INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2002-23

Bayesian Inference for the Mover-Stayer Model in Continuous Time with an Application to Labour Market Transition Data

D. FOUGERE[1]
T. KAMIONKA[2]

September 2002

[1] CNRS and CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France, CEPR, London, UK and IZA, Bonn, Germany.
[2] CNRS and CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France and IDEI, Toulouse, France.

Abstract. This paper presents bayesian inference procedures for the continuous time mover-stayer model applied to labour market transition data collected in discrete time. These methods allow to derive the probability of embeddability of the discrete-time modelling with the continuous-time one. A special emphasis is put on two alternative procedures, namely the importance sampling algorithm and a new Gibbs sampling algorithm. Transition intensities, proportions of stayers and functions of these parameters are then estimated with the Gibbs sampling algorithm for individual transition data coming from the French Labour Force Surveys collected over the period 1986-2000.

Résumé. Cet article présente deux procédures d'inférence bayésienne adaptées au modèle du *mover-stayer* (*mobile-stable*) en temps continu. La première procédure est un algorithme avec fonction d'importance, la seconde un échantillonneur de Gibbs. L'application concerne des données de transitions individuelles sur le marché du travail, observées en temps discret. Les méthodes développées permettent d'inférer la probabilité d'enchâssement (*embeddability*) du modèle en temps continu dans le modèle en temps discret. Les intensités de transition entre états de participation au marché du travail, les proportions d'individus stables, et plusieurs fonctions de ces paramètres sont ensuite estimées à l'aide de la procédure de Gibbs pour des données individuelles de transition provenant des panels des Enquêtes Emploi collectées par l'INSEE entre 1986 et 2000.

1. Introduction

The econometric literature on labour mobility makes often use of Markov chains to analyze individual transitions observed with discrete-time panel data. Data used in this field come generally from Labour Force Surveys recording individual labour market positions (such as employment, unemployment, or out-of-the labour force) at some given date. If such observations are repeated through time (e.g. at equally spaced dates) for the same individuals, the analyst has typically access to individual panel data. For example, in Labour Force Surveys which are yearly conducted by the French National Statistical Institute (INSEE), the third part of the sample is renewed each year, which implies that one third of the sample is surveyed three times, namely in March of three successive years. In such a case, we may observe that an individual is employed in March of the first sampled year, that he or she is unemployed one year later, and that he or she is employed at the last interview. This does not imply that this individual has experienced exactly one unemployment spell between two successive months of March. For instance, he or she could have been unemployed during a very brief period, say between June and October of the first year, before being hired in a job with a short-term labour contract finishing in January next year, so he or she is observed to be unemployed in March of the second year. Thus, in general, Labour Force Surveys do not provide observations of continuous labour market histories, and they do not allow to identify directly quantities such as the mean durations of individual employment and unemployment spells, or the probability to become unemployed at the end of an employment spell, which are crucial parameters for the analysis of labour market dynamics.[1]

One way to draw statistical inference on such parameters is to assume that the discrete-time mobility process is generated by a continuous-time homogenous Markov chain, whose parameters can be estimated through the quasi-Newton (or scoring) algorithm proposed by Kalbfleisch and Lawless (1985) and carried out by Fougère and Kamionka (1992a) on French data. However this model has a main drawback: generally it underestimates the probability of staying in the same state over a period longer than the sampling interval.[2] Another difficulty may appear: in some cases the discrete-time Markov chain cannot be represented by an underlying continuous-time Markov process. This problem is known as the embeddability problem which has been surveyed by Singer and Spilerman (1976a, 1976b). Geweke *et al.* (1986a) established a bayesian method to estimate the posterior mean of parameters associated to the Markov process and some functions of these parameters, using a diffuse prior defined on the set of stochastic matrices. Their procedure allows to determine the embeddability probability of the discrete-time Markov chain and to derive confidence intervals for its parameters under the posterior. To overcome the main disadvantage of the time-homogenous Markov chain model, one possible solution is to incorporate a very simple form of heterogeneity accross the individuals: this is done in the mover-stayer model, which is a stochastic process mixing two Markov chains. This modelling implies that the reference population consists of two types of individuals: the "stayers" permanently sojourning in a given state, and the "movers" evoluing between states according to a non-degenerate Markov process. Frydman (1984) has shown that the proportion of stayers in each state and the transition probabilities of movers can be identified with panel data containing *at*

[1] See, e.g., the recent literature on changes in job stability and job security in the United States, in particular papers by Gottschalk and Moffitt (1999), and Neumark *et al.* (1999).

[2] See Singer and Spilerman (1976b) for some empirical evidence on this point.

least three observation dates. Our paper focuses on three points:

(i) first, we propose a Bayesian procedure to estimate the continuous-time mover-stayer model from individual transition data observed on a discrete-time axis. This procedure extends the work by Geweke *et al.* (1986a), devoted to the estimation of the continuous-time Markov chain parameters from discrete-time panel data.

(ii) we develop both an independence sampling method (importance sampling)[3] and a Markov chain sampling method (Gibbs sampling) to estimate the model. A practical reason for using these methods is that, when the maximum likelihood estimates are on the boundary of the parameter region, it is not easy to obtain finite sample results. Both procedures are applied to individual transition data coming from the 1986-1988 French Labour Force panel survey, conducted by INSEE, the French National Statistical Institute. Both methods work quite well and their estimates are very similar.

(iii) Then the Gibbs algorithm is applied to French Labour Force surveys collected from 1986 to 2000.

As it is usual in that type of study, states that individuals can alternatively occupy at any point in time are unemployment, employment and out-of-the-labour-force states. Parameters of the mover-stayer model are then obtained by gender and age. Light and Ureta (1992) have emphasized the interest of such a distinction: according to these authors, women belonging to early U.S. birth cohorts appear to be "movers" more likely than men for unobserved reasons (i.e. more likely to quit jobs). This conclusion is reversed when more recent cohorts are considered: women's turnover behavior is changing, given that they are more and more attached to the labour force. Finally, "the evidence suggests that women do not constitute a homogeneous group characterized by sporadic labour force participation" (Light and Ureta, 1992, p.158-159). Although French data used in our paper are sampled differently (i.e. on a discrete-time axis rather than on a continuous-time one) and are analyzed inside a rather different model, our general conclusions confirm the ones obtained by Light and Ureta. Our main results are the following:

- in France, during a recent period (1986-2000), proportions of stayers in the usual labour states (employment, unemployment and non-participation) do not differ much for men and women;
- the probability for women to be stayers out of the labour force between 26 and 35 years old has significantly decreased over the whole period: it is estimated to be approximately equal to 0.2 in year 2000, while at the same date the estimate for men is about 0.4;
- the individuals exhibiting the highest probabilities to be stayers in unemployment are the adult men above 36 years old;
- younger workers, and particularly young women, are not "stayers" in unemployment, as stated by some dualist theories of the labour market.

[3]Importance sampling methods were introduced in econometrics by Kloek and Van Dijk (1978) and then developed by Van Dijk and Kloek (1980), Van Dijk *et al.* (1985) and Geweke (1989).

The next section introduces the continuous-time mover-stayer model and derives ML estimators of the model parameters. Importance sampling and Gibbs sampling procedures are presented in the third and fourth sections, respectively. The fifth section includes the presentation of the data source, and a discussion on the results. The last section concludes.

2. The Continuous-Time Mover-Stayer (CTMS) Model

2.1. Definitions and Notations

The mover-stayer model has been introduced by Blumen *et al.* (1955) for studying the mobility of workers in the labour market. Subsequently, Goodman (1961), Spilerman (1972) and Singer and Spilerman (1976a) have developed the statistical analysis of this model on the discrete-time axis. The mover-stayer model in continuous time is a stochastic process $\{X_t,\ t \in \mathbb{R}^+\}$, defined on a discrete state-space $E = \{1, \dots, K\}$, $K \in \mathbb{N}$, and resulting from the mixture of two independent Markov chains. The first of these two chains, denoted $\left\{X_t^1, t \in \mathbb{R}^+\right\}$ is degenerate, i.e. its transition probability matrix is the identity matrix, denoted I. The other chain, denoted $\left\{X_t^2\ ,\ t \in \mathbb{R}^+\right\}$ is characterized by a non-degenerate transition matrix $M(s,t) = \|m_{i,j}(s,t)\|$, $i,j = 1, \dots, K$, $0 \leq s \leq t$, where

$$m_{i,j}(s,t) = \Pr\left\{X_t^2 = j \mid X_s^2 = i\right\},\ i,j \in E,\ s,t \in \mathbb{R}^+,\ s \leq t \tag{2.1}$$

and $\sum_{j=1}^{K} m_{i,j}(s,t) = 1$. Moreover, the Markov chain $\left\{X_t^2, t \in \mathbb{R}^+\right\}$ is assumed to be time-homogeneous, i.e.

$$m_{i,j}(s,t) = m_{i,j}(0, t-s) \equiv m_{i,j}(t-s),\ 0 \leq s \leq t, \tag{2.2}$$

which is equivalent to

$$M(s,t) = M(0, t-s) \equiv M(t-s),\ 0 \leq s \leq t. \tag{2.2A}$$

This implies that transition intensities defined by

$$q_{i,j}(t) = \lim_{\Delta t \downarrow 0} m_{i,j}\left(t, t+\Delta t\right)/\Delta t,\ \Delta t \geq 0,\ i,j = 1, \dots, K,\ i \neq j, \tag{2.3}$$

are constant through time, i.e.

$$q_{i,j}(t) = q_{i,j}\,,\ t \geq 0,\ i,j = 1, \dots, K,\ i \neq j. \tag{2.4}$$

The $K \times K$ transition intensity matrix, which is associated to the time-homogeneous markovian process $\left\{X_t^2, t \in \mathbb{R}^+\right\}$, is denoted Q and has entries

$$
q_{i,j} = \begin{cases} q_{i,j} \in \mathbb{R}^+, \ j \neq i, \ i, \ j = 1, \dots, K, \\ \\ q_{i,i} = - \displaystyle\sum_{k=1, k \neq i}^{K} q_{i,k} \leq 0, \ j = i, \ i = 1, \dots, K. \end{cases} \tag{2.5}
$$

Let us denote $\mathcal{Q}$ the set of transition intensity matrices, i.e. the set of $(K \times K)$ matrices with entries verifying the conditions (2.5). It is well known (cf. Doob, 1953, p. 240 and 241) that the transition probability matrix over an interval of length T can be written

$$
M(0, T) = \exp(QT), \ T \in \mathbb{R}^+, \tag{2.6}
$$

where $\exp(A) = \sum_{k=0}^{\infty} A^k / k!$ for any $K \times K$ matrix A. The main properties of the time-homogeneous markovian process $\left\{X_t^2, t \in \mathbb{R}^+\right\}$ with state-space E, are the following:

- sojourn times in state i $(i \in E)$ are positive random variables, which are exponentially distributed with parameter $(-q_{i,i})$,
- if the time-homogeneous Markov process $\left\{X_t^2, t \in \mathbb{R}^+\right\}$ is ergodic, its equilibrium (or limiting) probability distribution is denoted $\Pi^{(m)} = (\pi_1^{(m)}, \dots, \pi_K^{(m)})'$ and defined as the unique solution to the linear system of equations

$$
Q'\Pi^{(m)} = 0, \ \text{with} \ \sum_{i=1}^{K} \pi_i^{(m)} = 1 \ \text{and} \ \pi_i^{(m)} \geq 0, \ i = 1, \dots, K. \tag{2.7}
$$

Now let us assume that the mixed process $\{X_t, t \in \mathbb{R}^+\}$ is observed at fixed and equally distant times $0, T, 2T, \dots, LT$, with $T > 0$ and $L \in \mathbb{N}$ $(L \geq 1)$. Transition probabilities for this process are given by the formulas

$$
p_{i,j}(0, kT) = \Pr\left(X_{kT} = j \mid X_0 = i\right), \ i, j \in E, \ k = 1, \dots, L \tag{2.8}
$$

$$
= \begin{cases} (1 - s_i) \left\{m_{i,j}(T)\right\}^{(k)}, & if \quad j \neq i, \\ \\ s_i + (1 - s_i) \left\{m_{i,i}(T)\right\}^{(k)}, & if \quad j = i, \end{cases}
$$

where $\left\{m_{i,j}(T)\right\}^{(k)}$ is the element (i, j) of the matrix $M(T)^k$, and $(s_i, 1 - s_i)$, with $s_i \in [0, 1]$, is a mixing measure for state $i \in E$. So, in the mover-stayer model, the reference population is composed of two kinds of individuals: the "stayers", permanently sojourning in the same state, and the "movers", who move from one state to another according to the time-homogeneous Markov chain with transition probability matrix $M(s, t)$, $s \leq t$, and with intensity matrix Q.

The proportion of "stayers" in state i ($i \in E$) is equal to s_i. We can also derive the equilibrium (or limiting) probability distribution for the mixed "mover-stayer" process $\{X_t, t \in \mathbb{R}^+\}$. For state i, the limiting probability, denoted π_i, is given by

$$\pi_i = s_i \eta_i + \pi_i^{(m)} \sum_{j=1}^{K} (1 - s_j)\eta_j, \; i \in E, \tag{2.9}$$

where $\eta = \{\eta_i, i \in E\}$ is the initial probability distribution (i.e. at date 0) for the process $\{X_t, t \in \mathbb{R}^+\}$,and $\pi_i^{(m)}$ is the limiting probability of "movers" in state i, given by equation (2.7).

2.2. The maximum-likelihood (ML) estimation of the CTMS model using discrete-time panel data

The ML estimation of the transition matrix $M(0,T)$ and of the mixing measure s, from a sample of N independent realizations of the process $\{X_t, t \in \mathbb{R}^+\}$, has been extensively treated by Frydman (1984) and then carried out by Sampson (1990), and Fougère and Kamionka (1992b). The method developed by Frydman relies on a simple recursive procedure, which will be rapidly surveyed. However, our presentation is more general than the one by Frydman, in the sense that it includes the case where some parameters s_i are zero, and it contains also the expression of the estimated asymptotic covariance matrix for the estimators of M and s. Because these results are mostly used as background for the importance sampling procedure, their proofs are moved to Appendix A.

First, let us recall that the form of the sample is

$$\left\{X_{0(n)}, X_{T(n)}, X_{2T(n)}, \ldots, X_{LT(n)};\; 1 \le n \le N\right\}$$

where $X_{kT(n)}$ ($k = 0, \ldots, L$) is the state of the process for the n_th realization at time kT, and $(L+1)$ is the number of equally spaced dates of observation. Let us denote $n_{i_0,\ldots,i_{LT}}$ the number of individuals for which the observed discrete path is $(i_0, \ldots, i_{LT})$, $n_i(kT)$ the number of individuals in state i at time kT, $n_{ij}(kT)$ the number of individuals who are in state i at time $(k-1)T$ and in state j at time (kT), n_i the number of individuals who have a constant path $i_0 = i_T = \ldots = i_{LT} = i \, (i \in E)$,[4] $n_{ij} = \sum_{k=1}^{L} n_{ij}(kT)$ the total number of observed transitions from state i to state j, $n_i^* = \sum_{k=0}^{L-1} n_{ij}(kT)$ the total number of visits to state i before time (LT), $\eta_i \ge 0$ the proportion of individuals initially (i.e. at date 0) in state i, $i \in E$, with $\sum_{i=1}^{K} \eta_i = 1$. The likelihood function for the sample is (Frydman, 1984, p. 633):

$$\begin{aligned} L = \;& \prod_{i=1}^{K} \eta_i^{n_i(0)} \prod_{i=1}^{K} \left\{ s_i + (1-s_i)\left[m_{ii}(0,T)\right]^L \right\}^{n_i} \\ & \times \prod_{i=1}^{K} (1-s_i)^{n_i(0)-n_i} \prod_{i=1}^{K} \left\{ \left[m_{ii}(0,T)\right]^{(n_{ii}-Ln_i)} \prod_{k=1,k\neq i}^{K} \left[m_{ik}(0,T)\right]^{n_{ik}} \right\} \\ = \;& \prod_{i=1}^{K} \eta_i^{n_i(0)} \prod_{i=1}^{K} L_i \end{aligned} \tag{2.10}$$

[4]Among the individuals permanently sojourning in state i, we must distinguish the "stayers" from the "movers"; indeed, the probability that a "mover" is observed to be in state i at each observation point is strictly positive and equal to $\{m_{ii}(0,T)\}^L$.

with

$$L_i = \left\{ s_i + (1-s_i)\left[m_{ii}(0,T)\right]^L \right\}^{n_i} (1-s_i)^{n_i(0)-n_i} \left[m_{ii}(0,T)\right]^{n_{ii}-Ln_i} \times \prod_{k=1,k\neq i}^{K} \left[m_{ik}(0,T)\right]^{n_{ik}}$$

where $(n_i(0) - n_i)$ is the number of individuals initially in state i, and who experience at least one transition in the L following periods, and n_{ik} is the total number of transitions from state i to state k. Maximizing the function (2.10) with respect to M and s is equivalent to maximize the K expressions L_i subject to the constraints:

$$s_i \in [0,1], i \in E, m_{ik}(0,T) \in [0,1], i,k \in E, \text{ and } \sum_{k=1}^{K} m_{ik}(0,T) = 1.$$

As long as true values of parameters belong to the interior of the set defined by these constraints, ML estimators given by Frydman (1984, p.634–635) are unrestricted MLE that are consistent and asymptotically normal.[5] In this case, the analytical expression of the estimated asymptotic covariance matrix for ML estimators $\widehat{M}$ and $\hat{s}$ is given in Appendix A. When at least one of the true values lies on the boundary of the parameter space, the ML estimator is the restricted MLE and its sampling distribution is truncated at 0. For instance, let us consider the case where $\hat{s}_i = 0$, that arises whenever $(n_i/n_i(0)) \leq \left[\widehat{m}_{ii}(0,T)\right]^L$ (see Frydman, 1984, p. 634). In that case, we show that $\widehat{m}_{ij}(0,T) = n_{ij}/n_i^*, \forall i,j = 1\ldots K$ (see Appendix A), which is the usual ML estimator for the probability of transition from i to j for a first-order Markov chain in discrete time,[6] and the sampling distribution of the restricted MLE of s_i, denoted $\hat{s}_i^c$, is no more normal (see Appendix A). Obviously, in such a case, the joint distribution of the restricted MLE $\left(\hat{s}^c, \widehat{M^c}\right)$ is difficult to obtain. As it is explained below in sections 3 and 4, the bayesian approach allows to reduce this difficulty.

From the ML estimator of the transition probability matrix $M(0,T)$, it is possible to obtain a ML estimator of the intensity matrix Q by resolving the matrix equation (2.6). Indeed, if the solution $\widehat{Q}$ to the equation

$$\widehat{M}(0,T) = \exp(\widehat{Q}T), \;\; T > 0, \tag{2.11}$$

belongs to the set $\mathcal{Q}$ of intensity matrices, then $\widehat{Q}$ is a ML estimator for Q. Nevertheless, two difficulties may appear:[7]

- the equation (2.11) can have multiple solutions $\widehat{Q} \in \mathcal{Q}$: this problem is known as the aliasing problem;
- none of the solutions $\widehat{Q}$ to the equation (2.11) belongs to the set $\mathcal{Q}$ of intensity matrices; in that case, the probability matrix $\widehat{M}(0,T)$ is said to be non-embeddable with a continuous-time Markov process.

[5] A referee pointed out that the problem of consistency and asymptotic normality of ML in this model is very similar to the case of models with regime switches (see, e.g., Kiefer, 1978).

[6] For example, see Anderson and Goodman (1957) or Billingsley (1961).

[7] A detailed analysis of these problems is developed in papers by Singer and Spilerman (1976a, 1976b).

Necessary conditions for the embeddability of the matrix $\widehat{M}(0,T)$ are recalled in Appendix A. If equation (2.11) has only one solution $\widehat{Q} \in \mathcal{Q}$, this solution is the ML estimator for the intensity matrix of the homogeneous continuous-time Markov process $\left\{X_t^2,\ t \in \mathbb{R}^+\right\}$. But it may happen that the solution $\widehat{Q}$ to the equation (2.11) does not belong to $\mathcal{Q}$, in particular because some of its extra-diagonal entries are negative. In that situation, bayesian inference is especially worthwhile, as shown by Geweke *et al.* (1986a) for the elementary Markov model. In the remaining part of our paper, we extend their approach to the continuous-time mover-stayer model, and we develop an alternative bayesian procedure based on the Gibbs sampling algorithm.

3. Bayesian inference with importance sampling

3.1. Definitions

To write the likelihood function and the expected value under the posterior of some functions of the parameters, additional notation is needed. Let M_K be the space of $K \times K$ stochastic matrices, including transition probability matrices $M(0,T)$:

$$M_K = \left\{ M = \parallel m_{ij} \parallel :\ m_{ij} \geq 0,\ \forall i,j \in E \ and \ \sum_{j=1}^{K} m_{ij} = 1,\ \forall\ i \in E \right\}.$$

Let us denote $\mathcal{Q}$ the space of intensity matrices :

$$\mathcal{Q} = \left\{ Q = \parallel q_{ij} \parallel :\ q_{ij} \geq 0, i,j \in E, i \neq j\ ,\ q_{ii} \leq 0, \forall i \in E \text{ and } \sum_{j=1}^{K} q_{ij} = 0 \right\}.$$

If $M(0,T)$ is embeddable, there exists at least one matrix $Q\ \in\ \mathcal{Q}$ defined by the equation $M(0,T) = \exp(QT)$. Let M_K^* the space of embeddable stochastic matrices:

$$M_K^* = \{\ M(0,T) \in M_K : \exists\ Q \in \mathcal{Q}, \exp(QT) = M(0,T)\}\,.$$

If $D_K = M_K \times [0,1]^K$ represents the parameter space of the model, then the space $D_K^* = M_K^* \times [0,1]^K$ denotes the set of embeddable parameters and $D_K^* \subset D_K$. Let us consider now the set of matrices $Q^{(k)} \in \mathcal{Q}$, solutions of the equation $Q^{(k)} = \log(M(0,T))/T$, for $k = 1\ , \ldots\ ,\ B(M)$, where $B(M)$ is the number of continuous-time underlying processes corresponding to the discrete-time Markov chain represented by $M(0,T) \in M_K$. We have $B(M) \in \mathbb{N}$ and $B(M) = 0$ if $M \notin M_K^*$. Finally, if $Q^{(k)}(M)$ denotes the intensity matrix that corresponds to the k-th solution of $\log(M)$, with $k = 1, \ldots, B(M)$, then it is clear that $Q^{(k)}$ is a function defined on M_K^* and valued in $\mathcal{Q}$. Now let $\mu(M,s)$ be a prior mapping D_K into $\mathbb{R}$ (the uniform prior is used in the application): $\mu(M,s)$ is defined for $M \in M_K$ and for a vector of mixing measures $s = \{s_i, i \in E\} \in [0,1]^K$. Then let $h^{(k)}(M)$ be a prior probability on the k-th solution of the equation $M(0,T) = \exp(QT)$, which verifies $\sum_{k=1}^{B(M)} h^{(k)}(M) = 1$. Without additional information on the parameter values, a prior probability $1/B(M)$ may be put over any candidate branch $Q^{(k)}$. However, when a complementary information is available from another data set (for instance, relative to the mean sojourn duration in a given state), the prior probability $h^{(k)}(M)$ can be chosen as

$$h^{(k)}(M) = \frac{\exp\left\{-d\left(f(Q^{(k)}), f(Q_0)\right)\right\}}{\sum_{k'=1}^{B(M)} \exp\left\{-d\left(f(Q^{(k')}), f(Q_0)\right)\right\}},$$

where $Q^{(k)}$ is the candidate intensity matrix, Q_0 represents the external information on the true intensity matrix, and $d(\cdot,\cdot)$ is a distance between the images of $Q^{(k)}$ and Q_0 under some given function f.[8]

Let $g(Q,s)$ be a function defined for $(Q,s) \in \mathcal{Q} \times [0,1]^K$. This function is such that the evaluation of its moments (in particular, the posterior mean and the posterior standard deviation) is a question of interest for the analyst. If $g(Q,s)$ is the indicator function of $\{M \in M_K^*\}$, then it is convenient to set in this case :

$$g(Q,s) = J(M) = \begin{cases} 1 & \text{if } M \in M_K^* \\ 0 & \text{elsewhere.} \end{cases}$$

Thus, the posterior probability that the transition probability matrix M is embeddable has the form :

$$\begin{aligned} \Pr\left[M \in M_K^* \mid (N,n)\right] &= E[J(M) \mid (N,n)] \\ &= \frac{\int_{D_K} J(M)L(M,s;N,n)\mu(M,s)d(M,s)}{\int_{D_K} L(M,s;N,n)\mu(M,s)d(M,s)} \\ &= \frac{\int_{D_K^*} L(M,s;N,n)\mu(M,s)d(M,s)}{\int_{D_K} L(M,s;N,n)\mu(M,s)d(M,s)} \end{aligned} \tag{3.1}$$

The posterior probability of aliasing can be obtained similarly, by setting $J(M) = 1$ if $B(M) > 1$, 0 elsewhere.

3.2. Bayesian inference with importance sampling

The likelihood function $L = L(M,s;N,n)$ up to the initial distribution of the process $\{X(t), t \geq 0\}$ is

$$L \propto \prod_{i=1}^{K} L_i \tag{3.2}$$

[8]This function f characterizes the nature of the additional available information about the intensity matrix: it can be either the identity function, either some mobility index (see Geweke, Marshall and Zarkin, 1986b, for definitions), or any other function of elements of Q, such as mean sojourn durations.

with L_i defined in (2.10). If $\Pr[M \in M_K^* \mid N, n] > 0$, then the expectation of the function of interest $g(Q, s)$ under the posterior, given that M is embeddable, is:

$$
\begin{aligned}
&E\left[g(Q,s) \mid (N,n); (M,s) \in D_K^*\right] \\
&= \frac{\int_{D_K^*} \sum_{k=1}^{B(M)} h^{(k)}(M) g(Q^{(k)}(M), s) L(M,s;N,n)\mu(M,s) d(M,s)}{\int_{D_K^*} L(M,s;N,n)\mu(M,s)d(M,s)}
\end{aligned}
\tag{3.3}
$$

To evaluate the integrals inside the expressions (3.1) and (3.3), an adaptation of the Monte-Carlo method will be used, because a closed form of $Q^{(k)}(M)$ or $B(M)$ when $K \geq 3$ has not been found yet. Let $I(M, s)$ be the importance function from which a sequence $\{M_i, s_i\}$ of parameters will be drawn. We suppose that $I(M, s) > 0$ and that $\mu(M, s)$ and $g(Q, s)$ are bounded above. Then, by applying directly theorem 1 in Geweke (1989), we have:

$$
\begin{aligned}
&\lim_{I \to +\infty} \frac{\sum_{i=1}^{I} J(M_i) L(M_i, s_i; N, n)\mu(M_i, s_i)/I(M_i, s_i)}{\sum_{i=1}^{I} L(M_i, s_i; N, n)\mu(M_i, s_i)/I(M_i, s_i)} \\
&\overset{a.s.}{=} \Pr\left[(M,s) \in D_K^* \mid N, n\right]
\end{aligned}
\tag{3.4}
$$

and

$$
\begin{aligned}
&\lim_{I \to +\infty} \frac{\sum_{i=1}^{I} \sum_{k=1}^{B(M)} h^{(k)}(M_i) g\{Q^{(k)}(M_i), s_i\} J(M_i) L(M_i, s_i; N, n)\mu(M_i, s_i)/I(M_i, s_i)}{\sum_{i=1}^{I} J(M_i) L(M_i, s_i; N, n)\mu(M_i, s_i)/I(M_i, s_i)} \\
&\overset{a.s.}{=} E\left[g(Q,s) \mid N, n, (M,s) \in D_K^*\right]
\end{aligned}
\tag{3.5}
$$

where $\Pr\left[(M,s) \in D_K^* \mid N, n\right]$ is the probability under the posterior that the discrete-time mover-stayer model is embeddable with the continuous-time one, and $E\left[g(Q,s) \mid N, n, (M,s) \in D_K^*\right]$ defines the posterior moments of the function of interest. For simplifying notations, let us denote $\hat{g}_I$ the ratio in the l.h.s. of equations (3.4) or (3.5). If we consider, for example, the estimator $\hat{g}_I$ of the posterior conditional expectation of $g(Q, s)$ given that M is embeddable, then a direct application of Theorem 2 in Geweke (1989) gives

$$
\sqrt{I}(\hat{g}_I - \bar{g}) \xrightarrow{d} N(0, \sigma^2) \tag{3.6}
$$

with

$$
\sigma^2 = \frac{\int_{D_K^*} \sum_{k=1}^{B(M)} \left(h^{(k)}(M)\left[g(Q^{(k)}(M), s) - \bar{g}\right]\right)^2 \frac{\left[L(M,s;N,n)\mu(M,s)\right]^2}{I(M,s)} d(M,s)}{\left[\int_{D_K^*} L(M,s;N,n)\mu(M,s)d(M,s)\right]^2}
$$

where $\xrightarrow{d}$ is the symbol for convergence in distribution, and $\bar{g} = E\left[g(Q,s) \mid N, n, (M,s) \in D_K^*\right]$. By applying the same theorem, we get

$$I\hat{\sigma}_I^2 \xrightarrow{a.s.} \sigma^2 \tag{3.7}$$

with

$$\hat{\sigma}_I^2 = \frac{\sum_{i=1}^{I}\sum_{k=1}^{B(M)} \left(h^{(k)}(M)\left[g(Q^{(k)}(M_i), s_i) - \hat{g}_I\right]\right)^2 \left(\frac{J(M_i)L(M_i, s_i; N, n)\mu(M_i, s_i)}{I(M_i, s_i)}\right)^2}{\left(\sum_{i=1}^{I} J(M_i)L(M_i, s_i; N, n)\mu(M_i, s_i)/I(M_i, s_i)\right)^2}$$

$\hat{\sigma}_I = \left(\hat{\sigma}_I^2\right)^{1/2}$ being the numerical standard error of $\hat{g}_I$. The choice of the importance sampling density is discussed in Appendix B. Appendix B presents also the algorithm and explains how to obtain posterior means of some functions of interest, such as transition intensities, mean durations and equilibrium distributions.

4. Bayesian inference using Gibbs sampling

Gibbs sampling represents an alternative bayesian procedure to make inference on the parameters of the mover-stayer model under the posterior densities. Previous works, including papers by Chib (1992) on the Tobit regression model, Mac Culloch and Rossi (1994) on the Multinomial Probit model, Albert and Chib (1993) and Mac Culloch and Tsay (1994) on Markov switching regressions, have found that the Gibbs sampler works very well in various econometric contexts.

Formally, the application of the Gibbs sampler to the mover-stayer model is analogous to its applications to other finite mixture distributions, which make use of missing data representations (see Robert, 1994, chapter 9). In the case of the mover-stayer model, the likelihood function for a sample $X = \{X_{(n)}; 1 \leq n \leq N\}$ is the result of the marginalization

$$L\left(X \mid s, M, X_0\right) = \prod_{n=1}^{N}\sum_{i=1}^{2} \mathcal{L}\left(X_{(n)} \mid s, M, X_{0(n)}, z_n = i\right) \Pr\left(z_n = i \mid s, M, X_{0(n)}\right)$$

where $\mathcal{L}$ is the conditional likelihood contribution of the n-th individual given his/her initial state $X_{0(n)}$ and his/her unobserved type z_n, and z_n is an unobservable or missing indicator taking value 1 if the individual is a stayer or value 2 if he/she is a mover[9]. This approach offers several advantages: first, the conditional likelihood given values of the unobservable component z_n is easier to manipulate than the marginal likelihood; the fact that the maximum likelihood estimator lies on the boundary of the parameter space has no more negative influence on the estimation procedure; finally, the Gibbs sampling approach allows to choose other priors than the uniform one. In particular, for Markov chains, it is known (see Martin, 1967) that the Dirichlet and matrix beta distributions are appropriate priors because they are closed under

[9]The Gibbs sampling algorithm can be implemented through the use of the posterior distribution of the parameters given the unobserved term z and of the conditional distribution of z given X and (s, M).

sampling. Precisely, the prior density on the parameters $\theta = (s, M)$ is taken to be the product of conjugate densities $\mu_1(s)$ and $\mu_2(M)$, where

$$\mu_1(s) = \prod_{j=1}^{K} \frac{\Gamma(a_j + b_j)}{\Gamma(a_j)\Gamma(b_j)} s_j^{a_j - 1}(1 - s_j)^{b_j - 1}$$

is the Dirichlet distribution with parameters $a_j > 0, b_j > 0, j = 1, \ldots, K$, and

$$\mu_2(M) = \left(\prod_{i=1}^{K} \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_{ik}\right)}{\prod_{k=1}^{K} \Gamma\left(\alpha_{ik}\right)} \right) \times \left(\prod_{i,j=1}^{K} m_{ij}^{\alpha_{ij} - 1} \right) \tag{4.1}$$

is the matrix beta distribution with parameters $\alpha_{ij} > 0, i, j = 1, \ldots, K$. Given the unobservable component z_n and the parameters $\theta = (s, M), X_{(n)}$ follows a Markov chain with transition probability matrix

$$P = (2 - z_n)I_K + (z_n - 1)M.$$

Furthermore,

$$z_n \mid \theta, X_{0(n)} \sim \mathcal{B}\left(1; s_{0(n)}\right)$$

where

$$\begin{aligned} s_{0(n)} = \Pr\left(z_n = 1 \mid \theta, X_{0(n)}\right) \quad &= \textstyle\prod_{j=1}^{K} s_j^{1\{X_{0(n)} = j\}} \\ &= 1 - \Pr(z_n = 2 \mid \theta, X_{0(n)}) \end{aligned}$$

Thus

$$z_n \mid \theta, X_{(n)} \sim \mathcal{B}\left(1; p(X_{(n)}; \theta)\right) \tag{4.2}$$

where

$$p(X_{(n)}; \theta) = \frac{\mathcal{L}(X_{(n)} \mid s, M, X_{o(n)}, Z_n = 1) \Pr[z_n = 1 \mid s, M, X_{o(n)}]}{\sum_{j=1}^{2} \mathcal{L}(X_{(n)} \mid s, M, X_{o(n)}, Z_n = j) \Pr[z_n = j \mid s, M, X_{o(n)}]}$$

Let us now derive the posterior distribution of $\theta = (s, M)$ given $X = (X_{(1)}, \ldots, X_{(N)})$ and $z = (z_1, \ldots, z_N)$. For that purpose, we need to combine the prior information on θ with the sample information, which gives

$$\mu(\theta \mid X) \propto \mu(\theta)p(X \mid \theta)$$

where

$$p(X|\theta) = \prod_{n=1}^{N} \left(\sum_{z_n=1}^{2} s_{0(n)}^{2-z_n} \left(1 - s_{0(n)}\right)^{z_n-1} \prod_{i,k=1}^{K} \left(\delta_{ik}^{2-z_n} m_{ik}^{z_n-1}\right)^{N_{ik}^{(n)}} \right)$$

δ_{ik} being the Kronecker indicator ($\delta_{ik} = 1$ if $k = i$, 0 elsewhere), $N_{ik}^{(n)}$ being the number of transitions from state i to state k experienced by the individual n during the observation period (with the convention $0^0 = 1$). After simple computations, we get

$$\begin{aligned} \mu(\theta \mid X) \propto \sum\nolimits_{z:z\in\{1,2\}^N} & \left(\prod\nolimits_{j=1}^{K} s_j^{(a_j+\sum_{n=1}^{N} i_j^{(n)}(2-z_n))-1} \times (1-s_j)^{(b_j+\sum_{n=1}^{N} i_j^{(n)}(z_n-1))-1}\right) \\ \times & \left(\prod\nolimits_{i,k=1}^{K} \delta_{i,k}^{\sum_{n=1}^{N}(2-z_n)N_{ik}^{(n)}} \times m_{ik}^{\left(\alpha_{ik}+\sum_{n=1}^{N}(z_n-1)N_{ik}^{(n)}\right)-1}\right) \end{aligned} \tag{4.3}$$

where $i_j^{(n)} = 1_{\{X_{0(n)}=j\}}$. Given X and Z:

$\sum_{n=1}^{N} i_j^{(n)}(2-z_n)$ is the number of stayers in state j ($j = 1, \dots, K$),

$\sum_{n=1}^{N} i_j^{(n)}(z_n-1)$ is the number of movers initially in state j ($j = 1, \dots K$),

$\sum_{n=1}^{N}(z_n-1)N_{jk}^{(n)}$ is the total number of transitions from state j to state k experienced by movers.

The formula (4.3) makes clear that

$$s_j \mid X, Z \sim \text{Dirichlet}\left(a_j + \sum_{n=1}^{N} i_j^{(n)}(2-z_n), b_j + \sum_{n=1}^{N} i_j^{(n)}(z_n-1)\right) \tag{4.4}$$

$$M \mid X, Z \sim \text{Matrix beta}\left(\alpha_{ik} + \sum_{n=1}^{N}(z_n-1)N_{ik}^{(n)};\ i,k = 1, \dots K\right) \tag{4.5}$$

Under the assumption that $Z_1, \dots, Z_N$ are independent, the knowledge of the conditional distributions (4.2), (4.4), (4.5) is sufficient to implement the Gibbs sampling algorithm. This algorithm works like this:

(i) start with an initial value $\theta^{(0)} = (s^{(0)}, M^{(0)})$, for instance the ML estimates of s and M;

(ii) update from $\theta^{(m)}$ to $\theta^{(m+1)}$ by:

1. generate $Z^{(m)}$ according to the conditional distribution (4.2), given $\theta = \theta^{(m)}$ and X,
2. generate $\theta^{(m+1)} = (s^{(m+1)}, M^{(m+1)})$ according to the conditional distributions (4.4) and (4.5), given $Z = Z^{(m)}$ and X.

Under general regularity conditions, irreducibility and aperiodicity of components (4.4) and (4.5)[10] implies that, for m large enough, the resulting random variable $\theta^{(m)}$ is distributed according to the stationary posterior distribution $\mu(\theta \mid X)$. Diagnostic tests for convergence of the Gibbs sampling algorithm could be proposed (see for instance, Geweke, 1992), but their implementation is beyond the scope of this research. Finally, drawings from the stationary posterior distribution $\mu(\theta \mid X)$ may be used to obtain posterior moments for functions of θ. The definition (3.5) then becomes:

$$\lim_{I\to\infty} \frac{\sum_{m=m_0}^{I} \sum_{k=1}^{B(M^{(m)})} h^{(k)}(M^{(m)})\; g[Q^{(k)}(M^{(m)}), s^{(m)}]\; J(M^{(m)})\; \mu(s^{(m)}, M^{(m)} \mid X)}{\sum_{m=m_0}^{I} J(M^{(m)})\; \mu(s^{(m)}, M^{(m)} \mid X)} \\ \overset{a.s.}{\to} E\left[g(Q^{(k)}(M^{(m)}), s^{(m)}) \mid X, (s^{(m)}, M^{(m)}) \in D_k^*\right] \tag{4.6}$$

where m_0 is a value of m for which the convergence has been obtained.

5. An application to individual labour market transitions in France

5.1. The data source

The French Labour Force Survey ("Enquête Emploi") is yearly conducted by INSEE, the French National Statistical Institute in geographic areas including about forty housings.[11] The sample must correspond to a 1/300 sampling rate and the third part of it is renewed each year. This renewal principle implies that the third part of the sample is surveyed three times, in March of three successive years.[12] Subsequently, the panelised samples coming from the INSEE Labour Force Surveys include all the housings and individuals surveyed in years N, $(N-1)$, $(N-2)$. The panel has two waves ($L = 2$), with length $T = 365$ days (the day is the time unit in our application).

5.2. Comparison of the two bayesian procedures

First, we use only the 1986-1988 panel to compare results obtained with bayesian procedures presented in sections 3 and 4. This data set is composed of $57,560$ individuals whose age is greater than 15. Our study uses the subsample including individuals who answer the question on their labour market situation at each interview and who are less than 65 years old. This subsample includes $27,647$ individuals. At each date, the persons can be in one of the three following states: employment (E), unemployment (U) and out-of-the labour force (OLF). Table 1 describes the stratification of this subsample.

[10] See Tierney (1994) or Robert and Casella, 1999, chapter 6).

[11] The samples are extracted from the results of the 1982 National Census for the surveys conducted between 1983 and 1990, and of the 1990 National Census for surveys conducted between 1991 and 2000.

[12] Let us recall that the identification of the mover-stayer model requires a panel data set with at least three observation dates, which is the case here.

Table I: Sizes of the subsamples
(Source: "Enquêtes Emploi", INSEE, 1986-1988)

	Age groups				
	15-25	26-35	36-50	51-65	Total
Males	3063	2555	4069	3856	13543
Females	2736	2846	4164	4358	14104
Total	5799	5401	8233	8214	27647

The individual age is computed at the time of the first survey. For each stratum, the importance sampling algorithm has been applied both with normal and split-normal importance functions: in each case, the total number of drawings is $I = 10,000$. The Gibbs sampler was run by choosing the first $10,000$ iterations to be "burn-in", in order to reach the stationary posterior distribution. Probably fewer than $10,000$ would be sufficient, but the execution time of $10,000$ iterations is very low. The Gibbs sampler was then run for $5,000$ additional iterations past burn-in. To investigate stationarity, numerous plots were made. None indicated an alarming pathology.

Estimates of posterior proportions of stayers are reported in Table 2. Normal and split-normal importance functions give very close estimates for parameters s_i and their standard errors. Nevertheless, the RNE is generally much higher for the split-normal density. Posterior mean proportions of stayers obtained with the Gibbs sampler are generally very similar to the ones obtained with normal and split-normal importance densities. However, they are higher in the unemployment state, namely in the case where the posterior mean proportions computed by importance sampling techniques are low (say, less than 15 per cent). This is particularly true for the group of women aged from 36 to 50 years old: for that stratum, the posterior mean proportion of stayers in unemployment is estimated to be zero with the importance sampling technique while it is estimated to be around 7 per cent (though with a posterior standard deviation equal to 3.8) with the Gibbs sampling algorithm. A kernel estimate of the posterior marginal density of the proportion of stayers in unemployment for that subgroup is drawn on Figure 1. This result makes clear that the Gibbs sampler allows to avoid a potential drawback of the importance sampling technique when the ML estimate lies on the boundary of the parameter space.

(Table 2 and Figure 1 around here)

Estimates of transition intensities and embeddability probabilities are given in Tables 3A and 3B. Once again, estimates obtained from normal and split-normal importance functions are rather identical, although RNE's for the split-normal density are generally higher (see Table 3A). Moreover, posterior mean transition intensities obtained with the Gibbs sampling algorithm are generally very similar to (though slightly higher than) the ones obtained with the importance sampling technique. Posterior probabilities of embeddability are 1 or very close to 1, indicating there exists almost surely a continuous-time mover-stayer model generating the discrete-time observations.

(Tables 3A and 3B around here)

Estimates of mean sojourn durations, limiting probabilities and mobility indices are reported in Tables 4A and 4B. Both techniques give similar posterior means for these variates, especially at younger ages. However, they exhibit some discrepancies at older ages, for instance in the cases of women between 36 and 50 years old and of men between 51 and 65 years old. In general, posterior mean durations obtained by the Gibbs sampling algorithm are slightly lower. On the whole, importance sampling and Gibbs sampling procedures give similar results for this subsample. However, the Gibbs sampling algorithm runs faster, and it overcomes some drawbacks of the importance sampling algorithm, in particular when one of the parameter lies on (or is close to) the boundary of the parameter space. For these reasons, the Gibbs sampling has been used to analyze the estimates of the continuous-time mover-stayer applied to the recent period 1986-2000 in France.

(Tables 4A and 4B around here)

5.3. Movers and stayers in France since 1986

Figure 2 presents the changes in the probability to be stayer in the different states, by gender and age. Each point (year) on the X axis corresponds to the first year of the panel; for example, the point corresponding to the value 92 on the X axis corresponds to the estimate of the probability to be stayer in the panel 1992-1994. At adult ages (namely, between 26 and 50), the probability to be stayer in employment is approximately constant through the period,[13] and it is slightly higher for men. At the beginning of the observation period, this probability was higher for younger and older women than it was in the same male subgroups. But it decreased for younger and older women over the period, while it increased for men of the same age groups. In the most recent years, the probabilities to be stayers in employment of young men and women (less than 25 years old) are almost equal. The probability to be stayer in unemployment depends also on age; it is generally lower for women than for men, and it is lower for young people (less than 25). Two exceptions must be noticed: the probability to be stayer in unemployment is higher for older women, and its estimated value is lower for women than for men at adult ages (between 26 and 50). Finally, the probability to be stayer out of the labour force is higher (and rather stable over the period) for older men and women, and it is lower for young adults (between 26 and 35 years old). Our main result is that, for young adult women, this probability decreased significantly over the period (its estimated value was 0.60 in 1986 but only 0.25 in 1998). These results confirm that women are more and more attached to the labour force, and that, as in the U.S., "women do not constitute a homogeneous group characterized by sporadic labour force participation" (Light and Ureta, 1992, p.158-159).

(Figure 2 around here)

Figure 3 presents the changes in the transition intensities between employment and unemployment for "movers", by gender and age. In general, these intensities are higher for men (at the exception of the intensity of transition from employment to unemployment which is higher for women at younger ages) and they decrease with age (it is particularly true for the intensity of transition from unemployment to employment). Generally these intensities of transition have

[13] However, this probability decreased during the recession of the early 90s, i.e. between 1991 and 1994.

increased over the period (except for older workers). Figure 4 shows that the estimated equilibrium probability to be employed, given by equation (2.9), is higher for men and stable at adult ages (around 90%). However it increased significantly for adult women. At younger ages (less than 25), this probability is lower (around 40%), comparable for men and women, but more correlated with the business cycle. However it is still lower for older workers (less than 30% for older men, around 20% for older women). The equilibrium probability to be unemployed decreases with age; but it has increased over the recent years for all the subgroups (except for older workers). This increase does not correspond to an increase of the probability to be stayer in unemployment (Figure 2 shows ups and downs with no clear trend for this probability); it is due to the continuous increase of the intensity of transition from employment to unemployment for the "mover" part of the population. For adult women, the estimated equilibrium probability to be out of the labor force decreased continuously between 1986 and 2000. This is in line with the continuous decline of the probability to be stayer in the non-participation state for this subgroup. In 1986, stayers in this state represented approximately 17.8% of the women between 26 and 35 years old, but they were only 5.8 % in 1998. The equilibrium probability to be out of the labour force is estimated to be very low for adult men (less than 5%), and it is very high but slightly decreasing for men and women between 51 and 65 years old. Finally, Figure 5 shows that the mobility index considered here (see equation B.13) is generally higher for men (at a given age), and lower for older workers. On the whole, the estimated values of this index indicate that labour market mobility has slightly decreased over the recent period in France.[14]

(Figures 3, 4 and 5 around here)

Finally, we have conducted a simple exercise to predict state occupation probabilities in March 2001 from the estimated parameters $\widehat{s}$ and $\widehat{M}$ obtained for the 1998-2000 panel data set.[15] The predicted probability to be in state j at date τ' (March 2001) is given by

$$\widehat{p}_j(\tau') = \sum_{i=1}^{3} \widehat{\Pr}\left(X_{\tau'} = j \mid X_\tau = i\right) \times \widehat{\eta}_i\left(\tau\right), \; i, j \in E,$$

where τ represents March 2000, $\widehat{\eta}_i\left(\tau\right)$ is the proportion of individuals in state i at time τ, and $\widehat{\Pr}\left(X_{\tau'} = j \mid X_\tau = i\right)$ is the estimated transition probability from state i to state j between these two dates (see equation 2.8). Results of this exercise are given in Table V. Predicted probabilities are quite good for males. Predictions are less precise for women, especially for unemployment at younger and older ages. In general, the predicted probability to be out-of-the labor force is higher than the observed probability for women.

[14]This result can be put together with the main findings obtained by Buchinsky *et al.*, 2001, who find that wage mobility has decreased in France over the period 1968-1999.

[15]We thank the referee who suggested us to do this exercise.

Table V: Predicted occupation probabilities in March 2001 (in percentage points) (Source: "Enquêtes Emploi", INSEE, 1999-2001)

		Age groups			
Males		15-25	26-35	36-50	51-65
Employment	Observed	36.29	90.04	91.74	41.97
	Predicted	36.46	89.76	90.87	40.75
Unemployment	Observed	8.38	6.79	4.73	4.68
	Predicted	8.12	7.11	5.20	4.18
OLF	Observed	55.33	3.16	3.54	53.35
	Predicted	55.42	3.14	3.94	55.07
Females		15-25	26-35	36-50	51-65
Employment	Observed	26.14	70.36	74.13	30.95
	Predicted	25.87	69.02	72.15	31.23
Unemployment	Observed	6.93	9.08	6.00	5.73
	Predicted	9.11	8.33	6.39	4.59
OLF	Observed	66.92	20.56	19.87	63.33
	Predicted	65.02	22.65	21.46	64.18

6. Conclusions

This paper has focused on bayesian estimation of parameters of a continuous-time mover-stayer model using discrete-time observations. The estimation procedure relies on the use of bayesian inference methods, namely an importance sampling procedure and a new Gibbs sampling algorithm. Both algorithms work quite well and give similar results with observational data coming from the 1986-1988 French Labour Force Surveys. But the Gibbs sampling algorithm runs faster, and it performs better than the importance sampling algorithm when one of the parameter is close to the boundary of the parameter space. Consequently we have implemented it on panel data coming from the French Labour Force Surveys collected by INSEE between 1986 and 2000. Results show that:

- the probabilities to be stayer in the different labour market states did not differ significantly for men and women over this period;
- one noticeable exception is the group of young adult women (between 26 and 35 years old) for whom the probability to be stayer out of the labour force decreased significantly through the last fifteen years;
- this decline explains the decrease of the equilibrium probability to be out of the labor force for this subgroup;
- the equilibrium probability to be unemployed has increased over the recent years for all the subgroups (except for older workers);
- this increase is due to the continuous increase of the intensity of transition from employment to unemployment for the "mover" part of the population;

- the probability to be stayer in unemployment is rather low, with the counter-intuitive exceptions of adult men and older women.

The mover-stayer model captures a particular form of duration dependence: at the aggregate level, i.e. in the whole population, this duration dependence (for example, for sojourn times in an unemployment spell) results exclusively from the presence of stayers. In the data sets analyzed in this paper, the length of the interval between two observation dates is relatively small compared to the mean sojourn durations in some states, and the number of observation dates is relatively small too (the persons are interviewed only three times). That means that the difficulty to distinguish between duration dependence and the presence of stayers should partly result from the sampling scheme. However the current form of the model does not allow transition intensities of movers to be duration dependent. In the conditional model for movers, duration dependence could arise, for example, from the aggregation of stochastic transition rates. It is well known that, even if individual transition rates of movers are constant through time but are affected by unobserved heterogeneity terms, their aggregated intensity rates may decrease. This phenomenon is present in the mover-stayer model at the aggregate level, because this model incorporates a particular form of unobserved heterogeneity. However the modelling may be improved by permitting conditional transition rates of movers to be time-dependent and to be affected by random terms (see Kamionka, 1998, for the estimation of duration-dependent transition models using discrete-time observations).

Consequently, further research will extend the current work in two ways. Firstly, it could be worthwhile to consider continuously distributed unobserved heterogeneity terms, as it is usually assumed in the econometric literature on duration data (see Heckman and Singer, 1984, or Lancaster, 1990), rather than dichotomous ones as in the elementary mover-stayer model. Secondly, the analysis could be greatly improved by making parameters s and Q dependent of explanatory variables, such as education or marital status. For that purpose, we may impose a logistic specification for the probabilities to be stayers and/or a proportional hazard specification for individual transition rates of movers.

APPENDIX A: ML ESTIMATION OF M AND s

Asymptotic covariance matrix

Here we set $m_{ij}(0,T) \equiv m_{ij}$ for simplifying notations. The derivation of the asymptotic covariance matrix for the ML estimators of the parameters M and s requires the computation of the hessian matrix of the log-likelihood function (2.10). Because M is a stochastic matrix, the model has only K^2 independent parameters, namely the $K(K-1)$ elements of the matrix M_{-K} obtained by dropping the last column of M, plus the K parameters s_i, $i = 1, \ldots, K$. Then the covariance matrix to be computed is a $K^2 \times K^2$ matrix. Let us notice that the hessian matrix is block-diagonal. The $i_$th block ($i = 1, \ldots, K-1$) consists of the elements:

$$\frac{\partial \ln L_i}{\partial m_{ii}^2}, \frac{\partial \ln L_i}{\partial s_i^2}, \frac{\partial^2 \ln L_i}{\partial s_i \partial m_{ii}}$$

$$\frac{\partial^2 \ln L_i}{\partial m_{ii} \partial m_{ij}}, \; j \neq i, \; j = 1, \dots, K-1$$

$$\frac{\partial^2 \ln L_i}{\partial m_{ij} \partial m_{ik}}, \; j \neq i, \; k \neq i, \; j,k = 1, \dots, K-1 \tag{A.1}$$

$$\frac{\partial^2 \ln L_i}{\partial s_i \partial m_{ij}}, \; j \neq i, \; j = 1, \dots, K-1.$$

The last (i.e. the K-th) block consists of the elements :

$$\frac{\partial^2 \ln L_K}{\partial m_{Kj} \partial m_{Kk}}, \; j,k = 1, \dots, K-1, \frac{\partial^2 \ln L_K}{\partial s_K^2}, \frac{\partial^2 \ln L_K}{\partial s_K \partial m_{Kj}}, \; j = 1, \dots, K-1.$$

Each of these blocks has dimension $(K \times K)$. The computation of the information matrix $R(M_{-K}, s)$ requires the knowledge of the conditional expectations of the variables n_i, n_{iK}, $(n_{ii} - Ln_i)$ and n_{ij} $(i, j \in E)$, given η (see equation 2.10). Given that the proportion of stayers in state i is equal to s_i, and that transition probabilities from state i to any other state j are m_{ij} $(i, j \in E)$, it is easily found that

$$E(n_i \mid \eta) = \eta_i N \left[s_i + (1 - s_i) m_{ii}^L \right], \; i \in E$$

$$E(n_{ij} \mid \eta) = m_{ij} \textstyle\sum_{l=0}^{L-1} \sum_{k=1}^{K} N \eta_k (1 - s_k) m_{ki}^{(l)}, \; j \neq i, \; i,j \in E,$$

$$E(n_{ii} \mid \eta) = LN\eta_i s_i + m_{ii} \textstyle\sum_{l=0}^{L-1} \sum_{k=1}^{K} N \eta_k (1 - s_k) m_{ki}^{(l)}, \; i \in E, \tag{A.2}$$

$$E(n_{ii} - Ln_i \mid \eta) = m_{ii} \textstyle\sum_{l=0}^{L-1} \sum_{k=1}^{K} N \eta_k (1 - s_k) m_{ki}^{(l)} - LN\eta_i (1 - s_i) m_{ii}^L, \; i \in E,$$

where $m_{ki}^{(l)}$ is the element (k, i) of the matrix M^l, with $M^0 = I_K$. The ratios $n_i(0)/N$, $i \in E$, are consistent estimators of η_i, $i \in E$. Estimated variances for the ML estimators of (M_{-K}, s) are given by the diagonal entries of the inverse of $R(M_{-K}, s)$, computed for the values of the ML estimators $(\widehat{M_{-K}}, \widehat{s})$. Let us notice that $R(M_{-K}, s)$ being block-diagonal with squared blocks, its inverse is also block- diagonal, each block being equal to the inverse of the corresponding one in $R(M_{-K}, s)$. Estimated variances for the $\widehat{m}_{iK}$, $i = 1, \dots, K$ are obtained by application of the formula

$$\widehat{\text{var}}(\widehat{m}_{iK}) = \sum_{k=1}^{K-1} \widehat{\text{var}}(\widehat{m}_{iK}) + 2 \sum_{\substack{k,j=1 \\ k<j}}^{K-1} \widehat{\text{cov}}(\widehat{m}_{ik}, \widehat{m}_{ij}), \; i \in E. \tag{A.3}$$

When $\widehat{s}_i = 0$, the estimation is conducted by setting $s_i = 0$ at each step of the procedure. In that case, the computation of first and second-order derivatives of the log-likelihood function is implemented after partitioning the state space E into two subsets, denoted E_1 and E_2, such as $\forall\ i \in E_1,\ s_i \neq 0,\ \forall\ j \in E_2\ ,\ s_j = 0$. Derivatives with respect to s_i $(i = 1, \ldots, K)$ are defined only for $i \in E_1$.

Then the model consists of two subsets of states, a subset E_2 of states from which individuals can move to any other state according to the transition probability matrix $M(0,T)$, and a subset E_1 of states in which two kinds of individuals coexist initially: the "stayers", permanently sojourning in the same state, and the "movers", moving from one state to another according to the matrix $M(0,T)$.

Computation of the ML estimators for transition probabilites when $\mathbf{s_i = 0}$

To simplify the notations, we set all along

$$m_{ij}(0,T) \equiv m_{ij},\ \ i,j \in E.$$

Let $\mathcal{L}_i = \ln L_i + \lambda_i s_i$, where the expression of L_i is given by (2.10). Setting $s_i = 0$ in the previous equation and using the fact that $\sum_{k=1}^{K} m_{ik} = 1$, we get

$$\begin{aligned} \mathcal{L}_i = \quad & n_i \ln(m_{ii}^L) + (n_{ii} - Ln_i)\ln m_{ii} + \textstyle\sum_{k=1;k\neq i}^{K-1} n_{ik}\ln m_{ik} \\ & + n_{iK}\ln\left(1 - \textstyle\sum_{k=1}^{K-1} m_{ik}\right) \end{aligned} \tag{A.4}$$

Firstly, we compute the derivative of this expression with respect to $m_{i,K-1}$, and we apply the first order condition to obtain :

$$m_{i,K-1} = n_{i,K-1}\left(1 - \sum_{k=1}^{K-2} m_{ik}\right) / \sum_{k=K-1}^{K} n_{ik}$$

This expression is then substituted into $\mathcal{L}_i$. The procedure is iterated, firstly by computing the derivative of $\mathcal{L}_i$ with respect to the next parameter (namely, $m_{i,K-2}, \ldots, m_{i,1}$), then by substituting into $\mathcal{L}_i$ the expression of this parameter deduced from the first-order condition. Thus, after computing the derivative of $\mathcal{L}_i$ with respect to the parameter m_{ij}, we get :

$$m_{ij} = n_{ij}\left(1 - m_{ii} - \sum_{k=1,k\neq i}^{j-1} m_{ik}\right) / \sum_{k=j,k\neq i}^{K} n_{ik} \tag{A.5}$$

for $i \neq j$ and $j \neq K$. This expression is introduced into $\mathcal{L}_i$, which is then shown to depend on the sole parameter m_{ii}:

$$\mathcal{L}_i = n_i Log\ m_{ii}^L + (n_{ii} - Ln_i) Log\ m_{ii} + \sum_{k=1,k\neq i}^{K} n_{ik}\ln(1 - m_{ii}) + C$$

for $i = 1, \ldots, K$, C being a constant. As $\sum_{k=1,k\neq i}^{K} n_{ik} = n_i^* - n_{ii}$, the condition $\dfrac{\partial \mathcal{L}_i}{\partial m_{ii}} = 0$ implies that

$$m_{ii} = n_{ii}/n_i^* \tag{A.6}$$

By mathematical induction based on formula (A.5), (A.6) implies :

$$\forall\ i, j = 1, \ldots, K,\ m_{ij} = n_{ij}/n_i^* \tag{A.7}$$

Sampling distribution of the restricted MLE $\hat{s}_i^c = 0$

In finite samples, $\hat{s}_i$ is a discrete random variable with possible values

$$\hat{s}_i = \frac{n_i - n_i(0)\left[\widehat{m}_{ii}(0,T)\right]^L}{n_i(0)\left[1 - \widehat{m}_{ii}(0,T)^L\right]},\ n_i = 0, \ldots, n_i(0).$$

So the conditional distribution of $\hat{s}_i$ can be directly derived from the binomial distribution of the number n_i of individuals permanently observed in state i:

$$n_i \sim B\left(n_{i(0)}, s_i + (1 - s_i) m_{ii}(0,T)^L\right)$$

Asymptotically, when $n_{i(0)}$ tends to ∞, this distribution may be approximated by one of the usual approximations of the binomial distribution, such as the Camp-Paulson approximation (see Johnson and Kotz, 1969, p.64). If $\hat{s}_i^c$ denotes the restricted MLE of s_i, then

$$\begin{aligned}\Pr\left(\hat{s}_i^c = 0\right) &= \Pr\left(\hat{s}_i < 0\right) \\ &= \Pr\left(n_i < n_{i(0)}\widehat{m}_{ii}(0,T)^L\right) \\ &= G\left(n_{i(0)}\widehat{m}_{ii}(0,T)^L\right)\end{aligned}$$

where G is the c.d.f. of the approximation for the binomial distribution of n_i. The density function of $\hat{s}_i^c$ on $\mathbb{R}^{+*}$ is the density $g(.)$ of this approximation. Then, following Gouriéroux and Monfort (1995, chapter 21) the law of $\hat{s}_i^c$ may be written as

$$P^{\hat{s}_i^c} = G\left(n_{i(0)}\widehat{m}_{ii}(0,T)^L\right) 1_{\hat{s}_i=0} \in_{(0)} + g\left(n_{i(0)}\left[\hat{s}_i + (1-\hat{s}_i)\widehat{m}_{ii}(0,T)^L\right]\right) 1_{\hat{s}_i>0}\lambda^+ \tag{A.8}$$

where $\in_{(0)}$ denotes the unit mass on 0 and λ^+ denotes the Lebesgue measure on $\mathbb{R}^+$.

Conditions for embeddability of the matrix $\widehat{M}(0,T)$

The unique necessary and sufficient condition for embeddability was given by Kendall, who proved that, when $K = 2$, the transition matrix $\widehat{M}(0,T)$ is embeddable if and only if the trace of $\widehat{M}(0,T)$ is strictly greater than 1. When $K \geq 3$, only necessary conditions are known; they are presented, for example, by Singer and Spilerman (1976a), and Geweke *et al.* (1986b). The solution to the equation (2.11) is given by the following theorem (Singer and Spilerman, 1976a):

If $\widehat{M}(0,T)$ has K distinct[16] *eigenvalues $(\lambda_1, \ldots, \lambda_K)$ and can be written $\widehat{M}(0,T) = A \times D \times A^{-1}$, where $D = diag(\lambda_1, \ldots, \lambda_K)$ and the eigenvector corresponding to λ_ℓ $(\ell = 1, \ldots, K)$ is contained in the ℓ_th column of the $(K \times K)$ matrix A, then:*

$$\log \widehat{M}(0,T) = \widehat{Q}T = A \times \begin{pmatrix} \log_{k_1}(\lambda_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \log_{k_K}(\lambda_K) \end{pmatrix} \times A^{-1} \tag{A.9}$$

where $\log_{k_\ell}(\lambda_\ell) = \log \mid \lambda_\ell \mid + (\arg \lambda_\ell + 2k_\ell \Pi)i, k_\ell \in \mathcal{Z}$, *is a branch of the logarithm of λ_ℓ, when $\lambda_\ell \in \mathcal{C}$.*[17]

Because equation (2.11) has as many solutions $\widehat{Q}$ as there are combinations of the form $(\log_{k_1}(\lambda_1), \ldots, \log_{k_K}(\lambda_K))$, the number of these solutions is infinite when the matrix $\widehat{M}(0,T)$ has at least two complex conjugate eigenvalues. However, an important implication of the necessary condition for embeddability given by Runnenberg (1962) is that only finitely many branches of $\log \widehat{M}(0,T)$ need to be checked for membership in $\mathcal{Q}$. In fact, this condition implies that the only branches to consider are such that

$$\forall \lambda_\ell, \; -L_\ell \leq k_\ell \leq U_\ell \tag{A.10}$$

with

$$\begin{aligned} U_\ell &= \text{intpt} \left| \frac{\log|\lambda_\ell| \; \tan\left\{(\frac{1}{2}+\frac{1}{K})\Pi\right\} - |\arg \lambda_\ell|}{2\Pi} \right|, \\ L_\ell &= \text{intpt} \left| \frac{\log|\lambda_\ell| \; \tan\left\{(\frac{3}{2}-\frac{1}{K})\Pi\right\} - |\arg \lambda_\ell|}{2\Pi} \right|, \end{aligned}$$

where the function "intpt" yields the integer part of a real number. So the number of branches of λ_ℓ which must be computed is $L_\ell + U_\ell + 1$, the last one corresponding to the main branch with $k_\ell = 0$. Then the number of solutions $\widehat{Q}$ that must be examined for membership in $\mathcal{Q}$ is denoted $k^*(\widehat{M})$ and is equal to

$$k^*(\widehat{M}) = \begin{cases} \prod_{j=1}^{v} \{L_j + U_j + 1\} & , \; if \; v \geq 1, \\ 1 & , \; if \; v = 0, \end{cases} \tag{A.11}$$

[16] The case of repeated eigenvalues arises very rarely in the empirical applications. For its treatment, the reader can consult the paper by Singer and Spilerman (1976a, p. 19-25).

[17] Let us recall that the logarithmic function is multiple valued in the complex set $\mathcal{C}$. If $z = a + ib$ $(z \in \mathcal{C})$, then : $\log_k(z) = \log \mid z \mid + i(\theta + 2k\Pi)$, $k \in \mathcal{Z}$, with $\mid z \mid = \sqrt{a^2+b^2}$, and $\theta = \arg(z) = \tan^{-1}(b/a)$. Each value for k generates a distinct value for $\log(z)$, which is called a branch of the logarithm.

where v denotes the number of complex conjugate eigenvalue pairs of the matrix $\widehat{M}(0,T)$. Let us remark that:

- for a real eigenvalue, only the principal branch of the logarithm must be examined: other branches (with $k_\ell \neq 0$) correspond to complex intensity matrices $\widehat{Q}$;
- each element of a complex conjugate eigenvalue pair has the same number of candidate branches; moreover, only combinations of branches involving the same k_ℓ in each element of the pair must be computed; all others correspond to complex intensity matrices; this fact explains why the computation of $k^*(\widehat{M})$ is based on the number of complex conjugate eigenvalue pairs, and why the number of branches needing to be checked for each pair j is equal to $L_j + U_j + 1$, but not to $\{L_j + U_j + 1\}^2$.

APPENDIX B: BAYESIAN INFERENCE WITH IMPORTANCE SAMPLING

Choice of the importance sampling density

The importance function $I(M,s)$ must be conveniently chosen in order to improve convergence properties of $\hat{g}_I$. Several contributions (Kloek and van Dijk, 1978, 1980; van Dijk *et al.*, 1985; Geweke, 1989) have emphasized the necessity for the tails of the importance function no to decay more rapidly than the tails of the posterior density. For instance, Geweke (1989) proposes a procedure for tailoring the importance sampling density more routinely and introduces a measure of relative numerical efficiency (RNE) to evaluate the adequacy of any importance sampling density.

Following the general definition given by Geweke (1989, p.1322), we can write the RNE of an importance sampling density $I(M,s)$ for the function of interest $g(Q(M),s)$ as

$$\begin{aligned} RNE &\equiv \frac{\int_{D_K^*} \sum_{k=1}^{B(M)} h^{(k)}(M) \left[g(Q^{(k)}(M),s) - \bar{g}\right]^2 J(M) L(M,s;N,n)\mu(M,s) d(M,s)}{\int_{D_K^*} \sum_{k=1}^{B(M)} h^{(k)}(M) \left[g(Q^{(k)}(M),s) - \bar{g}\right]^2 J(M)\omega(M,s) L(M,s;N,n)\mu(M,s) d(M,s)} \\ &\equiv var\left[g(Q(M),s)\right]/\sigma^2 \end{aligned} \tag{B.1}$$

where the weight function $\omega(M,s)$ is the ratio $P(M,s)/I(M,s)$ of the posterior distribution to the importance sampling density. In fact, the RNE is the ratio of the posterior variance of the interest function g when the importance sampling density is the posterior density itself, to the posterior variance of g when the importance sampling density is the function $I(M,s)$. If the RNE is less than 1, one should draw $(RNE)^{-1}$ times more replications with the function $I(M,s)$ than with the function $P(M,s)$ in order to obtain a certain numerical standard error. Geweke (1989, Theorem 3) shows that if the mean deviation of the function $g(Q(M),s)$ under the posterior density, denoted $\mathrm{md}(g(Q(M),s))$, is finite, then the importance sampling density that minimizes σ^2 has kernel $|g(Q(M),s) - \bar{g}|P(M,s)$, and for that choice, $\sigma^2 = \{\mathrm{md}(g(Q(M),s))\}^2$. Consequently, if the RNE is greater than 1, the function $I(M,s)$ is “closer” to that kernel than the posterior density itself. In fact, if we define $RNE^* \equiv var(g(Q(M),s))/\,\mathrm{md}(g(Q(M),s))$, then $RNE^* \leq 1$ (with $RNE^* = 1$ when the importance sampling density belongs to the previous kernel).

In order to obtain a better behavior of the function $I(M,s)$ away from the posterior mode and with a possible asymmetry of the posterior density, Geweke (1989) suggests the use of split-normal or split-Student densities as importance sampling densities. Let us denote $SN_k(\theta,T,q,r)$ and $ST_k(\theta,T,q,r,\nu)$ the k-variate split normal and split-Student distributions, respectively.

If $y = (y_1,\ldots,y_k)' \sim SN_k(\theta,T,q,r)$, then there exists a $(k \times 1)$ standard normal vector $\epsilon = (\epsilon_1,\ldots,\epsilon_k)', \epsilon \sim N(0,I_k)$, such as

$$y = \theta + Tx, \tag{B.2}$$

where $x = (x_1,\ldots,x_k)', x_i = \epsilon_i\left[q_i 1_{[\epsilon_i \geq 0]} + r_i 1_{[\epsilon_i<0]}\right], i = 1,\ldots,k$. In the application, θ is the posterior mode and T is an upper triangular matrix obtained from the Choleski decomposition of the negative inverse of the hessian matrix of the log posterior density computed at the posterior mode. The k-variate split normal p.d.f. evaluated at y is

$$\begin{aligned} f_k(y) &= (2\Pi)^{-k/2}\left|TD^2T'\right|^{-1/2}\exp\left\{-\tfrac{1}{2}(y-\theta)'(TD^2T')^{-1}(y-\theta)\right\} \\ &= (2\Pi)^{-k/2}\left|TT'\right|^{-1/2}\left\{\prod_{i=1}^{k}\left[q_i 1_{[\epsilon_i\geq 0]} + r_i 1_{[\epsilon_i<0]}\right]\right\}^{-1}\exp\left\{-\tfrac{1}{2}\epsilon'\epsilon\right\} \end{aligned} \tag{B.3}$$

where D denotes a diagonal matrix with entries $d_{ii} = \left[q_i 1_{[\epsilon_i\geq 0]} + r_i 1_{[\epsilon_i<0]}\right], i = 1,\ldots,k$.

A $(k \times k)$ random vector $y \sim ST_k(\theta,T,q,r,\nu)$ can be constructed following the same steps, setting now $x_i = \epsilon_i\left[q_i 1_{[\epsilon_i\geq 0]} + r_i 1_{[\epsilon_i<0]}\right](\xi/\nu)^{-1/2}$ with $\xi \sim \chi^2(\nu)$. The split-Student t distribution is derived from the multivariate t distribution with a common denominator (see Johnson and Kotz, 1972, for a general presentation). Consequently, the split-Student p.d.f. evaluated at y is

$$\begin{aligned} f_k(y) &= \frac{\Gamma((\nu+k)/2)}{(\Pi\nu)^{k/2}\Gamma(\nu/2)}\left|TT'\right|^{-1/2}\left\{1+\nu^{-1}(y-\theta)'(TD^2T')^{-1}(y-\theta)\right\}^{-\frac{(\nu+k)}{2}} \\ &\quad\times\left\{\prod_{i=1}^{k}\left[q_i 1_{[\epsilon_i\geq 0]} + r_i 1_{[\epsilon_i<0]}\right]\right\}^{-1} \\ &= \frac{\Gamma((\nu+k)/2)}{(\Pi\nu)^{k/2}\Gamma(\nu/2)}\left|TT'\right|^{-1}\left\{\prod_{i=1}^{k}\left[q_i 1_{[\epsilon_i\geq 0]} + r_i 1_{[\epsilon_i<0]}\right]\right\}^{-1} \\ &\quad\times(1+\nu^{-1}\epsilon'\epsilon)^{-(\nu+k)/2} \end{aligned} \tag{B.4}$$

Geweke (1989, p. 1325-1326) proposes to choose scalars q_i and $r_i (i = 1,\ldots,k)$ in order to fit the variance of the univariate normal or Student distribution to the slowest rate of decline of the posterior density along each axis. More precisely, q_i and r_i are defined by

$$q_i = \sup_{\delta>0}\ f_i(\delta) \text{ and } r_i = \sup_{\delta<0}\ f_i(\delta) \tag{B.5}$$

where $f_i(\delta)$ is given by

$$f_i(\delta) = |\delta| \left\{ 2 \left[\log P(\theta) - \log P(\theta + \delta T e^{(i)}) \right] \right\}^{-1/2} \tag{B.6}$$

for the split normal density, or

$$f_i(\delta) = \nu^{-1/2} |\delta| \left\{ \left[p(\theta)/p(\theta + \delta T e^{(i)}) \right]^{2/(\nu+k)} - 1 \right\}^{-1/2} \tag{B.7}$$

for the split Student density, where $e^{(i)}$ is a $(k \times 1)$ indicator vector, such as $e_i^{(i)} = 1$ and $e_j^{(i)} = 0$ for $j \neq i$. In practice, δ is chosen to vary through the range $\{0.1, 0.2, \ldots, 10\}$.

Considering analytical expressions of ML estimators $(\hat{M}, \hat{s})$ for parameters (M, s) of the mover-stayer model, it is possible to simplify slightly the replication procedure induced by formula (B.2), in which θ is in practice replaced by $(\hat{M}, \hat{s})$, and T is the Choleski factor of the covariance matrix $V(M, s)$ evaluated at $(\hat{M}, \hat{s})$ and defined by

$$V(M, s) = R^{-1}(M, s) = -E \left[\frac{1}{N} \frac{\partial^2 \ln(L(M, s; N, n))}{\partial\theta\partial\theta'} \right]^{-1}, \tag{B.8}$$

where $R(M, s)$ is the information matrix associated to $L(M, s; N, n)$, and

$$\ln(L) = \sum_{i=1}^{K} n_i(0) \ln(\eta_i(0)) + \sum_{i=1}^{K} \ln(L_i), \forall i \in E$$

Define now

$$V_i(M, s) = -E \left[\frac{1}{N} \frac{\partial^2 \ln(L_i(M, s; N, n))}{\partial\theta\partial\theta'} \right]^{-1} = R_i(M, s)^{-1}. \tag{B.9}$$

The covariance matrix V satisfies $V = diag(V_1, \ldots, V_K)$ (see appendix A) where V_i $(i \in E)$ is defined by equation (B.9). So V is block diagonal, consisting of blocks $V_i(M, s)$, $i \in E$. Consequently, to generate multivariate split-normal (or split-Student) vectors $(M, s)_{\ell=1,\ldots,I}$, we can generate independent multivariate split-normal (or split-Student) drawings $(M, s)_{\ell,i} = (s_i^\ell, m_{i1}^\ell, \ldots, m_{iK}^\ell)$ where ℓ is the index of the drawing, i is the index of the state, s_i^ℓ is the proportion of stayers in state i and m_{ij}^ℓ is the transition probability from state i to state j for the $\ell - th$ drawing. More precisely, to ensure that $(M, s)_\ell \in D_K$, firstly we draw vectors

$$\begin{aligned} (s_i^\ell, m_{i1}^\ell, \ldots, m_{iK-1}^\ell) \quad & \sim SN_K((\hat{s}_i, \hat{m}_{i1}, \ldots, \hat{m}_{i,K-1}), T_i-, q_i, r_i) \\ \text{or} \quad & \sim ST_K((\hat{s}_i, \hat{m}_{i1}, \ldots, \hat{m}_{i,K-1}), T_i-, q_i, r_i, \nu) \end{aligned} \tag{B.10}$$

where T_{i-} is the Choleski factor of $\hat{V}_{i-}$ obtained by dropping the last row and the last column of $\hat{V}_i = V_i(\hat{M}, \hat{s}) = R_i(\hat{m}, \hat{s})^{-1}$, q_i and r_i being $(K \times 1)$ vectors whose elements are given by (B.5), (B.6) and (B.7). Then we obtain m_{iK}^ℓ by setting $m_{iK}^\ell = 1 - \sum_{j=1}^{K-1} m_{ij}^\ell$.[18]

[18]To generate drawings $(M, s)_{\ell,i}$ verifying constraints $s_i \in [0, 1]$, $m_{ij} \in [0, 1]$ and $\sum_{j=1}^{K} m_{ij} = 1$, we apply a naive acceptance-rejection method which is known to be possibly inefficient. More efficient procedures are proposed by Geweke (1991) and Robert (1995).

When at least one of the true values of the parameters lies on the boundary of the parameter space, we have seen in Appendix A that the joint distibution of the restricted MLE is difficult to obtain. In that case, another way to construct the likelihood is to use a "normal approximation of prior" (see van Dijk and Kloek, 1980, pp.316-317). This procedure consists firstly in using an approximation of the uniform prior by a $N(\theta_0, \frac{1}{k}V_0)$ density, and then in constructing an "information contract curve" or "curve décolletage" by maximizing the expression

$$\ln L(x;\theta) - \frac{1}{2}k(\theta-\theta_0)'V_0^{-1}(\theta-\theta_0)$$

with respect to θ for a sequence of values for k.

In the case we consider, $L(x;\theta)$ is the likelihood function of the discrete time mover-stayer model and $\theta = (s, M)$. Moreover, we have to choose a value of k such that the approximation obtained for the posterior mode of θ is just inside the parameter space. The approximations obtained for the posterior mode and covariance matrix may then be used to construct the importance function. This method can be improved through a "two-step normal approximation of prior": the posterior mean and covariance matrix obtained as indicated before may be used as moments of a new normal prior (denoted θ_0' and V_0') for a second application of the procedure. The normal prior used at the first step becomes less informative in the sense that the information brought by the data is now used twice.

Algorithm

Let us consider drawings of vectors $(M, s)_{\ell=1,\dots,I}$ from the importance function $I(M, s)$.

Step 0: Compute the maximum likelihood estimates $(\hat{M}, \hat{s})$ and the estimated covariance matrix

$$\hat{V} = -\left[\frac{1}{N}\frac{\partial^2 \ln L(M, s; N, n)}{\partial\theta\partial\theta'}\right]^{-1}_{(M,s)=(\hat{M},\hat{s})}$$

The indices ℓ and i are relative to the drawing from the importance function and to the state for which the parameter vector $(s_i^\ell, m_{i1}^\ell, \dots, m_{iK-1}^\ell)$ is drawn, respectively.

For $\ell = 1, \dots, I$, and for $i = 1, \dots, K$, the following steps are examined:

Step 1: Compute scalars q_i and r_i according to formulas (B.5)-(B.7) and draw $(M, s)_{\ell,i}$ from the chosen importance sampling density.

Step 2:

- if $(M, s)_{\ell,i} \in D_K$ and $i < K$, then $i = i + 1$ and return to Step 1.
- if $(M, s)_{\ell,i} \in D_K$ and $i = K$, then go Step 3.
- if $(M, s)_{\ell,i} \notin D_K$ then return to Step 1.

Step 3: "Check for embeddability"

- if $(M, s)_\ell \in D_K^*$ then go to Step 4.
- if $(M, s)_\ell \notin D_K^*$ then go to Step 5.

Step 4: Compute $g(Q_\ell^{(k)}, s_\ell), k = 1, \dots, B(M_\ell)$. Then go to Step 5.
Step 5: If $\ell < I$, then return to step 1 with $\ell = \ell + 1$ and $i = 1$. Else go to Step 6.
Step 6: End.

Functions of interest

a) **Transition intensities**

To compute the posterior mean of q_{ij} (respectively s_i), we set $g(Q, s) = g_1(Q, s) = q_{ij}$ (respectively s_i) and we use equation (3.5). To compute the posterior standard deviation of q_{ij} (respectively s_i), we define functions $g(Q, s) = g_2(Q, s) = q_{ij}^2$ (respectively $g_1(Q, s) = s_i^2$), we use equation (3.5) and we form

$$\left[E\left[g_2(Q, s)|N, n; M \in M_K^*\right] - E\left[g_1(Q, s)|N, n; M \in M_K^*\right]^2\right]^{1/2}. \tag{B.11}$$

b) Posterior mean duration

The computation of the posterior mean sojourn duration of a mover in state i can be made by setting $g_1(Q, s) = -q_{ii}^{-1}$. For the posterior standard deviation, we consider $g_2(Q, s) = q_{ii}^{-2}$ and we use one of equations (B.11).

c) Equilibrium probabilities

Let us recall that the limiting probabilities for the mover-stayer model are given by the formula (2.9). This formula involves the equilibrium distribution $\Pi^{(m)}$ for the "mover"part of the population, which is given by equation (2.7). Let $\Pi^{(m)} = F(Q)$ denote the solution of the system of equations $Q' \, \Pi^{(m)} = 0$. To compute the posterior moments of $\Pi^{(m)}$, we set :

$$g_{1i}(Q, s) = F_i(Q) \text{ and } g_{2i} = (F_i(Q))^2, i \in E.$$

So, using the formula (2.9), we get the limiting probabilities of the mover-stayer model by setting :

$$g_{1i}(Q, s) = s_i \eta_i + F_i(Q) \sum_{j=1}^{K} (1 - s_j)\eta_j, \; i \in E, \tag{B.12}$$

and

$$g_{2i}(Q, s) = \{g_{1i}(Q, s)\}^2, i \in E.$$

The posterior moments of $\Pi^{(m)}$ and Π are obtained by application of formula (3.5).

d) Mobility indices

For the movers, the mobility index considered here is given by the formula :

$$M(Q) = -\log\left[\det M(0, T)\right]/K = -\operatorname{tr}(Q)/K. \tag{B.13}$$

This index satisfies the criteria of *monotonicity*, *strong immobility*, *velocity* and *freedom from aliasing* (see Geweke *et al.*, 1986b). Its posterior moments are estimated using formula (3.5).

ACKNOWLEDGEMENTS

We would like to thank Tony Lancaster, Michel Lubrano, Alain Monfort and Christian Robert for helpful comments and discussions. We are also grateful to the editors, M. Hashem Pesaran and Herman van Dijk, and two anonymous referees whose remarks and suggestions helped us to revise and improve greatly the paper. Finally we would thank the French National Statistical Institute (INSEE) for providing the data.

REFERENCES

Albert J, Chib S. 1993. Bayesian inference for autoregressive time series with mean and variance subject to Markov jumps. *Journal of Business and Economic Statistics* **11**: 1-15.

Anderson TW, Goodman LA. 1957. Statistical inference about Markov chains. *Annals of Mathematical Statistics* **28**: 89-110.

Billingsley P. 1961. *Statistical inference for Markov Processes.* The University of Chicago Press: Chicago, IL.

Blumen I, Kogan M, Mac Carthy PJ. 1955. *The industrial mobility of labour as a probability process.* Cornell Studies of Industrial and Labour Relations (Vol. 6), Cornell University Press: Ithaca, NY.

Buchinsky M, Fields G, Fougère D, Kramarz F. 2001. Ranks or francs? Earnings mobility in France: 1968-1999. Mimeo, CREST-INSEE.

Chib S. 1992. Bayes inference in the Tobit censored regression model. *Journal of Econometrics* **51**: 79-99.

Doob JL. 1953. *Stochastic processes.* Wiley: New-York.

Fougère D, Kamionka T. 1992a. Un modèle markovien du marché du travail. *Annales d'Economie et de Statistique* **27**: 149-188.

Fougère D, Kamionka T. 1992b. Mobilité et précarisation sur le marché français du travail: une analyse longitudinale pour les années 1986 à 1988. *Economie et Prévision* **102-103**: 157-178.

Frydman H. 1984. Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* **79**: 632-638.

Geweke J. 1989. Bayesian inference in econometric models using Monte-Carlo integration. *Econometrica* **57**: 1317-1339.

Geweke J. 1991. Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints. *Computer Science and Statistics:*

Proceedings of the 23rd Symposium on the Interface, 571-578. Interface Foundation of North America, Inc.: Fairfax, VA.

Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*, Vol. 4., Bernardo JM, Berger JO, Dawid DA, Smith AFM (eds). Oxford University Press: Oxford.

Geweke J, Marshall RC, Zarkin GA. 1986a. Exact inference for continuous time Markov chains. *Review of Economic Studies* **53**: 653-669.

Geweke J, Marshall RC, Zarkin GA. 1986b. Mobility indices in continuous-time Markov chains. *Econometrica* **54**: 1407-1423.

Goodman LA. 1961. Statistical methods for the mover-stayer model. *Journal of the American Statistical Association* **56**: 841-868.

Gottschalk P., Moffitt R. 1999. Changes in job instability and insecurity using monthly survey data. *Journal of Labor Economics* **17**: S91-S126.

Gouriéroux C, Monfort A. 1995. *Statistics and Econometric Models*. Cambridge University Press: Cambridge.

Heckman J, Singer B. 1984. Econometric duration analysis. *Journal of Econometrics* **24**: 63-132.

Kalbfleisch JD, Lawless JF. 1985. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**: 863-871.

Kamionka T. 1998. Simulated maximum likelihood estimation in transition models. *The Econometrics Journal* **1**: C129-C153.

Kiefer N. 1978. Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica* **46**: 427-434.

Kloek T, Van Dijk H. 1978. Bayesian estimates of equation system parameters: an application of integration by Monte-Carlo. *Econometrica* **46**: 1-20.

Lancaster T. 1990. *The econometric analysis of transition data*. Econometric Society Monographs. Cambridge University Press: New York.

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

Mac Culloch R, Rossi P. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64**: 207-240.

Mac Culloch R, Tsay R. 1994. Statistical analysis of macroeconomic time series via Markov switching models. *Journal of Time Series Analysis* **15**: 523-539.

Martin JJ. 1967. *Bayesian decision problems and Markov chains*. Wiley: New-York.

Neumark D., Polsky D., Hansen D. 1999. Has job stability decline yet? New evidence for the 1990s. *Journal of Labor Economics* **17**: S29-S64.

Robert C. 1994. *The Bayesian choice*. Springer-Verlag: New-York.

Robert C. 1995. Simulation of truncated normal variables. *Statistics and Computing* **5**: 121-125.

Robert C., Casella G. 1999. *Monte Carlo statistical methods.* Springer-Verlag: New-York.

Runnenberg JT. 1962. On Elfving's problem of imbedding a time-discrete Markov chain in a continuous-time one for finitely many states. *Proceedings, Koninklijke Nederlandse Akademic van Wetenschappen, serie A, Mathematical Sciences* **65**: 536-541.

Sampson M. 1990. A Markov chain model for unskilled workers and the highly mobile. *Journal of the American Statistical Association* **85**: 177- 180.

Singer B. and Spilerman S. 1976a. The representation of social processes by Markov models. *American Journal of Sociology* **82**: 1-54.

Singer B. and Spilerman S. 1976b. Some methodological issues in the analysis of longitudinal surveys. *Annals of Economic and Social Measurement* **5**: 447-474.

Spilerman S. 1972. Extensions of the mover-stayer model. *American Journal of Sociology* **78**: 599-626.

Tierney L. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**: 1701-1786.

Van Dijk HK, Kloek T. 1980. Further experience in bayesian analysis using Monte-Carlo integration. *Journal of Econometrics* **14**: 307-328.

Van Dijk HK, Kloek T, Boender CGE. 1985. Posterior moments computed by mixed integration. *Journal of Econometrics* **29**: 3-18.

Table II. Proportions of stayers (%)

	Age groups							
	15-25		26-35		36-50		51-65	
State	M	W	M	W	M	W	M	W
Employment								
Normal IS	42.63	53.49	87.43	68.68	82.84	81.02	28.93	58.40
	(4.70)	(4.30)	(2.26)	(5.88)	(3.32)	(2.65)	(11.64)	(3.88)
RNE	0.28	0.02	0.19	0.01	0.05	0.30	0.04	0.09
Split Normal IS	42.77	53.25	87.63	69.65	83.04	81.07	28.67	58.47
	(4.55)	(4.21)	(2.25)	(4.42)	(3.38)	(2.71)	(11.71)	(3.85)
RNE	0.70	0.37	0.35	0.46	0.08	0.26	0.03	0.23
Gibbs sampling	43.22	53.54	87.55	70.06	84.00	76.32	37.89	58.33
	(4.44)	(4.17)	(2.15)	(4.64)	(2.67)	(3.21)	(8.87)	(3.97)
Unemployment								
Normal IS	4.91	7.25	13.42	6.01	15.70	0.00	11.08	12.52
	(2.83)	(4.16)	(5.82)	(3.70)	(6.05)	–	(5.73)	(7.11)
RNE	0.69	0.26	0.41	0.13	0.15	–	0.14	0.51
Split Normal IS	4.72	6.90	13.27	6.04	14.96	0.00	10.77	12.35
	(2.81)	(4.15)	(5.81)	(3.79)	(6.12)	–	(5.79)	(7.06)
RNE	0.89	0.66	0.63	0.72	0.20	–	0.20	0.51
Gibbs sampling	6.47	9.53	16.06	8.36	17.55	6.93	13.61	16.10
	(2.78)	(3.96)	(5.17)	(3.72)	(5.39)	(3.78)	(5.55)	(6.60)
Out of the labour force								
Normal IS	59.06	56.16	30.34	56.91	68.02	68.20	89.15	86.66
	(2.04)	(5.07)	(6.79)	(4.16)	(5.74)	(3.42)	(2.75)	(2.03)
RNE	0.84	0.01	0.30	0.02	0.28	0.43	0.05	0.23
Split Normal IS	59.01	55.51	30.27	56.63	68.43	68.33	89.97	86.82
	(2.01)	(4.65)	(6.63)	(4.12)	(6.28)	(3.37)	(2.40)	(2.04)
RNE	0.74	0.25	0.57	0.42	0.14	0.58	0.24	0.37
Gibbs sampling	59.11	56.59	31.61	57.10	67.72	74.46	89.11	86.85
	(1.94)	(4.21)	(6.42)	(3.74)	(5.89)	(2.26)	(2.88)	(1.96)

Remarks on Table II:

- Data source: "Enquêtes Emploi" (INSEE)
- Posterior standard deviations are reported in parentheses below the posterior means; for importance sampling algorithms, the RNE of a parameter estimate is given below the posterior standard deviation.
- Abbreviations : M for men, W for women, IS for importance sampling.

Table IIIA. Transition intensities of "movers" ($\times 10^{-4}$) obtained by importance sampling

	Age groups							
	15-25		26-35		36-50		51-65	
Transition	M	W	M	W	M	W	M	W
E $\rightarrow$ U								
Normal	7.30	12.59	7.89	3.56	5.35	4.90	1.15	1.59
	(1.15)	(1.88)	(2.11)	(0.91)	(1.31)	(1.06)	(0.28)	(0.42)
RNE	0.45	0.06	0.22	0.01	0.13	0.40	0.05	0.29
Split Normal	7.35	12.55	8.02	3.58	5.42	4.89	1.15	1.59
	(1.14)	(1.90)	(2.26)	(0.78)	(1.40)	(1.01)	(0.29)	(0.42)
RNE	0.70	0.45	0.14	0.49	0.08	0.45	0.04	0.52
E $\rightarrow$ OLF								
Normal	8.30	4.91	2.95	4.61	1.93	6.41	5.39	13.65
	(1.13)	(1.08)	(1.12)	(1.05)	(0.58)	(1.29)	(1.09)	(1.96)
RNE	0.41	0.01	0.21	0.01	0.07	0.38	0.03	0.20
Split Normal	8.28	4.72	2.94	4.63	1.93	6.35	5.86	13.61
	(1.12)	(0.82)	(1.12)	(0.92)	(0.58)	(1.27)	(1.08)	(1.94)
RNE	0.71	0.54	0.25	0.41	0.10	0.42	0.04	0.31
U $\rightarrow$ E								
Normal	25.72	20.61	20.80	13.98	20.30	12.54	7.34	2.38
	(3.16)	(2.56)	(3.75)	(1.83)	(3.18)	(1.65)	(1.48)	(1.24)
RNE	0.73	0.47	0.50	0.13	0.22	0.65	0.13	0.20
Split Normal	25.75	20.64	21.00	13.98	20.17	12.45	7.34	2.28
	(3.14)	(2.59)	(3.94)	(1.86)	(3.16)	(1.63)	(1.43)	(1.21)
RNE	0.76	0.62	0.23	0.66	0.27	0.65	0.17	0.52
U $\rightarrow$ OLF								
Normal	10.63	5.67	8.18	13.75	5.96	7.16	15.14	18.43
	(2.11)	(1.11)	(2.83)	(2.06)	(1.68)	(1.36)	(2.37)	(3.21)
RNE	0.55	0.06	0.44	0.07	0.18	0.50	0.11	0.42
Split Normal	10.58	5.63	8.03	13.72	5.72	7.06	15.14	18.45
	(2.08)	(1.11)	(2.91)	(2.07)	(1.60)	(1.35)	(2.25)	(3.21)
RNE	0.78	0.76	0.74	0.71	0.37	0.49	0.13	0.44
OLF $\rightarrow$ E								
Normal	16.78	7.81	16.27	6.20	8.40	10.89	1.97	5.14
	(1.81)	(1.84)	(6.09)	(1.18)	(3.32)	(1.94)	(0.47)	(0.93)
RNE	0.68	0.004	0.22	0.02	0.19	0.54	0.07	0.18
Split Normal	16.88	7.55	15.99	6.15	8.28	10.89	1.93	5.05
	(1.82)	(1.22)	(5.92)	(1.13)	(3.31)	(1.91)	(0.41)	(0.87)
RNE	0.70	0.37	0.54	0.59	0.22	0.56	0.14	0.44
OLF $\rightarrow$ U								
Normal	6.88	4.28	25.61	11.08	14.68	5.98	1.21	2.12
	(1.15)	(0.95)	(8.98)	(1.89)	(4.72)	(1.25)	(0.32)	(0.47)
RNE	0.66	0.01	0.39	0.09	0.16	0.54	0.12	0.26
Split Normal	6.56	3.95	25.75	10.80	14.21	5.61	1.09	1.93
	(1.07)	(0.77)	(8.76)	(1.86)	(4.61)	(1.16)	(0.30)	(0.41)
RNE	0.98	0.54	0.65	0.70	0.25	0.87	0.19	0.70
Posterior probability of embeddability								
Normal	1.00	1.00	0.997	1.00	0.995	1.00	1.00	0.982
	(0.00)	(0.00)	(0.001)	(0.00)	(0.002)	(0.00)	(0.00)	(0.002)
Split Normal	1.00	1.00	0.996	1.00	0.998	1.00	1.00	0.984
	(0.00)	(0.00)	(0.001)	(0.00)	(0.001)	(0.00)	(0.00)	(0.001)

Table IIIB. Transition intensities of "movers" ($\times 10^{-4}$) obtained by Gibbs sampling

	Age groups							
	15-25		26-35		36-50		51-65	
Transition	M	W	M	W	M	W	M	W
E → U								
	7.51	12.98	8.23	3.75	5.88	3.58	1.39	1.61
	(1.18)	(2.04)	(2.23)	(0.87)	(1.37)	(0.71)	(0.33)	(0.42)
E → OLF								
	8.40	4.80	3.00	4.77	2.08	5.50	6.35	13.70
	(1.14)	(0.86)	(1.16)	(0.96)	(0.59)	(0.98)	(1.20)	(2.06)
U → E								
	26.37	21.52	21.99	14.53	21.14	10.55	7.90	2.80
	(3.28)	(2.81)	(3.96)	(1.93)	(3.12)	(1.42)	(1.59)	(1.42)
U → OLF								
	10.93	6.02	8.98	14.39	6.28	9.06	15.68	19.23
	(2.10)	(1.17)	(3.08)	(2.20)	(1.77)	(1.38)	(2.37)	(3.36)
OLF → E								
	16.86	7.72	15.80	6.18	8.35	11.00	2.03	5.22
	(1.78)	(1.21)	(5.98)	(1.16)	(3.44)	(1.57)	(0.46)	(0.91)
OLF → U								
	6.99	4.35	26.95	11.46	15.30	6.34	1.31	2.24
	(1.14)	(0.84)	(9.16)	(1.97)	(5.01)	(1.09)	(0.38)	(0.50)
Posterior probability of embeddability	1.00	1.00	0.994	1.00	0.996	1.00	1.00	0.985
	(0.00)	(0.00)	(0.077)	(0.00)	(0.065)	(0.00)	(0.00)	(0.121)

Remarks on Tables IIIA and IIIB:

- See remarks for Table II.
- Data source: "Enquêtes Emploi" (INSEE)
- Below the estimate of the posterior probability of embeddability is its numerical standard error, computed with the formula given by Geweke *et al.* (1986a, p. 658).
- Abbreviations : E for employment, U for unemployment, OLF for out of the labour force state.

Table IVA. Mean sojourn durations (in days), limiting occupation probabilities (%) and mobility indices (%) estimated with the importance sampling procedure (with a split-normal importance function)

	Age groups							
	15-25		26-35		36-50		51-65	
States	M	W	M	W	M	W	M	W
Employment								
$\bar{T}^{(m)}$	647.98	588.95	967.23	1256.17	1441.95	918.80	1599.94	671.12
	(73.85)	(77.89)	(235.37)	(225.69)	(353.30)	(168.29)	(322.58)	(95.90)
RNE	0.71	0.38	0.28	0.47	0.09	0.35	0.04	0.24
$\Pi^{(m)}$	56.40	43.86	64.44	52.86	68.56	50.81	26.24	23.86
	(2.11)	(2.86)	(4.46)	(4.12)	(4.73)	(3.95)	(4.75)	(3.07)
RNE	0.71	0.30	0.25	0.46	0.09	0.43	0.08	0.31
Π	40.59	35.39	91.30	65.56	91.20	67.05	27.56	25.48
	(0.84)	(1.11)	(0.51)	(0.86)	(0.48)	(0.66)	(3.38)	(0.74)
RNE	0.77	0.36	0.43	0.70	0.19	0.61	0.03	0.35
Unemployment								
$\bar{T}^{(m)}$	277.64	384.96	354.74	364.98	394.54	516.46	452.53	494.71
	(25.78)	(40.47)	(59.73)	(37.62)	(57.47)	(46.27)	(59.24)	(78.17)
RNE	0.79	0.64	0.45	0.71	0.22	0.63	0.17	0.47
$\Pi^{(m)}$	16.31	25.47	26.02	18.05	20.23	20.71	4.68	8.21
	(1.12)	(1.88)	(3.51)	(1.83)	(3.18)	(2.12)	(0.93)	(1.34)
RNE	0.75	0.54	0.31	0.53	0.11	0.45	0.08	0.55
Π	8.50	13.11	6.00	7.40	4.63	5.54	2.55	2.30
	(0.48)	(0.80)	(0.42)	(0.47)	(0.34)	(0.45)	(0.43)	(0.27)
RNE	0.78	0.44	0.55	0.70	0.17	0.69	0.15	0.56
Out of the labour force								
$\bar{T}^{(m)}$	430.21	887.79	250.76	602.36	469.81	620.91	3411.71	1469.08
	(39.37)	(129.37)	(53.09)	(86.95)	(116.11)	(97.48)	(619.41)	(235.65)
RNE	0.73	0.31	0.47	0.54	0.13	0.61	0.14	0.48
$\Pi^{(m)}$	27.29	30.67	9.54	29.09	11.21	28.48	69.08	67.93
	(1.74)	(3.06)	(1.62)	(3.19)	(2.37)	(3.18)	(4.83)	(3.51)
RNE	0.72	0.31	0.25	0.48	0.10	0.53	0.10	0.35
Π	50.91	51.50	2.69	27.04	4.17	27.41	69.89	72.22
	(0.85)	(1.35)	(0.27)	(0.75)	(0.34)	(0.62)	(3.28)	(0.73)
RNE	0.76	0.36	0.46	0.65	0.19	0.53	0.03	0.36
Mobility index	0.25	0.18	0.27	0.18	0.19	0.16	0.11	0.14
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
RNE	0.76	0.49	0.49	0.65	0.20	0.56	0.14	0.39

Table IVB. Mean sojourn durations (in days), limiting occupation probabilities (%) and mobility indices (%) obtained by Gibbs sampling

	Age groups							
	15-25		26-35		36-50		51-65	
States	M	W	M	W	M	W	M	W
Employment								
$\bar{T}^{(m)}$	636.45	572.71	939.80	1213.90	1313.83	1131.61	1337.79	667.26
	(72.24)	(77.58)	(219.26)	(229.87)	(283.70)	(193.49)	(248.85)	(99.51)
$\Pi^{(m)}$	56.25	44.09	64.57	52.64	67.58	54.46	24.25	24.46
	(2.12)	(2.71)	(4.20)	(4.30)	(4.16)	(3.73)	(4.65)	(3.05)
Π	40.46	35.14	91.14	65.45	91.20	63.69	30.39	25.57
	(0.84)	(1.05)	(0.49)	(0.85)	(0.45)	(0.58)	(2.62)	(0.74)
Unemployment:								
$\bar{T}^{(m)}$	270.43	367.51	331.65	349.65	371.62	514.42	431.42	464.20
	(25.31)	(40.06)	(53.37)	(36.16)	(51.06)	(47.51)	(56.61)	(69.78)
$\Pi^{(m)}$	16.42	25.49	25.54	18.22	20.75	18.56	5.30	8.62
	(1.10)	(1.75)	(3.29)	(1.83)	(2.86)	(1.91)	(1.10)	(1.39)
Π	8.64	13.09	6.08	7.54	4.66	5.50	2.70	2.48
	(0.48)	(0.76)	(0.42)	(0.47)	(0.32)	(0.38)	(0.45)	(0.29)
Out-of-labor-force								
$\bar{T}^{(m)}$	422.54	844.76	244.18	578.17	445.87	586.22	3115.80	1376.93
	(37.43)	(117.28)	(49.83)	(80.27)	(106.93)	(74.72)	(643.73)	(228.75)
$\Pi^{(m)}$	27.33	30.42	9.89	29.13	11.67	26.97	70.45	66.92
	(1.75)	(2.91)	(1.59)	(3.28)	(2.21)	(2.74)	(4.96)	(3.49)
Π	50.91	51.76	2.78	27.01	4.14	30.82	66.92	71.95
	(0.83)	(1.26)	(0.27)	(0.73)	(0.32)	(0.50)	(2.54)	(0.73)
Mobility index	0.26	0.19	0.28	0.18	0.20	0.15	0.12	0.15
	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)

Remarks on Tables IVA and IVB :

- See remarks for Table II
- Data source: "Enquêtes Emploi" (INSEE)
- Abbreviations : $\bar{T}^{(m)}$: mean sojourn duration of movers, $\Pi^{(m)}$: limiting occupation probability for movers, Π : limiting occupation probability for the whole population.

Figure 1 : Kernel estimates of the posterior marginal densities of the proportions of stayers in unemployment (Epanechnikov kernels)

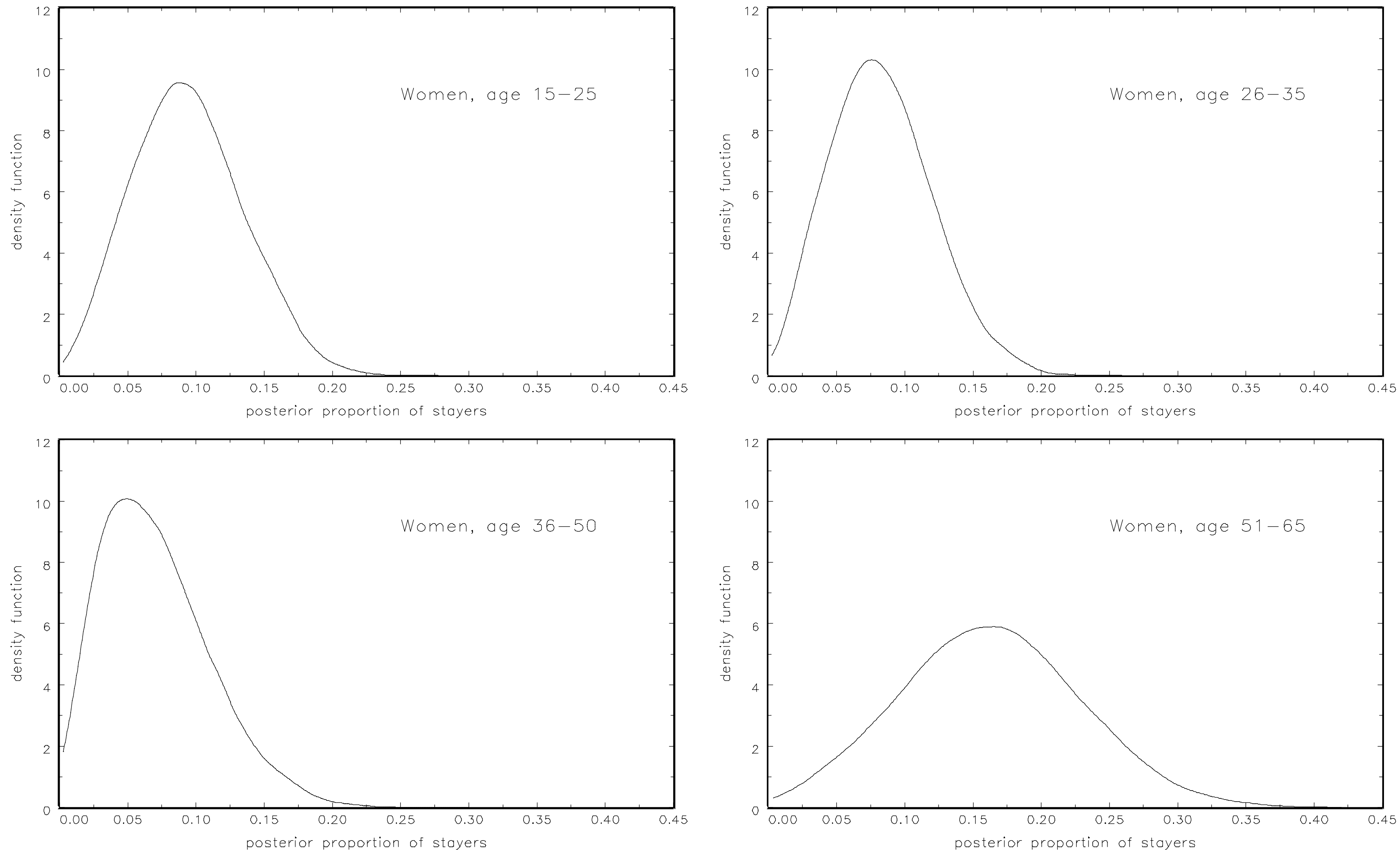

Figure 2: Proportion of stayers, 1986–1998

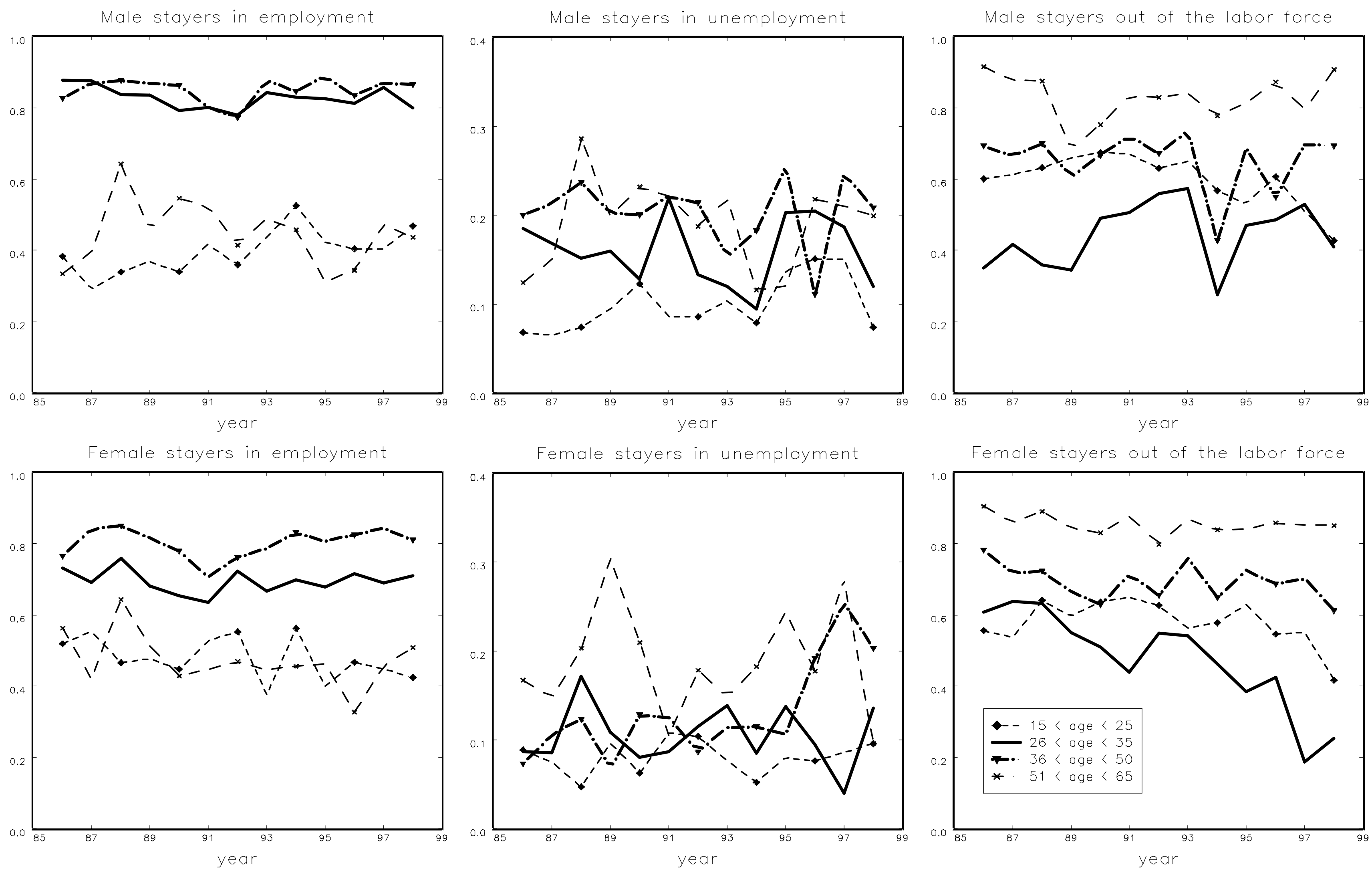

Figure 3: Transition intensities for movers, 1986–1998

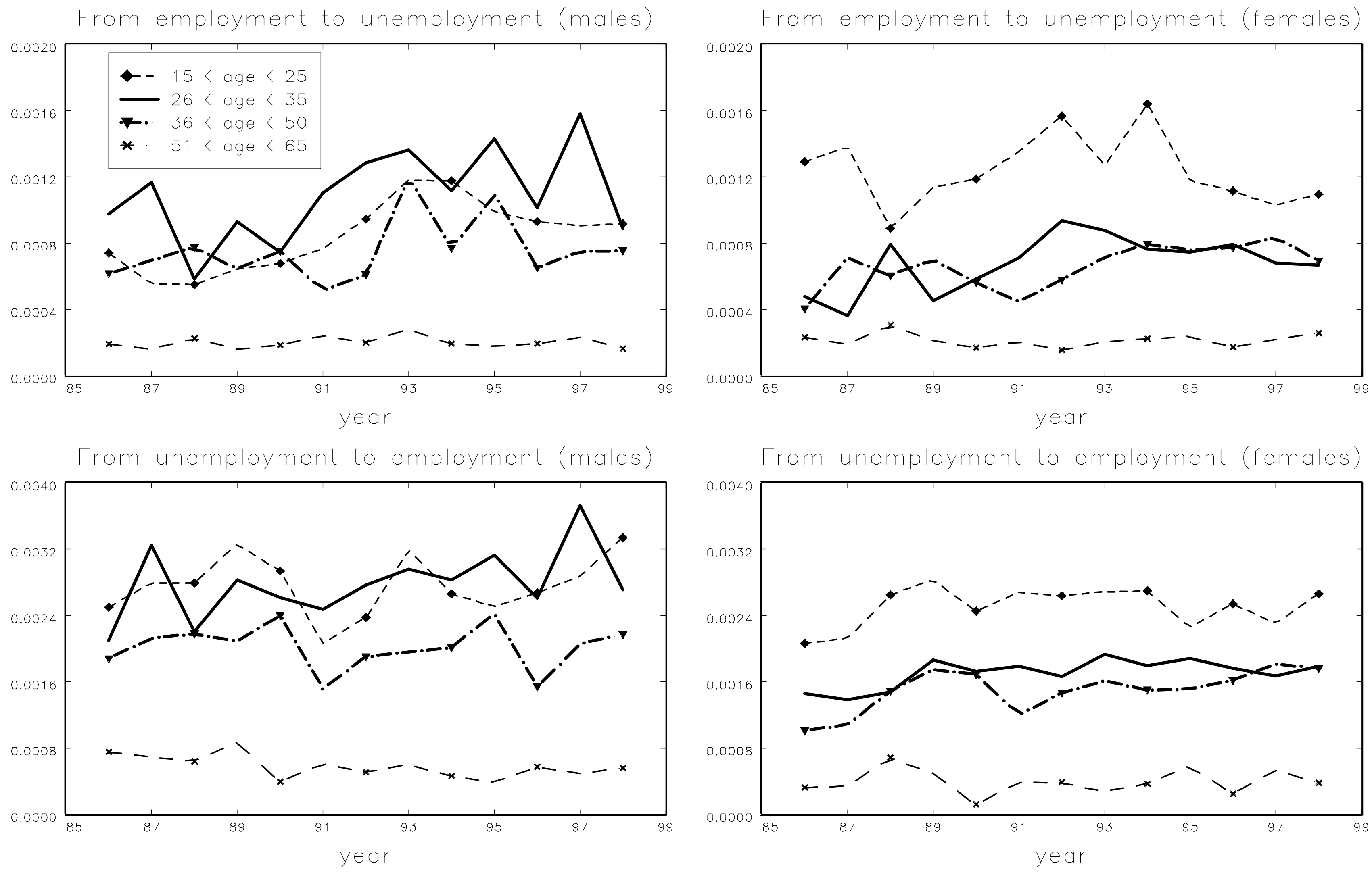

Figure 4: Equilibrium distributions, 1986–1998

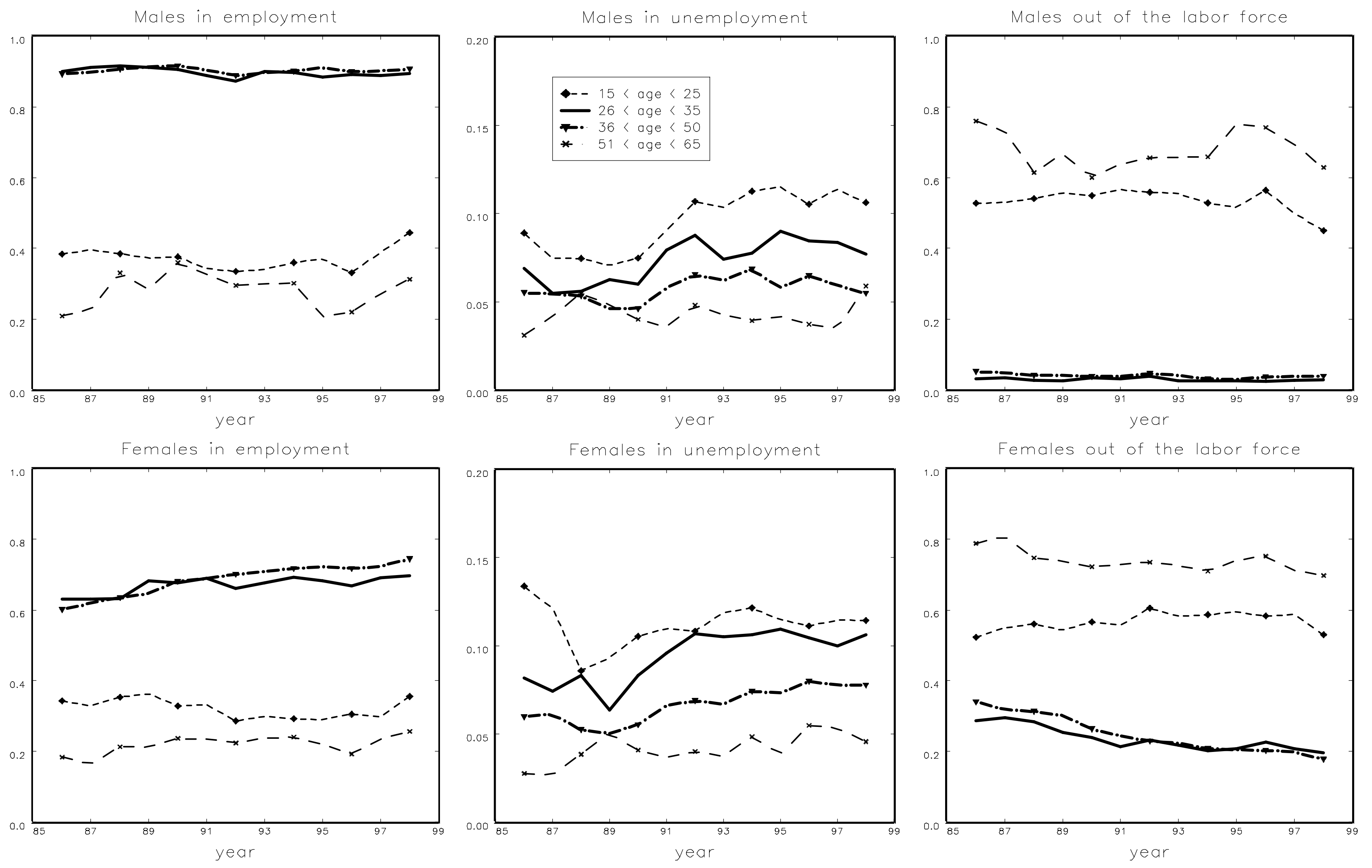

Figure 5: Mobility indices, 1986–1998