to the one reported by AKM, who found in their study a standard deviation of 0.40, for Men, and 0.38, for Women (AKM, Table IV).

[19] As a matter of fact, the procedure generated seven additional groups: Group 2, with 9 workers and 4 firms; Group 3, with 5 workers and 2 firms; Group 4, with 3 workers and 3 firms; and Groups 5, 6, 7, and 8, with 2 workers and 2 firms per group.

The correlation between observed and unobserved ability is given in Table 4. The high correlation coefficients shown in this table confirms of course that the hypothesis of orthogonality between unobservable and observable characteristics is unrealistic. This is particularly visible when we look at the correlation between schooling and $\hat{\theta}_i$, at 0.41. Clearly, individuals with above average unobserved ability have higher levels of schooling, which of course implies that estimates from regressions that do not control for unobserved heterogeneity will seriously overestimate the impact of schooling on earnings. Although to a lesser degree, training is positively correlated with ability, which also confirms that there is also self-selection into training. In turn, the correlation between $\hat{\theta}_i$ and labor market experience is negative. One interpretation for this result is that younger workers have perhaps higher innate abilities.[20] As expected, the unobservable component of human capital is highly positively correlated with earnings, at 0.59.[21]

In a separate exercise (not reported in the Table), we computed the correlation between $\hat{\theta}_i$ and the difference between each individual (observable) attribute and the corresponding firm average. For the schooling variable, for example, this correlation is equal to 0.47. The correlation in the case of other attributes is also very similar to the ones reported in Table 4.

4.2 Unobservable characteristics of firms

Table 5 contains the summary statistics of firm unobserved heterogeneity, $\hat{\psi}_j$, which, as described in Section 2, includes two components, $\bar{\alpha}_j$ and ϕ_j. As it can be seen, both the mean and the median of $\hat{\psi}_j$ are negative at worker level. The corresponding statistics at firm level are even more negative (not presented in the Table) which suggests that workers are slightly concentrated in firms with a higher than average unobserved ability.

[20] The negative correlation between unobserved ability and experience is also reported by Abowd, Lengrmann and McKinney (2003).
[21] In AKM this coefficient is slightly higher, at 0.73. This is not surprising as their set of observables attributes is much narrower than ours.

$\hat{\psi}_j$ is expected to be correlated with observable firm attributes, $\overline{X}$ and Z. The corresponding correlation coefficients are presented in Table 6. Since, by definition, $\hat{\psi}_j$ contains the average innate abilities of workers in the firm, it should not be surprising to find that unobserved ability of firms is positively correlated with schooling, training and skills, for example, and negatively correlated with experience.[22]

The correlation between $\hat{\theta}_i$ and unobservable characteristics of firms, $\hat{\psi}_j$, is slightly negative, at -0.004 (it is statistically different from zero, at a significance level of 0.01 or better). We note that this negative relation does not imply that high-ability workers are in low-ability firms, which would be counter-intuitive. The following example illustrates our point. Let us suppose, for example, that worker *i*, employed in a firm with a high $\overline{\alpha}_j$, has a high α_i. In this case, the difference between α_i and $\overline{\alpha}_j$ (or θ_i) although positive is presumably small. Assuming ϕ_j is also high, we have then a large ψ_j and a small θ_i, and therefore a negative correlation between θ_i and ψ_j, while α_i and ϕ_j are actually positively correlated.

The relation between the earnings firm average and observed and unobserved characteristics is presented on Table 7. The critical component in terms of wage determination seems to be the average unobservable characteristics of firms, with a correlation coefficient of 0.72. Time-invariant firm (observed) characteristics are also very important at 0.65, while the correlation between observable characteristics of workers and firm average wages is somewhat lower at 0.46 (row 2).

4.3 Unobserved ability in the case of switchers

Now we consider the case of switchers. We recall that by looking at this particular sample, we want to obtain the innate ability $\hat{\alpha}_i$, and not just $\hat{\theta}_i$. Another implication is that the unknown firm fixed effect, $\hat{\phi}_j$, can also be obtained.

[22] The sign of the corresponding correlation coefficient is the same as in Table 4.

The summary statistics for the sample of switchers is given in Table 8. Clearly, the standard deviation of the unobservable worker heterogeneity in this subsample is larger than in Table 3. This is an expected result since the standard deviation of $\hat{\theta}_i$ measures, by definition, how individual innate attributes deviate from the firm average, while the standard deviation of $\widehat{\alpha}_i$ measures simply the dispersion in individual unobserved ability.

Table 9 gives the correlation between $\widehat{\alpha}_i$ and worker characteristics, X_i. The absolute value of the correlation coefficient between the unobservable heterogeneity and the observable characteristics in this subsample is generally higher than those reported in Table 4.

In Table 10 we compute the correlation between $\hat{\theta}_i$ (and $\hat{\psi}_j$) and worker and firm specific effects, $\widehat{\alpha}_i$ (and $\hat{\phi}_j$), using the sample of switchers. In the first place, it seems that $\hat{\theta}_i$ captures most of the individual unobserved effect $\widehat{\alpha}_i$ as the correlation coefficient, in the third cell of the table (column 1), is equal to 0.8417. We observe though a weaker relation between unobserved ability of switchers and unobservable characteristics of firms as the coefficient of correlation between $\widehat{\alpha}_i$ and $\hat{\phi}_j$ is 0.2315. But this relation is stronger if the unobservable firm effect includes $\overline{\alpha}_j$.[23] In this case, the correlation between $\widehat{\alpha}_i$ and $\widehat{\psi}_j$ is equal to 0.3807.

As can be observed in Table 11, the correlation between the firm average wage and unobserved attributes are very similar to those obtained using the sample of stayers, although the correlation with the observable characteristics seems to be higher for switchers: 0.84 (for workers characteristics) and 0.34 (for firms characteristics) – Table 11 – and 0.46 and 0.15, respectively – Table 7.

Since we computed $\hat{\phi}_j$ using only workers who change jobs (switchers) it is possible that, for some firms, this parameter is being obtained using too few observations per firm. To circumvent this problem, we followed AKM and pooled all firms with less than 10 observations in a single entity. The resulting sample comprised 176 firms (115 firms in the case of AKM, just to keep our exercise in perspective). The results from this experiment are reported on Appendix Tables A2 and A3. There is, in the first place, less dispersion in $\hat{\phi}_j$ -

[23] Which means there is also a positive relation between unobserved heterogeneity of worker *i* and innate attributes of his/her co-workers.

the standard deviation is now 0.17 (in Table A2, column 2) rather than 0.28 (in Table 8, column 2). Summary statistics of $\hat{\alpha}_i$ and the correlation with $\overline{X}$ are practically the same. In turn, the first column of Appendix Table A3 reproduces the results of model (4.2), while the second column presents results of model (4.3) using the 'pooled' data. As it can be seen, controlling for unobserved heterogeneity further reduces the rate of return to schooling and training, while tenure and experience somewhat increased their impact.

4.4 Labor market return rates after controlling for unobserved worker and firm effects

Tables 12 and 13 contain the results of models (4.2) and (4.3), in columns (1) and (2), respectively. In Table 12 the two models are applied to Sample 1 (stayers), whereas in Table 13 we report the results from Sample 2 (switchers). In the first column of both tables there is no control for unobserved ability, while in the second we account for unobserved worker and firm effects. In this context, the difference in parameter estimates between columns (1) and (2) gives an indication of the magnitude of the bias in standard OLS earnings equations.

Firstly, in column (1) of Table 12, there is confirmation of the familiar result (Becker, 1962), that the investment in human capital, either general or specific, does pay off, as higher levels of schooling, labor market experience, tenure, and training result in higher wages. Interestingly enough, even after controlling for a wide array of worker and firm observable attributes (a total of 47 regressors), the model without control for unobserved ability still shows a substantial gender gap of approximately 15%. The overall fitness of the model, given by the $\overline{R}^2$, is 74%.

Column (2) gives the corresponding parameter estimates after controlling for the unobserved components estimated in sections 4.1 and 4.2. Not surprisingly, taking into account the unobserved heterogeneity yields a substantial increase in the explanatory power of the model (the $\overline{R}^2$ is equal to 0.91 in Sample 1).

In Table 12, the obvious change from column (1) to column (2) is the reduction on the rate of return to standard indicators of general and specific human capital – schooling, training, experience, skills, and tenure. The first conclusion to draw is therefore that the standard earnings regression in column (1) crudely overestimates the return rate to standard measures of human capital. In particular, OLS estimates imply an overestimation of the returns to schooling of approximately 80%. In other words, 80% of the impact of the marginal rate of return to schooling (in column (1)) is not due to schooling *per se*.[24]

The results reported in Table 13, in turn, suggest that worker mobility generates somewhat different rates of return to observed characteristics. But most conclusions drawn from Table 12 hold rather well. In particular, it confirms that the acquisition of human capital matters and that standard OLS estimates are greatly overstated. Another interesting aspect is that workers seem to have more incentive to change jobs if they have higher innate abilities than their co-workers. In fact, and although we do not report this result in our tables, the results from a fairly parsimonious probit model show that the probability of a worker being a switcher is higher if he/she has lower tenure and a higher $\hat{\theta}_i$.

4.5. Robustness

4.5.1 Outliers, homoscedasticity, and omission of relevant variables

As a first pass, we investigate the presence of outliers by comparing actual and predicted wages (using model 4.2). Differences between predicted and actual earnings were within an interval which is smaller than five times the value of the standard deviation. Using this criterion (see also AKM) there seems to be no indication that the presence of outliers is an issue in our regressions.

To test for the presence of heteroscedasticity, we run the model

$$\left(\hat{\varepsilon}_{it}\right)^2 = \alpha + \varphi H_{it} + e_{it}, \tag{5.1}$$

[24] The schooling level is generally considered a good predictor to employers in terms of expected worker effort and motivation; and individuals with higher innate attributes are expected to select themselves into higher levels of schooling. The coefficient of schooling in the first column of Table 12 reflects of course this self-selection effect. It is worthwhile to note however that only a fraction of the observed reduction is due to worker unobserved ability. Indeed, if we ignore firm unobserved effects, the schooling coefficient reduces only to 0.025, versus 0.010 when both effects are controlled for.

where H denotes the set of explanatory variables (including their cross-products and squares). Then, we compute the statistic nR^2, which follows (asymptotically) a Chi-Square with 85 degrees of freedom. (This is the *White* test, which is a particular case of the *Breusch-Pagan* test, Greene, p. 222-223.) For model (4.2), the corresponding Chi-Square (with 85 d.f.) is equal to $\chi_i^2 = 623.85$ ($P > \chi_i^2 = 0.000$), which means that the null is comfortably rejected and, hence, that the variance of the error term depends on the values of the explanatory variables. This result should not be surprising as unobservable heterogeneity of workers and firms are expected to be correlated with the observable variables included in the regression. For model (4.3), we have $\chi_i^2 = 91.55$ ($P > \chi_i^2 = 0.025$), which means that the hypothesis of homoscedasticity cannot be easily rejected.

As a further checking we played with the robust option available in STATA to relax the assumption of errors being IID in model (4.3), and we virtually obtained the same statistical significance on the estimated coefficients.

Regarding the omission of relevant variables in our specifications, the *Ramsey Reset* test for model (4.2) gives $F_{(1;8,102)} = 208.53$ ($P_F > F_{est} = 0.0000$), while for model (4.3) we obtain $F_{(1;8,102)} = 0.56$ ($P_F > F_{est} = 0.6405$). This result indicates that our treatment of the unobserved effects was successful in removing the omitted variable problem. No evidence of multicolinearity was detected as the variance inflation statistical test is in most cases lower than 5 and, except for sectoral dummies (which, in principle, are correlated to each other), always lower than 10.

4.5.2 Replicating AKM

It also seems to be appropriate at this stage to replicate the original AKM methodology using our data. The starting point is a model as the one formulated in section 2.2 above (equation 2.1). Thus, let us take the model[25]

$$Y = X\beta + D\theta + F\psi + \varepsilon, \tag{6.1}$$

[25] For convenience, we will use the AKM notation. We note, however, that in AKM the variables Y and X denote deviations from the grand mean.

where θ and ψ denote worker and firm unobserved effects. Then, making $F\psi = (P_Z + M_Z)F\psi = Z(Z^T Z)^{-1} Z^T F\psi + M_z F\psi$ and setting $\lambda \equiv (Z^T Z)^{-1} Z^T F\psi$, we have

$$Y = X\beta + D\theta + Z\lambda + M_z F\psi + \varepsilon, \tag{6.2}$$

where Z is an artificial set of regressors obtained by interacting observed worker and firm characteristics; P_Z denotes the matrix that provides an orthogonal projection in Z, and $M_Z = I - P_Z$. Finally, under $X^T M_z F = 0$ and $D^T M_z F = 0$, equation (6.2) can be given by:

$$Y = X\beta + D\theta + Z\lambda + \varepsilon. \tag{6.3}$$

In practice, this approach amounts to use Z to capture the correlation between the unobservable effects of firms and all observed and unobserved variables. Assuming then that the unobservable heterogeneity of workers is time-invariant, the fixed effects approach is applied to obtain $\widehat{\beta}$ and $\widehat{\lambda}$, being the associated estimator of θ given by:[26]

$$\hat{\theta} = \left(D^T D\right)^{-1} D^T (Y - X\widehat{\beta} - Z\widehat{\lambda}). \tag{6.4}$$

Finally, to compute firm effects, AKM provide two alternative estimation methods: 'the order-independent' and 'the order-dependent'. (The acronym is due to the fact that, in the former, worker and firm effects are estimated separately, while in the latter worker effects are estimated before firm effects or vice-versa.)

In the order-independent estimation case, AKM use the assumption that the correlation across the independent variables of model (6.1) is captured by the matrix *Z,* to then compute firm unobserved effects, $\widehat{\psi}$, using the model

$$Y = F\psi + Z\pi + \xi, \tag{6.5}$$

where $\widehat{\pi}$ is computed via the orthogonal projection of variables from *Z* and the dependent variable on the null space of *F*. Pre-multiplying equation (6.5) by M_F with $M_F \equiv I - F(F^T F)^{-1} F^T$, we get

$$\widehat{\pi} = (Z^T M_F Z)^{-1} Z^T M_F Y, \tag{6.6}$$

[26] One limitation of this approach is that the $\hat{\theta}$ will contain the unobserved worker ability and any time-invariant worker attribute (e.g. schooling). A remedy is to use feasible GLS to separate the unobserved ability from the schooling effect.

and therefore we have:

$$\widehat{\psi}_j = (F^T F)^{-1} F^T \left(Y - Z\widehat{\pi} \right). \tag{6.7}$$

In turn, worker effects are estimated independently using (6.4).

The 'approach of order dependent' in its 'worker first' version uses the parameter estimates obtained in (6.3) and (6.4) and sets $Y - X\hat{\beta} - D\hat{\theta} = F\psi + \omega$. The associated estimator of ψ is then given by $\hat{\psi} = (F^T F)^{-1} F^T (Y - X\hat{\beta} - D\hat{\theta})$. (An alternative 'firms first' method can also be applied.)

Table 14, column (1), presents the summary statistics for $\widehat{\alpha}_i$ and $\widehat{\psi}_j$, obtained by applying the AKM methodology to our data. Column (2) simply reproduces the AKM results (their Table IV). The reported estimates are for the male and female sub-samples.[27] As it can be seen, the standard deviation of worker and firm unobserved effects (via the order-independent method) are roughly of the same order of magnitude (rows 1 to 4). In contrast, the "order dependent" approach (rows 5 and 6) produces substantially lower dispersion in unobserved firm ability. In any case, differences on the distribution of unobserved heterogeneity across gender are small. The major difference between columns (1) and (2) is on the mean of $\widehat{\psi}_j$ (in the order-dependent case), which tends to be slightly higher in column (1). This can be due to a higher share of large firms in our sample. Presumably, larger firms tend to concentrate a higher proportion of highly-skilled individuals.[28]

One interesting aspect to mention is that the correlation between earnings and the estimated unobserved ability, $\widehat{\alpha}_i$ and $\widehat{\psi}_j$, is very similar in the two sets of results – columns (1) and (2) of Table 15). Schooling, for example, presents a higher correlation in our data at 0.58 vs. 0.41 in AKM (Table 15, row 4, columns (1) and (2), respectively).

The significance of the unobservable effects on the performance of firms was also analysed in AKM. In our case, although the impact of unobservable heterogeneity of workers

[27] In our replication, we selected a representative sample of stayers and switchers to obtain a total of approximately 8,000 individuals.

[28] It is also instructive to compute the correlation between firm effects arising from each approach (order-independent versus order-dependent cases). In our sample, the correlation is equal to 0.155, which of course suggest some degree of non-robustness across the two methods.

is stronger, either worker or firm effects have a positive impact on productivity. Firm fixed effects are also associated with a more intensive utilization of physical capital, while the worker effect seems to have a greater impact on firm operating income.

We finally note that although our dataset has a much smaller number observations, the number of variables in our case is much larger than in AKM. The quality of the fit is therefore higher in our case – excluding the contribution of the unobservable effects, our model is able to explain more than 70% of the wage variation, while in AKM the quality of the fit does not exceed 30%. On the other hand, the fact that the set of worker attributes in AKM is restricted to experience and schooling imposes serious limitations on the estimation of the unobserved effects. The richness of our data is therefore an important advantage as it allows the implementation of a modelling strategy which does not require computation of *Z* to estimate the parameters of interest. In any case, and despite having a much larger set of worker and firm characteristics, the null of the Hausman test**,** necessary to guarantee that *Z* captures the covariance between observed and unobserved worker characteristics and firm effects, is still rejected comfortably. Indeed, the corresponding χ_i^2 statistic is equal to 11,176.40 (in AKM, the corresponding statistic is 21,000, p. 300).

4.6. Parameter robustness using Monte Carlo simulation and bootstrapping

Using simulation techniques, in this section we want to know how the reported point estimates $\hat{\beta}$ and $\hat{\gamma}$ in Tables 12 and 13 differ from the simulation mean and, in particular, how sensitive are the reported standard errors and the corresponding confidence intervals to various types of assumptions.

We will use two alternative routes: the Monte Carlo simulation and the bootstrap. The former requires the full specification of the data generating process – that is, the knowledge of all explanatory variables, the unobserved effects $\hat{\theta}_i$ and $\hat{\psi}_j$, and the distribution of the error term; the bootstrapping uses an estimated *DGP* based on the sample distribution (Davidson and MacKinnon, 2004, Ch. 4). In our procedure, we will use the non-parametric bootstrapping to relax the assumption of the error distribution, which amounts to estimate model (4.3) multiple times by resampling observations from the original data. These observations are selected given a certain probability and such that the structure of the panel is preserved.

For the Monte Carlo simulation, the estimated parameters are obtained from model (4.3), while the explanatory variables, *X* and *Z*, are assumed fixed. Firstly, we generate a random variable for the error term, u_{it}, assuming (for stayers): $u_{it} \sim N(0;0.015)$. Secondly, we generate *N* (sample size) values for the dependent variable, $Ln\ w_{it} - \hat{\theta}_i - \widehat{\psi}_j - u_{it}$, and estimate $\hat{\beta}^m$ and $\hat{\gamma}^m$, which of course will be conditional on $\hat{\theta}_i$ and $\hat{\psi}_j$.[29] By repeating the process *B* times, we then compute the average of $\hat{\beta}^m$ and $\hat{\gamma}^m$, which, in turn, by comparing with 'observed' $\hat{\beta}$ and $\hat{\gamma}$ from model (4.3), allow us to compute the magnitude of the bias in $\hat{\beta}$ and $\hat{\gamma}$.[30] The sample variance obtained from the Monte Carlo simulations can also be used to generate confidence intervals.

Another alternative to evaluate the sensitivity of $\hat{\beta}$ and $\hat{\gamma}$ is to consider that $\hat{\theta}_i$ follows a normal distribution, while, at the same time, there is correlation between $\hat{\theta}_i$ and schooling. In this case, $\hat{\theta}_i$ will have the following distribution:

$$\left(\hat{\theta}_i \middle| Schooling_i\right) \sim N\left(M_{\hat{\theta}_i} + \frac{\hat{\sigma}_{\hat{\theta}S}}{\hat{\sigma}_S^2}\left(Schooling_i - M_S\right); \hat{\sigma}_{\hat{\theta}_i}^2 - \frac{\hat{\sigma}_{\hat{\theta}S}^2}{\hat{\sigma}_S^2}\right),$$

where *M* denotes the statistical average, $\hat{\sigma}_{\hat{\theta}S}$ is the covariance between $\hat{\theta}_i$ and schooling, and $\hat{\sigma}^2$ the estimated variance. Using the statistics obtained from the sample of stayers, this distribution is given by $\left(\hat{\theta}_i \middle| Schooling_i\right) \sim N\left(0.033\left(Schooling_i - 7.91\right); 0.092\right)$. To generate *N* values for the dependent variable, we then make $Ln\ w_{it} - \hat{\theta}_i - \widehat{\psi}_j - u_{it}$ and use the procedures described previously to obtain $\hat{\beta}^m$ and $\hat{\gamma}^m$.

The results of the Monte Carlo simulation, considering $\hat{\theta}_i$ fixed and the sample of stayers, are reported in Table 16. Clearly, neither $\hat{\beta}$ nor $\hat{\gamma}$ seem to be sensitive to the selected parameter perturbation as the computed mean in column (1) is virtually identical to the point

---

[29] $\hat{\theta}_i$ and $\hat{\psi}_j$ are assumed fixed and equal to the parameters obtained in models (2.13) and (2.18).

[30] The number of replications, *B,* must satisfy the rule $\alpha(B+1)=N, N \in \mathbb{N}$, where α is the selected confidence level (Davidson and Mackinnon, 2004, Ch. 4). In our simulations, we set *B = 999.*

estimate reported in column (5). Moreover, the standard deviation in column (2) is in general identical to the standard error in column (5).

As it can be seen in Table 17, the alternative modelling, in which $\hat{\theta}_i$ is assumed to be correlated with schooling, yields approximately the same conclusions. In particularly, the results in column (1) of Table 17 and column (2) of Table 12 are very similar.

The results of the simulation process applied to the sample of switchers are presented in Table 18. We again conclude that under $\widehat{\alpha}_i$ and $\hat{\phi}_j$ fixed the difference between the point estimate in column (5) and the corresponding sample mean is very small. The reported standard errors also seem to be quite robust to the assumed parameter perturbation.

In Table 19 we present the results from a simulation procedure where $\widehat{\alpha}_i$ is now correlated with schooling and follows $\left(\hat{\alpha}_i \middle| Schooling_i\right) \sim N\left(0.076\left(Schooling_i - 9.95\right); 0.116\right)$. We again conclude that the results obtained from assuming a non-fixed $\widehat{\alpha}_i$ are very similar to those reported in Table 13.

The main weakness of the Monte Carlo simulation is that the assumed DGP may be a too strong assumption. To relax this assumption, we use an alternative non-parametric bootstrapping method (with replacement) in order to compute $\hat{\beta}^b$ and $\hat{\gamma}^b$ and the corresponding confidence intervals (*b=1, 2, …B*) for the parameters of the observed characteristics. Then, based on $\bar{\hat{\beta}}$ and $\bar{\hat{\gamma}}$ (i.e. the bootstrap sampling average of $\hat{\beta}^b$ and $\hat{\gamma}^b$), we compute a possible measure of the bias, given by $\hat{\beta}_j - \bar{\hat{\beta}}$, and the corresponding standard deviation.

As shown in Table 20, the bootstrap technique yields roughly the same results as the Monte Carlo. The bias is very small (see columns (1) and (5)). Using Efron's criterion (Efron, 1979), there is no evidence of any problematic bias as the computed bias is always smaller than 25%. We note that the 95 percent bootstrapped confidence intervals in columns (3) and (4) were obtained via three different methods: the first one (in row 1) considers that the distribution of the parameters is a normal distribution; the second (in row 2) is obtained by finding the 2.5 and 97.5 percentiles; and the third (in row 3) is computed in a similar way of the second, but taking into account the median bias. Clearly, the point estimates in column (6) fall within the estimated intervals reported in columns (3) and (4).

To sum up the results contained in Tables 16-20, we can conclude that either $\hat{\beta}$ and $\hat{\gamma}$ are not expected to diverge visibly from the estimates reported in Tables 12 and 13. Assuming that the hypothesis implicit either in the Monte Carlo and bootstrapping simulations are valid, there seems to be therefore not much evidence suggesting that $\hat{\beta}$ and $\hat{\gamma}$ in Tables 12 and 13 are biased or inefficient.

5. Conclusions

Chief among the critical determinants of individual earnings is unmeasured ability of workers and firms. Given that unobserved heterogeneity is expected to be highly correlated with typical covariates in standard human capital earnings functions, proper control of worker and firm effects is crucial to avoid misleading inference on the role of human capital acquisition on earnings.

Using an original LEED dataset, obtained from matching two Portuguese datasets (*Quadros de Pessoal* and *Balanço Social*), we develop in this paper a new approach which tries to take full advantage of a comprehensive array of longitudinal worker and firm characteristics available in our database, including detailed information on firm-provided training.

Our modelling strategy assumes that i) the firm unobserved effect contains the worker average unobserved ability, plus a firm-specific effect; and ii) the gap between individual and firm average wages, unexplained by differences in observable characteristics, gives the extent to which the unobserved ability of a given individual deviates from the average of unobserved worker ability in the firm. Our procedure then enables us to evaluate the bias in standard OLS earnings regressions and analyse the relationship between unmeasured (innate) human capital and observable characteristics of workers and firms.

As expected, the standard human capital earnings function covariates (e.g. schooling, experience, and training) are highly correlated with worker and firm unobserved attributes. The main consequence of the correlation between observed and unobserved attributes is the existence of significant bias associated with selectivity effects. According to our estimates, ignoring worker and firm unobserved effects implies a substantial upward shift in the OLS 'return to education'. We were also able to confirm the negative correlation between unobserved ability and labour market experience obtained in other studies (e.g. Abowd, Lengrmann and McKinney, 2003). Not surprisingly, there is evidence that the correlation between wages and unobserved worker ability is much higher than the correlation between unobserved firm ability and wages.

Our analysis is conducted using two separate sets of individuals (stayers and switchers), and, despite obvious differences between these two sub-samples in terms of group composition (worker mobility is clearly non-random), the main results with respect to the role of unobserved ability on individual earnings seem to hold rather well, although apparently in a different order of magnitude. An interesting finding is that workers with above firm-average innate abilities seem to be the ones that actually change jobs.

Evidence from Monte Carlo simulation and bootstrapping shows that our estimated rates of return to human capital do not seem to be sensitive to parameter perturbation. On the whole, our study does provide therefore further evidence that a comprehensive set of individual and firm characteristics is critical to understanding the role of human capital variables on individual earnings.

REFERENCES

Abowd, J., Kramarz, F. and Margolis, D. (1999). High Wage Workers and High Wage Firms. Econometrica, 67: 251-334.

Abowd, J., Creecy, R. and Kramarz, F. (2002). Computing Person and Firm Effects Using Linked Employer-Employee Data. Technical Paper 2002-06, U.S. Census Bureau, April.

Abowd, J., Lengrmann, P. and McKinney, K. (2003). The Measurement of Human Capital in the U.S. Economy. Technical Paper, TP-2002-09. U.S. Census Bureau, Longitudinal Employer Household Dynamics.

Addison, J., Teixeira, P. and Zwick, T. (2009). Works Councils and the Anatomy of Wages. Industrial and Labor Relations Review. (Forthcoming.)

Andrews, M., Schank, T. and Upward, R. (2006). Practical Fixed-Effects Estimation Methods for the Three-Way Error-Components Model. Stata Journal, 2006, 6, (4), 461-481.

Becker, G. (1962). Investment in Human Capital: A Theoretical Analysis. Journal of Political Economy, 70: 9-49.

Card, D. (1999). The Casual Effect of Education on Earnings. In O. Ashenfelter and D. Card (Eds.), Handbook of Labour Economics. North-Holland.

Davison, R. and Mackinnon, J. (2004). Econometric Theory and Methods. New York: Oxford University Press.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7 (1), 1-26.

Greene, W. (2002). Econometric Analysis (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Johnston, J. and Dinardo, J. (1997). Econometric Methods (4th ed.). New York: McGraw-Hill.

Heckman, J.J. (2008). Skills, Schools, and Synapses. Economic Inquiry 46(3): 289-324.

Mincer, J. (1974). Schooling, Experience and Earnings. New York: Columbia University Press.

Wooldridge, J. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Table 1: Worker and firm level means, all workers, 1998-99

	Worker level (1)	Firm level (2)
(log) Earnings	1.59 (0.60)	1.35 (0.42)
Age (years)	40.69 (10.28)	38.55 (5.23)
Fraction male	0.6460	0.6054
Schooling (years)	7.93 (4.08)	7.35 (2.49)
Tenure (years)	14.02 (9.99)	10.96 (5.84)
Distribution by occupation level:		
Top managers and professionals	0.0641	0.0580
Other managers and professionals	0.0504	0.0577
Foremen and supervisors	0.0674	0.0621
Highly skilled and skilled personnel	0.5583	0.4309
Semiskilled personnel	0.1699	0.2267
Unskilled personnel	0.0837	0.1228
Fraction of trainees	0.5315	0.4110
Distribution by location:		
Norte	0.311	0.367
Centro	0.086	0.141
Lisboa e Vale do Tejo	0.586	0.460
Alentejo	0.012	0.013
Algarve	0.006	0.016
Foreign ownership	0.280	0.234
Proportion of full-time workers in the firm	0.9085	0.8779
(log) Productivity	2.89 (1.04)	2.49 (0.88)
Number or workers	401,258	401,258
Number of firms	1,792	1,792

Standard deviations in parenthesis.

Note: The description of the variables is presented in the Appendix Table A1. The sample comprises only workers who are observed in consecutive years (i.e. in 1998 and 1999).

Table 2: Worker level means in Sample 1 and Sample 2 (estimation samples, 1998-1999)

	Sample 1 (Stayers) (1)	Sample 2 (Switchers) (2)
(log) Earnings	1.59 (0.60)	1.47 (0.65)
Age (years)	40.75 (10.26)	32.16 (8.31)
Fraction male	0.6521	0.6611
Schooling (years)	7.93 (4.05)	9.97 (4.04)
Tenure (years)	14.26 (9.90)	2.47 (3.55)
Distribution by occupation level:		
Top managers and professionals	0.0635	0.0839
Other managers and professionals	0.0519	0.0570
Foremen and supervisors	0.0661	0.0430
Highly skilled and skilled personnel	0.5704	0.4724
Semiskilled personnel	0.1647	0.1337
Unskilled personnel	0.0780	0.2100
Fraction of trainees	0.4998	0.5230
Distribution by location:		
Norte	0.311	0.257
Centro	0.084	0.042
Lisboa e Vale do Tejo	0.589	0.685
Alentejo	0.012	0.007
Algarve	0.005	0.009
Foreign ownership	0.286	0.367
Proportion of full-time workers in the firm	0.9106	0.8773
(log) Productivity	2.89 (1.04)	2.68 (1.01)
Number or workers	357,081	4,069
Number of firms	1,475	802

Standard deviation in parenthesis.

Note: See Table 1.

Table 3: Summary statistics of $\hat{\theta}_i$, Sample 1 (Stayers)

Minimum	-1.699
Maximum	2.987
Mean	0
Median	-0.068
Standard deviation	0.331
Number of observations	714,162

Note: $\hat{\theta}_i$ was obtained using model (2.13).

Table 4: Correlation between observable attributes, X_i, and unobservable worker ability, $\hat{\theta}_i$. Sample 1 (Stayers)

	Coefficient
Schooling	0.4052
Tenure	-0.0378
Experience	-0.1007
Training	0.2227
Top managers and professionals	0.4966
Other managers and professionals	0.2141
Foremen and supervisors	0.1350
Highly skilled and skilled personnel	-0.1969
Unskilled personnel	-0.1765
Unskilled workers	-0.1478
Gender (Male)	0.1160
(log) Earnings	0.5940

Table 5: Summary statistics of unobserved firm effects, $\widehat{\psi}_j$, Sample 1 (Stayers)

Minimum	-0.792
Maximum	1.159
Mean	-0.001
Median	-0.006
Standard deviation	0.272
Number of observations	714,162

Note: This table reports the unobserved firm effect after excluding observable, time-invariant characteristics (see model (2.18)).

Table 6: Correlation between $\widehat{\psi}_j$ and firm average characteristics, $\overline{X}$ and Z, Sample 1 (Stayers)

	Coefficient
Schooling	0.4706
Tenure	0.0022
Experience	-0.1334
Training	0.2370
Gender (Male)	0.1513
Top managers and professionals	0.3924
Other managers and professionals	0.3354
Foremen and supervisors	0.0438
Highly skilled and skilled personnel	0.1049
Unskilled personnel	-0.1589
Unskilled workers	-0.2037
Foreign ownership	0.2336
Productivity bonus	0.0412
Medium/large firm	0.0223
Proportion of full-time workers	0.1765
Unobservable Characteristics (workers)	-0.0039

Table 7: Correlation between firm average wages and observed and unobserved attributes, Sample 1 (Stayers)

	Correlation
$\widehat{\psi}_j$	0.7246
$\overline{X}\widehat{\beta}$	0.4579
$Z_{jt}\widehat{\gamma}$	0.1530
$Z_j{}'\widehat{\kappa}$	0.6495

Table 8: Summary statistics of $\widehat{\alpha}_i$ and $\hat{\phi}_j$, Sample 2 (Switchers)

	$\widehat{\alpha}_i$	$\hat{\phi}_j$
Minimum	-1.219	-0.786
Maximum	2.446	1.512
Mean	0	0.015
Median	-0.074	-0.008
Standard deviation	0.452	0.275
Number of observations	8,052	8,052

Table 9: Correlation between $\widehat{\alpha}_i$ and worker characteristics, X_i, Sample 2 (Switchers)

	Coefficient
Schooling	0.6699
Tenure	0.1100
Experience	-0.4427
Training	0.4696
Top managers and professionals	0.4713
Other managers and professionals	0.2710
Foremen and supervisors	0.0215
Highly skilled and skilled personnel	0.0362
Unskilled personnel	-0.2799
Unskilled workers	-0.2937
Gender (Male)	0.1789
Earnings	0.8404

Table 10: Correlation across unobserved effects, Sample 2 (Switchers)

	$\widehat{\alpha}_i$	$\hat{\phi}_j$	$\hat{\theta}_i$	$\widehat{\psi}_j$
$\widehat{\alpha}_i$	1			
$\hat{\phi}_j$	0.2315	1		
$\hat{\theta}_i$	0.8417	0.4912	1	
$\widehat{\psi}_j$	0.3807	0.4948	0.0799	1

Table 11: Correlation between firm average wage and observed and unobserved attributes, Sample 2 (Switchers)

	Correlation
$\hat{\psi}_j$	0.6878
$\hat{\phi}_j$	0.5190
$\overline{X}\hat{\beta}$	0.8355
$Z_{jt}\hat{\gamma}$	0.3351
$Z_j'\hat{\kappa}$	0.6548

Table 12: Earnings regressions, Sample 1 (Stayers)

Variables	Coefficients	
	Without control for unobserved worker and firm effects (1)	With control for unobserved worker and firm effects (2)
Worker characteristics:		
Schooling	0.048	0.010
	(296.99)	(147.57)
Tenure	0.014	0.011
	(306.24)	(565.86)
Experience	0.010	0.004
	(181.37)	(203.24)
Gender (Male)	0.151	0.052
	(177.08)	(151.09)
Top managers and professionals	0.864	0.121
	(387.54)	(134.74)
Other managers and professionals	0.600	0.117
	(269.45)	(130.80)
Foremen and supervisors	0.420	0.108
	(207.94)	(132.43)
Highly skilled and skilled personnel	0.220	0.066
	(148.49)	(110.34)
Semiskilled personnel	0.087	0.022
	(52.76)	(33.34)
Training	0.086	-0.007
	(80.35)	(-17.14)
Firm Characteristics:		
Productivity bonus	0.206	0.020
	(95.79)	(23.31)
Proportion of full-time workers	0.141	0.029
	(34.21)	(17.21)
Proportion of fixed-term contract workers	-0.071	0.083
	(-24.34)	(70.54)
Foreign ownership	0.076	0.011
	(80.63)	(27.79)
Medium/large firm	0.023	-0.018
	(22.49)	(-38.83)
Norte	-0.065	-0.086
	(-66.91)	(-220.50)
Centro	-0.104	-0.095
	(-69.44)	(-156.29)
Alentejo	-0.041	-0.037
	(-10.93)	(-24.44)
Algarve	-0.004	-0.197
	(-0.67)	(-90.67)
Number of observations	714,162	714,162
F − Statístic	44,044.22	.
$\bar{R}^2$	0.7435	0.9132

t-statistics in parenthesis.

Notes: Columns (1) and (2) present the estimates from models (4.2) and (4.3), respectively. The description of variables is presented in the Appendix Table A1. The model includes a constant, 27 industry dummies, and 2 dummies flagging the legal status of the firm.

Table 13: Earnings regressions. Sample 2 (Switchers)

Variables	Coefficients: Without control for unobserved worker and firm effects (1)	Coefficients: With control for unobserved worker and firm effects (2)
Worker characteristics:		
Schooling	0.060	0.029
	(35.31)	(44.54)
Tenure	0.020	0.021
	(16.57)	(45.81)
Experience	0.016	0.025
	(29.53)	(122.21)
Gender	0.149	-0.008
	(17.15)	(-2.37)
Top managers and professionals	0.905	0.267
	(46.20)	(35.67)
Other managers and professionals	0.676	0.258
	(33.24)	(33.20)
Foremen and supervisors	0.457	0.269
	(21.81)	(33.50)
Highly skilled and skilled personnel	0.229	0.145
	(19.75)	(32.59)
Semiskilled personnel	0.036	0.095
	(2.41)	(16.85)
Training	0.115	0.013
	(10.89)	(3.18)
Firm characteristics:		
Productivity bonus	0.087	0.067
	(5.79)	(11.75)
Proportion of full-time workers	0.164	0.015
	(5.28)	(1.28)
Proportion of fixed-term contract workers	0.012	0.039
	(0.68)	(6.00)
Foreign ownership	0.015	0.062
	(1.73)	(18.78)
Medium-large firm	-0.040	-0.022
	(-4.02)	(-5.74)
Norte	-0.033	0.037
	(-3.27)	(9.78)
Centro	-0.211	0.032
	(-10.25)	(4.12)
Alentejo	-0.044	0.009
	(-0.92)	(0.52)
Algarve	0.064	-0.035
	(1.54)	(-2.18)
Number of observations	8.052	8.052
$F - Statístic$	496.31	23,826.05
$\bar{R}^2$	0.7346	0.9925

t-statistics in parenthesis
Note: See Table 12.

Table 14: Summary statistics of the unobservable heterogeneity of workers and firms, AKM methodology

	AKM applied to our data (1)		AKM (1999, Table IV, p. 293) (2)	
	Mean	Standard deviation	Mean	Standard deviation
$\widehat{\alpha}_i$ (M – OI)	0	0.469	0	0.405
$\widehat{\alpha}_i$ (F – OI)	0	0.461	0	0.377
$\widehat{\psi}_j$ (M – OI)	-0.064	0.341	-0.036	0.464
$\widehat{\psi}_j$ (F – OI)	0.050	0.645	0.067	0.512
$\widehat{\psi}_j$ (M – OD)	0.260	0.108	0.003	0.069
$\widehat{\psi}_j$ (F – OD)	0.004	0.065	-0.004	0.057

Notes: *M* denotes male and *F* female; $\widehat{\alpha}_i$ is unobserved ability and $\widehat{\psi}_j$ is the unobserved fixed effect of firms. *OI* and *OD* are the acronyms for the 'order-independent' and the 'order-dependent' methods, respectively.

Table 15: Correlation between wages and the unobserved effects

	AKM applied to our data (1)	AKM (1999, Table VI, p. 295) (2)
$\hat{\theta}_i$	0.935	0.931
$\widehat{\alpha}_i$	0.742	0.733
$\widehat{\psi}_j$	0.246	0.213
Schooling	0.579	0.414

Note: Correlations obtained using the 'order-dependent' method, Men and Women.

Table 16: Estimates of $\hat{\beta}$ and $\hat{\gamma}$ by Monte Carlo, Sample 1 (Stayers; $\hat{\theta}_i$ fixed)

	Monte Carlo Simulation				
	Mean (1)	S. deviation (2)	Minimum (3)	Maximum (4)	Table 12, col. (2) (5)
Schooling	0.0099994	0.0000632	0.0098084	0.0102151	0.0097 (0.000066)
Tenure	0.0109999	0.0000172	0.0109377	0.0110581	0.0106 (0.000019)
Experience	0.0039994	0.0000204	0.0039188	0.0040642	0.0040 (0.000022)
Top managers and professionals	0.1210147	0.0008851	0.1178679	0.1240937	0.1210 (0.000898)
Other managers and professionals	0.1170213	0.0008769	0.1141494	0.1201373	0.1171 (0.000895)
Foremen and supervisors	0.1079901	0.0007832	0.1052811	0.1099861	0.1078 (0.000814)
High-skilled and skilled personnel	0.0659933	0.0005535	0.0643694	0.0676667	0.0658 (0.000597)
Semiskilled personnel	0.0219792	0.0006279	0.0201308	0.0244636	0.0221 (0.000662)
Gender	0.0520007	0.0003402	0.0509779	0.0530827	0.0520 (0.000344)
Training	-0.0070151	0.0004171	-0.0082837	-0.0056537	-0.0074 (0.000433)
Productivity bonus	0.0199772	0.0008218	0.0173335	0.0224428	0.0202 (0.000868)
Proportion of full-time workers	0.0290180	0.0016542	0.0234428	0.0334407	0.0286 (0.001659)
Proportion of fixed-term contracts	0.0830548	0.0011286	0.0795282	0.0874164	0.0827 (0.001173)
Foreign ownership	0.0109932	0.0003592	0.0098937	0.0120874	0.0106 (0.000381)
Medium/large firm	-0.0160171	0.0004006	-0.0174517	-0.0148467	-0.0157 (0.000405)
Norte	-0.0859986	0.0003721	-0.0870803	-0.0845966	-0.0859 (0.000390)
Centro	-0.0950122	0.0005931	-0.0966797	-0.0930806	-0.0946 (0.000605)
Alentejo	-0.0370439	0.0013972	-0.0416618	-0.0321754	-0.0368 (0.001507)
Algarve	-0.1969498	0.0019868	-0.2026307	-0.1907141	-0.1973 (0.002176)

Note: The dependent variable in the simulation is given by $Ln\ w - \hat{\theta}_i - \hat{\psi}_j - u_{it}$.

Table 17: Estimates of $\hat{\beta}$ and $\hat{\gamma}$ by Monte Carlo, Sample 1 (Stayers; $\hat{\theta}_i$ with normal distribution)

	Monte Carlo Simulation				
	Mean (1)	S. deviation (2)	Minimum (3)	Maximum (4)	Table 12, col. (2) (5)
Schooling	0.0100002	0.0000685	0.0097766	0.0102089	0.0097 (0.000066)
Tenure	0.0110003	0.0000182	0.0109350	0.0110626	0.0106 (0.000019)
Experience	0.0040000	0.0000216	0.0039295	0.0040600	0.0040 (0.000022)
Top managers and professionals	0.1209664	0.0009362	0.1179862	0.1240024	0.1210 (0.000898)
Other managers and professionals	0.1169887	0.0008896	0.1140173	0.1198012	0.1171 (0.000895)
Foremen and supervisors	0.1079642	0.0008392	0.1053727	0.1103254	0.1078 (0.000814)
High-skilled and skilled personnel	0.0659721	0.0005878	0.0641359	0.0676391	0.0658 (0.000597)
Semiskilled personnel	0.0219646	0.0006746	0.0200118	0.0239103	0.0221 (0.000662)
Gender	0.0520009	0.0003530	0.0509562	0.0530952	0.0520 (0.000344)
Training	-0.0070100	0.0004290	-0.0084507	-0.0057893	-0.0074 (0.000433)
Productivity bonus	0.0200177	0.0008566	0.0178788	0.0276767	0.0202 (0.000868)
Proportion of full-time workers	0.0290211	0.0016997	0.0230692	0.0337262	0.0286 (0.001659)
Proportion of fixed-term contracts	0.0830782	0.0011580	0.0796214	0.0868951	0.0827 (0.001173)
Foreign ownership	0.0109964	0.0003896	0.0098830	0.0120868	0.0106 (0.000381)
Medium/large firm	-0.0160064	0.0004076	-0.0172912	-0.0146968	-0.0157 (0.000405)
Norte	-0.0859831	0.0003863	-0.0871866	-0.0847645	-0.0859 (0.000390)
Centro	-0.0950116	0.0006026	-0.0968601	-0.0925035	-0.0946 (0.000605)
Alentejo	-0.0370362	0.0015524	-0.0417850	-0.0317029	-0.0368 (0.001507)
Algarve	-0.1969216	0.0021338	-0.2033090	-0.1896298	-0.1973 (0.002176)

Note: See Table 16.

Table 18: Estimates of $\widehat{\beta}$ and $\widehat{\gamma}$ by Monte Carlo, Sample 2 (Switchers; $\widehat{\alpha}_i$ fixed)

	Monte Carlo Simulation				Table 13, col. (2)
	Mean (1)	S. deviation (2)	Minimum (3)	Maximum (4)	(5)
Schooling	0.0290104	0.0006613	0.0266604	0.0307493	0.0287 (0.000646)
Tenure	0.0210146	0.0004691	0.0190620	0.0224663	0.0212 (0.000463)
Experience	0.0249907	0.0002182	0.0241830	0.0257162	0.0245 (0.000201)
Training	0.0127988	0.0042236	-0.0001219	0.0251638	0.0128 (0.004049)
Top managers and professionals	0.2668372	0.0086775	0.2431742	0.2923803	0.2674 (0.007498)
Other managers and professionals	0.2578592	0.0081529	0.2328640	0.2828239	0.2583 (0.007779)
Foremen and supervisors	0.2691041	0.0082291	0.2448734	0.2947112	0.2688 (0.008024)
High-skilled and skilled personnel	0.1448386	0.0044957	0.1298593	0.1587391	0.1445 (0.004435)
Semiskilled personnel	0.0947854	0.0057602	0.0761082	0.1119815	0.0955 (0.005664)
Productivity bonus	0.0670428	0.0058549	0.0498623	0.0855237	0.0673 (0.005726)
Proportion of full-time workers	0.0151590	0.0121499	-0.0310917	0.0570088	0.0150 (0.011728)
Proportion of fixed-term contracts	0.0387175	0.0067440	0.0871570	0.0562703	0.0387 (0.006456)
Foreign ownership	0.0618605	0.0034178	0.0510272	0.0758158	0.0616 (0.003281)
Medium/large firm	-0.0220673	0.0039956	-0.0367898	-0.0100911	-0.0220 (0.003837)
Norte	0.0370223	0.0039155	0.0238178	0.0508251	0.0375 (0.003829)
Centro	0.0318642	0.0079522	0.0082244	0.0579691	0.0324 (0.007872)
Alentejo	-0.0002988	0.0182522	-0.0644768	0.0539856	0.0094 (0.018257)
Algarve	-0.0370223	0.0159225	-0.0870280	0.0174182	-0.0346 (0.015849)

Note: See Table 16.

Table 19: Estimates of $\widehat{\beta}$ and $\widehat{\gamma}$ by Monte Carlo, Sample 2 (Switchers; $\widehat{\alpha}_i$ with normal distribution)

	Monte Carlo Simulation				Table 13, col. (2)
	Mean (1)	S. deviation (2)	Minimum (3)	Maximum (4)	(5)
Schooling	0.0290449	0.0007355	0.0264497	0.0323887	0.0287 (0.000646)
Tenure	0.0210161	0.0004769	0.0195082	0.0223448	0.0212 (0.000463)
Experience	0.0250050	0.0002050	0.0244297	0.0256973	0.0245 (0.000201)
Training	0.0129883	0.0042671	-0.0013624	0.0261930	0.0128 (0.004049)
Top managers and professionals	0.2670442	0.0077661	0.2420791	0.2905999	0.2674 (0.007498)
Other managers and professionals	0.2578111	0.0078029	0.2326471	0.2824763	0.2583 (0.007779)
Foremen and supervisors	0.2693494	0.0080113	0.2444046	0.2941111	0.2688 (0.008024)
High-skilled and skilled personnel	0.1451267	0.0047463	0.1306952	0.1592565	0.1445 (0.004435)
Semiskilled personnel	0.0953967	0.0057803	0.0716830	0.1142916	0.0955 (0.005664)
Productivity bonus	0.0668537	0.0059313	0.0458505	0.0858605	0.0673 (0.005726)
Proportion of full-time workers	0.0151609	0.0123295	-0.0186483	0.0593766	0.0150 (0.011728)
Proportion of fixed-term contracts	0.0383938	0.0064554	0.0186516	0.0582680	0.0387 (0.006456)
Foreign ownership	0.0621491	0.0034299	0.0518939	0.0739912	0.0616 (0.003281)
Medium-large firm	-0.0221641	0.0040793	-0.0349302	-0.0082266	-0.0220 (0.003837)
Norte	0.0370282	0.0039563	0.0245599	0.0499139	0.0375 (0.003829)
Centro	0.0325014	0.0078785	0.0093549	0.0599668	0.0324 (0.007872)
Alentejo	0.0006147	0.0184734	-0.0564151	0.0544352	0.0094 (0.018257)
Algarve	-0.0354781	0.0165162	-0.0877203	0.0189340	-0.0346 (0.015849)

Note: See Table 16.

Table 20: Estimates of $\widehat{\beta}$ and $\hat{\gamma}$ by bootstrapping, Sample 1 (Stayers)

	Bootstrap					Table.12, col. (2)
	Bias[1] (1)	S. deviation (2)	95 percent confidence interval[2] (3)	(4)	Magnitude of bias[3] (5)	(6)
Schooling	-1.35E-06	6.99E-05	0.0095469 0.0095464 0.0095528	0.0098212 0.0098258 0.0098304	-1.93%	0.0097
Tenure	-6.63E-07	1.94E-05	0.0105849 0.0105836 0.0105863	0.0106612 0.0106617 0.0106663	-3.42%	0.0106
Experience	4.42E-07	2.14E-05	0.0044262 0.0044237 0.0044230	0.0045103 0.0045084 0.0045067	2.07%	0.0045
Top managers and professionals	3.32E-05	0.001011	0.1190461 0.1189971 0.1189414	0.1230121 0.1231345 0.1229785	3.29%	0.1210
Other managers and professionals	-1.9E-05	0.001040	0.1150532 0.1149566 0.1150434	0.1191361 0.1190609 0.1190690	-1.84%	0.1171
Foremen and supervisors	6.08E-07	0.000810	0.1061811 0.1061830 0.1061306	0.1093582 0.1093485 0.1093058	0.08%	0.1078
Highly skilled and skilled personnel	-1.7E-05	0.000611	0.0646410 0.0645903 0.0646456	0.0670369 0.0670281 0.0671172	-2.82%	0.0658
Semiskilled personnel	1.36E-06	0.000599	0.0208985 0.0208456 0.0208178	0.0232475 0.0232419 0.0231956	0.23%	0.0221
Gender	2.36E-05	0.000349	0.0513468 0.0513408 0.0512959	0.0527148 0.0527576 0.0527129	6.77%	0.0520
Training	0.000011	0.00045	-0.0083030 -0.0082815 -0.0082815	-0.0065360 -0.0065568 -0.0065584	2.44%	-0.0074
Productivity bonus	-8.51E-06	0.000921	0.0184123 0.0182485 0.0182103	0.0220250 0.0221213 0.0220236	-0.92%	0.0202
Proportion of full-time workers	4.49E-05	0.001813	0.0249988 0.0249815 0.024888	0.0321158 0.0322694 0.0321682	2.48%	0.0286
Proportion of fixed-term contract workers	5.87E-05	0.001347	0.0800967 0.0800572 0.0798839	0.0853836 0.0853901 0.0852447	4.36%	0.0827
			0.0097320	0.0114259		

Foreign ownership	3.04E-06	0.000432	0.0097461 0.0096859	0.0114317 0.0113918	0.70%	0.0106
Medium/large firm	-1.7E-05	0.000364	-0.0164254 -0.0164667 -0.0164406	-0.0149964 -0.0150080 -0.0149839	-4.78%	-0.0157
Norte	5.81E-06	0.000419	-0.0867246 -0.0867085 -0.0867085	-0.0850793 -0.0850329 -0.0850523	1.39%	-0.0859
Centro	1.68E-05	0.000461	-0.095477 -0.0954153 -0.0954398	-0.0936684 -0.0936383 -0.0936556	3.65%	-0.0946
Alentejo	1.39E-05	0.001074	-0.0389497 -0.0390008 -0.0390695	-0.0347336 -0.0348883 -0.0349240	1.29%	-0.0368
Algarve	1.96E-05	0.002008	-0.2012195 -0.2012323 -0.2012595	-0.1933379 -0.1931443 -0.1932159	0.98%	-0.1973

Notes: The reported values were computed considering model (4.3).

[1] The bias is given by $\hat{\beta}_j - \bar{\hat{\beta}}_j$, where $\bar{\hat{\beta}}_j$ denotes the average of β_j^b, obtained using the bootstrap samples.

[2]The 95 percent bootstrapping confidence interval is obtained using three alternative methods: the first one (row 1) considers that the distribution of the parameters is a normal distribution, the second (row 2) is obtained by finding the 2.5 and 97.5 percentiles, and the third (row 3) is identical to the second one but includes a correction for the bias.

[3] The relative magnitude of the bias is obtained dividing the bias in column (1) by the corresponding sample standard deviation.

Figure 1: Unobserved worker ability ($\hat{\theta}_i$) density function

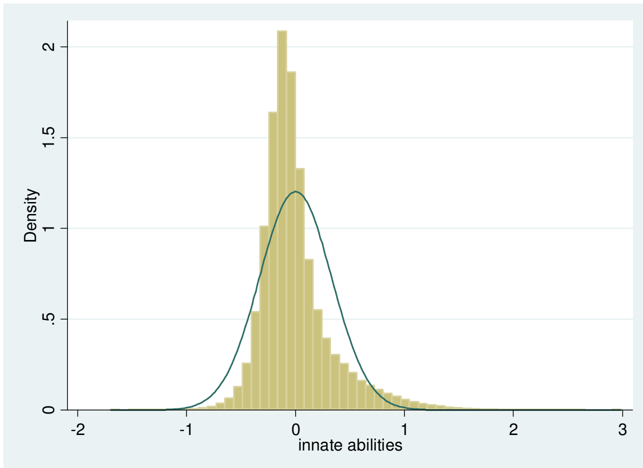
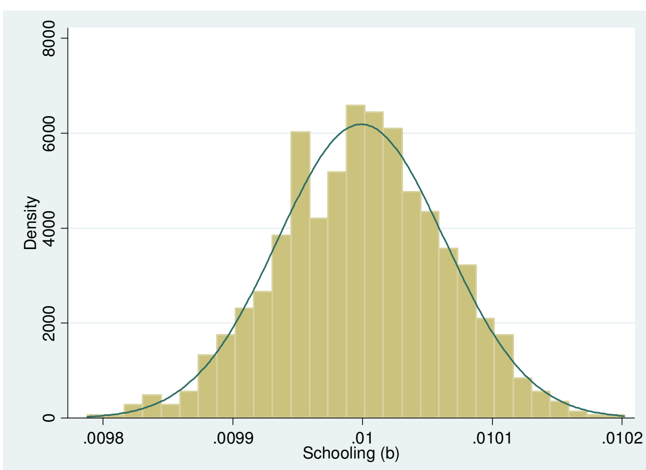


Figure 2: Distribution of the coefficient on schooling, Monte Carlo simulation



Appendix Table A1: Description of Variables

Variable	Definition
Earnings	Hourly (log) gross earnings. This variable is obtained dividing total monthly earnings (in euros) by the number of monthly hours worked.
Schooling	Schooling level in years.
Tenure	Number of years in the current firm.
Experience	Labor market potential experience excluding the experience in the current job. It is defined as Age-6-Schooling-Tenure.
Gender (Male)	Dummy: 1 if the worker is male; 0 otherwise.
Top managers and professionals	Dummy: 1 if the worker is *Quadro Superior*; 0 otherwise.
Other managers and professionals	Dummy: 1 if the worker is *Quadro Médio*; 0 otherwise.
Foremen and supervisors	Dummy: 1 if the worker is *Encarregado, contramestre, mestre ou chefe de equipa*; 0 otherwise.
Highly skilled and skilled personnel	Dummy: 1 if the worker is *Profissional altamente qualificado e profissional qualificado*; 0 otherwise.
Semiskilled personnel	Dummy: 1 if the worker is *Profissional semi-qualificado*; 0 otherwise.
Unskilled personnel	Dummy: 1 if the worker is *Profissional não-qualificado*; 0 otherwise.
Training	Dummy: 1 if the worker has participated in firm provided training; 0 otherwise.
Norte/Centro/Lisboa e Vale do Tejo/Alentejo/Algarve	Dummy: 1 if the firm is located in the North/Centro/Lisboa e Vale do Tejo/Alentejo/Algarve; 0 otherwise.
Productivity bonus (firm average)	Ratio between non-standard compensation and basic earnings
Proportion of full-time workers	Percentage of full-time employees in the firm.
Proportion of fixed-term contract workers	Percentage of fixed-term contract workers in the firm.
Foreign ownership	Dummy: 1 if the firm is owned partial or totally by foreigners; 0 otherwise.
Medium/large firm	Dummy: 1 if the number of employees is more than 250; 0 otherwise.
Productivity	Ratio between gross value added and total hours worked

Note: The training variable at worker level was obtained using an imputation model that draws on the training information at firm level. (The imputation procedure is available upon request from the authors.)

Appendix Table A2: Summary Statistics of $\widehat{\alpha}_i$ and $\hat{\phi}_j$, Sample 2 (Switchers)

	$\widehat{\alpha}_i$	$\hat{\phi}_j$
Minimum	-1.285	-0.360
Maximum	2.524	0.670
Mean	0	0.015
Median	-0.056	0.016
Standard Deviation	0.464	0.172
Number of observations	8,052	8,052

Appendix Table A3: Earnings regressions. Sample 2 (Switchers)

Variables	Coefficients	
	Table 13, column (1) (1)	With control for unobserved worker and firm effects (2)
Worker characteristics:		
Schooling	0.060	0.035
	(35.31)	(46.55)
Tenure	0.020	0.024
	(16.57)	(45.70)
Experience	0.016	0.029
	(29.53)	(125.74)
Gender (Male)	0.149	-0.004
	(17.15)	(-0.97)
Top managers and professionals	0.905	0.268
	(46.20)	(30.90)
Other managers and professionals	0.676	0.226
	(33.24)	(25.07)
Foremen and supervisors	0.457	0.237
	(21.81)	(25.51)
Highly skilled and skilled personnel	0.229	0.114
	(19.75)	(22.07)
Semiskilled personnel	0.036	0.071
	(2.41)	(10.81)
Training	0.115	0.057
	(10.89)	(12.05)
Firm characteristics:		
Productivity bonus	0.087	0.017
	(5.79)	(2.51)
Proportion of full-time workers	0.164	0.035
	(5.28)	(2.58)
Proportion of fixed-term contract workers	0.012	-0.004
	(0.68)	(-0.53)
Foreign ownership	0.015	0.086
	(1.73)	(22.05)
Medium/large firm	-0.040	0.030
	(-4.02)	(6.64)
Norte	-0.033	0.002
	(-3.27)	(0.44)
Centro	-0.211	-0.036
	(-10.25)	(-3.90)
Alentejo	-0.044	0.067
	(-0.92)	(3.18)
Algarve	0.064	-0.073
	(1.54)	(-4.00)
Number of observations	8.052	8.052
$F - Statístic$	496.31	23,826.05
$\bar{R}^2$	0.7346	0.9925

t-statistics in parenthesis
Note: See Table 12.

APPENDIX

Appendix A1:

Equation (2.2) in the text follows from the assumption that

$$\left(\frac{\sum_{i=1}^{N_{jt}} W_i}{h * N_{jt}}\right)_{jt} \equiv \overline{w}_{jt} = e^{\beta \overline{X}_{jt} + \gamma Z_{jt} + \varsigma_{jt}}. \tag{A1.1}$$

Further assuming that

$$\left(\frac{W}{h}\right)_{it} \equiv w_{it} = e^{\beta X_{it} + \gamma Z_{it} + \vartheta_{it}}, \tag{A1.2}$$

we have

$$\frac{w_{it}}{\overline{w}_{jt}} = \frac{e^{X_{it}\beta + Z_{jt}\gamma + \alpha_{it} + \phi_{jt} + \vartheta_{it}}}{e^{\overline{X}_{jt}\beta + Z_{jt}\gamma + \overline{\alpha}_{jt} + \phi_{jt} + \varsigma_{jt}}} \Leftrightarrow \frac{w_{it}}{\overline{w}_{jt}} = e^{\left(X_{it} - \overline{X}_{jt}\right)\beta + \left(\alpha_{it} - \overline{\alpha}_{jt}\right) + \left(\vartheta_{it} - \varsigma_{jt}\right)}, \tag{A1.3}$$

which yields, taking logarithms, equation (2.7) in the text, that is:

$$\left(Ln\ w_{it} - Ln\ \overline{w}_{j(i)t}\right) = \left(X_{it} - \overline{X}_{j(i)t}\right)\beta + \left(\alpha_{it} - \overline{\alpha}_{j(i)t}\right) + \left(\varepsilon_{it} - \omega_{j(i)t}\right), \tag{A1.4}$$

We note that by replacing (A1.2) by

$$\left(\frac{W}{h}\right)_{it} \equiv w_{it} = e^{\beta' X_{it} + \gamma Z_{it} + \vartheta_{it}}, \tag{A1.2'}$$

we get

$$\left(Ln\ w_{it} - Ln\ \overline{w}_{j(i)t}\right) = X_{it}\beta - \overline{X}_{j(i)t}\beta' + \left(\alpha_{it} - \overline{\alpha}_{j(i)t}\right) + \left(\varepsilon_{it} - \omega_{j(i)t}\right). \tag{A1.4'}$$

In our data, in spite of the evidence in favor of $\beta \neq \beta'$, the resulting correlation between $\hat{\theta}$ obtained from (A1.4) and (A1.4)' is extremely high, at 0.9719. Similarly for firm effects $\hat{\psi}$. All results reported in section 4 are based on equation (A1.2).

Appendix A2:

$\widehat{\sigma}_{\psi}^{2} = \left(F^T F\right)^{-1} \widehat{\sigma}_{w}^{2}$ (Proof)

According to equation (2.17), we have

$$\widehat{\psi}' = \left(F^T F\right)^{-1} F^T \left(L\overline{W}^{j} - \overline{X}^{j} \widehat{\beta} - Z^{j} \widehat{\gamma}\right),$$

which, using (2.15), yields

$$\widehat{\psi} = \left(F^T F\right)^{-1} F^T \left(F\psi + w\right), \tag{A2.1}$$

or

$$\widehat{\psi} = \left(F^T F\right)^{-1} F^T F\psi + \left(F^T F\right)^{-1} F^T \omega = \psi + \left(F^T F\right)^{-1} F^T \omega. \tag{A2.2}$$

Then, assuming $E(\hat{\psi}') = \psi$, we can obtain the variance of $\widehat{\psi}'$, that is,

$$\widehat{\sigma}_{\psi}^{2} = E\left(\widehat{\psi}' - \psi\right)^{2} = E\left[\left(F^T F\right)^{-1} F^T (w^T w) F \left(F^T F\right)^{-1}\right] = \widehat{\sigma}_{w}^{2}\left[\left(F^T F\right)^{-1} F^T F \left(F^T F\right)^{-1}\right] =$$

$$= \widehat{\sigma}_{w}^{2} \left(F^T F\right)^{-1}$$

It is then easy to prove that $\left(F^T F\right)^{-1}$ corresponds to a (*J* x *J*) diagonal matrix, with the (*j* x *j*)th element given by $\dfrac{1}{\sum_{t=1}^{2} N_{jt}}$.