Job mobility in Portugal: a Bayesian study with matched worker-firm data

Guillaume Horny[*], Rute Mendes[†], Gerard J. Van den Berg[‡]

April 25, 2007

Abstract

We study job mobility using a multivariate hazard model in discrete time. It involves two correlated random effects, one at the firm level and another at the worker level. Bayesian estimates are based on a Portuguese matched employer-employee dataset. Our results confirm the importance of unobserved heterogeneity at the individual level and at the firm level. Furthermore, the model performs better when allowing for an assortative matching mecanism in terms of employers' and employees' unobservables.

Keywords: Job transitions, assortative matching, Gibbs sampling, frailties, dynamic models, matched employer-employee data.

JEL Classification: C110, C150, C410, J200, J410, J620

Acknowledgements: We thank François Laisney for detailed comments on previous versions of this paper. Rute Mendes acknowledges the Portuguese Foundation for Science and Technology for their financial support. We are grateful to the Portuguese Ministry of Employment, Statistics Department, for the access to the data set.

[*]BETA, University Louis Pasteur, Strasbourg I, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France. horny@cournot.u-strasbg.fr

[†]Tinbergen Institute and Free University Amsterdam, Department of Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. mendes@mail.tinbergen.nl

[‡]Free University Amsterdam, IFAU-Uppsala, Tinbergen Institute, CEPR, IZA, and INSEE-CREST.

1 Introduction

Major features of the labor market regarding the dynamics of job mobility are well established: long-term employment relationships are common, most new jobs end early, and the probability of a job ending declines with tenure. Farber (1999) presents a survey of the literature that provides empirical evidence for these facts in the OECD countries and Europe.

Unobserved heterogeneity in the probabilities of job change can largely account for these stylized facts. Excluding true duration dependence, if different types of workers exist in terms of mobility propensities, the observed mobility rate at any point in time depends only on the proportions of those types. Higher mobility workers experience several short spells while lower mobility workers engage in fewer but longer employment relationships. The fact that most new jobs end early is explained by a sufficiently large proportion of high mobility workers. Furthermore, the fact that the probability of job ending is observed to decline with tenure is explained by sorting of the workers into different tenure groups: longer (shorter) tenure groups include a larger proportion of lower (higher) mobility workers.[1]

Since job mobility is a decision that respects the worker and the firm, it is plausible that transition rates are affected simultaneously by unobserved heterogeneity of workers and firms. Whereas the relevance of worker unobserved heterogeneity in job durations models is well established (see e.g. Farber, 1999, Bellmann *et al.* 2000, and Del Boca and Sauer, 2006), the empirical evidence on influence of firm heterogeneity is much more limited. To our knowledge, Abowd *et al.* (2006) are the only to include worker and firm unobserved heterogeneity in a job mobility model. They conclude that there is a daunting heterogeneity among firms and human resource policies.

We thus estimate a model of job transitions with a flexible specification for unobserved heterogeneity, allowing job transitions to be dependent on the combination of the unobserved heterogeneity of both the worker and the firm. The former encompasses the unobservable persistent propensity of the worker to change jobs and the later can reflect firm's preference to employ a more or less stable workforce. Furthermore, considering that the matching process between firms and workers may follow some assortative pattern, also in terms of unobservables, we allow the unobserved effects of matched firms and workers to be correlated. To our knowledge, this is the first study that allows for such flexible modeling in job mobility decisions. We specify both unobserved heterogeneity terms as random effects. They

[1] This pure heterogeneity model is a generalization of the mover-stayer model, introduced by Blumen *et al.* (1955) and extended by Goodman (1961) and Spilerman (1972).

act in a Mixed Proportional Hazard model (hereafter referred to as a MPH model, see Van den Berg, 2001 for a survey), explaining job duration. A firm is cross-sectionally and longitudinally connected to multiple workers, whereas a worker is longitudinally connected to multiple firms. Our approach allows thus for a flexible dependency structure as the effects are neither nested, nor independent. Because of the complex pattern relating the two random effects, the model is estimated using a Bayesian approach in line with Manda and Meyer (2005). Our paper thus contributes to the methodological literature by showing how to handle this complex unobserved heterogeneity structure. We also propose and apply a decomposition of the transition probability into the variation of each random effect and the variation of the explanatory variables.

Modeling unobserved heterogeneity in mobility decisions follows the recent availability of more complete data on labor market. Matched worker-firm datasets give the possibility to control for observables of both the worker and the firm. A recent body of literature on simultaneous estimation of wage and employment duration processes (e.g., Abowd *et al.* (2006), Buchinsky *et al.* 2005, Dostie, 2005, and Beffy *et al.* 2006) explores matched employer-employee data and presents the most flexible specifications considered up to now. All these studies include a worker-specific effect in the mobility equation, capturing the unobservables characteristics of the worker that impact the propensity to change jobs.

The paper is organized in 5 sections. The data are described in Section 2. Section 3 presents the MPH model in discrete time with two correlated random effects. Special cases are the MPH model with one random effect and the model without random effect, thereafter referred to as the MPH without unobserved heterogeneity. In Section 4, we discuss the choice of the prior distributions and estimation using Gibbs sampling. The results are discussed in Section 5.

2 Data

The study is based on *Quadros de Pessoal*, a longitudinal matched employer-employee data set gathered by the Portuguese Ministry of Labor and Solidarity. The data are collected through a report that all firms with registered employees are legally obliged to provide every year. The reported data is relative to all workers employed by the firm in the month when the survey is collected (March up to 1993, October since 1994). Coverage is low for the agricultural sector and non existent for public administration and domestic services. On the other hand, the manufacturing and private services sectors are almost fully covered.

An identification code is assigned to every firm when it enters the data set for the first time, and the identification code of the worker is a transformation of his social security number. We found some inconsistencies in the firm identification code in waves before 1994, and thus use the data covering the period 1994-2000. They comprise around 385 thousand firms and 4 million workers. Based on these identification numbers, one can match workers and firms and follow both over time to identify job-to-job transitions. We reconstruct the data as if they were collected using flow sampling by keeping only spells with observed entry to avoid the initial condition problem.

We extract a sample of the data with a few conditions on characteristics of workers, firms and spells. We discard firms that leave, temporarily or permanently, the market to exclude job transitions caused by the closure of the firm. Furthermore, we exclude workers who, at some point of the covered time period, are observed in a non-paid job or in self-employment. Spells with no observed entry or ended by a transition out of the labour market are also excluded. This results in a dataset covering around 338 thousand workers and 55 thousand firms. We refer hereafter to this dataset as "full sample".

In the full sample, 70% of the spells are still ongoing at the end of the observed time period. For the spells with observed transitions, the distribution of their length is depicted in Table 1.

Table 1: Observed transitions

Job spell duration	Percent
4 or more	7
3	9
2	20
1	64
Total	100

Note: durations are in years.

Observed job-to-job transitions occur at 84% in the first 2 years. We retrieve in Table 1 the stylized facts that new jobs end early, and that the transition rate decreases with tenure.

Due to the way in which the data are collected, we do not have details on the worker's labor history between two surveys, nor do we know the date at which the worker leaves the firm. We identify job-to-job transitions occurring in the interval of one year, without excluding occurrence of other short spells (job, unemployment or non-participation spells).

Table 2 summarises the number of spells per worker.

Table 2: Number of spells per worker

Number of spells	Percent
3 and more	2
2	8
1	90
Total	100

Most workers experience employment relationships with few firms: only 2 % experienced 3 or more job matches[2]. We are not investigating temporary jobs but instead jobs with contracts of at least one year.

To make computation easier, we extract a 3% subsample (on a worker basis) maintaining the proportions of each group of observations defined by worker, firm and job spell characteristics. We thus obtain a dataset, hereafter refered as "subsample", of almost 23 thousand of observations, corresponding to around 10 thousand spells, 9 thousand workers and 6.5 thousand firms.

We include in our model of job-to-job transition the following characteristics of the worker: age, gender, education and part-time job. Age may capture life-cycle effects. At an early age, 'job shopping' tends to take place while the worker is not aware of his own abilities or of the characteristics of the labor market (Johnson, 1978). Age is grouped into the categories: 16 - 25, 26 - 35 and 36 - 55 years old. Workers older than 55 were omitted in order to avoid considering also transitions to retirement. Different degrees of attachment to the labor market, for example differences in child care and family responsibilities, may result in gender differences in terms of job mobility. Therefore, we also include a female indicator. We also control for education, which is grouped into three categories: primary school, lower secondary, upper secondary and higher education. A part-time indicator is also included because one may think that firms facing negative demand shocks tend to first terminate part-time jobs in order to minimize the loss of specific human capital.

The observed firm characteristics included in our analysis are economic sector, location and an indicator for multiple plants. With these variables, we aim to capture the effect of characteristics of the labor market that may be specific to sectors and regions. The wage is handled specifically. In search models, it is seen as a firm characteristic and an exogenous variable that should be included in job transition equations. On the other hand, one may

[2] Figures in table 2 do not characterize the complete working life of workers. In our sample, workers can be observed for a maximum of seven years. In particular, he is observed for seven years if he starts a new job spell in the 1993/1994.

think that the wage is partly determined by job mobility decisions, and so should be kept out of the controls for its endogeneity. We thus estimate our model with and without the wage in the right hand side.

3 Model

Workers enter and leave companies at any time and durations are continuous. However, we only observe them in grouped form due to the sampling scheme. We thus specify a MPH model in discrete time, that is, we use the complementary log-log link function described in Kalbfleisch and Prentice (1980). An alternative approach is the logit link, but it is not well suited here because the underlying transition process can be treated as continuous and time discreteness is only due to the way the data are gathered. Furthermore, the logit model is sensitive to a change of time scale (Firth *et al.* 1999). Specifying a complementary log-log link function leads to the continuation ratio (Kalbfleisch and Prentice, 1980) and the grouped continuous models (MacCullagh, 1980), which differ by the baseline hazard parametrisation. Grilli (2005) shows that the two models give different results when extended to time-dependent variables, and the continuous ratio model achieves a more parsimonious and accurate representation of the hazard pattern when there are many covariates with non-proportional effect. We thus consider the random effects continuation ratio model in ou application.

We consider different models belonging to the MPH model family in discrete time. The simplest one accounts for observed heterogeneity only and, as commonly done in the literature, the second one allow for a worker random effect. We extend them to involve two random effects. The effects allow for realisations shared among several spells and capture dependency among durations.

In our application, a firm is cross-sectionally and longitudinally connected to multiple workers, but a worker is only longitudinally connected to multiple firms. There is thus no hierarchy in the sample: although a firm consists of multiple workers, these workers change between firms when they move to another job. We denote by $i = 1, ..., I$ the company index and by $j = 1, ..., J$ the worker index.

Discrete time duration models are described in Hosmer and Lemeshow (1999) among others. Let the time scale be divided into intervals $[a_{k-1}, a_k[$ where $0 = a_0 < a_1 < \ldots < a_K < \infty$. The discrete time duration t_{ijk} is in $\{1, \ldots, K\}$ and indicates a transition observed in $[a_{k-1}, a_k[$. Here, the hazard function is a conditional probability, contrary to the continuous time case where it is a rate, and can be written as:

$$\lambda\left[t_{ijk}|x_{ij}(t_{ij(k-1)}), v_i, w_j\right] = p[a_{k-1} < T < a_k|T \geq a_{k-1}, x_{ij}(t_{ij(k-1)}), v_i, w_j], \tag{1}$$

where $x_{ijk}(t_{ij(k-1)})$ are both worker and firm observed explanatory variables potentially time varying, v_i is a random effect at the company level and w_j a random effect at the worker level.[3]

The discrete time MPH model without unobserved heterogeneity is defined as:[4]

$$\lambda\left[t_{ijk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1\right] = 1 - \exp\Big(-\exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1]\Big), \tag{2}$$

where $\beta_{0(k-1)}$ is the baseline hazard over the time interval $[a_{k-1}, a_k[$. The model involving a worker effect is:

$$\lambda\left[t_{jk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1, w_j\right] = 1 - \exp\Big(-\exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1 + w_j]\Big). \tag{3}$$

A discrete time MPH model with two frailties is defined as:

$$\lambda\left[t_{ijk}|x_{ij}(t_{ij(k-1)}), \beta_0, \beta_1, v_i, w_j\right] = 1 - \exp\Big(-\exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1 + v_i + w_j]\Big). \tag{4}$$

Let us denote by λ_{ijk} the value of the hazard function (1) at time t_{ijk}. The departure of worker j from firm i at time t_{ijk} contributes to the likelihood as:

$$L^d_{ij}(t_{ijk}|\beta_0, \beta_1, v_i, w_j) = \lambda_{ijk}\prod_{s=1}^{k-1}(1 - \lambda_{ijs}). \tag{5}$$

[3]The time varying variables for spelle ijk are measured at time $t_{ij(k-1)}$ to ensure they constitute a predictable stochastic process. Their values are influenced only by events that have occurred just before time t_{ijk}, to avoid potential endogeneity (see Van den Berg, 2001, for a non technical discussion on predictable processes and Fleming and Harrington, 1991, for an exposure involving measure theory.)

[4]An alternative commonly used in applications is the Logit model (see Firth *et al.* 1999, Biggeri *et al.* 2001, Manda and Meyer, 2005, among others). MacCullagh and Nelder (1996, p. 107-110) explain that both models give similar results when the transitions probabilities are less than 0.15. The logit model is however not invariant to a change of the time interval length: modifying the time at which data are gathered or even changing the time scale influences the results (Rodriguez, 2001).

A censored spell of length t_{ijk} contributes to the likelihood as:

$$L_{ij}^{c}(t_{ijk}|\beta_0, \beta_1, v_i, w_j) = \prod_{s=1}^{k} (1 - \lambda_{ijs})\,. \tag{6}$$

The full likelihood is thus:

$$L(t|\beta_0, \beta_1, v, w) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \lambda_{ijk}^{\delta_{ijk}} (1 - \lambda_{ijk})^{1-\delta_{ijk}}, \tag{7}$$

where δ_{ijk} is a transition indicator. Likelihood (7) is equivalent to the one of a model treating the δ_{ijk} as Bernouilli draws. By omitting the absent effects from relation (5), we obtain the likelihood of the models with one frailty or less.

3.1 Specification of the correlated unobserved heterogeneity

The hazard function is conditional on the unobserved heterogeneity and we proceed by specifying the mixing distributions. They are assumed continuous, extending the dichotomous distribution from the original mover-stayer model of Blumen (1955).

Typically in discrete time, the random effects are assumed to be independent draws from a gaussian or log-gamma distribution in case of a single-frailty (Firth *et al.* 1999 and Conaway, 1990, respectively). Lindeboom and Van den Berg (1994) show that the mixing distribution affects the evolution of the hazard, and its choice is of importance to avoid too restricted time paths of the hazard. They also show that the duration marginal distributions are not restricted in case of a multivariate gaussian mixing distribution. We thus assume the worker effect of model (3) to be distributed as:

$$w_j \underset{\text{i.i.d}}{\sim} N\left(0, \sigma_w^2\right). \tag{8}$$

Each w_j is common to all worker j's spells. We estimate model (4) with two independent random effects, distributed as the univariate normal variables w_j and:

$$v_i \underset{\text{i.i.d}}{\sim} N\left(0, \sigma_f^2\right). \tag{9}$$

It is likely that both random effects are correlated. Indeed, Mendes *et al.* (2005) found positive assortative matching in terms of productivity in these data, that is, more productive firms tend to match with more productive workers.[5] Therefore, we assume a multivariate distribution allowing for correlations between both types of random effects. At a given date, correlation among firms depends on the covariates, and correlation among workers on the covariates and the firm effect. The multivariate mixing distribution is the product of the firm effect's marginal distributions with the distribution of the worker effect conditional on the firm log-frailties:

$$v_i \underset{\text{i.i.d}}{\sim} N\left(0, \sigma_f^2\right), \tag{10}$$

$$w_j | v_1, \ldots, v_I \underset{\text{i.i.d}}{\sim} N\left(\frac{\rho\sigma_w}{\sigma_f}\left(\sum_{l=1}^{I} \delta_{lj} v_l\right), \sigma_w^2\left(1 - \rho^2 \sum_{l=1}^{I} \delta_{lj}\right)\right), \tag{11}$$

where σ_w^2 is the variance of the marginal distribution of the worker effect, σ_f^2 the variance of the marginal distribution of the firm effect and ρ the correlation between w_j and v_i. The variable δ_{ij} is an indicator equal to 1 if worker j is at any time in firm i, 0 otherwise. A detailed justification of equation (11) is provided in Appendix A.

A worker who is followed over time can be matched consecutively with multiple firms, and suppose an employee is observed to work at many firms. A high correlation depicts a strong relation of the unobserved characteristics between the worker and the firms, where he is currently employed and where he previously worked. The current firms' unobserved characteristics are thus related with the previous firms' unobserved characteristics through the worker's history. A more general formulation would thus involve 3 correlations: between workers, between firms and across firms and workers. We have to draw from a joint normal distribution with a dimension equal to the sum of the number of workers and the number of firms, with a structured variance matrix keeping only two variances and three correlations. However, each element has to fill a number of constraints increasing with the matrix dimension to ensure its positiveness. With the number of workers and firms at hand, we conclude from numerical investigations that the three correlations are restricted in a range between 0 and 0.1. The model nearly involves two independent random effects with distributions (9) and (8).

The above mentioned problem of independence does not occur when the firm effect is assumed independent from the worker effect. If there is no link between the firm and worker unobserved characteristics, the v_i are independent (and the same for w_j). In order to keep the assumptions of independent

[5]Assortative matching in terms of productivity does not imply correlation between the unobservables of workers and firms affecting job duration. In the extreme case where wage fully captures productivity, the correlation between w_j and v_i, as obtained from a model that controls for wage, would be unrelated to assortative matching in terms of productivity.

v_i and independent w_j, we restrict the number of related firms and workers. The underlying idea is to prevent a correlation between the v_i (and the w_j) to emerge due to strong consecutive matchings. We draw a first subsample to ease computation, and a second subsample with no more than three job spells drawn for a given worker, and no more than three workers drawn for a given firm. In other words, the second subsample contains up to 3 spells per worker and 3 employees per firm.

`Compléter ici`

The unrestricted subsample comprises ? firms and ? workers, and the restricted one comprises 6577 firms and 7749 workers. Tables 6 and 7, in Appendix B, depict respectively the durations and the number of spells in all samples while Tables 9 and 8, in Appendix C, depict the covariates for employers and employees in all samples. Due to the small number of transitions, the restriction does not modify statistics of the workers characteristics. However, as the firm identifier in the data refers to a company and not to a specific plant, drawing no more than 3 workers per firm reduces the influence of large companies and especially those with multiple plants.

4 Bayesian Inference

The Bayesian approach augments the assumed model with the prior beliefs on the parameters. We choose proper but uninformative priors. Manda and Meyer (2005) specify a baseline hazard with steps, related through a first-order autocorrelated process, and Grilli (2005) uses a polynomial specification. Due to the sampling scheme, durations last less than 6 years, we specify a piecewise constant baseline hazard with unrelated coefficients over the small number of time intervals.[6] The coefficients are given independent gaussian priors with mean 0 and variance 1000.

The precision of each random effect (i.e. σ_f^{-2} and σ_w^{-2}) follows a gamma distribution, and we base our prior elicitation on descriptive statistics. The rate of transition per worker is about 3% for the 5th quantile of the duration distribution and 13% for the 95th quantile. For 90% of the population, there is at most a fourfold variation between the odds of two workers, which implies $\sigma_w = 2/3$. We set our prior for the precision σ_w^{-2} to a gamma distribution with expectation 9/4 and variance 9/16. Similarly, the rates of transition per firm are in a range from 3% to 10% for 90% of the population, implying a

[6]We also estimated models using polynomial specification and were led to a 6 degrees polynomial, that is, less parsimonious specifications than with a piecewise constant baseline hazard.

gamma prior with expectation and variance both equal to 4 for σ_f^{-2}. A uniform distribution over $[-1, 1]$ is specified for ρ, which is the least informative possible prior.

Let us denote by T the vector of durations and by M the number of covariates. The joint density of the data and parameters is:

$$f(T, \beta_0, \beta_1, v, w) = f(\sigma_f^2) f(\sigma_w^2) f(\rho) \left[\prod_{s=1}^{K-1} f(\beta_{0s})\right] \left[\prod_{m=1}^{M} f(\beta_{1m})\right]$$
$$\left[\prod_{i=1}^{I} f(v_i|\sigma_f^2)\right] \left[\prod_{j=1}^{J} f(w_j|\sigma_w^2, \rho)\right] \left[\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} L_{ijk}(t_{ijk}|\beta_0, \beta_1, v_i, w_j)\right] \quad (12)$$

The posterior is the ratio of (12) over its integral over the parameter space. Even with all priors being independents, it does not admit an analytical solution. However, we can construct a Markov chain with elements following the posterior distribution and approximate the Bayesian estimator using a Monte Carlo method.[7] Here, the quantities of interest are approached using Gibbs sampling (Gelfand and Smith, 1990), an MCMC method involving draws from the distributions of a given parameter conditional on the other relevant parameters.

5 Results

We run two chains for each model. On previous runs, we observed the Markov chains for the parameters σ_f^{-2} and σ_w^{-2} to converge more slowly than those for parameters β and ρ. The starting values for β are thus set at the maximum likelihood estimates in a model without unobserved heterogeneity for both chains. For σ_f^{-2} and σ_w^{-2}, they are set to 1 for the first chain and to 50 for the second one. We set the starting value of ρ to 0 for both chains. We run 50 000 iterations for the models with the two frailties. From convergence plots of the sampled values and Gelman and Rubin (1992) statistics, 20 000 iterations were sufficient for the burn in. The posterior statistics are computed from the post-convergence iterations.

The estimates on the restricted sample of the unobserved heterogeneity distributions are in Table 5. Results are similar for every models, that is, increasing the unobserved heterogeneity complexity by considering a further parameter does not affect the existing results.[8]

[7] Robert and Casella, 1999, provide a survey of Markov Chain Monte Carlo methods (MCMC).

[8] A similar remark is found in Horny *et al.* (2005) on a MPH model in continuous time

Table 3: Estimates of the standard-errors of the unobserved heterogeneity distributions on the restricted subsample

Type of heterogeneity	Parameter	Mean	2.5%	97.5%
Correlated frailties				
correlation	ρ	0.50	0.29	0.58
firm effect	σ_f	0.61	0.48	0.75
worker effect	σ_w	0.29	0.22	0.37
Independent frailties				
firm effect	σ_f	0.72	0.58	0.88
worker effect	σ_w	0.26	0.20	0.33
Single frailty				
worker effect	σ_w	0.29	0.22	0.44

The correlation between the worker and firm effects is estimated around 0.50 and significant.[9] It seems that workers and firms do not tend to match in a clear assortative way in terms of propensity to move from job-to-job and retention policies. Note that this correlation encompasses only unobservables of both sides of the match and completely disregards the observables. Thus, our result does not exclude the existence of an equilibrium pattern of assortative matching in job turnover decisions. This point clearly deserves further investigations.

The posterior means for the β coefficients together with information regarding their significance are reported in Table 11. The estimates of the β on the unrestricted sample are in Table 11, in Appendix D. Negative duration dependence is found to be significant, and pure heterogeneity models cannot fully explain the empirically observed inverse relationship between separation rates and job tenure.

Regarding controlled worker characteristics, we find that females tend to move less. This result contradicts the findings of many previous studies of job mobility. The main reason could be the fact that our data covers the time period 1994-2000, and the gender difference in terms of mobility rates is changing over time. Indeed, Light and Ureta (1992) find that women's turnover behavior is changing: women belonging to early US birth cohorts appeared to be more mobile than men but this conclusion is reversed when

and two random effects. In their study, maximum likelihood results are sensitive to a change in the unobserved heterogeneity structure.

[9]The estimates of the standard errors of the mixing distributions on the unrestricted sample are in Table 10, in Appendix D. The correlation is positive and significant, however, the assumption of independent v_i and independent w_j is likely to be violated on the unrestricted sample.

Table 4: Bayesian estimates on the restricted sample

Variable	**None**	**Worker**	**Random Effect(s)**	
			Independent	Correlated
Tenure				
2 years	**-0.59**	**-0.57**	**-0.38**	**-0.46**
3 years	**-0.92**	**-0.89**	**-0.61**	**-0.73**
4 years	**-1.42**	**-1.39**	**-1.06**	**-1.20**
5 years and more	**-2.11**	**-2.07**	**-1.70**	**-1.86**
Worker characteristics				
Female	**-0.31**	**-0.31**	**-0.35**	**-0.33**
Age:				
16 - 25	**0.65**	**0.65**	**0.75**	**0.71**
26 - 35	**0.37**	**0.37**	**0.44**	**0.41**
Education:				
primary school	**0.15**	**0.15**	**0.18**	**0.17**
lower secondary	**0.21**	**0.21**	**0.25**	**0.23**
Part-time	**0.65**	**0.65**	**0.69**	**0.67**
Wage	-0.04	-0.04	-0.04	-0.04
Firm characteristics				
Multiple plants	**0.30**	**0.30**	**0.35**	**0.32**
Region:				
Center	**0.14**	**0.15**	**0.18**	**0.17**
Lisbon and Tagus Valley	**0.38**	**0.39**	**0.46**	**0.43**
Alentejo, Algarve and Islands	**0.25**	**0.25**	**0.33**	**0.30**
Sector:				
Construction	**0.31**	**0.32**	**0.36**	**0.33**
Trade	**0.24**	**0.24**	**0.27**	**0.25**
Financial	**0.51**	**0.52**	**0.60**	**0.55**
Constant	**-2.33**	**-2.38**	**-2.87**	**-2.67**
Log-likelihood	-6015	-5915	-5100	-5400
DIC				
Number of workers	7749	7749	7749	7749
Number of firms	6577	6577	6577	6577

Note: coefficients in bold type are significant at the 5% level.

more recent cohorts are considered. The reason is that women are becoming more and more attached to the labor force.

The results for age are relative to the omitted category of workers with 36 to 55 years (the oldest age group considered in our study). Thus, they indicate higher transition probabilities for the younger workers. Notice that, controlling for education, age captures labor market experience and thus these estimates contradict the prediction of no-effect, typical from the pure heterogeneity models. Instead, these estimates can be interpreted under the light of on-the-job search models or models of job shopping. The first type of models predicts that, since the match quality is known ex-ante, more experienced workers are less mobile because they had already time/opportunity to move into high quality matches. Job shopping predicts that mobility decreases with age, as the worker becomes more aware of his own abilities and of the characteristics of the labor market.

Education does not affect mobility decisions significantly, a result in line with Buchinsky *et al.* (2005). They emphasize that this result is in accordance with human capital theory. Since education represents general human capital which can be carried by the worker from one employer to another, it should not influence mobility.

The impact of part-time job confirms one of the stylized facts of the empirical job duration literature: part-time job status has a strong positive effect on the probability of job separation.

As already mentioned, we estimate the model including correlated frailties with the wage as a regressor. The wage can be seen as an endogenous characteristic to the extent that it may be determined simultaneously along with job transition. On the other hand, under the framework of search theory, it can be seen as a firm characteristic and thus a variable that is exogenous to job mobility decisions. We thus estimate the models with and without wages. Estimated β differ by only one or two hundredth between the two specifications, and the wage coefficient is not significant.

Looking at the characteristics of the firms, we find some differences across economic sectors and across regions. The North (the reference category) is the region with the lowest job mobility, while Lisbon is at the other extreme. In terms of sectors, the financial sector exhibits the highest job turnover rates while manufacture (the omitted sector) has the lowest ones. It is also the only significant sector in the model with two frailties.

We decompose the variation of the hazard to separate the influences of three components: the variation due to the firm unobserved heterogeneity, the variation due to the worker unobserved heterogeneity and the variation due to the observed explanatory variables. Conditioning successively the variance of $\ln \lambda_{ijk}$, we have:

$$\begin{aligned}\mathrm{Var}(\ln \lambda_{ijk}) = &\left[\mathrm{E}_b(b_j)\mathrm{E}_{x|a,b}(c_{ijk})\right]^2 \mathrm{Var}_a(a) + \mathrm{E}_a(a^2)\mathrm{Var}_b(b)\left[\mathrm{E}_x(c_{ijk})\right]^2 \\ &+ \mathrm{E}_a(a^2)\mathrm{E}_b(b^2)\mathrm{Var}_x(c_{ijk}),\end{aligned} \tag{13}$$

where $a_i = \exp(v_i)$, $b_j = \exp(w_j)$ and $c_{ijk} = \exp[\beta_{0(k-1)} + x_{ij}(t_{ij(k-1)})'\beta_1]$. A detailed justification is in Appendix E. Table 5 reports the results of the decomposition.[10] The firm and worker effects have the same influence, be-

Table 5: Decomposition of total variation of the log-hazard

Source	**Random Effect(s)**	
	Independent	Correlated
firm unobserved effect	26 %	25 %
worker unobserved effect	26 %	25 %
observed effects	48 %	50 %

cause they enter the hazard in a symmetric way for both models and their estimated standard errors differ by only one hundredth. Both effects explain half of the log of the transition probability, and the firm and worker observed explanatory variables are clearly insufficient to capture the heterogeneity in job mobility decisions.

The estimates of the β on the unrestricted sample are in Table 11, in Appendix D. They are similar to the results on the restricted sample, except for the part-time indicator which turns on to be insignificant.

The models fit very differently as the log-likelihoods are much more important while the dependency structure becomes more detailed. To compare the models on a formal basis, we compute the Deviance Information Criterion (DIC, Spiegehalter *et al.* 2002). Defining what is an important difference in DIC, and more generally difference in information criteria, is a difficult task and we follow the rule of thumb proposed in Burnham and Anderson (1998) and Spiegelhalter *et al.* (2002). That is, the model with the smallest DIC is prefered, while the others deserve consideration with a difference within 1-2, and have considerably less support with a difference within 3-7. The difference is here of 4 in favor of the model with the correlated frailties, which would best predict a replicate dataset of the same structure as that currently observed.[11] From the DIC, we can conclude that there is a strong evidence in favor of the model allowing for two correlated frailties.

[10]The transition probability is not linear in a_i, b_j and c_{ijk}, and the outcome of the decomposition depends on the sequence in which we split up the total variance. Conditioning on v and w leads to weights of 23% and 27% respectively in the model with correlated frailties (23% and 29% in the model with independent frailties), while conditioning on w and v leads to 27% and 23% respectively (29% and 23% in the model with independent frailties). The observables contributes to 50% of the variance for all conditioning sequence in the model with crrelated frailties (48% in the model with independents frailties). Both random effects enter the transition probability in a symmetric way, and we report the averages over the conditioning sequences.

[11]Results on the unrestricted sample in Table 11, Appendix D, lead to a difference in the DIC of: 9088 - 9046= 42, which is considerably in favor of the model with correlated frailties.

The building of the Markov chains is computer intensive and we also estimated the models without correlated frailties using maximum likelihood, and adaptive Gauss-Hermite quadrature approximations of the mixing distribution. The gain of speed allow us to use the full sample and results are presented in details in Appendix F, Table 12. They are close to the Bayesian estimates but more coefficients are significant, as we use the complete dataset and thus all the information.

6 Conclusion

We estimate a Mixed Proportional Hazard model in discrete time using a Bayesian approach. It involves different structures of the unobserved heterogeneity, the most detailed accounting for two correlated random effects, one at the firm level, one at the individual level. The correlation captures a potential assortative matching in terms of variables that play a role in mobility decisions but are not observed by the econometrician. It is a complex unobserved heterogeneity pattern, as a firm is cross-sectionally and longitudinally connected to multiple workers, whereas a worker is only longitudinally connected to multiple firms. We show how to carry on inference using Gibbs sampling. We also propose and apply a decomposition of the transition probability into the variation of each random effect and the variation of the explanatory variables.

Our results confirm the importance of the unobserved heterogeneity at the individual level, and indicate an huge unobserved heterogeneity at the firm level. It is important, as only a few studies account for unobserved determinants at the firm level and none, as far as we know, accounts yet for two levels of heterogeneity in a reduced form approach. Modeling the unobserved heterogeneity underlying job transitions as coming only from worker unobservables, as commonly done, is insufficient. Intuitively, job transition behavior depends on the individual unobserved propensity to change jobs and on the unobserved retention policies of the firms. The first characteristic is very dispersed across workers as is the second one across firms. Even allowing for two effects does not depict precisely the complex interactions between firms and workers. Furthermore, the model fit increases when accounting for assortative matching in terms on the employers' and employees' unobservables. These findings give support to models of unobserved heterogeneity as an explanation for the stylized facts of the labor market, implying that the time elapsed in a company has only a side effect on job mobility. This explanation is partial but relevant, and has not yet been investigated in details.

However, our results do not encompass directly the influence of the observables. A correlation between worker and firm characteristics (both observed and unobserved) would be a more accurate indicator of assortative matching. This point clearly deserves further studies, using fixed effects or in a multiple spells setting.

A Computation of conditional expectations and variances

Let us denote by V the vector $(v_1, \ldots, v_I)'$. V and w_j are jointly gaussian. Thus, $\mathrm{E}(w_j|V)$ is the linear conditional expectation and $\mathrm{V}(w_j|V)$ is the partial variance (see Gouriéroux and Monfort, 1990). Let us denote by V the vector $(v_1, \ldots, v_I)'$. As $\mathrm{E}(w_{ij}) = \mathrm{E}(v_i) = 0$, we have:

$$\begin{aligned}
\mathrm{E}\left(w_j|V\right) &= \mathrm{cov}(w_j, V)\,\mathrm{var}(V)^{-1}V \\
&= \frac{1}{\sigma_f^2}\mathrm{E}\left(w_j V'\right)V \\
&= \frac{\rho\sigma_w}{\sigma_f}\left(\sum_{l=1}^{I}\delta_{lj}v_l\right).
\end{aligned} \tag{14}$$

>From the partial variance, we obtain:

$$\begin{aligned}
\mathrm{var}\left(w_j|V\right) &= \mathrm{var}(w_j) - \mathrm{cov}(w_j, V)\,\mathrm{var}(V)^{-1}\,\mathrm{cov}(V, w_j) \\
&= \sigma_w^2 - \frac{1}{\sigma_f^2}\mathrm{cov}(w_j, V)\,\mathrm{cov}(V, w_j) \\
&= \sigma_w^2\left(1 - \rho^2\sum_{l=1}^{I}\delta_{lj}\right).
\end{aligned} \tag{15}$$

B Summary statistics of the durations

Table 6: Observed transitions

Job spell duration	Full Sample	Subsample	Restricted Subsample
4 and more	7	4	4
3	9	8	8
2	20	20	20
1	64	68	68
Total	100	100	100

Note: durations are in years.

Table 7: Number of spells per worker

Number of spells	Full Sample	Subsample	Restricted Subsample
3 and more	2	2	1
2	8	8	7
1	90	90	92
Total	100	100	100

Note: durations are in years.

C Summary statistics of the explanatory variables

Table 8: Firms characteristics

Variable	Full Sample		Subsample		Restrict. Sample	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Multiple plants	0.39	0.49	0.38	0.48	0.29	0.45
Sector:						
Mining	0.01	0.08	0.01	0.05	0.01	0.05
Manufacturing	0.37	0.48	0.38	0.49	0.41	0.49
Electricity, gas, water	0.01	0.04	0.01	0.03	0.01	0.03
Construction	0.15	0.35	0.14	0.35	0.15	0.35
Trade, hotels, restaurants	0.30	0.46	0.31	0.46	0.32	0.46
Transport, communication	0.06	0.23	0.05	0.22	0.04	0.19
Finance, insurance and real estate	0.12	0.33	0.12	0.32	0.09	0.28
Region:						
North	0.35	0.48	0.36	0.48	0.38	0.49
Center	0.14	0.35	0.14	0.34	0.15	0.36
Lisbon, Tagus Valley	0.41	0.49	0.43	0.49	0.38	0.49
Alentejo, Algarve	0.06	0.23	0.05	0.21	0.05	0.22
Islands	0.05	0.21	0.03	0.17	0.03	0.17
Number of firms	55 325		6 582		6577	

Table 9: Worker characteristics

Variable	Full Sample		Subsample		Restricted Sample	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Female	0.38	0.49	0.37	0.48	0.36	0.48
Age:						
16 - 25	0.34	0.47	0.34	0.47	0.31	0.46
26 - 35	0.40	0.49	0.40	0.49	0.39	0.49
36 - 55	0.26	0.44	0.26	0.44	0.29	0.45
Education:						
primary school	0.31	0.46	0.30	0.46	0.33	0.47
lower secondary	0.43	0.49	0.44	0.50	0.44	0.49
upper secondary and university	0.27	0.44	0.26	0.44	0.23	0.42
Part-time	0.08	0.27	0.06	0.24	0.06	0.23
Wage	703.80	449.63	699.49	441.88	670.36	419.93
Number of workers	338 445		9222		7749	

D Bayesian estimates on the unresricted sample

Table 10: Estimates of the standard-errors of the unobserved heterogeneity distributions on the unrestricted subsample

Type of heterogeneity	Parameter	Mean	2.5%	97.5%
Correlated frailties				
correlation	ρ	0.51	0.34	0.58
firm effect	σ_f	0.76	0.65	0.89
worker effect	σ_w	0.29	0.22	0.38
Independent frailties				
firm effect	σ_f	0.87	0.76	0.98
worker effect	σ_w	0.26	0.20	0.33
Single frailty				
worker effect	σ_w	0.30	0.22	0.41

Table 11: Bayesian estimates on the unrestricted sample

Variable	None	Worker	Random Effect(s)	
			Independent	Correlated
Tenure				
2 years	**-0.59**	**-0.57**	**-0.26**	**-0.39**
3 years	**-0.95**	**-0.92**	**-0.49**	**-0.67**
4 years	**-1.34**	**-1.31**	**-0.79**	**-1.00**
5 years and more	**-2.11**	**-2.07**	**-1.49**	**-1.73**
Worker characteristics				
Female	**-0.28**	**-0.28**	**-0.35**	**-0.32**
Age:				
16 - 25	**0.55**	**0.56**	**0.75**	**0.68**
26 - 35	**0.31**	**0.32**	**0.41**	**0.36**
Education:				
primary	**0.26**	**0.26**	**0.18**	**0.21**
lower secondary	**0.28**	**0.28**	**0.23**	**0.19**
Part-time	**0.56**	**0.57**	**0.66**	**0.62**
Wage	**-0.08**	**-0.08**	**-0.04**	**-0.03**
Firm characteristics				
Multiple plants	**0.21**	**0.21**	**0.32**	**0.27**
Region:				
Center	0.12	0.12	0.19	0.19
Lisbon and Tagus Valley	**0.37**	**0.38**	**0.46**	**0.41**
Alentejo, Algarve and Islands	**0.19**	**0.20**	**0.31**	**0.29**
Sector:				
Construction	**0.38**	**0.39**	**0.42**	**0.42**
Trade	**0.32**	**0.33**	**0.34**	**0.30**
Financial	**0.68**	**0.70**	**0.73**	**0.61**
Constant	**-2.41**	**-2.47**	-3.06	-2.75
Log-likelihood	- 7695	- 7565	- 5100	- 5400
DIC				
Number of workers	9222	9222	9222	9222
Number of firms	6582	6582	6582	6582

Note: coefficients in bold type are significant at the 5% level.

E Variance decomposition

F Frequentist estimates

Estimates in column 2 suggest that including worker random effects hardly affects the coefficient estimates. Only duration dependence becomes lower, as expected. Around 12% of the total variance is contributed by the worker random effects variance. Likelihood is improved with the inclusion of worker random effects.

The model in column 3 includes both worker and firm random effects, imposing them to be independent of each other. The likelihood is close to those of the model with the worker frailty, and may be a signal that the assumption of independence between firm and worker random effects is too strong.

Table 12: Frequentist estimates (full sample)

Variable	**Unobserved heterogeneity**		
	None	Worker	Firm and worker
Tenure			
2 years	**-0.31**	**-0.29**	**-0.26**
3 years	**-0.50**	**-0.46**	**-0.41**
4 years	**-0.81**	**-0.76**	**-0.71**
5 years or more	**-1.22**	**-1.15**	**-1.08**
Worker characteristics			
Female	**-0.28**	**-0.29**	**-0.29**
Age:			
16-25	**0.47**	**0.48**	**0.49**
26-35	**0.29**	**0.29**	**0.30**
Education:			
primary school	**0.09**	**0.09**	**0.09**
lower secondary	**0.08**	**0.08**	**0.09**
Part time	**0.19**	**0.20**	**0.20**
Firm characteristics			
Multiple plants	**0.14**	**0.15**	**0.15**
Region:			
Center	**0.11**	**0.11**	**0.12**
Lisbon and Tagus Valley	**0.22**	**0.23**	**0.23**
Alentejo and Algarve	**0.27**	**0.28**	**0.28**
Islands	0.03	0.02	0.02
Sector:			
Construction	**0.17**	**0.18**	**0.18**
Trade	**0.21**	**0.21**	**0.22**
Transports	-0.01	-0.01	-0.01
Financial	**0.41**	**0.41**	**0.43**
Constant	**-2.93**	**-3.06**	**-3.19**
Log-likelihood	-388034	-387719	-387701
σ_w	-	0.48	0.51
σ_f	-	-	0.48
Number of workers	756 120	756 120	756 120
Number of firms	77 603	77 603	77 603

References

Abowd, J., F. Kramarz, and S. Roux (2006): "Wages, mobility and firm performance: Advantages and insights from using matched worker-firm data," *The Economic Journal*, 116(512), F245–F285.

Beffy, M., M. Buchinsky, D. Fougère, T. Kamionka, and F. Kramarz (2006): "The returns to seniority in France (and why are they lower than in the United States?)," Discussion Paper Series 5486, CEPR, London.

Bellmann, L., S. Bender, and U. Hornsteiner (2000): "Job tenure of two cohorts of young German men 1979 - 1990: An analysis of the (West-)German Employment Statistic Register Sample concerning multivariate failure times and unobserved heterogeneity," Discussion Paper Series 106, IZA.

Biggeri, L., M. Bini, and L. Grilli (2001): "The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job," *Journal of the Royal Statistical Society. Series A*, 164, 293–305.

Blumen, I., M. Kogen, and P. McCarthy (1955): *The industrial mobility of workers as a probability process*, vol. 6 of *Cornell studies in industrial and labor relations*. Cornell University, Ithaca, NY.

Buchinsky, M., D. Fougère, F. Kramarz, and R. Tchernis (2005): "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.," Discussion Paper Series 1521, IZA.

Burnham, K., and D. Anderson (1998): *Model Selection and Inference*. Springer, New York.

Conaway, M. (1990): "A random effect model for binay data," *Biometrics*, 46, 317–328.

Del Boca, D., and R. Sauer (2006): "Life cycle employment and fertility across institutional environments," mimeo.

Dostie, B. (2005): "Job turnover and the returns to seniority," *Journal of Business and Economic Statistics*, 23(2), 192–199.

Farber, H. (1999): "Mobility and stability: The dynamics of job change in labor markets," in *Handbook of Labor Economics*, vol. 3, chap. 37, pp. 2439–2483. O. Ashanfelter and D. Card (eds.), Elsevier, Amsterdam.

Firth, D., C. Payne, and J. Payne (1999): "Efficacy of programmes for the unemployed: discrete time modelling of duration data from a matched-comparison study," *Journal of the Royal Statistical Society. Series A*, 162, 11–120.

Fleming, T., and D. Harrington (1991): *Counting processes and survival analysis*. Wiley, New-York.

Gelfand, A., and A. Smith (1990): “Sampling based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., and D. Rubin (1992): “Inference from iterative simulation using multiple sequences,” *Statistical Science*, 7, 457–472.

Goodman, L. (1961): “Statistical methods for the mover-stayer model,” *Journal of the American Statiscal Association*, 56, 841–868.

Gouriéroux, C., and A. Monfort (1990): *Statistiques et modèles économétriques*, vol. 1-2. Economica, Paris, 2 edn.

Grilli, L. (2005): “The random-effects proportional hazards model with grouped survival data: a comparison between the grouped continuous and continuation ration versions,” *Journal of the Royal Statistical Society. Series A*, 168, 83–94.

Horny, G., B. Boockmann, D. Djurdjevic, and F. Laisney (2005): “Bayesian estimation of Cox models with non-nested random effects: An application to the ratification of ILO conventions by developing countries,” Discussion Paper 05-23, ZEW.

Hosmer, D., and S. Lemeshow (1999): *Applied Survival Analysis*. John Wiley, New York.

Johnson, W. (1978): “A theory of job shopping,” *Quarterly Journal of Economics*, 92(2), 261–278.

Kalbfleisch, J., and R. Prentice (1980): *The Statistical analysis of failure time data*. John Wiley, New York.

Light, A., and M. Ureta (1992): “Panel estimates of male and female job turnover behavior: can female nonquitters be identified?,” *Journal of Labor Economics*, 10, 156–181.

Lindeboom, M., and G. Van den Berg (1994): “Heterogeneity in models for bivariate survival: The importance of the mixing distribution,” *Journal of the Royal Statistical Society. Series B*, 56, 49–60.

MacCullagh, P. (1980): “Regression models for ordinal data,” *Journal of the Royal Statistical Society*, 42, 109–142.

MacCullagh, P., and J. Nelder (1996): *Generalized Linear Models*. Chapman Hall, London, UK, 2nd edn.

Manda, S., and R. Meyer (2005): “Age at first marriage in Malawi: a Bayesian multilevel analysis using a discrete time model,” *Journal of the Royal Statistical Society. Series A*, 168, 439–455.

Mendes, R., G. van den Berg, and M. Lindeboom (2005): “An empirical assessment of assortative matching in the labor market,” in *Econometric society world congress*.

Robert, C., and G. Casella (1999): *Monte Carlo statistical methods*. Springer.

Rodriguez, G. (2001): “Survival models,” Lecture Notes, Chapter 7.

Spiegelhalter, D., N. Best, B. Carlin, and A. Van der Linde (2002): “Bayesian Measures of model complexity and fit,” *Journal of the Royal Statistical Society*, 64, 583–639.

Spilerman, S. (1972): “Extensions of the mover-stayer model,” *American Journal of Sociology*, 78, 599–626.

Van den Berg, G. (2001): “Duration models: specification, identification and multiple durations,” in *Handbook of Econometrics*, vol. 5, chap. 55, pp. 3381–3463. J. J. Heckman and E. Leamer (eds.), Elsevier, Amsterdam.