González de San Román, Ainara; Rebollo-Sanz, Yolanda F.

Working Paper

An estimation of worker and firm effects with censored data

Economics Discussion Papers, No. 2014-28

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: González de San Román, Ainara; Rebollo-Sanz, Yolanda F. (2014) : An estimation of worker and firm effects with censored data, Economics Discussion Papers, No. 2014-28, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/98704

Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2014-28 | July 04, 2014 | http://www.economics-ejournal.org/economics/discussionpapers/2014-28

An Estimation of Worker and Firm Effects with Censored Data

Ainara González de San Román and Yolanda F. Rebollo-Sanz

Abstract
In this paper, the authors develop a new estimation method that is suitable for censored models with two high-dimensional fixed effects and that is based on a sequence of least squares regressions, yielding significant savings in computing time and hence making it applicable to frameworks in which standard estimation techniques become unfeasible. The authors analyze its theoretical properties and evaluate its practical performance in small samples through a detailed Monte Carlo study. Finally, using a longitudinal match employer-employee dataset from Spain, they show that the biases encountered when ignoring censored issues can be significant to the role of firms in terms of wage dispersion: individual heterogeneity explains more than 60% of wage dispersion.

JEL I21 J16 J31
Keywords Fixed effects; algorithm; wage decomposition; censoring; simulation; assortative matching

Authors
Ainara González de San Román, ✉ IE University, Castellón de la Plana, 8, 28006, Madrid, Spain, ainara_pfpi@hotmail.com
Yolanda F. Rebollo-Sanz, Universidad Pablo de Olavide, Crta. Utrera, Km. 1, 41013, Sevilla, Spain

The authors are grateful to CEMFI's faculty for their suggestions and helpful comments, and especially to Stéphane Bohonomme, Manuel Arellano and Manuel Bagues. They also acknowledge the comments from the seminar participants at Pablo Olavide University and ESPE Congress.

Citation Ainara González de San Román and Yolanda F. Rebollo-Sanz (2014). An Estimation of Worker and Firm Effects with Censored Data. Economics Discussion Papers, No 2014-28, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2014-28

Received June 27, 2014 **Accepted as Economics Discussion Paper** July 1, 2014 **Published** July 4, 2014

1. Introduction

Wage heterogeneity is an intrinsic feature of the labour market that cannot be exclusively related to worker characteristics such as education, race, gender and labour market experience, among others. Until recently, all observational data on the labour market have been comprised of individual surveys, household surveys or population censuses, making it impossible to link firms' characteristics to any specific worker. Recently, the advent of linked employer-employee panel datasets has made it possible to investigate the role of firms in individual wage variation and to account for the role of unobservables. The development of suitable econometric methods that take advantage of these new data structures was initiated in the seminal paper of Abowd et al. (1999, 2002), who presented an iterative algorithm leading to the exact solution for the least squares estimation of the model with two high-dimensional fixed effects. New approaches to the estimation of wage equations with two high-dimensional fixed effects can also be found in the studies by Andrews et al. (2008, 2012), Corneliseen, (2008) and Torres et al. (2012). What none of these previous methods takes into account, however, is that in many instances, wages are censored because most of these new longitudinal databases come from administrative earnings reports[1]. The wage information provided by a typical administrative register is based on the employer's social security contributions, which are often top- or bottom-coded for employees whose wages are above or below certain limits, respectively[2]. Thus, it is not possible to observe the true wages of these employees. In this context, if the censored wage is used as the dependent variable, the ordinary least squares method yields inconsistent estimates. In particular, a typical concern raised by these studies is the correlation between individual and firm time-invariant effects. However, given the way in which these time-invariant firm and individual effects are computed, this correlation will most likely be misreported when wages are censored. Although several methods have been developed in the relevant literature to address censoring problems[3], none of them has been used to estimate models with two high-dimensional fixed effects.

In this paper, we propose an estimation method that enables the estimation of models with high-dimensional fixed effects after controlling for censoring. This new estimator is an easily implementable least squares estimation method that is based on an iterative algorithm with intuitive theoretical properties and good large sample properties as shown in a Monte Carlo

[1] Examples of this type of data include the *Declarations Annuelles de Salaires* (DAS) for France, the *Longitudinal Working Life Sample* (Muestra Continua de Vidas Laborales) for Spain, *Quadros de Pessoal* for Portugal, Austrian Social Security Data (ASSD) and the *Beschaftigtenstatistik* for Germany.

[2] This contribution limit is defined by the government and is renewed every year.

[3] The most popular censored estimator is the standard Tobit regression model (Tobin, 1956). An alternative that is attributable to Powell (1986) is known as *"Trimmed least squares"* and is based on censoring the dependent variable in such a way as to restore the symmetry of its distribution. Some other papers in the literature treat the censoring difficulty as a missing data problem using multiple imputation techniques developed by Rubin (1996).

study. We refer to this estimation procedure as *Fill-in Iterated Least Squares* (FILS). This censoring adjustment procedure has two main advantages. Firstly, it is much faster and easier to implement than Tobit. Secondly, and more importantly, it is able to provide an answer even in complex settings when no strategy is available, such as censoring models with high-dimensional fixed effects. In the estimation of these models, Tobit may be unfeasible due to the large scale of the problem.

This paper also contributes to the empirical literature on wage dispersion by providing the first decomposition of wages for Spain that is fully adjusted for censoring and takes into account both firm and individual effects. For this purpose, we apply our FILS estimator. Given these consistent estimates, we can further investigate the importance of firms in explaining wage differences across individuals in the Spanish economy. We use actual matched employer-employee data from the Spanish *Longitudinal Working Live Sample* (LWLS) in which wages are censored. The period under analysis runs from 1996 to 2012, and we focus on a sample of full-time workers.

Our findings show that the analysis of wage determination can be misleading when wages are censored. Firstly, some parameter estimates associated with job qualifications or educational attainment levels are notably biased. Secondly, the role of firm wage policies in wage dispersion is overestimated by more than ten percentage points, while the role of time-invariant individual characteristics is underestimated by fifteen percentage points. Hence, controlling for censored wages appears to reinforce the idea that when explaining individual wage dispersion, what workers "are" is more important than what workers "do". Thirdly, it turns out that our adjustment for censored wages raises the coefficient of correlation between individual and firm time-invariant effects from -9% to -1.2% in a context where top- and bottom-coded wages represent only approximately 19% and 3% of the observations, respectively.

The rest of the paper is organised as follows. The next section describes the method proposed for a simple econometric model in a wage equation context. Under this scenario, Section 3 summarises the theoretical properties of the algorithm. The Monte Carlo results of the simulation exercises are presented in Section 4. Section 5 extends the method to a panel data context in which both worker- and firm-fixed effects may be present. An illustration using real Spanish data are presented in Section 6. The final section concludes with a summary of our findings.

2. The algorithm: Fill-in iterated least squares

We begin by considering the following econometric model,

$$y_i^* = \alpha + \varepsilon_i \tag{1}$$

where y_i^* is the log wage of individual *i*, $i = 1,\ldots, N$ and ε_i is the stochastic error term, which follows a normal distribution $\varepsilon_i \sim N(0, \sigma^2)$. The simplicity of this model enables the derivation of easily interpretable analytical expressions. The set of parameters to be estimated is denoted by θ, which is the (2×1) vector $(\alpha, \sigma)'$. Generally, wages could be top- and bottom-coded, which means that y_i^* is the latent variable, but instead, y_i is observed. Denote with c_T and c_B top- and bottom-coding, respectively. Then,

$$y_i = \begin{cases} c_T & if \quad y_i^* \geq c_T \\ y_i^* & if \quad c_B < y_i^* < c_T \\ c_B & if \quad y_i^* \leq c_B \end{cases}$$

Unlike Tobit, FILS constitutes an alternative way of estimating $\hat{\theta}$ consistently where the underlying idea is very simple and intuitive. Ideally, if we could observe the true wage, we would merely use the conventional OLS regression of the true wage onto a constant. Unfortunately, the observed wage is censored, which implies the inconsistency of the OLS estimator if y_i is used as the dependent variable. The proposed algorithm works as follows. First, we need to construct an initial simulated wage $\tilde{y}_i^{(0)}$ that works well as a proxy for the true wage. Given the observed data, the simulated variable is constructed in the following way: it must be equal to the observed wage whenever the true wage is observed and will be filled by random numbers from a truncated normal distribution whenever the observed wage is equal to the coding. An OLS regression is run in which the simulated dependent variable and the estimators are saved. This is performed iteratively as a sequence of least squares regressions until the estimators converge to the true values of the parameters. In each iteration *k,* the estimated parameters change, as does the simulated wage distribution, which converges further towards the true wage distribution as the iterations proceed.

We now formalise the approach. We require an initial vector of the estimated parameters denoted by $\hat{\theta}^{(0)} = (\hat{\alpha}^{(0)}, \hat{\sigma}^{(0)})'$, which may be obtained by regressing the OLS estimator of y_i onto a constant, that is,

$$\hat{\alpha}^{(0)} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad , \quad \hat{\sigma}^{(0)} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{\alpha}^{(0)})^2\right]^{\frac{1}{2}}$$

As shown in the next section, we could have followed any other criterion for establishing the starting values because it makes no difference in terms of convergence. A crucial assumption when deriving the algorithm is the normality of the errors, which allows us to use the properties of the truncated normal distribution. To make this point clear, equation (1) can be written alternatively as follows, $y_i^* = \alpha + \sigma u_i$ where $u_i \sim N(0, 1)$. Given y_i, $\hat{\theta}^{(0)}$ and a vector of uniform random numbers U_i (which varies at each iteration), we can start the iterative process. The three possible ranges for the observed wage are taken into account when constructing the simulated wage, $\tilde{y}_i^{(0)}$:

(a) If $y_i = c_B$ then $\dfrac{\Phi\left(\dfrac{y_i^* - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)}{1 - \Phi\left(\dfrac{c_B - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)} = U_i$. Solving for $\tilde{y}_i^{(0)}$, we obtain

$$\tilde{y}_i^{(0)} = \hat{\alpha}^{(0)} + \hat{\sigma}^{(0)}\Phi^{-1}\left[\Phi\left(\frac{c_B - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)U_i\right]$$

where $\Phi(\cdot)$ is the cdf of a standard normal distribution.

(b) If $c_B \le y_i \le c_T$ we observe the true values, and therefore, $\tilde{y}_i^{(0)} = y_i = y_i^*$

(c) If $y_i = c_T$ then $\dfrac{\Phi\left(\dfrac{\tilde{y}_i^{(0)} - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right) - \Phi\left(\dfrac{c_T - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)}{1 - \Phi\left(\dfrac{c_T - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)} = U$. Solving for $\tilde{y}_i^{(0)}$,

$$\tilde{y}_i^{(0)} = \hat{\alpha}^{(0)} + \hat{\sigma}^{(0)}\Phi^{-1}\left[\Phi\left(\frac{c_T - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right) + \left(1 - \Phi\left(\frac{c_T - \hat{\alpha}^{(0)}}{\hat{\sigma}^{(0)}}\right)\right)U\right]$$

Once $\tilde{y}_i^{(0)}$ has been constructed, a new set of parameters $\hat{\theta}^{(1)} = (\hat{\alpha}^{(1)}, \hat{\sigma}^{(1)})'$ can be estimated using the simulated wage as the dependent variable. If $\hat{\theta}^{(1)} \approx \hat{\theta}^{(0)}$, then the iteration can be stopped. If not, we simulate $\tilde{y}_i^{(1)}$ given $\hat{\theta}^{(1)}$ and U, estimate $\hat{\theta}^{(2)}$ and compare $\hat{\theta}^{(2)}$ with $\hat{\theta}^{(1)}$. The stopping rule for convergence is to iterate until $\hat{\theta}^{(k)} \approx \hat{\theta}^{(k-1)}$, where k is the number of iterations.

The consistency of FILS is guaranteed, as shown in the next section. Note also that FILS is easier to implement than Tobit because it requires only a series of least squares estimations. As shown below, in practice, this algorithm can result in enormous savings in computing time.

3. Properties of FILS

In this section, we discuss the theoretical properties of FILS. We start with the numerical convergence of the algorithm, after which we establish the moment conditions, provide the proof of consistency, and derive a robust formula for the standard errors. For the sake of analytical simplicity, the simple econometric model from the previous section is used for the particular case in which we have only right censoring, denoted by *c*. The properties derived also apply to the general case of both-side coding and to more general models.

3.1. Numerical convergence

Clearly, there is no obvious reason why the procedure should provide us with estimators that converge to the true values of the parameters. Nevertheless, we next derive some analytical expressions that help test the likelihood of convergence. Take the k^{th} iteration and focus on the vector of estimated parameters one iteration ahead. By the analogy principle[4], the consistent estimators for the intercept, $\hat{\alpha}^{(k+1)}$ and for the variance of the error term, $\hat{\sigma}^{2(k+1)}$, are the sample mean and the sample variance of $\tilde{y}_i^{(k)}$ respectively, that is,

$$\begin{pmatrix} \hat{\alpha}^{(k+1)} \\ \hat{\sigma}^{2(k+1)} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \tilde{y}_i^{(k)} \\ \frac{1}{N}\sum_{i=1}^{N} \left(\tilde{y}_i^{(k)} - \hat{\alpha}^{(k+1)}\right)^2 \end{pmatrix}$$

Denote by *S* the number of simulations of the algorithm undertaken. Starting with the estimator of the intercept and taking the average across simulations, we can write,

$$\hat{\alpha}^{(k+1)} = \frac{1}{SN}\sum_{i=1}^{N}\sum_{s=1}^{S} \tilde{y}_i^{(s,k)} = \frac{1}{SN}\left(S\sum_{i,y_i \le c} y_i + \sum_{i,y_i > c}\sum_{s=1}^{S}\left[\hat{\alpha}^{(k)} + \hat{\sigma}^{(k)} u_i^{(s,k)}\right]\right)$$

$$= \frac{1}{N}\sum_{i,y_i \le c} y_i + \frac{N_1}{N}\hat{\alpha}^{(k)} + \hat{\sigma}^{(k)}\frac{1}{N}\sum_{i,y_i > c}\left(\frac{1}{S}\sum_{s=1}^{S} u_i^{(s,k)}\right)$$

[4] The analogy principle for choosing an estimator is to turn the population object into its sample counterpart. See Goldberger (1968) for a detailed theoretical description.

where $N_1 = \#\{i, y_i > c\}$, $\frac{N_1}{N}$ is the relative frequency of censored observations and $u_i^{(s,k)}$ follows a standard normal distribution with a truncation point $\frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}$. When the number of observations tends to infinity, by the law of large numbers,

$$p\lim \frac{1}{S}\sum_{s=1}^{S} u_i^{(s,k)} \xrightarrow[S\to\infty]{} E\left[u_i^{(k)} \mid u_i^{(k)} > \frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right]$$

To find this limiting distribution, we need to derive $E\left[u_i^{(k)} \mid u_i^{(k)} > \frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right]$. To that end, the following fact regarding the normal distribution is needed: if $z \sim N(0,1)$, then for any constant h, $E[z \mid z > h] = \frac{\phi(h)}{1-\Phi(h)} = \lambda(h)$, where $\phi(\cdot)$ is the probability density function of the standard normal distribution, $\Phi(\cdot)$ is the cumulative distribution function and $\lambda(\cdot)$ is the inverse Mill's ratio. Applying this property to our particular case,

$$E\left[u_i^{(k)} \mid u_i^{(k)} > \frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right] = \frac{\phi\left(\frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right)}{1-\Phi\left(\frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right)} = \lambda\left(\frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}\right)$$

Finally, substituting into the equation for $\hat{\alpha}^{(k+1)}$ yields,

$$\hat{\alpha}^{(k+1)} = \frac{1}{N}\sum_{i,y_i<c} y_i + \frac{N_1}{N}\left[\hat{\alpha}^{(k)} + \hat{\sigma}^{(k)}\lambda(A)\right] \qquad \text{where} \quad A = \frac{c-\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}}$$

Exactly the same procedure can be applied to the estimated variance of the k^{th} iteration.

$$\hat{\sigma}^{2(k+1)} = \frac{1}{SN}\sum_{i=1}^{N}\sum_{s=1}^{S}\left(\tilde{y}_i^{(s,k)}\right)^2 - \left(\hat{\alpha}^{(k+1)}\right)^2$$

$$\text{where } \frac{1}{SN}\sum_{i,s}\left(\tilde{y}_i^{(s,k)}\right)^2 = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N}\sum_{i,y_i<c} y_i^2 + \frac{1}{N}\sum_{i,y_i>c}\left[\hat{\alpha}^{(k)} + \hat{\sigma}^{(k)}u_i^{(s,k)}\right]^2\right)$$

Following the same argument as before, when the number of simulations tends to infinity, by the law of large numbers,

$$p\lim \frac{1}{S}\sum_{s=1}^{S}\left(u_i^{(s,k)}\right)^2 \xrightarrow[S\to\infty]{} E\left[\left(u_i^{(k)}\right)^2 \mid u_i^{(k)} > A\right]$$

At this point, to derive $E\left[\left(u_i^{(k)}\right)^2 \mid u_i^{(k)} > A\right]$, we need the second moment property of the truncated normal distribution: if $z \sim N(0, 1)$, then for any constant *h*, $Var[z \mid z > h] = 1 - [\lambda(h)(\lambda(h) - h)]$. If we also take into account the fact that $Var[z \mid z > h] = E[z^2 \mid z > h] - E[z \mid z > h]^2$, then

$$E\left[\left(u_i^{(k)}\right)^2 \mid u_i^{(k)} > A\right] = 1 - [\lambda(A)(\lambda(A) - A)] + \lambda(A)^2 = \delta(A)$$

Given the aforementioned information, a final expression for $\hat{\sigma}^{2(k+1)}$ is obtained,

$$\hat{\sigma}^{2(k+1)} = \frac{1}{N}\sum_{i,y_i<c} {y_i}^2 + \frac{N_1}{N}\left[\hat{\alpha}^{2(k)} + \hat{\sigma}^{2(k)}\delta(A) + 2\hat{\alpha}^{(k)}\hat{\sigma}^{(k)}\lambda(A)\right] - \left(\hat{\alpha}^{(k+1)}\right)^2$$

Now, take the expressions already derived for $\hat{\alpha}^{(k+1)}$ and $\hat{\sigma}^{2(k+1)}$ together to obtain the bivariate function $G\left(\hat{\alpha}^{(k)}, \hat{\sigma}^{(k)}\right) = G$

$$G = \begin{pmatrix} \frac{1}{N}\sum_{i,y_i<c} y_i + \frac{N_1}{N}\left[\hat{\alpha}^{(k)} + \hat{\sigma}^{(k)}\lambda(A)\right] \\ \frac{1}{N}\sum_{i,y_i<c} {y_i}^2 + \frac{N_1}{N}\left[\hat{\alpha}^{2(k)} + \hat{\sigma}^{2(k)}\delta(A) + 2\hat{\alpha}^{(k)}\hat{\sigma}^{(k)}\lambda(A)\right] - \left(\hat{\alpha}^{(k+1)}\right)^2 \end{pmatrix}$$

As a simple test for determining whether the algorithm is likely to converge, note that it will converge if the above mapping is a contraction[5]. As stated by the contraction-mapping theorem, if the function determining a given iteration is contractive, then there exists a fixed point. A sufficient condition for the function to be contractive can be stated as follows: if the eigenvalues of the Jacobian of *G* are less than one in absolute value for all $\hat{\alpha}^{(k)}$ and $\hat{\sigma}^{(k)}$, then it follows that *G* is a contraction map and a fixed point therefore exists. The Jacobian is computed next.

[5] A contraction *f* defined in a metric space (*M, d*) as an operator such that $d(f(x),f(y)) \leq d(x,y),\ \forall x,y \in M$

$$J\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)=\begin{pmatrix}\dfrac{\partial G_1\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\alpha}^{(k)}} & \dfrac{\partial G_1\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\sigma}^{(k)}}\\ \dfrac{\partial G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\alpha}^{(k)}} & \dfrac{\partial G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\sigma}^{(k)}}\end{pmatrix}$$

where

$$\frac{\partial G_1\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\alpha}^{(k)}}=\frac{N_1}{N}\left[1-\lambda'(A)\right]$$

$$\frac{\partial G_1\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\sigma}^{(k)}}=\frac{N_1}{N}\left[\lambda(A)-\lambda'(A)A\right]$$

$$\frac{\partial G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\alpha}^{(k)}}=\frac{N_1}{N}\left[2\hat{\alpha}^{(k)}\left[1-\lambda'(A)\right]-\hat{\sigma}^{(k)}\left[\lambda(A)-\lambda'(A)A\right]\right]+2\hat{\sigma}^{(k)}\lambda(A)$$
$$-2\hat{\alpha}^{(k+1)}\frac{N_1}{N}\left[1-\lambda'(A)\right]$$

$$\frac{\partial G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{\partial\hat{\sigma}^{(k)}}=\frac{N_1}{N}\left[2\hat{\sigma}^{(k)}\left[1+\lambda(A)A\right]+2\hat{\alpha}^{(k)}\left[\lambda(A)-\lambda'(A)A\right]\right]$$
$$-\frac{N_1}{N}(c-\hat{\alpha}^{(k)})\left[\lambda(A)+\lambda'(A)A\right]-2\hat{\alpha}^{(k+1)}\frac{N_1}{N}\left[\lambda(A)+\lambda'(A)A\right]$$

Following Olsen (1978), we use a reparameterisation of the function *G,* which has the computational advantage of imposing concavity. In addition, the new parameterisation is a monotonic transformation of the old one using $\left(\frac{\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}},\frac{1}{\hat{\sigma}^{(k)}}\right)$ as arguments of the function rather than $\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)$. That is, if we denote the parameterised function by $\widetilde{G}$, we can write,

$$\widetilde{G}\left(\frac{\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}},\frac{1}{\hat{\sigma}^{(k)}}\right)=\begin{pmatrix}\widetilde{G}_1\left(\dfrac{\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}},\dfrac{1}{\hat{\sigma}^{(k)}}\right)\\ \widetilde{G}_2\left(\dfrac{\hat{\alpha}^{(k)}}{\hat{\sigma}^{(k)}},\dfrac{1}{\hat{\sigma}^{(k)}}\right)\end{pmatrix}=\begin{pmatrix}\dfrac{G_1\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}{G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}\\ \dfrac{1}{G_2\left(\hat{\alpha}^{(k)},\hat{\sigma}^{(k)}\right)}\end{pmatrix}$$

This transformation is especially appealing if the starting values are located in a part of the parameter space for which the function using the traditional parameterisation is convex[6]. The simulation results reported in the next section confirm that the parameterised function is contractive because convergence occurs most of the time.

3.2. Consistency

The system of two GMM conditions can be easily derived when the number of iterations tends to infinity $p\lim\hat{\theta}^{(k+1)} \xrightarrow[k\to\infty]{} \hat{\theta}_\infty$ where,

$$\hat{\alpha}_\infty = \frac{1}{N}\sum_{i,y_i<c} y_i + \frac{N_1}{N}\left[\hat{\alpha}_\infty + \hat{\sigma}_\infty \lambda\left(\frac{c-\hat{\alpha}_\infty}{\hat{\sigma}_\infty}\right)\right]$$

$$\hat{\sigma}_\infty^2 = \frac{1}{N}\sum_{i,y_i<c} y_i^{\,2} + \frac{N_1}{N}\left[\hat{\alpha}_\infty^{\,2} + \hat{\sigma}_\infty^2 \delta\left(\frac{c-\hat{\alpha}_\infty}{\hat{\sigma}_\infty}\right) + 2\hat{\alpha}_\infty\hat{\sigma}_\infty\lambda\left(\frac{c-\hat{\alpha}_\infty}{\hat{\sigma}_\infty}\right)\right] - \hat{\alpha}_\infty^{\,2}$$

and taking limits when N tends to infinity $p\lim\hat{\theta}_\infty \xrightarrow[N\to\infty]{} \theta_\infty$ where,

$$\alpha_\infty = E\left[1\{y_i < c\}y_i\right] + P(y_i \geq c)\left[\alpha_\infty + \sigma_\infty \lambda(A_\infty)\right]$$

$$\sigma_\infty^2 = E\left[1\{y_i < c\}y_i^2\right] + P(y_i \geq c)\left[\alpha_\infty^2 + \sigma_\infty^2\delta(A_\infty) + 2\alpha_\infty\sigma_\infty\lambda(A_\infty)\right] - \alpha_\infty^{\,2}$$

where $A_\infty = \dfrac{c-\alpha_\infty}{\sigma_\infty}$ and $1\{z\}$ is the indicator of event z happening.

We have a just-identified system: two moment conditions and two parameters to estimate.

$$E\begin{bmatrix} 1\{y_i < c\}y_i + 1\{y_i \geq c\}\left[\alpha_\infty + \sigma_\infty\lambda(A_\infty)\right] - \alpha_\infty \\ 1\{y_i < c\}y_i^2 + 1\{y_i \geq c\}\left[\alpha_\infty^2 + \sigma_\infty^2\delta(A_\infty) + 2\alpha_\infty\sigma_\infty\lambda(A_\infty)\right] - \alpha_\infty^{\,2} - \sigma_\infty^{\,2} \end{bmatrix} = 0$$

The *Identification Assumption* reads as follows,

$$E\left[\psi(y_i;\alpha_\infty,\sigma_\infty)\right] = 0 \qquad \Leftrightarrow \qquad \begin{aligned} \alpha_\infty &= \alpha \\ \sigma_\infty &= \sigma \end{aligned}$$

[6]This same parameterisation can also be used to show that for the Tobit model, there is a unique maximum to the likelihood function for given values.

FILS provides consistent estimates of the parameters. This follows from the previous identification assumption and from taking σ as known[7]. The proof is given in the Appendix.

3.3. Standard errors

From the moment conditions of the just-identified system described above and applying standard GMM theory, the asymptotic variance-covariance matrix of the estimator can be easily obtained. Let $\theta = (\hat{\alpha}, \hat{\sigma})'$ be the consistent estimator of the true parameter values θ and let,

$$\psi = \begin{pmatrix} 1\{y_i < c\}y_i + 1\{y_i \geq c\}[\alpha + \sigma\lambda(A)] - \alpha \\ 1\{y_i < c\}y_i^2 + 1\{y_i \geq c\}[\alpha^2 + \sigma^2\delta(A) + 2\alpha\sigma\lambda(A)] - \alpha^2 - \sigma^2 \end{pmatrix}$$

Asymptotic normality will imply $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega)$

It follows from classical arguments (e.g., Newey and McFadden, 1994) that $\Omega = E\left(\frac{\partial\psi}{\partial\theta'}\right)^{-1} E(\psi\psi')E\left(\frac{\partial\psi}{\partial\theta'}\right)^{-1}$. This is well known in popular terms as the *sandwich* formula. The element $E(\psi\psi')$ is given by the following (2×2) symmetric matrix:

$$E(\psi\psi') = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

where

$$a_{11} = \frac{1}{N}\sum_{i, y_i < c} y_i^2 + \frac{N_1}{N}[\alpha + \sigma\lambda(A)]^2 - \alpha^2$$

$$a_{12} = \frac{1}{N}\sum_{i, y_i < c} y_i^3 + \frac{N_1}{N}[\alpha + \sigma\lambda(A)][\alpha^2 + \sigma^2\delta(A) + 2\alpha\sigma\lambda(A)] - \alpha(\alpha^2 + \sigma^2)$$

$$a_{22} = \frac{1}{N}\sum_{i, y_i < c} y_i^4 + \frac{N_1}{N}[\alpha^2 + \sigma^2\delta(A) + 2\alpha\sigma\lambda(A)]^2 - (\alpha^2 + \sigma^2)^2$$

To obtain the whole expression for Ω, we construct the matrix:

[7] A sufficient condition for consistency is the numerical convergence of the algorithm, which is already discussed. Thus, given convergence, the identification assumption might not be needed.

$$E\left(\frac{\partial \psi}{\partial \theta'}\right) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

where

$$b_{11} = \frac{N_1}{N}\left[1 - \lambda'(A)\right]$$

$$b_{12} = \frac{N_1}{N}\left[\lambda(A) - \lambda'(A)A\right]$$

$$b_{21} = \frac{N_1}{N}\left[2\alpha + \sigma\lambda(A) - \lambda'(A)(c-\alpha) - 2\alpha\lambda'(A)\right] - 2\alpha$$

$$b_{22} = \frac{N_1}{N}\left[2\alpha\left[1 + \lambda(A)(1+A) - \lambda'(A)A\right] - (c-\alpha)\left[\lambda'(A)A + \lambda(A)\right]\right]$$

It can be shown that the FILS method's standard errors are actually larger than those of Tobit, which are simply computed as the inverse of the Hessian of the log-likelihood function. The next section compares the two sets of standard errors through a simulation exercise from a non- analytical perspective.

4. Simulation experiment

Having established the limiting behaviour of the FILS estimator, it is useful to proceed by considering the degree to which these large sample results apply to finite sample sizes. First, in an attempt to validate the convergence of the algorithm to the true values of the parameters, we run an experiment using the model from the previous sections. The sample size is set to *N* = 1000. A range for both $\hat{\theta}$ and *c* is established, in particular, $\hat{\alpha} \in [-10, 10]$, $\hat{\sigma} \in [0.01, 5]$ and $c \in [-10, 10]$. Our experience confirms that the algorithm converges overwhelmingly. We test for this in two ways. First, we check that given a certain population, the algorithm converges to the same fixed value irrespective of the initial conditions. Second, the Jacobian is computed numerically for all possible combinations of $\hat{\theta}$ and *c* in the ranges already defined, and the corresponding eigenvalues are calculated for each of those possible combinations.

Figure 1 depicts the eigenvalues of the Jacobian where each point corresponds to a different combination of $\hat{\alpha}$, $\hat{\sigma}$ and *c*. If the application is contractive, both eigenvalues should be lower than one in terms of absolute value. This happens for 99.93% of the combinations. There is a very small region in which one of the eigenvalues appears to be greater than another in terms of absolute value. It coincides with a combination of the parameters such that $\hat{\alpha}$, $\hat{\sigma}$ and *c* are all very close to zero.

[Insert Figure 1]

The second objective of the experiment is to compute the magnitude of bias caused by $\hat{\theta}_{FILS}$ in the estimation of θ and to determine the efficiency of $\hat{\theta}_{FILS}$ compared to $\hat{\theta}_{Tobit}$. To that end, we carry out a Monte Carlo simulation. We choose a population distribution that depends on the bidimensional vector $\theta = (\alpha,\ \sigma)'$ and use different assumptions for sample size, *N* and the degree of censoring *c*. We set the values of $\theta = (1,\ 2)'$. We next draw a random sample of size *N* from this distribution and generate the artificial y_i^*. Finally, we establish different values for *c* to generate the truncated variable y_i.

We compute the estimate of θ by FILS using the first random sample and the counterpart estimated by a Tobit regression. These first estimators are saved. Then, a new random sample is drawn and another pair of estimates of θ is computed. The process is repeated for several simulations, say *S*. Let $\hat{\theta}_{FILS}^{(s)}$ and $\hat{\theta}_{Tobit}^{(s)}$ be the estimates of θ of FILS and Tobit, respectively, which are based on the s^{th} simulation. Given that $\left\{\hat{\theta}_{FILS}^{(s)}, \hat{\theta}_{Tobit}^{(s)} : s = 1, 2, ..., S\right\}$, we can compute the sample mean and sample variance over Monte Carlo simulations to estimate $E\left(\hat{\theta}\right)$ and $Var\left(\hat{\theta}\right)$, respectively.

Table 1 reports the results for different sample sizes and degrees of censoring. For this experiment, 1000 simulations were run.

[Insert Table 1]

Several conclusions can be drawn from the above table. The bias of the estimation provided by FILS is not significantly greater than that provided by Tobit. The convergence of the algorithm appears to be quite fast, and the estimation improves with the size of the sample. Another unsurprising feature of the results is the loss in efficiency of FILS relative to Tobit, as reflected in the higher standard errors of the former. Of course, as the proportion of censoring decreases, so does the efficiency loss; in fact, in the limit where the level of censoring tends to zero, the relative efficiency tends to one, as both estimators reduce to the classic least squares estimators. To provide visual evidence of the asymptotic normality of the FILS estimator, Figure 2 plots the kernel density estimates for the distributions of $\hat{\alpha}_{FILS}$ and $\hat{\sigma}_{FILS}$ for the first design and *N* = 5000.

[Insert Figure 2]

Finally, we use the "sandwich" formula derived in the previous section to find how well the Monte Carlo standard deviations of $\hat{\theta}_{FILS}$ approximate the computed asymptotic standard errors. Table 2 reports both sets of standard errors for different censoring levels and $N = 5000$.

[Insert Table 2]

As can be observed in the table, the lower the level of censoring, the better the standard errors of FILS across Monte Carlo simulations approximate the asymptotic standard errors.

5. Applications to panel data models

By using the simple econometric model from previous sections, we have been able to compare the estimation results implied by FILS and Tobit, but in more general settings, Tobit can be computationally complex or worse still, unfeasible, such as in wage equations with fixed effects when the individual unit of interest is too large. This largeness of scale requires the inclusion of a large set of dummies in the Tobit regression, which could make the estimation unfeasible. The approach proposed here for estimating censored models can be generalised in a number of ways. In this section, we present two applications where the feasibility of the algorithm is guaranteed, even when Tobit might be impossible to implement.

5.1. Model with worker-fixed effects only

In studies of wage determination, it is well known that not only the observed characteristics of workers but also their time-invariant unobserved heterogeneity are crucial to explaining individual wages. Thus, consider the following panel data model:

$$y_{it}^{*} = \alpha_i + x_{it}'\gamma + \varepsilon_{it} \tag{2}$$

where y_{it}^{*} is the log wage of individual i in period t, x_{it} is a vector of K time-varying exogenous covariates regarding the characteristics of individual i, α_i is the unobserved worker effect and ε_{it} is the stochastic error term, which is assumed to be normally distributed[8], $\varepsilon_{it} \sim N\left(0, \sigma^2\right)$. True wages are right-censored, so the observed wage is y_{it} where $y_{it} = y_{it}^{*}$ if $y_{it}^{*} < c$ and $y_{it} = c$ elsewhere. Equation (2) can be rewritten in matrix notation,

$$y^{*} = D\alpha + X\Gamma + \xi \tag{3}$$

[8] The use and properties of the algorithm could be extended to errors that are not normally distributed.

where y^* is the $(NT \times 1)$ vector of wages, D is the $(NT \times N)$ matrix of indicators for individual $i = 1,...., N$, X is the $(NT \times K)$ matrix of observable time-varying characteristics and ξ is the vector of error components, which is assumed to be $\xi \sim N(0, \sigma^2 I_{NT})$ where I_{NT} is the identity matrix of size $(NT \times NT)$. The parameters of equation (3) to be estimated are: α, the $(N \times 1)$ vector of individual effects; Γ, the $(K \times 1)$ vector of the coefficient on time-varying personal characteristics; and finally, the error variance σ^2.

We next list the steps followed to generate the data based on the model described by equation (3), show how FILS is applied in this context and finally describe how the Monte Carlo simulation study was carried out.

1. Data-generating process

Step 1 (a). Constructing balanced panel data[9]: Draw a sample of N = 1000 individuals. These are observed at the baseline for T = 10 periods. (A second design will increase the number of periods to T = 20).

Step 2 (a). Constructing the true wage, y_{it}^*: Each component of equation (2) needs to be created.

α_i: The $(N \times 1)$ vector of worker-fixed effects is set to be a vector of uniformrandom numbers over the interval [0, 2] with a subsequent mean of 1 and variance of 1/3.

$x_{it}\gamma$: We create a single time-varying regressor (K = 1) to represent an individual's age. At $t = 1$, the initial age is represented as a $(N \times 1)$ vector of uniform random integers over [18,55]. At $t = 2$, a vector of ones is added to the initial age, and so on up to $t = T$. The coefficient of the regressor, γ, interpreted as returns to age, is set to 0.02.

ε_{it}: A $(NT \times 1)$ vector of normally distributed random numbers is drawn with mean 0 and standard deviation $\sigma = 2$.

[9] By "balanced panel data" we refer to a sample in which consecutive observations of individual units are available and the number of time periods is identical from unit to unit. The simulation exercise could be easily extended to account for unbalanced panels.

Step 3 (a). Constructing the censored wage, y_{it}: Setting a value *c* for the top such that 25% of the observations are censored.

2. Application of FILS

Step 1 (a). Estimation strategy for each iteration: There is a debate in the literature regarding whether α_i should be treated as a *fixed effect* or a *random effect*. In particular, it will be a random effect when it is treated as a random variable and a fixed effect when it is treated as a parameter to be estimated for each observation *i*, which means that arbitrary correlation between α_i and x_{it} is allowed[10]. In our estimation strategy, α_i is treated as a fixed effect because age is likely to be correlated with the unobserved components taken to represent skills.

Step 2 (a). Initial set of estimators: We run a fixed-effect regression of y on age to obtain $\left(\hat{\gamma}_{FE}^{(0)}, \hat{\sigma}_{FE}^{(0)}\right)$. Using the least squares first-order conditions, $\hat{\alpha}_i$ can be shown to be $\hat{\alpha}_i^{(0)} = \bar{y}_i - \bar{x}_i \hat{\gamma}_{FE}^{(0)}$, and therefore, $\hat{\alpha}_{FE}^{(0)}$ is the $(N \times 1)$ vector of $\hat{\alpha}_i^{(0)}$ for *i* =1,…, *N*. We obtain the initial $(N + 2 \times 1)$ vector $\hat{\theta}_{FE}^{(0)} = \left(\hat{\alpha}_{FE}^{(0)'}, \hat{\gamma}_{FE}^{(0)}, \hat{\sigma}_{FE}^{(0)}\right)'$.

Step 3 (a). Simulation of the wage $\tilde{y}^{(0)}$: Given $\hat{\theta}_{FE}^{(0)}$ and a vector of uniform random numbers, we simulate $\tilde{y}^{(0)}$, which will be the new dependent variable to be used in the iterative process. Once again, a fixed-effects regression of $\tilde{y}^{(0)}$ is run on age and the worker-fixed effects are recovered as before. This gives the new set of estimated parameters $\hat{\theta}_{FE}^{(1)} = \left(\hat{\alpha}_{FE}^{(1)'}, \hat{\gamma}_{FE}^{(1)}, \hat{\sigma}_{FE}^{(1)}\right)'$.

Step 4 (a). Checking convergence: Compare $\hat{\theta}_{FE}^{(1)}$ and $\hat{\theta}_{FE}^{(0)}$. If they are close enough, stop iterating. If not, go back to step 3.

3. Monte Carlo simulation:

We repeat the process already described for *S* = 100 number of simulations. At each simulation, we save the FILS vector of estimators, which is equal to $\hat{\theta}_{FE}^{(k)}$, where *k* is the number of iterations it took for the algorithm to converge. For the sake of comparability, Tobit estimators

[10] A fixed-effects analysis has a drawback if time-constant explanatory variables are to be included because there is no way to disentangle the effects of such time-invariant observables from the unobserved heterogeneity, which is also constant over time.

are also implemented and saved in each simulation. Finally, the mean and the standard deviation of both the FILS and Tobit estimators across Monte Carlo simulations are computed. The results are shown in Table 3 for the baseline case of $T = 10$ and for the alternative design with $T = 20$.

[Insert Table 3]

A very interesting feature can be observed in the table. The gains made by FILS relative to Tobit in terms of computing time are striking. A look at the benchmark case ($T = 10$) reveals that it takes nearly one minute for FILS to carry out the Monte Carlo simulation, while Tobit takes over six hours. This is because the implementation of Tobit requires the inclusion of the $N = 1000$ individual dummies in the regression. It can also be observed that the loss of efficiency of FILS relative to Tobit is still present. However, the biases in the estimation of γ and σ of using FILS are smaller relative to Tobit.

At this point, a question of crucial importance arises: *Are the worker-fixed effects estimated accurately?* There is an important difference between $\hat{\alpha}_i$ and $\left(\hat{\gamma}_{FE}, \hat{\sigma}_{FE}\right)$: It is known that $\left(\hat{\gamma}_{FE}, \hat{\sigma}_{FE}\right)$ are consistent with fixed T as N tends to infinity, but this is not the case for $\hat{\alpha}_i$. Unfortunately, when the time series dimension T is smaller than the cross-sectional dimension N, those estimates can be severely biased. This is the so-called incidental parameter problem. The evidence in Table 3 (throughout the first row) reflects this phenomenon: notice that the larger the value of T, the more accurate the estimation of the mean of $\hat{\alpha}_i$ across cross-sectional units.

5.2. Model with worker-and firm-fixed effects

A good understanding of wage determination requires an in-depth examination not only of the characteristics of the workers but also of the particular features of the firms that employ them. It has been argued that unobserved heterogeneity at both the worker and firm level plays a crucial role in determining wages. In close consideration of this literature, we consider a panel data model,

$$y_{it}^{*} = \alpha_i + \beta_{j(i,t)} + x_{it}'\gamma + \varepsilon_{it} \tag{4}$$

where y_{it}^{*}, x_{it}, α_i and ε_{it} are defined as in the model given by equation (2), and the new term $\beta_{j(i,t)}$ is the unobserved firm effect for a firm employing worker i at time t. Equation (4) can be written in matrix notation,

$$y^* = D\alpha + F\beta + X\Gamma + \xi \tag{5}$$

where y^*, D, X and ξ are the same matrices defined by equation (3) and F is the $(NT \times J)$ matrix of indicators for the firm effect for a firm employing worker i at time t. The parameters of equation (5) to be estimated are α, the $(NT \times 1)$ vector of individual effects; β, the $(J \times 1)$ vector of firm effects; Γ, the $(K \times 1)$ vector of the coefficient on the time-varying personal characteristics; and the error variance σ^2.

Remarks on Identification: An essential question at this point is whether the parameters of the model are identified or not. The *between-firm mobility* of workers is crucial for the identification of the statistical model; otherwise, it is not possible to identify the fixed effects α_i and $\beta_{j(i,t)}$ separately. This means that $\beta_{j(i,t)}$ must vary over time for some cross-sectional units, that is, there may be people who never change firms, but there must be a fraction of movers in the sample of workers.

Remarks on Estimation: The usual computational methods for the least squares estimation of the parameter vector $(\alpha', \beta', \Gamma')'$ are not feasible because the normal equations for least squares estimations are very large when the model includes more than one level of fixed effects. To solve this problem, AKM (1999, 2002) obtain the least squares solution by means of a within-groups regression of equation (4) including dummies for the employer firms. Unfortunately, the computational complexity of the estimation prevents censoring problems from being properly addressed because this methodology does not work in the presence of such problems, as discussed above.

Two different simulation experiments are presented next to show that the use of the algorithm developed in this paper provides a solution to the censoring problem even when faced with complex models such as the one presented here. In the first experiment, the parameters are arbitrarily set, whereas the second uses a more sophisticated calibration procedure.

5.2.1. Experiment 1: *Arbitrary values for the parameters*

In this experiment, the values of the parameters are set arbitrarily. Because the model described by equation (5) differs from the one described by equation (3) only in the inclusion of the new term $F\beta$, we only need to add a few comments to the description provided earlier of the steps involved in generating the data, applying FILS in this context, and running the Monte Carlo simulation. Furthermore, in this setting, it is also important to discuss how workers should be assigned to firms when simulating the data.

1. Data-generating process

Step 1 (b). In addition to step 1 (a), here we need to construct the sample of firms: in particular, a sample of $J = 20$ firms is drawn.

Step 2 (b). Constructing the true wage, y_{it}^{*}: In addition to the terms created in step 2 (a), it is necessary to create the new term $\beta_{j(i,t)}$.

$\beta_{j(i,t)}$: The $(J \times 1)$ vector of firm-fixed effects is set to be a vector of uniform random numbers over the interval [0, 2] with a mean of 1 and variance of 1/3 – similar to the vector of worker-fixed effects. Note that the value for the fixed effect of a particular firm *j* is time-invariant and identical for all workers employed by that firm.

Step 3 (b). Constructing the censored wage, y_{it}: Setting a value *c* for the top, such that 25% of the observations are censored, as performed previously.

2. How are workers assigned to firms?

We assume that the correlation between worker and firm effects is high and positive. This matching is created as follows: at time $t = 1$, workers with high α s are assigned to firms with high β s and vice versa. From period $t = 2$ onwards, this process is subject to an assignment noise $v_{it} \sim N(0, \tau^2)$ and hence, the assignment variable is $\alpha_i + v_{it}$ instead of α_i. It is important to note that the variance of v_{it} uniquely determines the mobility index. An easy way to understand this point is to imagine the extreme case in which $\tau = 0$; because there is no noise in the assignment process and both α s and β s are time-invariant, an individual starts working for a given firm and stays with that firm forever. Hence, $\tau = 0$ is associated with zero mobility and no identification can be made of the two fixed effects separately. Different values for τ are considered. Benchmark: $\tau = 0.1$.

3. Application of FILS

Step 1 (b). Estimation strategy for each iteration: We follow the methodology of Abowd et al. (1999) (henceforth, AKM) to estimate this type of model, which is to run a within-groups regression of equation (4) including dummies for the employer firms[11].

Step 2 (b). Initial set of estimators: We perform a within-groups regression of y onto age and the firm dummies, and we obtain $\left(\hat{\gamma}_{AKM}^{(0)}, \hat{\sigma}_{AKM}^{(0)}, \hat{\beta}_{j(i,t)AKM}^{(0)'}\right)$. $\hat{\alpha}_i$ can be shown to be $\hat{\alpha}_i^{(0)} = \bar{y}_i - \hat{\beta}_{j(i,t)AKM} - \bar{x}_i\, \hat{\gamma}_{AKM}^{(0)}$. This results in the initial $\left(N + J + 2 \times 1\right)$ vector of estimated parameters, $\hat{\theta}_{AKM}^{(0)} = \left(\hat{\alpha}_{AKM}^{(0)'}, \hat{\gamma}_{AKM}^{(0)}, \hat{\sigma}_{AKM}^{(0)}, \hat{\beta}_{j(i,t)AKM}^{(0)'}\right)'$.

Step 3 (b). Simulation of the wage $\tilde{y}^{(0)}$: Given $\hat{\theta}_{AKM}^{(0)}$ and a vector of uniform random numbers, we obtain a simulation of $\tilde{y}^{(0)}$, which is the new dependent variable to be used in the iterative process. Once again, we run a within-groups regression of $\tilde{y}^{(0)}$ onto age including the firm dummies and recover the worker-fixed effects as before. Thus, we obtain $\hat{\theta}_{AKM}^{(1)} = \left(\hat{\alpha}_{AKM}^{(1)'}, \hat{\gamma}_{AKM}^{(1)}, \hat{\sigma}_{AKM}^{(1)}, \hat{\beta}_{j(i,t)AKM}^{(1)'}\right)'$

Step 4 (b). Checking convergence: Compare $\hat{\theta}_{AKM}^{(1)}$ and $\hat{\theta}_{AKM}^{(0)}$. If they are close enough, stop iterating. If not, go back to step 3.
Monte Carlo simulation: We use S = 100 simulation replications.

Table 4 reports the mean and standard deviation of both the FILS and Tobit estimators across Monte Carlo simulations for the two different values of *T*.

[Insert Table 4]

One main conclusion can be extracted: Regardless of which estimation method is chosen, the fixed effects estimates are more biased and less accurate than $\hat{\gamma}$ and $\hat{\sigma}$. The cause of this is not intrinsically related to our algorithm but rather to the incidental parameter phenomenon already mentioned and to the degree of mobility present in the sample.
The benchmark results presented in Table 4 use $\tau = 0.1$, which implies a mobility index of approximately 50%, where this index is defined as the total percentage of worker mobility across firms over the whole time interval. Alternative results using different values for τ are presented in Table 1.A in the Appendix. There is an important conclusion to be drawn from

[11] A multicollinearity problem arises when the firm dummies are included in the regressions. To overcome this technical difficulty, we perturb the firm dummies by adding to each one a uniform random number with variance 0.0001.

these comparative statistics: the lower the value of τ, and hence the less mobility there is in the sample, the less accurate the fixed effects estimates will be, as seen by the increase in the standard deviation from $\tau = 0.1$ to $\tau = 0.025$. The reason is simple: as mobility decreases, the more difficult it becomes to separate worker effects from firm effects.

A curious finding regarding computing time can be observed in Table 4. Although the differences between FILS and Tobit are still striking, the reason why the Tobit takes so much longer is obviously is due to the inclusion of worker-specific dummies (see Table 3), of which there are 1000, versus only 20 firm dummies. Lastly, the estimated correlation between $\hat{\alpha}_i$ and $\hat{\beta}_{j(i,t)}$ is affected by a strong downward bias, which rapidly diminishes as *T* increases.

In relation to this evidence, we also find it important to analyse how FILS behaves asymptotically, so it becomes necessary to extend the sample size[12]. We therefore set *N* = 5000 and run the simulation for different combinations of *T* and *J*. Table 5 presents the results of these estimations.

[Insert Table 5]

We observe that consistency is still guaranteed for any combination of sample sizes. The most notable differences are found in the standard deviations. If we compare the first column of Table 5 with the third column of Table 4 (which differs only in the number of workers), we can observe that the estimation with larger *N* presents lower standard deviations both for the fixed effects and also for the remaining estimated parameters. It is interesting to note that the efficiency of the estimation of the fixed effects diminishes considerably as the number of firms increases, as shown in the second column of Table 5. Moreover, the percentage of worker mobility across firms over the whole time interval reaches 83%, as can be expected considering that the probability of mobility increases with the number of firms.

The last additional finding to be drawn from the data in the last two columns discussed above is the improved efficiency of the estimators for all the parameters, especially those of the fixed effects, as the time horizon increases.

5.2.2. Experiment 2: *Calibration of the parameters to* AKM

In the previous experiment, the data-generating process involved the use of arbitrary fixed parameter values. For a more realistic experiment, this time, the parameter values are calibrated

[12] The limitation of increasing the sample sizes in this way is that the comparison with Tobit becomes almost impossible. In some cases, it would require weeks of computing time and, in the most ambitious cases, it proves unfeasible.

based on the data given in Tables IV-VI in the seminal paper by Abowd et al. (1999, 2002). Instead of describing all of the details and steps followed above, we now merely discuss the changes with respect to the previous experiment.

Both types of fixed effects are set to follow a normal distribution with zero mean and $\sigma_\alpha = \sigma_\beta = 0.45$. Age is used as a proxy for experience, one of the time-varying explanatory variables used in the original paper, which found a coefficient of 0.04 for the returns to experience. Therefore, we set $\gamma = 0.04$. The variance of the error term is calibrated to match the difference between the variance of log wages and the estimated variance of the observed components. Thus, we set $\sigma_\varepsilon = 0.5$. Lastly, they had a 10-year panel, so we set $T = 10$.

In assigning workers to firms, the previous experiment assumed positive assortative matching. However, in studying France and the US, first Abowd et al. (1999) and later Abowd et al. (2004) found that the estimated correlation between the estimated worker and firm effects is negative or even zero. They show that this lack of positive assortative matching is not explained by estimation biases due to a lack of mobility in the data. Furthermore, they examine the pattern of inter-firm movements between all samples of individuals and find that 40.6% of the individuals in the sample change employer at least once.

To calibrate the characteristics of the mobility found by Abowd et al. (1999), it is more convenient to assign workers to firms by random rather than assortative matching. Random matching works as follows: at $t = 1$, workers are assigned to a firm randomly. From $t = 2$ onwards, there is a constant probability of moving denoted by p. Each individual is hit by a shock represented by a random number between 0 and 1. If the number is greater than p, the individual moves to another firm. If it is lower than p, the worker remains at the current firm. We set $p = 0.15$ constant for all i and t to match the mobility index of 40.6%. Table 6 summarises the results of this simulation experiment. Column (2) shows the Monte Carlo results of applying the algorithm, column (3) reports the estimation results of following the AKM methodology in spite of censoring, and the last column shows the original estimation by Abowd et al. (1999) with no censoring[13].

[Insert Table 6]

By comparing columns (3) and (4) of Table 6, it can be concluded that using AKM in the presence of censoring provides biased estimators of the parameters, whereas AKM with no censoring is consistent, as expected. The main difference in the results of the FILS estimator,

[13] In order to implement this procedure with our simulated data, for column (3), we conduct a within-groups analysis including dummies for the firms using censored wages, and for column (4), we do the same using the true wage as the dependent variable.

given in column (2) of Tables 4 and 6, lies in the accuracy of the estimation of worker- and firm-fixed effects and their correlation values. In particular, the accuracy of the estimates is much higher in the calibration experiment that relies on the random matching of workers and firms (column (2) of Table 6). The efficiency is much closer to that found in the simulations with larger sample sizes shown in Table 5.

6. An Illustration for real data from Spain

Next, we carry out an empirical exercise using real data on the Spanish economy. The source of data used for this illustration is the *Muestra Continua de Vidas Laborales* or *Longitudinal Working Lives Sample* (LWLS), a large-scale employer-employee matched database drawn from the Spanish Social Security administrative register. The LWLS is compiled annually and consists of a sample of over one million individuals, making up a representative sample of the population related to the Spanish Social Security system - either as employees or as unemployed receiving unemployment benefits - in each reference year. This paper uses all available waves of the LWLS (from 2005 to 2012)[14]. This longitudinal linked dataset contains observations about individuals and their employers linked through a working history including information about the jobs held by each individual with each employer. This dataset is based on mandatory reports to the employer about the gross earnings of each employee subject to payroll taxes. Because it is compulsory, it does not suffer from the non-response problems that often plague household and firm surveys. Hence, the LWLS reproduces the complete labour market histories of the individuals, where each observation corresponds to a unique employee-contract-firm combination. In particular, this database provides highly detailed information about workers, including monthly wages, Social Security contribution groups, the type of contract and several characteristics of the hiring firms, such as size, age, ownership, location and sector of activity[15]. Individual characteristics such as age, gender, level of education, province and nationality are also present in the database.

The final estimation sample is obtained through the following selection process. We restrict the analysis to full-time workers in regular employment aged 18 to 60 years over the period 1996-2012. From this database, we build monthly panel data where the unit of analysis is each worker-month combination. Using this selection criterion, the initial sample includes 44,676,551 observations for 619,613 individuals. On this initial database, we need to impose

[14] The initial simple is composed of workers having had any link with social security as a contributor, as a recipient of unemployment benefits or as a pensioner on at least one day in the period 2005-2012.
[15] Nevertheless, firm characteristics are accurately measured only for the period 2005-2012.

additional sample restrictions (see Abowd et al. (2002) or Andrews (2004) for a detailed explanation of these identification conditions).

Firstly, as stated above, inter-firm mobility is critical to this type of analysis because firm effects are identified by the number of movers in each firm. The estimated correlation between firms and individual effects is biased downwards when there are few movers in the data (i.e ***limited mobility bias***[16]**)**. We restrict the analysis to firms with a minimum of five different observed employees[17] to reduce the limited mobility bias in our application. Secondly, we need to identify connected groups of workers and firms because firm-fixed effects are only identified for each connected group. Once all of the connected groups are identified, we keep only the largest group[18]. Hence, the sample used for the estimation contains all workers who have ever worked for any of the firms in that group and all of the firms that have employed them. The selected connected group accounts for 91.7% of observations. Once these sample restrictions are applied, our final database comprises 11,401,929 observations with 494,210 individuals[19] and 68,119 firms (see Table 7). In this dataset, job movers represent 41.01% of the individuals, and their average number of firm movements is 2.71. Finally, 25.30% of the observations involve establishments with five different observed employees and 16.61% with six different observed employees.

[Insert Table 7]

6.1 Model Estimation

The idea is to measure the importance of Spanish firms in the determination of individual wages looking at the contribution of firms to overall wage variance. Hence, we propose to estimate a wage equation similar to equation (4):

$$y_{it}^{*} = \alpha_i + \beta_{ji} + x_{1it}'\gamma_1 + x_{2j(i,t)}'\gamma_2 + \varepsilon_{it} \qquad (6)$$

[16] Andrews et al. (2008) supply a formula that establishes this proposition in a simple, stylized dataset. Ultimately, the size of the bias is an empirical issue and should be computed for every application of employer–employee matched data.

[17] This sample selection criterion has mainly affected firm size characteristics.

[18] It is not possible to identify all of the coefficients for the fixed effects. Abowd et al. (2002) showed that one needs to impose one restriction on the coefficients for each mobility group with two fixed effects. Alternatively, the analysis could exploit all of the connected groups by making a normalisation within each group. However, as the worker and firm effects are measured relative to *G* different normalisations in the fixed-effect model, a comparison of worker and firm effects across the connected groups in the labor market is impossible. It turns out that the largest identified group contains the vast majority of observations, so the loss incurred by ignoring the remaining groups is arguably negligible.

[19] The average number of observations per employee is 39.6.

where there are N workers i =1,.,.N,; J firms j=1,.,.J; and T_i[20] months t=1,..,T_i. The endogenous variable y_{it}^* represents real wages in logarithms; x_{1it} is a vector of observable time-varying worker covariates, and $x_{2j(i,t)}$ is a vector of observable time-varying firm covariates. The terms α_i and β_{ji} are time-invariant individual and firm heterogeneities that could be correlated not only with each other but also with x_{1it} and $x_{2j(i,t)}$. It is common to assume that α_i and β_{ji} are correlated with observed variables from the same side of the market. This means that random effects methods are biased and inconsistent, and thus that two-way fixed-effects methods are needed to consistently estimate the parameters of interest. To circumvent the estimation of N worker effects and J firm effects, Abowd et al. (2004) noted that explicitly including firm heterogeneity dummies but sweeping out the worker heterogeneity algebraically gives exactly the same solution as the Least Squared Dummy Variable estimator. However, when faced with censoring problems (i.e., a Tobit model), it is not possible to apply this solution because the estimated model needs to be linear. The advantage of the algorithm proposed in this paper is that it simultaneously allows the application of the solution proposed by Abowd et al. (2004) and an adjustment for censoring.

In our particular empirical exercise, we take one step further and use the new estimation methods proposed by Carneiro et al. (2008) and later extended by Guimaraes et al. (2009). The advantage of their estimation methods is that they enable the estimation of wage equations even when both the dimensions of J and N are in the order of hundreds of thousands. The explicit introduction of dummy variables is not an option in this case because the number of units either of firms or of individuals is too large[21].

Summing up, to estimate our wage equation, while simultaneously dealing with two high-dimensional fixed effects and adjusting for censoring, we proceed in two stages. In the first stage, we estimate our wage equation using the observed censored wage as our endogenous variable and obtain the coefficient estimates for the individual and firm effects using the partitioned iterative algorithm mentioned above[22]. To speed up this algorithm, we sweep out the individual effects by subtracting the individual's group means from all the variables. Then,

[20] Both workers and firms are assumed to enter and exit the panel, which means that we have an unbalanced panel with Ti observations per worker.

[21] They showed that with a full Gauss-Seidel iterative algorithm, it is possible to obtain least squares solutions for linear regression models with two or more high-dimensional fixed effects, imposing minimal memory constraints. This procedure is computationally intensive but converges steadily albeit slowly.

[22] This algorithm produces stable but slow iteration depending on the correlation of the parameters (Smyth, 1996). However, it does not require the explicit calculation of the inverse of the X'X matrix.

using these initial estimates, we correct the observed censored wage using the FILS algorithm. These two steps are repeated iteratively until convergence of the parameters is reached[23].

6.1.1 Person Effects

In this framework, the worker-fixed effects include both the worker's unobserved and observed but time-invariant characteristics. Similarly, the firm effects include both unobserved and observed time-invariant firm characteristics. The next stage, therefore, is to decompose these two estimated effects into their respective observed and unobserved components using regression techniques. This decomposition can help to clarify the sources of the biases incurred when censored wages are not properly considered in the analysis. In our empirical exercise, we provide a decomposition only of person effects because firms' time-invariant characteristics could not be accurately observed for the chosen period of analysis[24]. The person effect covers both the effects of observable time-invariant personal characteristics and unobserved personal heterogeneity. We can decompose these two components of the pure person effect as,

$$\alpha_i = \theta_i + \pi_i \eta \qquad (7)$$

where θ_i is the unobservable personal heterogeneity, π_i is a vector of time-invariant personal characteristics and η is a vector of the effects associated with time-invariant personal characteristics. An important feature of the decomposition of equation (7) is that an estimation of the person effects can proceed without a direct estimation of η (see Abowd et al., 2006).

6.2 Construction of main variables

The dataset contains information on monthly earnings, which are censored at the upper and lower social security contribution limits. We can estimate our wage equation controlling for censoring, given that the censoring point is known[25]. The monthly wage is deflated using the consumer price index for each period. This is the dependent variable for the analysis (in logarithms). Table 8 presents the main descriptive statistics. The proportion of censored observations in the initial sample is approximately 18.9% (right-censored) and 3.2% (left-censored). Controlling for censoring in this context is important because the share of censored wages is unequally distributed across individuals, as can be observed in Table 8.

[23] The convergence criteria are defined in terms of parameters of interest and the sum of squared residuals. In the Gauss-Siegel iterative algorithm, the convergence is defined in terms of coefficients of the firm effects as well as the error term.
[24] The dataset includes information on the firms, such as their sector of activity or size, for the period 2005-2012.
[25] The upper and lower social security contribution limits are fixed by the government each year.

[Insert Table 8]

According to equation (6), there are four components that explain wage variability: the observed time-varying characteristics of workers, the observed time-varying characteristics of firms, the observed time-varying characteristics of the economy, worker heterogeneity or worker-fixed effects, firm heterogeneity or firm-fixed effects, and an error component assumed to follow the conventional assumptions. There has been much debate on whether variables that control for job characteristics, industry or occupation ought to be included in the specification. We keep our analysis as simple as possible and follow the approach found in Abowd et al. (1999, 2004) and adopted in various recent papers, such as Lalive et al. (2009). In these papers, the wage equation specification uses variables related with individual human capital characteristics as covariates. We decided to adhere to this simple approach to offer a general idea of the biases incurred when censoring issues are ignored[26].

We include in the wage equation the following time-varying covariates: age, tenure, labour market experience, previous unemployment duration (all in logs), the type of contract (permanent contract), job qualifications (high skill job), the regional unemployment rate and the national GDP growth rate. We define previous work experience as the number of accumulated months actually worked since the employee's first employment experience. Tenure reflects the number of months the worker has stayed with the same firm despite possible spells of unemployment between two consecutive jobs at the same firm. Labour market experience and tenure are modelled as second-degree polynomials[27]. We introduce the duration of previous unemployment as a proxy to control for the type of job mobility. Our analysis implicitly assumes strict exogeneity. This implies that workers' mobility decisions, conditional on covariates, are independent of ε_{it}.

6.3 Results

This section reports the estimation results of the wage equation. We should inform the reader that we estimated the same wage equation for alternative sample schemes to assess the generality of the results. As Andrews et al. (2008, 2012) stated, the estimation results notably depend on sample characteristics, such as the period of analysis, the sample size of the firms and

[26] Unfortunately, the Spanish data do not include firm accounting data. Thus, we are unable to control for variables such as capital, value added per worker, etc.

[27] We also estimated the model with polynomials of higher orders, and the results did not change – they are available upon request.

the degree of inter-firm mobility. For instance, the estimation was also run for subsamples composed of firms with at least two and firms with at least ten observed employees. In both cases, the results resemble those presented in this section. Hence, for the sake of conciseness, we chose to omit those results and mention them only when significant differences appear. The results are reported in Table 9 and are displayed for two alternative estimators depending on whether wages are corrected for censoring (FILS) or not[28].

[Insert Table 9]

Observe that the R^2 is considerably higher than it is in standard wage regressions. Worker- and firm-fixed effects, together with worker, firm and aggregate covariates, explain more than 95% of the variability in real wages. The coefficient estimates of the regressors and those of the variance of the error component are reasonable. All estimated coefficients are of the expected sign and statistically significant at the 1% level. This is remarkable taking into account that the standard errors of the estimation with the algorithm are higher than the corresponding ones in the Tobit estimation. The estimates show that both human capital accumulation and mobility seem to be important determinants of observed wage growth over the workers' careers. Wages are increasing with age, tenure and experience and permanent and high-skill jobs are better than the rest. There is a wage penalty for workers with spells of unemployment. Finally, the higher (lower) the regional unemployment rate (GDP), the lower the wage rate.

However, coefficient estimates differ between the two alternative models presented. More specifically, when censored wages are not corrected, the wage penalty for spells of unemployment is overestimated whereas the wage returns for high-skill jobs are underestimated. For instance, while our FILS estimator finds that high-skill jobs command approximately 17% higher monthly wages than other types of jobs, this return drops to 13% when wages are not corrected for censoring. The wage penalty due to unemployment is estimated to be 8.6% for the not corrected model whereas it increases to 9.4 with the FILS. The differences in the coefficient estimates for labor market experience and tenure are also important (Figures 4 and 5).

[Insert Figure 4]

[Insert Figure 5]

[28] Given the sample characteristics, we cannot offer results for the Tobit estimator.

Next, we focus on the estimated correlation between individual and firm time-invariant characteristics because this term has traditionally been used to test for the hypothesis of positive assortative matching. It is important to check for positive assortative matching because of its implications for equity and efficiency in the labour market. Firstly, under positive assortative matching, workers with higher skills will earn higher wages not only because of their higher innate productivity but also because they tend to be employed in better firms. This means that wage inequality will be higher than the underlying inequality in worker abilities. Secondly, positive assortative matching may be evidence of complementarities in production[29] between worker ability and firm productivity. This implies that if workers and firms are optimally matched, total output is higher than it would be under random matching, for example. As a consequence, job search frictions would have negative output effects and policies to prevent search frictions would be important.

Table 9 also shows that the correlation coefficient between time-invariant individual and fixed effects is sensitive to censoring issues. This correlation coefficient increases from -9% to -1.2% when censoring is accounted for, which is a non-negligible difference given the amount of censored observations in the data. Nonetheless, a 600% movement in the correlation represents a sizable bias. Hence, the relationship between firms' compensation policies –proxied by the firm-fixed effects, and the quality of their workforces –proxied by the employee-fixed effect—is shown to be close to zero in Spain[30]. Abowd *et al.* (2004) reported correlations of −0.24 for French data and 0.02 for data fromWashington State, whereas Goux and Maurin (1999), using different French data, found a correlation ranging from −0.32 to 0.01 depending on the time period that was chosen. Gruetter and Lalive (2004) reported a correlation of −0.27 for Austrian data. All of these are weaker than Barth and Dale-Olsen's (2003) correlations of between −0.47 and −0.55 for Norwegian data. In other words, when focusing on unobserved components, low wage workers tend to work in high wage firms, and vice versa. This seems counterintuitive in the light of theories of assortative matching. Andrews et al for German data obtain that their preferred estimate of the correlation of −0.066.

6.3.1 Wage Variance Decomposition: the importance of firms in wage determination

Once we have the estimated results, we can further decompose the variance of wages in relation to the covariates defined in our wage equation. This exercise enables us to assess the role of

[29] Structural labor market search models advise extreme caution when interpreting this correlation as evidence of complementarities in the production function.

[30] When we estimate the model using firms with ten observed employees, the estimated correlation turns out to be positive. This last result is in line with those presented by Andrews et al (2012), as this second sample with larger inter-firm mobility better enabled us to identify the role of firms in individual wage determination.

firms in wage setting via the contribution of firms to overall wage variance. From equation (6), we know that total wage variance $\left\{Var(y_{it}^{*})\right\}$ can be decomposed into a firm (time-varying and time-invariant) component and into an individual (time-varying and time-invariant) component. In particular, we propose to decompose wage variance into the following components: (i) the covariance of wages with time-varying individual covariates; (ii) the covariance of wages with time-varying firm covariates (i.e., the firm's wage policy), (iii) the covariance of wages with the time-invariant individual term; (iv) the covariance of wages with the time-invariant firm term; and (v) the covariance of wages with the transitory component of wages. In equation form, this reads as follows:

$$Var\ (y_{it}^{*}) = Cov\ (y_{it}^{*}, x_{1it}^{'}\gamma_1) + Cov\ (y_{it}^{*}, \alpha_i) + Cov\ (y_{it}^{*}, \beta_{ji}) + Cov\ (y_{it}^{*}, x_{2jit}^{'}\gamma_2) + Cov\ (y_{it}^{*}, \varepsilon_{it}) \quad (8)$$

Equation (8) shows that firms can play an important role in wage determination for two reasons. The first is that firms might generate strong wage differentials if motivated to pay different wages to observationally equivalent workers. This is mainly measured by the firm-specific component. The second is that firms can also amplify the effect by attracting more productive workers with steeper wage-tenure profiles. This second argument can be tested by measuring the relative variances of the tenure and other job characteristic components of the firm's contribution. Table 10 reports the wage variance decomposition from our estimated model. The second column presents the variance decomposition when the FILS estimator is used to account for censoring, and the first displays the corresponding decomposition when censoring is ignored.

[Insert Table 10]

In general, we observe that the contribution of firms' wage policies to the individual wage structure is lower than that of individual effects. More importantly, the results also indicate that failing to adjust for censoring could mislead research on the respective roles of firms and workers in explaining wage dispersion by overestimating the role of firms. This error results primarily from the overestimation of firm time-invariant effects[31]. With FILS, we obtain that firm and worker characteristics explain approximately 31% and 63% of the observed wage dispersion, respectively, whereas with the traditional approach, these figures increase to 42% for firm effects and decrease to 39% for individual effects. Hence, we obtain that controlling for

[31] Andrews et al. (2004) pointed that if the estimates of the worker and firm dummy variables are estimated with error, it is possible that the estimated correlation between them is biased downwards. It is not immediately obvious why this is so, but an overestimate of a worker effect leads to, on average, an underestimate of a firm effect.

censored wages reinforces the idea that when explaining individual wage dispersion, what workers *are* is more important than what workers *do*. This result is line with previous empirical literature such as Woodcock (2008), Bagger et al. (2010), le Maire and Scheuer (2011) and Sørensen and Vejlin (2012). They also find that firm effects are much less important than individual effects in explaining wage dispersion.

Hence, the total contribution of firms to wage variance is approximately 31%, of which only eight percentage points correspond to firm policies concerning the remuneration of tenure (a proxy for accumulated specific human capital skills), contract type or qualification (proxies for general human capital skills). The rest of the firm contribution consists of time-invariant firm effects (23%). The total contribution of individual effects amounts to 64%. Again, time-varying observable characteristics play a secondary role. For instance, work experience and age account for only 8.7% and 5.5%, respectively. The contribution of individual time-invariant effects amounts to 49%. Finally, unobserved time-varying characteristics account for 4.94% of the variance.

- Decomposition of Individual Fixed Effects

Finally, this section investigates the decomposition of the individual fixed effects into an observed part and an unobserved part. Table 11 reports the estimation results for the worker-fixed effects regressions for the model with censored wages (first column) and for the FILS estimator (second column). The estimated fixed effects are regressed on gender and education, and we allow for heterogeneous effects of education by gender.

[Insert Table 11]

In general, the parameter estimates are similar except for the case of male employees with high educational attainment levels. In this case, the model that ignores wage censoring predicts that workers with high educational attainment levels earn less than lower-educated workers, whereas our FILS estimator predicts the opposite. These differences reflect the biases incurred when omitting adjustments for censoring. In general, our estimates show that there is an increasing *premium* associated with educational attainment. Notice that these results are pure effects in the sense that they were obtained by estimating the dependent variable (the worker-fixed effect) in a regression controlling simultaneously for worker and firm time-varying characteristics and firm heterogeneity.

[Insert Table 12]

An orthogonal decomposition of the worker effect into time-invariant observed characteristics (gender and education dummies) and an unobserved effect shows that the

unobserved part is the most important, although 9% of the wage variation is due to observed differences, i.e., gender and education (see Table 12). Woodcock (2008) and Sørensen and Vejlin (2012) also find that observable differences are able to explain only a small part of wage variation.

7. Conclusions

The relatively recent availability of employer-employee matched longitudinal datasets of considerable length has added a new dimension to the analysis of worker wage dispersion because it allows the simultaneous identification of worker and firm heterogeneity. Nevertheless, although there has been an intense improvement in the estimation of wage equations with two high-dimensional fixed effects, none of the new methods considers the possibility that wages might be censored.

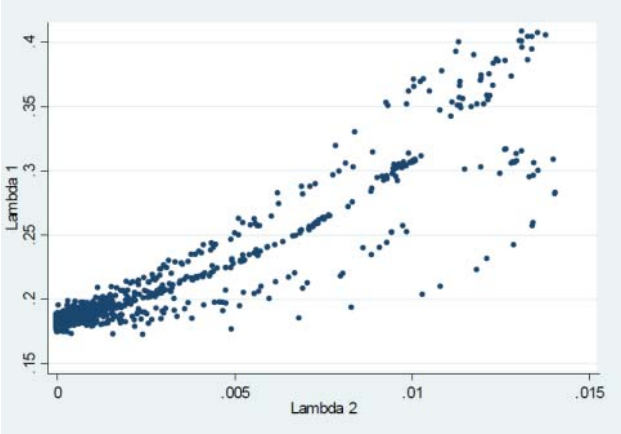
This paper discusses the estimation of general censoring models using an iterative algorithm based on a series of least squares regressions. The results of a Monte Carlo simulation designed to assess the practical performance of the procedure reveal negligible biases in the estimators and substantially greater variances than the Tobit estimates. Although it is less efficient, the algorithm enables huge time savings, and numerical experiments confirm that it is surprisingly fast and stable. The standard errors of FILS across Monte Carlo simulations provide a good approximation of the asymptotic standard errors. Thus, the proposed algorithm constitutes a powerful econometric tool for future research in this area.

In an attempt to increase the external validity of the procedure developed in this paper, we test it empirically using real data on the Spanish economy. Due to the nature of the dataset used to apply the algorithm, an initial assessment can be obtained of the importance of Spanish firms in wage setting while controlling for wage censoring. Our findings indicate that firms play a less important role than workers in explaining wage variance: person-specific effects account for 49% of the cross-employee wage variance, while firm-specific effects only account for approximately 22%. We also find evidence of misreporting issues when the model estimator does not control for censoring. The time-invariant person-specific and firm-specific effects are the most seriously misreported, but the problem also affects other person-specific parameters and job characteristics, such as required skills and education. For the time-invariant firm- and person-specific effects, we find that the role of firms in wage determination might be overestimated, whereas the opposite might be true for the role of person-specific time-invariant characteristics.

On the whole, the results indicate that the relationship between Spanish firms' wage policies and the quality of the selected workforce is close to zero and that factors other than wage policies undoubtedly intervene in explaining the distribution of high-wage workers across firms.

Figures

Figure 1. Eigenvalues of the Jacobian



Note: 99.93 % of the combinations of parameters are depicted.

Figure 2. Kernel density estimates for $\hat{\alpha}_{FILS}$ and $\hat{\sigma}_{FILS}$ ($N = 5000$)

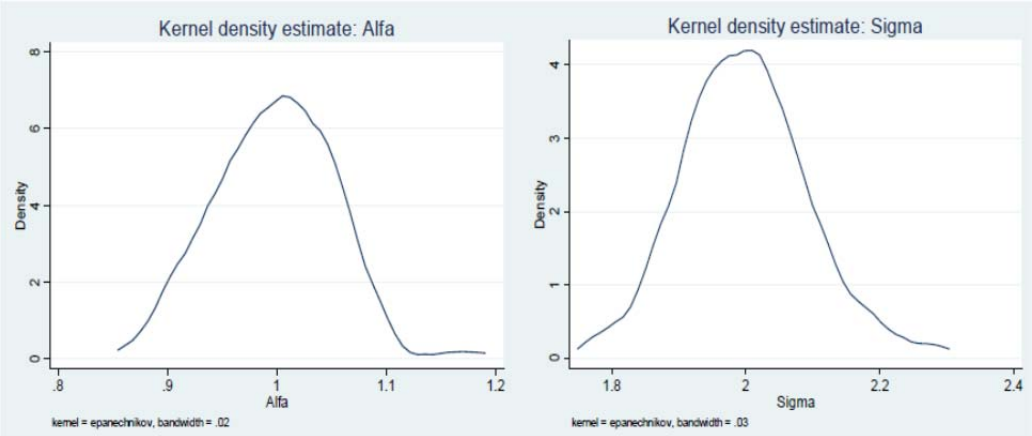


Figure 4: Effect of tenure on monthly wages

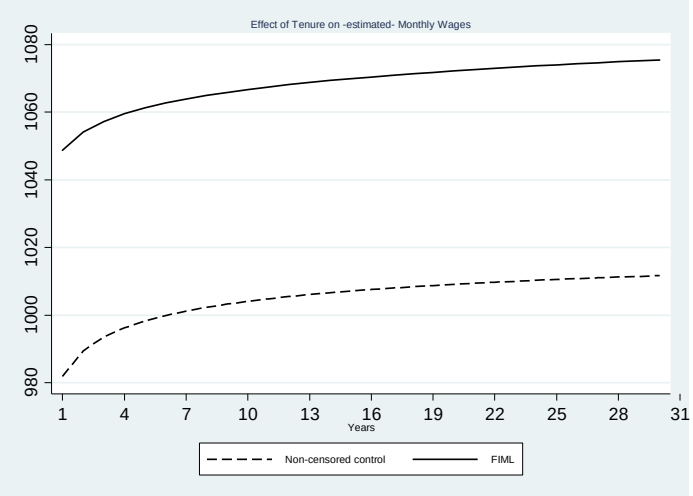
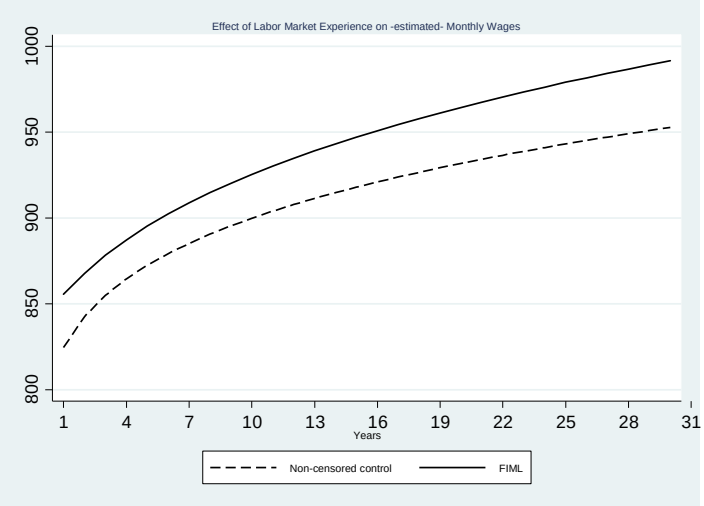


Figure 5: Effect of labor market experience on monthly wages



Tables

Table 1. Monte Carlo simulation: FILS vs Tobit

	$N = 1000$		$N = 5000$	
	FILS	**Tobit**	**FILS**	**Tobit**
Design 1	$(\alpha,\ \sigma) = (1,\ 2)$		*Censoring = 50%*	
Mean				
$\hat{\alpha}$	1.0024	1.0019	1.0018	1.0012
$\hat{\sigma}$	2.0068	2.0025	2.0025	2.007
Std Deviation				
$\hat{\alpha}$	0.1246	0.0790	0.0543	0.0367
$\hat{\sigma}$	0.1683	0.0952	0.0662	0.0418
Design 2	$(\alpha,\ \sigma) = (1,\ 2)$		*Censoring = 25%*	
Mean				
$\hat{\alpha}$	0.9948	0.9968	1.0013	1.0015
$\hat{\sigma}$	1.9977	1.9996	2.0016	2.0007
Std Deviation				
$\hat{\alpha}$	0.0786	0.0723	0.0324	0.0298
$\hat{\sigma}$	0.0911	0.0867	0.0414	0.0348

Table 2. Comparison of std errors of $\theta = (\alpha,\ \sigma)'$

Censoring	**FILS**	**Asymptotic**
0%	(0.0279, 0.0193)	(0.0281, 0.0193)
25%	(0.0339, 0.0337)	(0.0296, 0.0268)
50%	(0.0494, 0.0633)	(0.0350, 0.0319)
75%	(0.1110, 0.1223)	(0.0460, 0.0377)

Table 3. Simulation with worker dummies: FILS vs Tobit

	$T = 10$		$T = 20$	
	$(\overline{\alpha}_i, \gamma, \sigma) = (1, 0.02, 2),\ N = 1000,\ Censoring = 25\%$			
	FILS	**Tobit**	**FILS**	**Tobit**
Mean				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.8660	1.0645	1.0283	1.0521
$\hat{\gamma}$	0.0220	0.0176	0.0193	0.0198
$\hat{\sigma}$	2.0025	1.8911	1.9897	1.9451
Std Deviation				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.3123	0.3631	0.1266	0.0822
$\hat{\gamma}$	0.0077	0.0071	0.0028	0.0013
$\hat{\sigma}$	0.0667	0.0230	0.0228	0.0118
Computing Time	**58 seconds**	**396 min**	**2.03 min**	**512 min**

Table 4. Simulation with worker and firm dummies: FILS vs Tobit

	$T = 10$		$T = 20$	
$(\overline{\alpha}_i, \overline{\beta}_{j(i,t)}, \gamma, \sigma) = (1, 1, 0.02, 2),\ N = 1000,\ Censoring = 25\%,\ \tau = 0.1$				
	FILS	**Tobit**	**FILS**	**Tobit**
Mean				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	1.0694	0.8064	1.0607	0.9891
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.9119	0.8297	0.9005	0.9699
$Corr(\hat{\alpha}_i, \hat{\beta}_{j(i,t)})$	0.5217	0.5656	0.6858	0.7135
$\hat{\gamma}$	0.0203	0.0215	0.0203	0.0206
$\hat{\sigma}$	1.9633	1.8938	1.9867	1.9478
Std Deviation				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.4548	0.3791	0.1528	0.1268
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.3114	0.2423	0.1471	0.1515
$Corr(\hat{\alpha}_i, \hat{\beta}_{j(i,t)})$	0.0648	0.0421	0.0474	0.0370
$\hat{\gamma}$	0.0078	0.0068	0.0022	0.0017
$\hat{\sigma}$	0.0882	0.0173	0.0256	0.0145
Computing Time	**3.30 min**	**512 min**	**6.25 min**	**590 min**

Table 5. FILS Simulation ($S = 100$) with worker and firm dummies for different combinations of sample sizes: *(N, J, T)*

	***T* = 20**		***T* = 50**	
	***J* = 20**	***J* = 100**	***J* = 20**	***J* = 100**
Mean				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	1.0389	1.0820	1.0192	1.0041
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.9663	0.9741	0.9587	0.9402
$Corr(\hat{\alpha}_i, \hat{\beta}_{j(i,t)})$	0.7387	0.7267	0.8303	0.8432
$\hat{\gamma}$	0.0195	0.0193	0.0193	0.0191
$\hat{\sigma}$	1.9697	1.9528	1.9504	1.9452
Std Deviation				
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.1401	0.4261	0.0883	0.2417
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.1140	0.4314	0.0874	0.2432
$Corr(\hat{\alpha}_i, \hat{\beta}_{j(i,t)})$	0.0193	0.0217	0.0087	0.0095
$\hat{\gamma}$	0.0013	0.0011	0.0002	0.0003
$\hat{\sigma}$	0.0075	0.0101	0.0068	0.0044
Mobility index (%)	58.35	83.40	67.03	91.55

Note: $\left(\overline{\alpha}_i, \overline{\beta}_{j(i,t)}, \gamma, \sigma\right) = \left(1, 1, 0.02, 2\right)$, $N = 5000$, $Censoring = 25\%$, $\tau = 0.1$

Table 6. Simulation with data calibrated from AKM

$\left(\overline{\alpha}_i, \overline{\beta}_{j(i,t)}, \gamma, \sigma\right) = \left(0, 0, 0.4, 0.5\right)$, $N = 1000$, $p = 0.15$, $T = 10$

	FILS	**AKM with cens**	**AKM original**
Mean			
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.0007	0.0031	0.0012
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.1029	0.2990	0.1029
$Corr\left(\hat{\alpha}_i, \hat{\beta}_{j(i,t)}\right)$	-0.1272	-0.3342	-0.0967
$\hat{\gamma}$	**0.0397**	**0.0299**	**0.0402**
$\hat{\sigma}$	**0.4920**	**0.4308**	**0.5001**
Std Deviation			
$\frac{1}{N}\sum_i \hat{\alpha}_i$	0.0274	0.0226	0.0259
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.0180	0.0206	0.0204
$Corr\left(\hat{\alpha}_i, \hat{\beta}_{j(i,t)}\right)$	0.0232	0.0312	0.0217
$\hat{\gamma}$	0.0007	0.0004	0.0004
$\hat{\sigma}$	0.0037	0.0028	0.0023
Censoring	25%	25%	0%

Table 7. Main Sample Characteristics

Nº of Individuals	494210
Nº of Firms	68119
Nº of Time Periods (months)	32
Total Nº of Observations	11401.929
% of Movers	41.01%
Average Firms Per Workers (movers)	2.71
Nº of workers per firm	11.34

Table 8. Main descriptive statistics of the sample for different censoring levels

	Uncensored	Left-Censored	Right-Censored
Real Wage (Euros)	931	414	1813
Age (years)	36.7	29.6	42.5
Effective Labor Market Experience (Months)	93.68	61.64	131.8
Men	62.89%	45.14%	77.6%
Bachelor Degree	7.64%	6.5%	24.45%
Tenure (Months)	45.31	15.36	100.75
Unemployment Duration (months)	2.14	7.50	2.21
Permanent Contract	58.99%	16.31%	90.21%
Firm Size (employees)	364	112	561
Sector: Services	44.64%	61.5%	53.45%
N %	77.8%	3.19%	18.95%

Table 9. Model Estimation Results: FILS vs Censoring ignored

	Censoring:ignored		FILS	
	Coefficient	Std Error	Coefficient	Std Error
Age (log)	0.4715^{***}	0.0015	0.5427^{***}	0.0010
Labour Experience (log)	0.0283^{***}	0.0003	0.0144^{***}	0.0002
Labour Experience squared	0.0041^{***}	0.0001	0.0085^{***}	0.0002
Unemployment Duration (log)	-0.0086^{***}	0.0000	-0.0094^{***}	0.0000
Tenure (log)	0.0116^{***}	0.0001	0.0074^{***}	0.0001
Tenure squared	-0.0008^{***}	0.0000	0.00001^{***}	0.0000
Permanent Contract	0.0708^{***}	0.0001	0.0632^{***}	0.0001
High Skill Job	0.1348^{***}	0.0002	0.1693^{***}	0.0002
Regional Unemployment Rate	-0.0001^{***}	0.0000	-0.00002^{***}	0.0000
National GDP	0.0070^{***}	0.0000	0.0042^{***}	0.0000
Constant	4.9237^{***}	0.0024	4.7104^{***}	0.0003
Mean(α)	-0.0011		-0.0303	
Mean(β)	-0.0869		-0.0987	
$Corr(\hat{\alpha}_i, \hat{\beta}_{ji})$	-9.0%		-1.2%	
R^2	91.87%		95.06%	

Note: *** $p < 0.01$.

Table 10. Individual Wage Variance Decomposition

Contribution of	**Censoring:ignored**	**FILS**
Firms		
TV (x_{2jit})	**10.99%**	**8.38%**
Tenure	0.83%	0.65%
Permanent Contract	3.14%	1.83%
High skill Job	7.01%	5.89%
Fixed (β_{ij})	**31.48%**	**22.88%**
Workers		
TV (x_{1it})	**17.92%**	**14.49%**
Age	11.17%	8.72%
Experience	6.50%	5.59%
Unemployment Duration	0.25%	0.18%
Fixed (α_i)	**31.49%**	**49.31%**
Agg. Vbles.	**0.21%**	**0.19%**
Error Term	**8.14%**	**4.94%**
Total	100%	100%

Table 11. Regression of individual effects: *Dep Var*: Individual Fixed Effects

	Censoring:ignored		***FILS***	
	Coefficient	Std Error	Coefficient	Std Error
Male	0.0743***	0.002	0.1493***	0.004
Educational Attainment Levels				
Elementary	0.0458***	0.002	0.0879***	0.003
High School Degree	0.1616***	0.002	0.2818***	0.003
College or Bachelor's degree	0.3188***	0.002	0.5138***	0.003
Interactions				
Male*Elementary	0.0205***	0.002	0.0104***	0.003
Male*High School	-0.0089***	0.002	-0.0088*	0.005
Male*College	-0.0509***	0.002	-0.0242***	0.005
Constant	-0.1647***	0.001	-0.3279***	0.002
R^2	0.1269		0.0889	

Note: *** $p < 0.01$.

Table 12. Variance Decomposition of the individual effects (of the Different Components of the Individual Wage)

	Censoring:ignored	***FILS***
Gender	0.81%	0.92%
Education	13.04%	8.23%
Gender*Education	-1.15%	-0.53%
Error Term	87.31%	91.11%
Total	100%	100%

Appendix

A. Consistency

(a) $E(y_i) = \alpha$ Given the identification assumption and taking σ as known, the first moment condition can be written as follows,

$$E\left[1\{y_i < c\}y_i + 1\{y_i \geq c\}\left[\alpha + \sigma E\left[u_i \mid u_i \geq \frac{c-\alpha}{\sigma}\right]\right]\right] = \alpha$$

Since $y_i = \alpha + \sigma u_i$, then $E\left[u_i \mid u_i \geq \frac{c-\alpha}{\sigma}\right] = E[u_i \mid y_i \geq c]$

At the same time, an indicator function can be rewritten as a probability mass,

$$P(y_i < c)E[y_i \mid y_i < c] + P(y_i \geq c)[\alpha + \sigma E[u_i \mid y_i \geq c]] = \alpha$$

Taking into account that $\sigma E[u_i \mid y_i \geq c] = E[y_i \mid y_i \geq c] - \alpha$, the following condition emerges:

$$P(y_i < c)E[y_i \mid y_i < c] + P(y_i \geq c)[\alpha + E[y_i \mid y_i \geq c] - \alpha] = \alpha$$

It follows directly that:

$$P(y_i < c)E[y_i \mid y_i < c] + P(y_i \geq c)\,E[y_i \mid y_i \geq c] = E(y_i) = \alpha$$

q.e.d.

(b) $Var(y_i) = E(y_i^2) - E(y_i)^2 = \sigma^2$

By the same reasoning, the second moment condition can be written as

$$E\left[1\{y_i < c\}y_i^2 + 1\{y_i \geq c\}\left[\alpha^2 + \sigma^2 E[u_i^2 \mid u_i \geq A] + 2\alpha\sigma E[u_i \mid u_i \geq A]\right] - \alpha^2\right] = \sigma^2$$

Since $y_i^2 = \alpha^2 + \sigma^2 u_i^2 + 2\alpha\sigma u_i$, then

$$\sigma^2 E[u_i^2 \mid y_i \geq c] = E[y_i^2 \mid y_i \geq c] - \alpha^2 - 2\alpha\sigma E[u_i \mid y_i \geq c]$$

Substituting this into the previous expression and using the law of iterated expectations yields the following:

$$P(y_i < c)E[y_i^2 \mid y_i < c] - \alpha^2 + P(y_i \geq c)[\alpha^2 + E[y_i^2 \mid y_i \geq c] - \alpha^2 - 2\alpha\sigma E[u_i \mid y_i \geq c] + 2\alpha\sigma E[u_i \mid y_i \geq c]\ ] = \sigma^2$$

It follows directly that

$$P(y_i < c)E\left[y_i^2 \mid y_i < c\right] + P(y_i \geq c)E\left[y_i^2 \mid y_i \geq c\right] - \alpha^2 = E(y_i^2) - E(y_i)^2 = \sigma^2$$

q.e.d.

B. Simulation - Comparative statistics: different mobility indexes

Table 1.A Monte Carlo Simulation of FILS for different values of τ

	$\tau = 0.025$	Benchmark: $\tau = 0.1$	$\tau = 0.35$
Mean			
$\frac{1}{N}\sum_i \hat{\alpha}_i$	1.0641	1.0607	1.0601
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	0.7308	0.9005	0.8013
$Corr\left(\hat{\alpha}_i, \hat{\beta}_{j(i,t)}\right)$	0.4102	0.6858	0.4465
$\hat{\gamma}$	0.0191	0.0203	0.0195
$\hat{\sigma}$	1.9709	1.9867	1.9714
Std Deviation			
$\frac{1}{N}\sum_i \hat{\alpha}_i$	**0.5296**	0.1528	0.1821
$\frac{1}{J}\sum_j \hat{\beta}_{j(i,t)}$	**0.5292**	0.1471	0.1756
$Corr\left(\hat{\alpha}_i, \hat{\beta}_{j(i,t)}\right)$	**0.2256**	0.0474	0.0291
$\hat{\gamma}$	0.0034	0.0024	0.0022
$\hat{\sigma}$	0.0258	0.0256	0.0187
Mobility index	**25%**	**50%**	**75%**

Note: we consider here $N = 1000$, $J = 20$, $T = 20$ and a censoring rate of 25%.

References

[1] Abowd, J. & Kramarz, F. & Margolis, F. D. (1999): "High wage workers and high wage firms". *Econometrica*, Vol. 67 No.2.

[2] Abowd, J. & Kramarz, F. (1999a): "The Analysis of Labour Markets using Matched Employer-Employee Data". *Handbook of Labour Economics*, chapter 26, Vol. 3B.

[3] Abowd, J. & Kramarz, F. (1999b): "Econometric analyses of linked employer-employee data". *Labour Economics*, 6, pp 53-74.

[4] Abowd, J. & Kramarz, F. & Lengermann, P. & Perez-Duarte, S.: (2004): "Are Good Workers Employed by Good Firms? A test of a simple assortative mating model for France and the United States". *Mimeo*

[5] Abowd, J. & Woodcock, S. (2004): "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Dat"· in *J. Domingo-Ferrer and V. Torra* (eds.), pp. 290-297.

[6] Abowd, J. & Creecy, F. & Kramarz F. (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

[7] Amemiya, T. (1973): "Regression analysis when the dependent variable is truncated normal" *Econometrica* 41 (6), pp 997-1016.

[8] Andrews, M. J. & Gill, L. & Schank, T. & Upward R., (2008): "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, vol. 171(3), pages 673-697.

[9] Andrews, M. J. & Gill, L. & Schank, T. & Upward R., (2012):"High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, Elsevier, vol. 117(3), pages 824-827.

[10] Carneiro, A. & Guimaraes, P. & Portugal, P. (2009): "Real Wages and the Business Cycle: Accounting for Worker and Firm Heterogeneity," IZA Discussion Papers 4174, Institute for the Study of Labor (IZA).

[11] Cornelissen, T. (2008): "The Stata Command felsdvreg to fit a linear model with two high-dimensional fixed effects". *The Stata Journal* 8, Nº2 pp170-189.

[12] Guimaraes P. & Portugal, P. (2009): "Estimating high-dimensional fixed-effects models," DC09 Stata Conference 2, Stata Users Group.

[13] Lalive, R. & Gruetter, M. (2009): "The Importance of Firms in Wage Determination". *Labour Economics* 16 (2), pp 149-160.

[14] Newey W. & McFadden, D. (1994): "Large Sample Estimation and Hypothesis Testing" in *Handbook of Econometrics* Vol 4, pp 2111-245

[15] Olsen, R.J. (1978): "Note on the uniqueness of the maximum likelihood estimator for the Tobit model". *Econometrica* 46, 1211.1214.

[16] Powell, J.L.: (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models". *Econometrica* 54 (6), 1435-1460.

[17] Rubin, D.B. (1996): "Multiple imputation after 18+ years". *Journal of the American Statistical Association*, 91, 473-489.

[18] Sørensen, K.L. & Vejlin, R. (2012): "Return to Experience and Initial Wage Level: Do Low Wage Workers Catch Up?" Economics Working Papers 2012-02, School of Economics and Management, University of Aarhus.

[19] Tobin, J. (1956): "Estimation of Relationships for Limited Dependent Variables". *Econometrica*, 26, 24-36.

[20] Torres, S. & Portugal, P. & Addison, J.T. & Guimaraes, P. (2012): "The Sources of Wage variation: A three way high dimensional fixed effects model" *IZA Discussion Paper* Nº 7276.

[21] Woodcock, S.D. (2008): "Wage differentials in the presence of unobserved worker, firm, and match heterogeneity" *Labour Economics*, Elsevier, vol. 15(4), pp.7