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

n° 2005-24

Econometrics of Individual Labor Market Transitions

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

[1] CNRS and CREST-INSEE, CEPR and IZA. fougere@ensae.fr
[2] CNRS and CREST-INSEE, Paris. kamionka@ensae.fr

ECONOMETRICS OF INDIVIDUAL LABOR MARKET TRANSITIONS

Denis FOUGERE
CNRS and CREST-INSEE, CEPR and IZA.
fougere@ensae.fr

Thierry Kamionka
CNRS and CREST-INSEE, Paris.
kamionka@ensae.fr

October 2005

Abstract

This survey is devoted to the modelling and the estimation of reduced-form transition models, which have been extensively used and estimated in labor microeconometrics. The first section contains a general presentation of the statistical modelling of such processes using continuous-time (event-history) data. It also presents parametric and nonparametric estimation procedures, and focuses on the treatment of unobserved heterogeneity. The second section deals with the estimation of markovian processes using discrete-time panel observations. Here the main question is whether the discrete-time panel observation of a transition process is generated by a continuous-time homogeneous Markov process. After discussing this problem, we present maximum-likelihood and bayesian procedures for estimating the transition intensity matrix governing the process evolution. Particular attention is paid to the estimation of the continuous-time mover-stayer model, which is the more elementary model of mixed Markov chains.

Keywords: labor market transitions, markovian processes, mover-stayer model, unobserved heterogeneity.

Résumé

Cette revue de la littérature est relative à la modélisation et l'estimation de modèles de transition en formes réduites. Ces modèles ont été largement utilisés en microéconométrie du marché du travail. La première section contient une présentation générale de ces processus lorsque ceux-ci peuvent être observés à partir de données en temps continu. Nous traitons aussi dans cette section des procédures d'estimation paramétriques et nonparamétriques, et de la question de l'hétérogénéité non observable. La seconde section est consacrée à l'estimation de processus observés à partir d'observations réalisées en temps discret. Ici, la question principale est de déterminer si les observations en temps discret du processus de transition ont pu être générées par un processus markovien en temps continu homogène. Après avoir discuté de ce problème, nous présentons des procédures bayésiennes et par maximum de vraisemblance permettant d'estimer la matrice des intensités de transition gouvernant le processus étudié. Une attention particulière est portée au cas de l'estimation d'un processus de type mover-stayer en temps continu, qui constitue le modèle le plus simple de mélange de chaînes de Markov.

Mots Clés : transitions sur le marché du travail, processus markoviens, modèle de mover-stayer, hétérogénéité non observable.

JEL Codes: C41, C51, J64.

1 Introduction

During the last twenty years, the microeconometric analysis of individual transitions has been extensively used for investigating some problems inherent in the functioning of contemporary labor markets, such as the relations between individual mobility and wages, the variability of flows between employment, unemployment and non-employment through the business cycle, or the effects of public policies (training programs, unemployment insurance, ...) on individual patterns of unemployment. Typically, labor market transition data register sequences of durations spent by workers in distinct states, such as employment, unemployment and non-employment. When individual participation histories are completely observed through panel or retrospective surveys, the econometrician then disposes of continuous-time realizations of the labor market participation process. When these histories are only observed at many successive dates through panel surveys, the available information is a truncated one; more precisely it takes the form of discrete-time observations of underlying continuous-time processes. Our presentation of statistical procedures used for analysing individual transition or mobility histories is based on the distinction between these two kinds of data.

Statistical models of labor market transitions can be viewed as extensions of the single-spell unemployment duration model (see Chapter 14, this volume). Theoretically, a transition process is a continuous-time process taking its values in a finite discrete state space whose elements represent the main labor force participation states, for example employment, unemployment and non-employment.

The goal is then to estimate parameters which capture effects of different time-independent or time-varying exogenous variables on intensities of transition between states of participation. Here transition intensities represent conditional instantaneous probabilities of transition between two distinct states at some date. Typically, the analyst is interested in knowing the sign and the size of the influence of a given variable, such as the unemployment insurance amount or the past training and employment experiences, on the transition from unemployment to employment for example, and more generally in predicting the effect of such variables on the future of the transition process. For this purpose, she can treat these variables as regressors in the specification of transition intensities. Doing that, she estimates a reduced-form model of transition. Estimation of a more structural model requires the specification of an underlying dynamic structure in which the participation state is basically the choice set for a worker and in which parameters to be estimated influence directly individual objective functions (such as intertemporal utility functions) which must be maximized under some revelant constraints inside a dynamic programming setup. Such structural models have been surveyed by Eckstein and Wolpin (1989) or Rust (1994).

Our survey focuses only on reduced-form transition models, which have been extensively used and estimated in labor microeconometrics. The first section contains a general presentation of the statistical modelling of the transition process for continuous-time (event-history) data. The first section briefly recalls the useful mathematical definitions, essentially the ones characterizing the distribution of the joint sequence of visited states and of sojourn durations in these states. It also presents parametric and nonparametric estimation procedures, and ends with the question of the unobserved heterogeneity treatment in this kind of process.

The second section deals with inference for a particular class of transition processes, namely markovian processes or simple mixtures of markovian processes, using discrete-time panel observations. Here the main problem is the embeddability of the discrete-time Markov chain into a continuous time one. In other words, the question is whether or not the discrete-time panel observations of a transition process are generated by a continuous-time homogeneous Markov process. After a discussion of this problem, the second section presents maximum-likelihood and bayesian procedures for estimating the transition intensity matrix governing the evolution of the continuous-time markovian process. Particular attention is paid to the estimation of the continuous-time mover-stayer model, which is the more elementary model of mixed Markov processes.

The conclusion points out some extensions.

2 Multi-Spell Multi-State Models

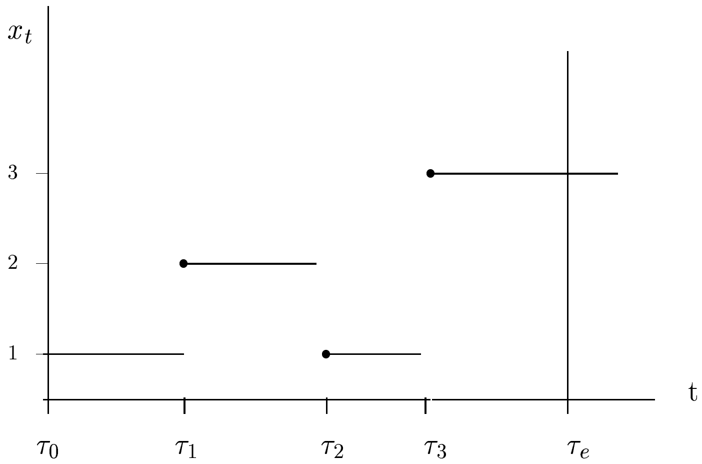
2.1 General framework

2.1.1 Notations

Let us consider a cadlag[1] stochastic process X_t, $t \in I\!R^+$, taking its value in a finite discrete-state space denoted $E = \{1, \ldots, K\}$, $K \in I\!N$ and $K \geq 2$. In other words, K represents the total number of states for the process, and X_t is the state occupied at time t by the individual (so $X_t \in E, \forall t \in I\!R^+$). Let

[1]"cadlag" means right-continuous, admitting left limits. For the definition of a cadlag process, see chapter 15, section II.1, this volume.

Figure 1



$\{x_t,\, t \in I\!R^+\}$ be a realization of this process. We suppose that all the individual realizations of this process are identically and independently distributed: to simplify the notations, we can then omit the index for individuals.

As an illustration we consider the case of a labor force participation process describing the state occupied by a worker at time t. In order to simplify, we set:

$$X_t = \begin{cases} 1 & \text{if the individual is employed at time t} \\ 2 & \text{if the individual is unemployed at time t} \\ 3 & \text{if the individual is out of the labour force at time t} \end{cases} \tag{1}$$

Now we suppose that each individual process is observed from the date of entry into the labor market, denoted τ_0 for the individual, up to an exogenously fixed time τ_e ($\tau_e > \tau_0$). An example of realization of process X_t is represented in Figure 1.

This figure shows that the individual is first employed from time τ_0 up to time τ_1, then unemployed from time τ_1 up to time τ_2, then employed once again from time τ_2 up to time τ_3, and finally out of the labor force (from time τ_3 on) when the observation stops at time τ_e. If we denote:

$$u_\ell = \tau_\ell - \tau_{\ell-1}\ ,\ \ \ell = 1, 2, \ldots \tag{2}$$

the sojourn duration in state $x_{\tau_{(\ell-1)}}$ reached by the individual at time $\tau_{(\ell-1)}$ (before a transition to state x_{τ_ℓ} at time τ_ℓ), the process x_t can be equivalently characterized by the sequences $\{ (\tau_\ell,\ x_{\tau_\ell}) \, ; \ell \in I\!N\}$ or $\{(u_\ell,\ x_{\Sigma_{k=0}^{\ell} u_k})_\ell \ ; \ \ell \in I\!N\}$ with $u_0 = \tau_0$.

Now suppose that process X_t is observed from the exogenous date τ_s, with $\tau_s \in]\tau_0,\ \tau_1[$, up to time τ_e and that the date of entry into the state occupied at time τ_s (i.e. the date of entry into the labor market, τ_0) is unknown to the analyst. Then, the sojourn duration in state $x_{\tau_s} = x_{\tau_0}$ is said to be *left-censored*. Symmetrically, for the example in Figure 1, the sojourn duration in state $x_{\tau_e} = x_{\tau_3}$ is said to be *right-censored*, because the couple $(\tau_4,\ x_{\tau_4})$ is not observed.

We restrict now our attention to non left-censored samples, i.e. such that $\tau_s \ = \ \tau_0$, for all individuals.[2] We define the *event-history* corresponding to process X_t for the observation period $[\tau_0,\ \tau_e]$ as:

$$\omega \ = \ \{\tau_0,\ x_{\tau_0},\ \tau_1,\ x_{\tau_1}, \ldots,\ \tau_n,\ x_{\tau_n}\} \tag{3}$$

where n is the number of transitions, i.e. the number of modifications, of the studied process during the period $[\tau_0,\ \tau_e]$. This event-history can be equivalently defined as:

$$\omega = \left\{\tau_0, u_1, x_{\tau_0+u_1}, u_2, x_{\tau_0+u_1+u_2}, \ldots, u_n), x_{\tau_0+\ \Sigma_{\ell=1}^{n} u_\ell}\right\} \tag{4}$$

This realization of the process from time τ_0 to time τ_e can be written:

$$\omega = ((\tau_0, x_{\tau_0}), (u_1, x_{\tau_1}), \ldots, (u_n, x_{\tau_n}), (u_{n+1}, 0)) \tag{5}$$

where $u_{n+1} = \tau_e - \tau_n$ is the duration of the last observed spell. The last spell is right-censored. Indeed, τ_{n+1} and x_{n+1} are not observed. Consequently, we fix $x_{n+1} = 0$ in order to signify that the last duration is at least equal to u_{n+1}. This realization of the process can be rewritten

$$\omega = (y_0, y_1, \ldots, y_n, y_{n+1}) \tag{6}$$

where

$$y_k = \begin{cases} (\tau_0, x_{\tau_0}) \text{ if } k = 0 \\ (\tau_k, x_{\tau_k}) \text{ if } 1 \leq k \leq n \\ (\tau_{n+1}, 0) \text{ if } k = n+1 \end{cases}$$

Let us define a spell as a period of time delimited by two successive transitions. The history of the process is a sequence of variables $y_k = (u_k, x_{\tau_k})$, where u_k is the length of spell k and x_{τ_k} is the state occupied by the individual at time τ_k.

[2]The statistical treatment of left-censored spells has been considered by Heckman and Singer (1984), Ondrich (1985) and Amemiya (2001).

2.1.2 Distributions of spell durations.

Suppose now that the process enters state $x_{\tau_{\ell-1}}$ ($x_{\tau_{\ell-1}} \in \{1, \ldots, K\}$) at time $\tau_{\ell-1}$ ($\ell = 1, \ldots, n+1$). Let us examine the probability distribution of the sojourn duration in state $x_{\tau_{\ell-1}}$ entered after the $(\ell-1)$-th transition of the process. For that purpose, we assume that this sojourn duration is generated by a conditional probability distribution P given the event-history $(y_0, \ldots, y_{\ell-1})$ and a vector of exogenous variables z, defined by the cumulative distribution function

$$\begin{aligned} F(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) &= \Pr\left[U_\ell \leq u \mid y_0, \ldots, y_{\ell-1}; z; \theta\right] \\ &= 1 - S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) \end{aligned} \tag{7}$$

where θ is a vector of unknown parameters. Here U_ℓ denotes the random variable corresponding to the duration of the $\ell - th$ spell of the process, starting with its $(\ell-1) - th$ transition. $S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta)$ is the *survivor function* of the sojourn duration in the $\ell - th$ spell. If the probability distribution P admits a density f with respect to the Lebesgue measure, then:

$$F(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \int_0^u f(t \mid y_0, \ldots, y_{\ell-1}; z; \theta)\ dt \tag{8}$$

and

$$\begin{aligned} f(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) &= \frac{d}{du} F(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) \\ &= -\frac{d}{du} S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) \end{aligned} \tag{9}$$

If the function $f(u \mid y_0, \ldots, y_{\ell-1}; z; \theta)$ is cadlag, then there exists a function, called the *hazard function* of the sojourn duration in the $\ell - th$ spell, defined as

$$\begin{aligned} h(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) &= \frac{f(u \mid y_0, \ldots, y_{\ell-1}; z; \theta)}{S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta)} \\ &= -\frac{d}{du} \log\ S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) \end{aligned} \tag{10}$$

or equivalently as

$$h(u|y_0, \ldots, y_{\ell-1}; z; \theta)\ du = \lim_{du \downarrow 0} \frac{\Pr\left[u \leq U_\ell < u + du \mid U_\ell \geq u; y_0, \ldots, y_{\ell-1}\right]}{du} \tag{11}$$

From (9), it follows that:

$$\begin{aligned} -\log\ S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) &= \int_0^u h(t \mid y_0, \ldots, y_{\ell-1}; z; \theta)\ dt \\ &= H(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) \end{aligned} \tag{12}$$

The function $H_\ell(u \mid y_0, \ldots, y_{\ell-1})$ is called the conditional *integrated hazard function* of the sojourn in the $\ell - th$ spell, given the history of the process up to time $\tau_{\ell-1}$.

Reduced-form statistical models of labour-market transitions can be viewed as extensions of competing risks duration models or multi-states multi-spells duration models. These concepts will now be specified.

2.1.3 Competing risks duration models

Let us assume that the number of states K is strictly greater than 2 ($K > 2$) and that, for each spell, there exists $(K-1)$ independent latent random variables, denoted $U^\star_{k,\ell}$ ($k \neq x_{\tau_{\ell-1}}$; $k \in E$). Each random variable $U^*_{k,\ell}$ represents the latent sojourn duration in state $x_{\tau_{\ell-1}}$ before a transition to state k ($k \neq x_{\tau_{\ell-1}}$) during the $\ell - th$ spell of the process.

The observed sojourn duration u_ℓ is the minimum of these $(K-1)$ latent durations:

$$u_\ell \;=\; \inf_{k \neq x_{\tau_{\ell-1}}} \left\{ u^*_{k,\ell} \right\} \tag{13}$$

Then, for any $\tau_{\ell-1} \in \omega$:

$$S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \prod_{\substack{k=1 \\ k \neq j}}^{K} S(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta) \tag{14}$$

where $S(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \Pr(U^*_{k,\ell} \geq u \mid y_0, \ldots, y_{\ell-1}; z)$ is the conditional survival function of the sojourn duration in state $x_{\tau_{\ell-1}}$ before a transition to state k during the $\ell - th$ spell of the process, given the history of the process up to time $\tau_{\ell-1}$.

Let $g(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta)$ be the conditional density function of the latent sojourn duration in state $x_{\tau_{\ell-1}}$ before a transition to state k, and $h_k(u \mid y_0, \ldots, y_{\ell-1}; z; \theta)$ the associated conditional hazard function. Then we have the relations:

$$h_k(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \frac{g(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta)}{S(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta)} \tag{15}$$

and

$$S(u, k \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \exp\left(- \int_0^u h_k(t \mid y_0, \ldots, y_{\ell-1}; z; \theta)\; dt \right) \tag{16}$$

Let us remark (14) and (16) imply:

$$S(u \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \exp\left(- \int_0^u \sum_{k \neq x_{\tau_{\ell-1}}} h_k(t \mid y_0, \ldots, y_{\ell-1}; z; \theta) \; dt \right) \tag{17}$$

Thus the conditional density function of the observed sojourn duration in state j during the $\ell - th$ spell of the process, given that this spell starts at time $\tau_{\ell-1}$ and ends at time $\tau_{\ell-1} + u$ by a transition to state k, is:

$$\begin{aligned} f(u, k \mid y_0, \ldots \quad y_{\ell-1}; z; \theta) &= h_k(u \mid y_0, \ldots, y_{\ell-1}; z; \theta), \\ &\times \; \exp\left(- \int_0^u \sum_{\substack{k'=1 \\ k' \neq x_{\tau_{\ell-1}}}}^{K} h_{k'}(t \mid y_0, \ldots, y_{\ell-1}; z; \theta) \; dt \right) \end{aligned} \tag{18}$$

This is the likelihood contribution of the $\ell - th$ spell when this spell is not right-censored (i.e. when $\tau_\ell = \tau_{\ell-1} + u \leq \tau_e$). When the $\ell - th$ spell lasts more than $\tau_e - \tau_{\ell-1}$, the contribution of this spell to the likelihood function is:

$$S(\tau_e - \tau_{\ell-1} \mid y_0, \ldots, y_{\ell-1}; z; \theta) = \Pr(U_\ell > \tau_e - \tau_{\ell-1} \mid y_0, \ldots, y_{\ell-1}; z)$$

2.1.4 Multi-spells multi-states duration models

These models are the extension of the preceding independent competing risks model, which treats the case of a single spell (the $\ell - th$ spell) with multiple destinations. In the multi-spells multi-states model, the typical likelihood contribution has the following form:

$$\mathcal{L}(\theta) = \prod_{\ell=1}^{n+1} f(y_\ell \mid y_0, \ldots, y_{\ell-1}; z; \theta) \tag{19}$$

where $f(y_\ell \mid y_0, \ldots, y_{\ell-1}; \theta)$ is the conditional density of Y_ℓ given $Y_0 = y_0, Y_1 = y_1, \ldots, Y_{\ell-1} = y_{\ell-1}, Z = z$ and θ is a vector of parameters. Definition (18) implies that:

$$\begin{aligned} \mathcal{L}(\theta) = \; & \prod_{\ell=1}^{n} f(\tau_\ell - \tau_{\ell-1}, x_{\tau_\ell} | y_0, \ldots, y_{\ell-1}; z; \theta) \\ & \times \; S_{n+1}(\tau_e - \tau_n | y_0, \ldots, y_n; z; \theta) \end{aligned} \tag{20}$$

The last term of the right-hand side product in (20) is the contribution of the last observed spell, which is right-censored. References for a general presentation of labor market transition econometric models can be found in surveys by Flinn and Heckman (1982a, b, 1983a) or in the textbook by Lancaster (1990a).

2.2 Nonparametric and parametric estimation

2.2.1 Nonparametric estimation

2.2.1.1 The Kaplan-Meier estimator In the elementary duration model, a nonparametric estimator of the survivor function can be obtained using the Kaplan-Meier estimator for right-censored data. Let us suppose that we observe I sample paths (i.i.d. realizations of the process X_t) with the same past history $\omega[\tau_0, \tau_{n-1}]$. Let $I^\star$ be the number of sample paths such that $\tau_{n,i} \leq T_2$ and $I - I^\star$ the number of sample paths for which the n-th spell duration is right-censored, i.e. $\tau_{n,i} > T_2$ (or $n(\tau_0, T_2) < n$), i denoting here the index of the process realization ($i = 1, \ldots, I$). If $\tau_{n,1}, \ldots, \tau_{n,I^\star}$ are the $I^\star$ ordered transition dates from state $X_{\tau_{n-1}}$ (i.e. $\tau_{n,1} \leq \ldots \leq \tau_{n,I^\star} \leq T_2$), the Kaplan-Meier estimator of the survivor function $S_n(t \mid \omega[\tau_0, \tau_{n-1}])$ is:

$$\begin{aligned} \hat{S}_n(t \mid \omega[\tau_0, \tau_{n-1}]) &= \prod_{i:\tau_{n,i} \leq t} \left(1 - \frac{d_i}{r_i}\right) \\ & i = 1, \ldots, I^\star, \ t \in]\tau_{n-1}, T_2] \end{aligned} \tag{21}$$

where r_i is the number of sample paths for which the transition date from state $X_{\tau_{n-1}}$ is greater than or equal to $\tau_{n,i}$ and d_i is the number of transition times equal to $\tau_{n,i}$. An estimator for the variance of the survivor function estimate is given by the Greenwood's formula:

$$\begin{aligned} & \mathrm{Var}\left[\hat{S}_n(t \mid \omega[\tau_0, \tau_{n-1}])\right] \\ & \quad \simeq \left\{\hat{S}_n(t \mid \omega[\tau_0, \tau_{n-1}])\right\}^2 \times \sum_{i:\tau_{n,i} \leq t} \frac{d_i}{r_i(r_i - d_i)} \end{aligned} \tag{22}$$

This estimator allows to implement nonparametric tests for the equality of the survivor functions of two different subpopulations (such as the Savage and log-rank tests).

In the case of multiple destinations (i.e. competing risks models), we must restrict the set of sample paths indexed by $i \in \{1, \ldots, I^\star\}$ to the process realizations experiencing transitions from the state $X_{\tau_{n-1}}$ to some state k ($k \neq X_{\tau_{n-1}}$). Transitions to another state than k are considered as right-censored durations. If we set $X_{\tau_{n-1}} = j$, then the Kaplan-Meier estimator of the survivor function $S_{jk}(t \mid \omega[\tau_0, \tau_{n-1}])$ is given by the appropriate application of formula (21), and an estimator of its variance is given by formula (22).

2.2.1.2 The Aalen estimator The function $H_\ell(u \mid \omega[\tau_0, \tau_{\ell-1}])$, defined in equation (12) and giving the integrated hazard function of the sojourn duration in the $\ell - th$ spell, can be estimated nonparametrically using the Aalen estimator (Aalen, 1978):

$$\hat{H}_\ell(u \mid \omega[\tau_0, \tau_{\ell-1}]) = \sum_{i:\tau_{\ell-1} \le \tau_{\ell,i} < u} \frac{d_i}{r_i} \tag{23}$$

$\hat{H}_\ell(u \mid \omega[\tau_0, \tau_{\ell-1}])$ is an unbiased estimator of $H_\ell(u \mid \omega[\tau_0, \tau_{\ell-1}])$, and an estimator of its variance is given by:

$$var\left[\hat{H}_\ell(u \mid \omega[\tau_0, \tau_{\ell-1}])\right] = \sum_{i:\tau_{\ell-1} \le \tau_{\ell,i} < u} \frac{d_i}{r_i(r_i - d_i)} \tag{24}$$

In the competing risks model, equation (12) is equivalent to:

$$\begin{aligned} -\log\ S_{jk}(u \mid \omega[\tau_0, \tau_{\ell-1}]) &= \textstyle\int_0^u h_{jk}(t \mid \omega[\tau_0, \tau_{\ell-1}])\ dt \\ &= H_{jk}(u \mid \omega[\tau_0, \tau_{\ell-1}]) \end{aligned} \tag{25}$$

where $H_{jk}(u \mid \omega[\tau_0, \tau_{\ell-1}])$ is the integrated intensity (or hazard) function for a transition from state j to state k ($k \neq j$) during the $\ell - th$ spell of the process, and given the past history $\omega[\tau_0, \tau_{\ell-1}]$ of the process. The Aalen estimator of this function can be derived from the formula (24) by considering indexes i corresponding to transitions from state j to state k during the $\ell - th$ spell of the process; indexes corresponding to other types of transition from state j are now considered as right-censored durations. The Aalen estimator can be used to implement nonparametric tests for the equality of two or more transition intensities corresponding to distinct transitions.

2.2.2 Specification of conditional hazard functions

2.2.2.1 The Markov model In a markovian model, the hazard functions $h_k(t \mid y_0, \ldots, y_{\tau_{\ell-1}}; z; \theta)$ depend on t, on states $x_{\tau_{\ell-1}}$ and on k, but are independent of the previous history of the process. More precisely:

$$h_k(t \mid y_0, \ldots, y_{\tau_{\ell-1}}; z; \theta) = h_k(t \mid x_{\tau_{\ell-1}}; z; \theta), \quad k \neq x_{\tau_{\ell-1}} \tag{26}$$

and

$$h_j(t \mid y_0, \ldots, y_{\tau_{\ell-1}}; z; \theta) = 0, \quad \text{if } j = x_{\tau_{\ell-1}}$$

When the Markov model is time-independent, it is said to be time-homogeneous. In this case:

$$h_k(t \mid x_{\tau_{\ell-1}}; z; \theta) = h_k(x_{\tau_{\ell-1}}; z; \theta) = h_{x_{\tau_{\ell-1}},k}(z; \theta)\ ,\ k \neq x_{\tau_{\ell-1}},\ \forall t \in I\!R^+ \tag{27}$$

The particular case of a continuous-time markovian model observed in discrete-time will be extensively treated in the following subsection (this Chapter). Let us now consider two simple examples of markovian processes.

Example 1:

Consider the case of a time-homogeneous markovian model with two states ($K = 2$) and assume that:

$$h_k(t \mid x_{\tau_{\ell-1}}; \theta) = \begin{cases} \alpha \ \text{ if } x_{\tau_{\ell-1}} = 1 \text{ and } k = 2 \\ \beta \ \text{ if } x_{\tau_{\ell-1}} = 2 \text{ and } k = 1 \\ 0 \text{ otherwise} \end{cases} \tag{28}$$

with $\theta = (\alpha, \beta)$. The parameter $\alpha > 0$ is the instantaneous rate of transition from state 1 (for instance, the employment state) to state 2 (for instance, the unemployment state). Reciprocally, $\beta > 0$ is the instantaneous rate of transition from state 2 to state 1.

Durations of employment (respectively, unemployment) are independently and identically distributed according to an exponential distribution with parameter α (respectively, with parameter β). If $p_1(t_0)$ and $p_2(t_0)$ denote occupation probabilities of states 1 and 2 at time t_0 respectively, then occupation probabilities at time t ($t > t_0$) are respectively defined by:

$$\begin{aligned} p_1(t) = &\ \frac{\beta}{\alpha+\beta} + \left\{ p_1(t_0) - \frac{\beta}{\alpha+\beta} \right\} e^{-(\alpha+\beta)t} \\ p_2(t) = &\ \frac{\alpha}{\alpha+\beta} + \left\{ p_2(t_0) - \frac{\alpha}{\alpha+\beta} \right\} e^{-(\alpha+\beta)t} \end{aligned} \tag{29}$$

Let $(p_1^\star,\ p_2^\star)$ denote the stationary probability distribution of the process. Then it is easy to verify from (29) that:

$$p_1^\star = \frac{\beta}{\alpha+\beta} \ \ and \ \ p_2^\star = \frac{\alpha}{\alpha+\beta} \tag{30}$$

In the economic literature, there are many examples of stationary job search models generating such a markovian time-homogeneous model with two states (employment and unemployment): see, for instance, the survey by Mortensen (1986). Extensions to three-states models (employment, unemployment and out-of-labor-force states) have been considered, for example, by Flinn and Heckman (1982a) and Burdett et al. (1984a, b). Markovian models of labor mobility have been estimated, for instance, by Tuma and Robins (1980), Flinn and Heckman (1983b), Mortensen and Neuman (1984), Olsen, Smith and Farkas (1986) and Magnac and Robin (1994).

Example 2:

Let us consider now the example of a non-homogeneous markovian model with two states (employment and unemployment, respectively denoted 1 and 2). Let us assume that the corresponding conditional hazard functions verify

$$h_k(t \mid x_{\tau_{\ell-1}};\theta) = \begin{cases} h_2(t;\theta) & \text{if } x_{\tau_{\ell-1}} = 1 \text{ and } k=2 \\ h_1(t;\theta) & \text{if } x_{\tau_{\ell-1}} = 2 \text{ and } k=1 \\ 0 \text{ otherwise} & \end{cases} \tag{31}$$

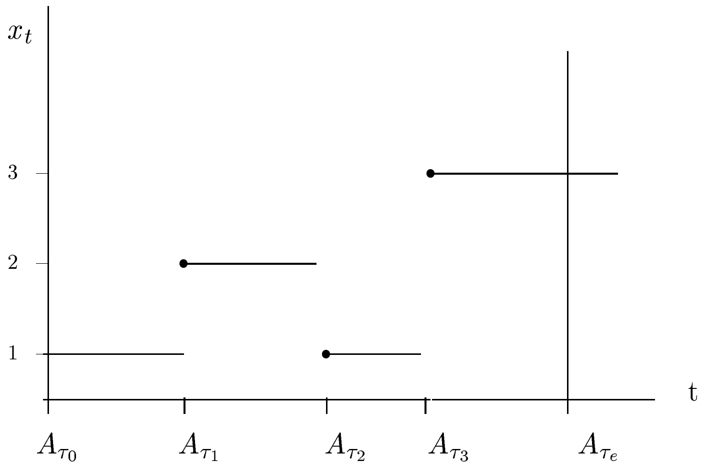
Let $p(0) = (p_1(t_0), p_2(t_0))'$ denote the initial probability distribution at time t_0. The distribution of state occupation probabilities at time t, denoted $p(t) = (p_1(t), p_2(t))'$, is given by:

$$\begin{aligned} p_1(t) = \quad & \exp\left\{ -\int_{t_0}^t [h_1(s;\theta)+h_2(s;\theta)]\ ds \right\} \\ & \times\ \left[p_1(t_0) + \int_{t_0}^t h_1(s;\theta)\ \exp\left\{ \int_{t_0}^s (h_1(u;\theta)+h_2(u;\theta))\ du \right\}\ ds \right] \end{aligned} \tag{32}$$

and $p_2(t) = 1 - p_1(t)$ (see Chesher and Lancaster, 1983).

Non-homogeneous markovian models are often used to deal with processes mainly influenced by the individual age at the transition date. For example, let us consider a transition process $\{X_t\}_{t\geq 0}$ with state-space $E = \{1,2,3\}$, and for which the time scale is the age (equal to A_t at time t). If the origin date of the process (i.e. the date of entry into the labor market) is denoted A_{τ_0} for a given individual, then a realization of the process $\{X_t\}_{t\geq 0}$ over the period $[A_{\tau_0},\ \tau_e]$ is depicted in Figure 2.

Figure 2



Now let us suppose that transition intensities at time t depend only on the age attained at this time and are specified such as:

$$\begin{aligned} h_k\,(t \mid y_0, \ldots, y_{\ell-1}; A_{\tau_0}; \theta) &= h_k(A_t; x_{\tau_{\ell-1}}; \theta) \\ &= \exp\left(\alpha_{x_{\tau_{\ell-1}},k} + \beta_{x_{\tau_{\ell-1}},k}\, A_t\right) \end{aligned} \tag{33}$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ ($j, k \in E \times E$ and $k \neq j$) are parameters to be estimated. In formula (33), the individual index is omitted for simplifying notations. By noting that:

$$A_t = A_{\tau_{\ell-1}} + (A_t - A_{\tau_{\ell-1}}) = A_{\tau_{\ell-1}} + u_{t\ell} \tag{34}$$

where $u_{t\ell}$ denotes the time already spent in the $\ell - th$ spell at date t, it is possible to write again transition intensities as:

$$h_k(t \mid y_0, \ldots, y_{\ell-1}; A_{\tau_0}; \theta) = \exp\left(\alpha_{x_{\tau_{\ell-1}},k} + \beta_{x_{\tau_{\ell-1}},k}\, A_{\tau_{\ell-1}} + \beta_{x_{\tau_{\ell-1}},k}\, u_{t\ell}\right) \tag{35}$$

and to deduce the survivor function of the sojourn duration in the $\ell - th$ spell which has the form:

$$\begin{aligned} &S(u \mid y_0, \ldots, y_{\ell-1}; A_{\tau_0}; \theta) \\ &\quad = \exp\{-\sum_{k \neq x_{\tau_{\ell-1}}} \int_{A_{\tau_{\ell-1}}}^{A_{\tau_{\ell-1}}+u} \exp(\alpha_{x_{\tau_{\ell-1}},k} + \beta_{x_{\tau_{\ell-1}},k}\, A_{\tau_{\ell-1}} + \beta_{x_{\tau_{\ell-1}},k}\, u_{t\ell})\; dt\,\} \end{aligned} \tag{36}$$

where $\ell \geq 1$. By setting $u_{t\ell} = t - A_{\tau_{\ell-1}}$ in expression (36), it follows that:

$$S\{u \mid y_0, \ldots, y_{\ell-1}; A_{\tau_0}; \theta\} =$$
$$\exp\left(-\sum_{k \neq x_{\tau_{\ell-1}}} \frac{\exp(\alpha_{x_{\tau_{\ell-1}},k})}{\beta_{x_{\tau_{\ell-1}},k}} \left[\exp(\beta_{x_{\tau_{\ell-1}},k}\,(A_{\tau_{\ell-1}}+u)) - \exp(\beta_{x_{\tau_{\ell-1}},k}\,A_{\tau_{\ell-1}}) \right] \right) \tag{37}$$

if $\beta_{x_{\tau_{\ell-1}},k} \neq 0$. Then the likelihood contribution of the $\ell - th$ spell beginning at age $A_{\tau_{\ell-1}}$ with a transition to state $x_{\tau_{\ell-1}}$ and ending at age A_{τ_ℓ} with a transition to state x_{τ_ℓ} is:

$$\begin{aligned} L_\ell &= f(A_{\tau_\ell} - A_{\tau_{\ell-1}}, x_{\tau_\ell} \mid y_0, \ldots, y_{\tau_{\ell-1}}; A_{\tau_0}; \theta) \\ &= h_{x_{\tau_\ell}}(\tau_\ell \mid y_0, \ldots, y_{\tau_{\ell-1}}; A_{\tau_0}; \theta)\; S(A_{\tau_\ell} - A_{\tau_{\ell-1}} \mid y_0, \ldots, y_{\tau_{\ell-1}}; A_{\tau_0}; \theta) \\ &= \exp\left(\alpha_{x_{\tau_{\ell-1}},k} + \beta_{x_{\tau_{\ell-1}},k}\,A_{\tau_\ell}\right) \\ &\times \exp\left(-\sum_{k' \neq x_{\tau_{\ell-1}}} \frac{\exp(\alpha_{x_{\tau_{\ell-1}},k'})}{\beta_{x_{\tau_{\ell-1}},k'}} \left[\exp(\beta_{x_{\tau_{\ell-1}},k'}\,A_{\tau_\ell}) - \exp(\beta_{x_{\tau_{\ell-1}},k'}\,A_{\tau_{\ell-1}}) \right] \right) \end{aligned} \tag{38}$$

Non-homogeneous markovian models of transitions between employment and unemployment have been estimated, for example, by Ridder (1986) and Trivedi and Alexander (1989).

2.2.2.2 Semi-Markov models In semi-Markov models, hazard functions depend only on the currently occupied state (denoted $x_{\tau_{\ell-1}}$ for spell ℓ), on the destination state (denoted k), on the sojourn duration in state $x_{\tau_{\ell-1}}$ and on the time of entry into the currently occupied state. If the spell corresponding to the currently occupied state is the $\ell - th$ spell of the process, then hazard functions of the semi-Markov model have two alternative representations:

$$h_k(t \mid y_0, \ldots, y_{\ell-1}; \theta) = h_k(t \mid \tau_{\ell-1}; x_{\tau_{\ell-1}}; \theta) \tag{39}$$

or

$$h_k(u \mid y_0, \ldots, y_{\ell-1}; \theta) = h_k(u \mid \tau_{\ell-1}; x_{\tau_{\ell-1}}; \theta) \tag{40}$$

where $u = t - \tau_{\ell-1}$ is the time already spent in the current state (i.e. in the $\ell - th$ spell of the process). When the hazard functions do not depend on the date $\tau_{\ell-1}$ of the last event, but depend only on the time already spent in the current state, then the semi-Markov model is said to be time-homogeneous. In this case, hazard functions defined in (40) are such that:

$$h_k(u \mid \tau_{\ell-1}; x_{\tau_{\ell-1}}; \theta) = h_k(u \mid x_{\tau_{\ell-1}}; \theta),\ u \in \mathbb{R}^+ \tag{41}$$

In this model, the mean duration of a sojourn in state $x_{\tau_{\ell-1}}$ can be calculated using definitions of hazard and survivor functions, and thus it is given by:

$$E(U_\ell \mid x_{\tau_{\ell-1}}; \theta) = \int_0^\infty u\ S(u \mid x_{\tau_{\ell-1}}; \theta) \Big\{ \sum_{k \neq x_{\tau_{\ell-1}}} h_k(u \mid x_{\tau_{\ell-1}}; \theta) \Big\}\ du \tag{42}$$

where U_ℓ is the random variable representing the duration of a spell ℓ and

$$S(u \mid x_{\tau_{\ell-1}}; \theta) = \exp(-\int_0^u \sum_{k \neq x_{\tau_{\ell-1}}} h_k(s \mid x_{\tau_{\ell-1}}; \theta)\ d\,s) \tag{43}$$

This conditional expectation can be obtained using the following property:

$$E(U_\ell \mid x_{\tau_{\ell-1}}; \theta) = \int_0^\infty S(u \mid x_{\tau_{\ell-1}}; \theta)\ d\,u, \tag{44}$$

(see, for instance, Klein and Moeschberger, 2003). Semi-markovian models of transition between two or three states have been estimated by Flinn and Heckman (1982b), Burdett, Kiefer and Sharma (1985), Bonnal, Fougère and Sérandon (1997), and Gilbert, Kamionka and Lacroix (2001).

2.3 Unobserved Heterogeneity

Here heterogeneity is supposed to cover individual observable and unobservable characteristics. Once again, we will omit the individual index.

2.3.1 Correlation between spells

Let us assume that the conditional model is time-homogeneous semi-markovian and

$$h_k(u \mid y_0, \ldots, y_{\ell-1}; z; v; \theta) = h_k(u_\ell \mid x_{\tau_{\ell-1}}; z; v_{x_{\tau_{\ell-1}},k}; \theta_{x_{\tau_{\ell-1}},k}) \tag{45}$$

where v is a vector of individual unobserved heterogeneity terms and θ is the vector of parameters to be estimated.

Let $h_k(u_\ell \mid x_{\tau_{\ell-1}}; z; v_{x_{\tau_{\ell-1}},k}; \theta_{x_{\tau_{\ell-1}},k})$ denote the conditional hazard function for the sojourn duration in the $\ell - th$ spell of the participation process, when the currently occupied state is state $x_{\tau_{\ell-1}}$ and the destination state is k. Here z is a vector of exogenous variables, possibly time-dependent, $v_{(j,k)}$ is an heterogeneity random term, which is unobserved, and λ_{jk} is a vector of parameters. The preceding hazard function is often supposed to be equal to:

$$h_k(u_\ell \mid x_{\tau_{\ell-1}}; z, v_{x_{\tau_{\ell-1}},k}, \theta_{x_{\tau_{\ell-1}},k}) = \exp\left[\varphi(z; u_\ell; \theta_{x_{\tau_{\ell-1}},k}) + v_{x_{\tau_{\ell-1}},k}\right] \quad (46)$$

Several assumptions can be made concerning the unobserved random terms $v_{j,k}$. Firstly, $v_{j,k}$ can be supposed to be specific to the transition from j to k, so

$$v_{j,k} \neq v_{j',k'} \quad \text{for any } (j,k) \neq (j',k').$$

It can be also specific to the origin state, in which case :

$$v_{j,k} = v_j \quad \text{for any } k \neq j.$$

Finally, $v_{j,k}$ can be supposed to be independent of states j and k and thus to be fixed over time for each individual, i.e.

$$v_{j,k} = v \quad \text{for any } (j,k) \in E \times E,\ k \neq j.$$

This last assumption will be made through the remaining part of our presentation. Let us remark that a fixed heterogeneity term is sufficient to generate some correlation between spells durations. If we assume that v has a probability density function with respect to the Lebesgue measure denoted $g(v \mid \alpha)$, where α is a parameter, then we can deduce that the marginal survivor function of the sojourn duration in the $\ell - th$ spell of the process, when current state is $x_{\tau_{\ell-1}}$, has the form:

$$\begin{aligned} S(u_\ell \mid x_{\tau_{\ell-1}}; z; \theta_{x_{\tau_{\ell-1}}}) &= \int_{D_G} S(u_\ell \mid x_{\tau_{\ell-1}}; z; v; \theta_{x_{\tau_{\ell-1}}})\ g(v \mid \alpha)\, dv \\ &= \int_{D_G} \exp\left\{ -\exp(v) \int_0^{u_\ell} \left(\sum_{k \neq x_{\tau_{\ell-1}}} \exp(\varphi(z; t; \theta_{x_{\tau_{\ell-1}},k})\right) dt \right\}\ g(v \mid \alpha)\, dv \end{aligned} \quad (47)$$

where $\theta_{x_{\tau_{\ell-1}}} = \left\{(\theta_{x_{\tau_{\ell-1}},k})_{k \neq x_{\tau_{\ell-1}}}, \alpha\right\}$ and D_G is the support of the probability distribution of the random variable v.

Such formalizations of heterogeneity have been used for estimation purposes by Heckman and Borjas (1980), Butler et al. (1986, 1989), Mealli and

Pudney (1996), Bonnal, Fougère and Sérandon (1997), Gilbert, Kamionka and Lacroix (2001), and Kamionka and Lacroix (2003).

• Example

To illustrate the treatment of unobserved heterogeneity in transition processes, let us consider a realization of a two state time-homogeneous Markov process. More precisely, let us assume that this realization generates a complete spell in state 1 over the interval $[0, \tau_1]$ and a right-censored spell in state 2 over the interval $[\tau_1, \tau_e[$. Transition intensities between the two states are given by:

$$h_k(t \mid x_{\tau_{\ell-1}}; v_{x_{\tau_{\ell-1}}}; \lambda_{x_{\tau_{\ell-1}}}) = \lambda_{x_{\tau_{\ell-1}}} + v_{x_{\tau_{\ell-1}}} \tag{48}$$

where $k \in \{1, 2\}$, $\lambda_{x_{\tau_{\ell-1}}} > 0$ and $t \in I\!R^+$, λ_1 and λ_2 are two positive parameters, and v_1 and v_2 are two random variables supposed to be exponentially distributed with a density function $g(v \mid \alpha) = \alpha \ \exp(-\alpha \ v)$, $\alpha > 0$. We want to deduce the likelihood function for this realization of the process when v_1 and v_2 are supposed to be spell-specific and independent ($v_1 \neq v_2$ and $v_1 \perp\!\!\!\perp v_2$) or fixed over time ($v_1 = v_2 = v$). In the first case ($v_1 \neq v_2$ and $v_1 \perp\!\!\!\perp v_2$), the conditional likelihood function is:

$$\begin{aligned} L_v(\lambda) &= f(\tau_1, x_{\tau_1} \mid x_0; v; \lambda) \ S(\tau_e - \tau_1 \mid x_{\tau_1}; v; \lambda), \\ &= (\lambda_1 + v_1) \ \exp\{-(\lambda_1 + v_1)\tau_1\} \ \exp\{-(\lambda_2 + v_2)(\tau_e - \tau_1)\} \end{aligned} \tag{49}$$

where $v = (v_1, v_2)'$, $\lambda = (\lambda_1, \lambda_2)'$, $x_0 = 1$ and $x_{\tau_1} = 2$. Because v_1 and v_2 are unobserved, we must deal with the following marginalized likelihood function:

$$\begin{aligned} L(\alpha; \lambda) &= \int_0^\infty \int_0^\infty L(v_1, v_2, \lambda_1, \lambda_2) \ g(v_1 \mid \alpha) \ g(v_2 \mid \alpha) \ d v_1 \ d v_2 \\ &= f(\tau_1, x_{\tau_1} \mid x_0; \alpha; \lambda) \ S(\tau_e - \tau_1 \mid x_1; \alpha; \lambda) \end{aligned} \tag{50}$$

where

$$\begin{aligned} & f(\tau_1, x_{\tau_1} \mid x_0; \alpha; \lambda) = \exp(-\lambda_1 \ \tau_1) \left(\frac{\alpha}{\tau_1 + \alpha}\right) \left(\lambda_1 + \frac{1}{\tau_1 + \alpha}\right) \\ \text{and} \quad & S(\tau_e - \tau_1 \mid x_{\tau_1}; \alpha; \lambda) = \exp(-\lambda_2 \ (\tau_e - \tau_1)) \ \left(\frac{\alpha}{(\tau_e - \tau_1) + \alpha}\right) \end{aligned} \tag{51}$$

are the marginalized density and survivor functions of sojourn durations τ_1 and $(\tau_e - \tau_1)$ in the first and second spells respectively.

When the heterogeneity term is fixed over time ($v_1 = v_2 = v$), then the marginal likelihood contribution is:

$$\begin{aligned} L(\alpha, \lambda) &= \int_0^{\infty} (\lambda_1 + v) \ \exp\{-(\lambda_1 \tau_1 + \lambda_2 (\tau_e - \tau_1) + v\, \tau_e)\}\, \alpha \ \exp(-\alpha\, v) \ d\, v, \\ &= \exp\{-\lambda_1 \tau_1 - \lambda_2(\tau_e - \tau_1)\} \frac{\alpha}{\alpha + \tau_e} \left\{ \lambda_1 + \frac{\alpha}{\alpha + \tau_e} \right\} \end{aligned} \tag{52}$$

which is obviously not equal to the product of the marginalized density and survivor functions of the sojourn durations in the first and second spells as in the case where $v_1 \neq v_2$. $\star$

Now, let us assume that there exists a function ψ defining a one-to-one relation between v and some random variable ν, such as:

$$v = \psi(\nu, \alpha) \tag{53}$$

For instance, ψ can be the inverse of the c.d.f. for v, and ν can be uniformly distributed on $[0, 1]$. Then:

$$S(u_\ell \mid x_{\tau_{\ell-1}}; z; \theta_{x_{\tau_{\ell-1}}}) = \int_0^1 S(u_\ell \mid x_{\tau_{\ell-1}}; z; \psi(\nu, \alpha); \theta_{x_{\tau_{\ell-1}}}) \ \phi(\nu) \ d\nu \tag{54}$$

where $\phi(.)$ is the density function of ν. The marginal hazard function for the sojourn in the $\ell - th$ spell can be deduced from equation (54) as:

$$h(u_\ell \mid x_{\tau_{\ell-1}}; z; \theta_{x_{\tau_{\ell-1}}}) = -\frac{d}{du_\ell} S(u_\ell \mid x_{\tau_{\ell-1}}; z; \theta_{x_{\tau_{\ell-1}}}) \tag{55}$$

Using definitions (54) and (55), the individual contribution to the likelihood function can be easily deduced and maximized with respect to θ, either by usual procedures of likelihood maximization if the integrals (40) and (41) can be easily calculated, or by simulation methods (see, e.g., Gouriéroux and Monfort, 1997) in the opposite case.

For instance, let us consider the case of a semi-markovian model where the individual heterogeneity term is fixed over time, i.e. $v_{j,k} = v$ for any $(j, k) \in E \times E$. From (20) and (46)-(47), the typical likelihood contribution in the present case is:

$$\begin{aligned} L_v(\theta) &= \prod_{\ell=1}^{n} h_{x_{\tau_\ell}} \left(\tau_\ell - \tau_{\ell-1} \mid x_{\tau_{\ell-1}}; z; v; \theta_{x_{\tau_{\ell-1}}, x_{\tau_\ell}} \right) \\ &\times \prod_{\ell=1}^{n+1} \exp \left\{ - \int_{\tau_{\ell-1}}^{\tau_\ell} \sum_{k \neq x_{\tau_{\ell-1}}} h_k(t \mid x_{\tau_{\ell-1}}; z; v; \theta_{x_{\tau_{\ell-1}}, k}) \ dt \right\} \end{aligned} \tag{56}$$

with $\tau_{n+1} = \tau_e$ by convention. Using relation (53), the marginalized likelihood contribution obtained by integrating out ν is:

$$\mathcal{L}(\theta) = \int_0^1 L_{\psi(\nu,\alpha)}(\theta) \ \phi(\nu) \ d\nu \tag{57}$$

When the integral is not analytically tractable, simulated ML estimators of parameters α and $(\theta_{jk})_{k\neq j}$ can be obtained by maximizing the following simulated likelihood function with respect to α and $(\theta_{jk})_{k\neq j}$:

$$L_N(\theta) = \frac{1}{N} \sum_{n=1}^{N} L_{\psi(\nu_n,\alpha)}(\theta) \tag{58}$$

where ν_n is drawn from the distribution with density function $\phi(.)$, which must be conveniently chosen (for asymptotic properties of these estimators, see Gouriéroux and Monfort, 1997).

2.3.2 Correlation between destination states

Let us assume that the conditional hazard function for the transition into state k is given by the expression

$$h_k(u \mid y_0, \ldots, y_{\ell-1}; z; v; \lambda) = h_k^0(u;\gamma) \ \varphi(y_0, \ldots, y_{\ell-1}; z; \beta) \ \zeta_k \tag{59}$$

where $\varphi(.)$ is a positive function depending on the exogenous variables and the history of the process, ζ_k an unobserved heterogeneity component specific to the individual ($\zeta_k > 0$), β and γ are vectors of parameters, $h_k^0(u;\gamma)$ is a baseline hazard function for the transition to state k ($k \in \{1, \ldots, K\}$). Let us assume that (see Gilbert et al., 2001)

$$\zeta_k = \exp(a_k \ v_1 + b_k \ v_2) \tag{60}$$

where a_k and b_k are parameters such that $a_k = 1\!\!1[k \geq 2]$ for $k = 1, \ldots, K$ and $b_1 = 1$. The latent components v_1 and v_2 are assumed to be independently and identically distributed with a p.d.f. denoted $g(v;\alpha)$, where α is a parameter and $v_s \in D_G$, $s = 1, 2$.

In this two factor loading model, the correlation between $log(\zeta_k)$ and $log(\zeta_{k'})$, $\rho_{k,k'}$, is given by the expression

$$\rho_{k,k'} = \frac{a_k \ a_{k'} + b_k \ b_{k'}}{\sqrt{a_k^2 + b_k^2}\sqrt{a_{k'}^2 + b_{k'}^2}} \tag{61}$$

where $k, k' = 1, \ldots, K$. The contribution to the conditional likelihood function of a given realization of the process $w = (y_1, \ldots, y_n, y_{n+1})$ is:

$$\mathcal{L}(\theta) = \int_{D_G} \int_{D_G} \prod_{\ell=1}^{n+1} f(y_\ell | y_0, \ldots, y_{\ell-1}; z; v_1, v_2; \lambda) \; g(v_1; \alpha) \; g(v_2; \alpha) \; dv_1 \; dv_2 \tag{62}$$

where

$$\begin{aligned} f(u, k \mid y_0, \ldots, y_{\ell-1}; z; v_1, v_2; \lambda) &= h_k(u \mid y_0, \ldots, y_{\ell-1}; z; v_1, v_2; \lambda)^{\delta_k} \\ &\times \exp\{-\int_0^u \sum_{j \neq x_{\tau_{\ell-1}}} h_j(t \mid y_0, \ldots, y_{\ell-1}; z; v_1, v_2; \lambda) \, dt\} \end{aligned} \tag{63}$$

and the conditional hazard function is given by expression (59). The exponent δ_k is equal to 1 if $k \in \{1, \ldots, K\}$, and to 0 otherwise. λ is a vector of parameters and $\theta = (\alpha, \lambda)$. As the last spell is right-censored, the corresponding contribution of this spell is given by the survivor function

$$f(y_{n+1} | y_0, \ldots, y_n; z; v_1, v_2; \lambda) = \exp\{-\int_0^{u_{n+1}} \sum_{j \neq x_{\tau_n}} h_j(t | y_0, \ldots, y_n; z; v_1, v_2; \lambda) \, dt\} \tag{64}$$

where $y_{n+1} = (u_{n+1}, 0)$ (state 0 corresponds to right-censoring).

Bonnal et al. (1997) contains an example of a two factor loading model. Lindeboom and van den Berg (1994), Ham and Lalonde (1996) and Eberwein et al. (1997, 2002) use a one factor loading model in order to correlate the conditional hazard functions. A four factor loading model has been proposed by Mealli and Pudney (2003). Let us remark that, in the case of bivariate duration models, association measures were studied by Van den Berg (1997). Discrete distributions of the unobserved heterogeneity component can be alternatively used (see, for instance, Heckman and Singer (1984), Gritz (1993), Baker and Melino (2000)).

This way to correlate the transition rates using a factor loading model is particularly useful for program evaluation on nonexperimental data. In this case, it is possible to characterize the impact on the conditional hazard functions of previous participation to a program by taking into account entry selectivity phenomena.

3 Markov Processes Using Discrete-Time Observations

The econometric literature on labor mobility processes observed with discrete-time panel data makes often use of two elementary stochastic processes describing individual transitions between a finite number of participation states.

The first one is the continuous-time Markov chain, whose parameters can be estimated through the quasi-Newton (or scoring) algorithm proposed by Kalbfleisch and Lawless (1985). This kind of model allows to calculate stationary probabilities of state occupation, the mean duration of sojourn in a given state, and the intensities of transition from one state to another.

A main difficulty can appear in this approach: in some cases the discrete-time Markov chain cannot be represented by a continuous-time process. This problem is known as the embeddability problem which has been surveyed by Singer and Spilerman (1976a, b) and Singer (1981, 1982). However, some non-embeddable transition probability matrices can become embeddable after an infinitesimal modification complying with the stochastic property. This suggests that the embeddability problem can be due to sampling errors.

Geweke et al. (1986a) established a bayesian method to estimate the posterior mean of the 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.

The second frequently used modelization incorporates a very simple form of heterogeneity among the individuals: this is the mover-stayer model, which was studied in the discrete-time framework by Frydman (1984), Sampson (1990) and Fougère and Kamionka (2003). The mover-stayer model 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" moving between states according to a non-degenerate Markov process.

These two modelizations will be successively studied in the following subsection.

3.1 The time-homogeneous markovian model

Let us consider a markovian process $\{X_t,\ t \in I\!R^+\}$ defined on a discrete state-space $E = \{1, \ldots, K\}$, $K \in I\!N$, with a transition probability matrix $P(s,t)$ with entries $p_{j,k}(s,t)$, $(j,k) \in E \times E$, $0 \leq s \leq t$, where:

$$p_{j,k}(s,t) = \Pr\{X_t = k \mid X_s = j\} \tag{65}$$

and $\sum_{k=1}^{K} p_{j,k}(s,t) = 1$. If this markovian process is time-homogeneous, then:

$$p_{j,k}(s,t) = p_{j,k}(0, t-s) \equiv p_{j,k}(t-s), \quad 0 \leq s \leq t \tag{66}$$

or equivalently:

$$P(s,t) = P(0, t-s) \equiv P(t-s), \quad 0 \leq s \leq t \tag{67}$$

This implies that transition intensities defined by:

$$h_{j,k} = \lim_{\Delta t \downarrow 0} p_{j,k}(t, t+\Delta t)/\Delta t, \quad \Delta t \geq 0,\ (j,k) \in E \times E,\ j \neq k \tag{68}$$

are constant through time, i.e.:

$$h_k(t \mid x_{\tau_{\ell-1}}; \theta) = h_{j,k}(t \mid \theta) = h_{j,k}\ ,\ t \geq 0,\ (j,k) \in E \times E,\ j \neq k \tag{69}$$

where $x_{\tau_{\ell-1}} = j$. These transition intensities are equal to the hazard functions previously defined in equations (26) and (27). The $K \times K$ transition intensity matrix, which is associated to the time-homogeneous markovian process $\{X_t, t \in I\!R^+\}$, is denoted Q and has entries:

$$q(j,k) \ = \ \begin{cases} h_{j,k} \in I\!R^+ \ if \ j \neq k,\ (j,k) \in E \times E \\ -\displaystyle\sum_{\substack{m=1 \\ m \neq j}}^{K} h_{j,m} \leq 0 \ if \ j = k,\ j \in E \end{cases} \tag{70}$$

Let us denote $\mathbb{Q}$ the set of transition intensity matrices, i.e. the set of $(K \times K)$ matrices with entries verifying the conditions (70). 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:

$$P(0,T) \ = \ \exp(QT),\ T \in I\!R^+ \tag{71}$$

where $\exp(A) = \sum_{k=0}^{\infty} A^k/k!$ for any $K \times K$ matrix A.

Main properties of the time-homogeneous markovian process $\{X_t,\ t \in I\!R^+\}$ with state-space E, are the following:

• sojourn times in state j ($j \in$ E) are positive random variables, which are exponentially distributed with parameter $-q(j,j)$:

$$u_j \sim \exp(-q(j,j)),\ j = 1, \ldots, K \tag{72}$$

with $E[u_j] = var[u_j]^{1/2} = -q(j,j)^{-1}$,

• the probability of a transition to state k given that the process is currently in state j ($k \neq j$) is independent of the sojourn time in state j, and is found to be:

$$r_{j,k} = -q(j,k)/q(j,j),\ k \neq j,\ (j,k) \in E \times E \tag{73}$$

• if the time-homogeneous Markov process $\{X_t\}$ is ergodic, its equilibrium (or limiting) probability distribution is denoted $P^* = (p_1^*, \ldots, p_K^*)'$ and defined as the unique solution to the linear system of equations:

$$Q'P^* = 0 \ \ , \text{ with } \ \sum_{i=1}^{K} p_i^* = 1 \tag{74}$$

3.1.1 Maximum likelihood estimator of the matrix P using discrete-time (multiwave) panel data

Let us suppose now that we observe η independent realizations of the process $\{X_t\}$ at equally spaced times $T_0,\ T_1, \ldots,\ T_L$ ($L > 1$) such as: $T_\ell - T_{\ell-1} = T$, $\ell = 1, \ldots, L$. Let us denote:

• $n_{j,k}(\ell)$ the number of individuals who were in state j at time $T_{\ell-1}$ and who are in state k at time T_ℓ,

• $n_j(\ell-1)$ the number of individuals who were in state j at time $T_{\ell-1}$.

Maximizing the conditional likelihood function given the initial distribution at T_0:

$$\begin{aligned} L(P(0,T)) &= \prod_{\ell=1}^{L} \ \prod_{j,k=1}^{K} \{p_{j,k}(T_{\ell-1}, T_\ell)\}^{n_{j,k}(\ell)} \\ &= \prod_{j,k=1}^{K} \{p_{j,k}(0,T)\}^{\Sigma_{\ell=1}^{L} n_{j,k}(\ell)} \end{aligned} \tag{75}$$

with $\sum_{k=1}^{K} p_{j,k}(0,T) = 1$, gives the (j,k) entry of the MLE $\widehat{P}(0,T)$ for $P(0,T)$:

$$\widehat{p}_{j,k}(0,T) = \left(\sum_{\ell=1}^{L} n_{j,k}(\ell)\right) / \left(\sum_{\ell=1}^{L} n_j(\ell-1)\right) \tag{76}$$

(see Anderson and Goodman, 1957). If the solution $\widehat{Q}$ to the equation:

$$\widehat{P}(0,T) = \exp(\widehat{Q}T),\ T > 0 \tag{77}$$

belongs to the set $\mathbb{Q}$ of intensity matrices, then $\widehat{Q}$ is a MLE estimator for Q. Nevertheless, two difficulties may appear:[3]

• the equation (77) can have multiple solutions $\widehat{Q} \in \mathbb{Q}$: this problem is known as the aliasing problem;[4]

• none of the solutions $\widehat{Q}$ to the equation (77) belongs to the set $\mathbb{Q}$ of intensity matrices; in that case, the probability matrix $\widehat{P}(0,T)$ is said to be non-embeddable with a continuous-time Markov process.

3.1.2 Necessary conditions for embeddability

The unique necessary and sufficient condition for embeddability was given by Kendall, who proved that, when $K = 2$, the transition matrix $\widehat{P}(0,T)$ is embeddable if and only if the trace of $\widehat{P}(0,T)$ is strictly greater than 1. When $K \geq 3$, only necessary conditions are known; they are the following:[5]

<u>1st necessary condition</u> (Chung, 1967):

• if $\widehat{p}_{j,k}(0,T) = 0, then\ \widehat{p}^{(n)}_{j,k}(0,T) = 0,\ \forall n \in I\!N$, where $\widehat{p}^{(n)}_{j,k}(0,T)$ is the entry (j,k) of the matrix $[\widehat{P}(0,T)]^n$,

• if $\widehat{p}_{j,k}(0,T) \neq 0$, then $\widehat{p}^{(n)}_{j,k}(0,T) \neq 0,\ \forall n \in I\!N$;

<u>2nd necessary condition</u> (Kingman, 1962): $det\left[\widehat{P}(0,T)\right] > 0$,

<u>3rd necessary condition</u> (Elfving, 1937):

• no eigenvalue λ_i of $\widehat{P}(0,T)$ can satisfy $\mid \lambda_i \mid= 1$, other than $\lambda_i = 1$;

• in addition, any negative eigenvalue must have even algebraic multiplicity;

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

[4] The aliasing problem has also been considered by Phillips (1973).

[5] Singer and Spilerman (1976a) and Geweke, Marshall and Zarkin (1986b) survey this problem.

4th necessary condition (Runnenberg, 1962): the argument of any eigenvalue λ_i of $\widehat{P}(0,T)$ must satisfy:

$$(\frac{1}{2}+\frac{1}{K})\Pi \leq \arg(\log\ \lambda_i) \leq (\frac{3}{2}-\frac{1}{K})\Pi$$

This last condition plays an important role in the remainder of the analysis.

3.1.3 Resolving the equation $\hat{P}(0,T) = \exp(\hat{Q}T)$

The proof of the following theorem can be found in Singer and Spilerman (1976a):

If $\widehat{P}(0,T)$ has K distinct[6] *eigenvalues $(\lambda_1,\ldots,\lambda_K)$ and can be written $\widehat{P}(0,T) = A \times D \times A^{-1}$, where $D = diag(\lambda_1,\ldots,\lambda_K)$ and the eigenvector corresponding to λ_i $(i = 1,\ldots,K)$ is contained in the i_th column of the $(K \times K)$ matrix A, then:*

$$log(\widehat{P}(0,T)) = \widehat{Q}T = A \times \begin{pmatrix} log_{k_1}(\lambda_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & log_{k_K}(\lambda_K) \end{pmatrix} \times A^{-1} \tag{78}$$

where $log_{k_i}(\lambda_i) = log \mid \lambda_i \mid +(arg\lambda_i + 2k_i\Pi)i, k_i \in \mathbb{Z}$, is a branch of the logarithm of λ_i, when $\lambda_i \in \mathcal{C}$.[7]

Since equation (77) 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 $\widehat{P}(0,T)$ has at least two complex conjugate eigenvalues. However, an important implication of the fourth necessary condition for

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

[7] 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: $\log_k(z) = \log \mid z \mid +i(\theta+2k\Pi)$, $k \in \mathbb{Z}$, 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.

embeddability is that only finitely many branches of $\log(\widehat{P}(0,T))$ need to be checked for membership in $\mathbb{Q}$. Indeed, this condition implies:

$$\forall \lambda_i, \quad -L_i(K) \leq k_i \leq U_i(K) \tag{79}$$

$$\text{where} \qquad U_i(K) = intpt \left| \frac{log \mid \lambda_i \mid tan\{(\frac{1}{2} + \frac{1}{K})\Pi\} - \mid arg\ \lambda_i \mid}{2\Pi} \right|$$

$$L_i(K) = intpt \left| \frac{log \ \mid \lambda_i \mid tan\{(\frac{3}{2} - \frac{1}{K})\Pi\} - \mid arg\ \lambda_i \mid}{2\Pi} \right|$$

the function "*intpt*" being the integer part of a real number. So the number of branches of λ_i which must be computed is equal to $L_i(K)+U_i(K)+1$, the last one corresponding to the main branch (with $k_i = 0$). Then the number of solutions $\widehat{Q}$ that must be examined for membership in $\mathbb{Q}$ is denoted $k^*(\widehat{P})$ and is equal to:

$$k^*(\widehat{P}) = \begin{cases} \prod_{j=1}^{v} \{L_j(K) + U_j(K) + 1\} & \text{if } v \geq 1 \\ 1 & \text{if } v = 0 \end{cases} \tag{80}$$

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

• for a real eigenvalue, only the principal branch of the logarithm must be examined: other branches (with $k_i \neq 0$) correspond to complex intensity matrices $\widehat{Q}$;

• each element of a complex conjugate eigenvalue pair has the same number of candidate branches (see (79)); moreover, only combinations of branches involving the same k_i in each element of the pair must be computed; all others correspond to complex intensity matrices; this fact explains why the calculation of $k^*(\widehat{P})$ 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 $L_j(K)+U_j(K)+1$ rather than $\{L_j(K) + U_j(K) + 1\}^2$.

If equation (77) has only one solution $\widehat{Q} \in \mathbb{Q}$, this solution is the MLE for the intensity matrix of the homogeneous continuous-time Markov process $\{X_t,\ t \in I\!R^+\}$; an estimator for the asymptotic covariance matrix of $\widehat{Q}$ has been given by Kalbfleisch and Lawless (1985).

3.1.4 The scoring procedure

Kalbfleisch and Lawless (1985) have proposed to maximize with respect to θ the conditional likelihood function (75), i.e.

$$L(\theta) = \prod_{i,j=1}^{K} \{\exp(QT)\}_{(i,j)}^{\Sigma_{\ell=1}^{L} n_{i,j}(\ell)}, \quad Q \in \mathbb{Q} \tag{81}$$

through a scoring algorithm. In this expression, $\{\exp(QT)\}_{i,j}$ is the entry (i,j) of the matrix $\exp(QT) = P(0,T)$ and θ is the vector of extra diagonal elements of the matrix Q ($\theta \equiv \theta(Q)$). If it is assumed that matrix Q has K distinct eigenvalues, denoted $(d_1, \cdots, d_K)$, matrices Q and $P(0,T)$ can be written as:

$$\begin{aligned} Q \quad &= A\ D_Q A^{-1} = A \text{ diag } (d_1, \cdots, d_K) A^{-1} \\ \text{and } P(0,T) \quad &= \exp(QT) = A \exp(D_Q T) A^{-1} \\ &= A \text{ diag}(e^{d_1 T}, \cdots, e^{d_K T}) A^{-1} = A \text{ diag}(\lambda_1, \cdots, \lambda_K) A^{-1} \end{aligned} \tag{82}$$

These formulae lead to a convenient expression of the score (or gradient) vector, which is:

$$S(\theta) = \left\{ \frac{\partial \ \log L(Q)}{\partial \ q_{k\ell}} \right\} = \left\{ \sum_{i,j=1}^{K} \sum_{\ell=1}^{L} n_{i,j}(\ell) \ \frac{\partial \{\exp(QT)\}_{(i,j)} / \partial q_{k\ell}}{\{\exp(QT)\}_{(i,j)}} \right\} \tag{83}$$

where

$$\begin{aligned} \frac{\partial \{\exp(QT)\}}{\partial q_{k\ell}} \quad &= \sum_{s=1}^{\infty} \left(\frac{\partial Q^s}{\partial q_{k\ell}}\right) \frac{T^s}{s!} = \sum_{s=1}^{\infty} \sum_{r=0}^{s-1} Q^r \frac{\partial Q}{\partial q_{k\ell}} \cdot Q^{s-1-r} \cdot \frac{T^s}{s!} \\ &= A V_{k\ell} A^{-1} \end{aligned}$$

the matrix

$$V_{k\ell} \quad = \sum_{s=1}^{\infty} \sum_{r=0}^{s-1} D_Q^r \left(A^{-1} \frac{\partial Q}{\partial q_{k\ell}} A\right) D_Q^{s-1-r} \frac{T^s}{s!} \text{ having elements:}$$

$$\begin{cases} (G_{k\ell})_{(i,j)} \dfrac{e^{d_i t} - e^{d_j t}}{d_i - d_j} & , i \neq j, \\ (G_{k\ell})_{(i,j)} t \ e^{d_i t} & , i = j, \end{cases}$$

where $(G_{k\ell})_{(i,j)}$ is the entry (i,j) of the matrix $G_{k\ell} = A^{-1} \dfrac{\partial Q}{\partial q_{k\ell}} A$.

The information matrix, which has the form

$$E\left[-\frac{\partial^2 \log L(\theta)}{\partial q_{k\ell}\partial q_{k'\ell'}}\right] = \left\{\sum_{\ell=1}^{L}\sum_{i,j=1}^{K} \frac{E[N_i(\ell-1)]}{p_{i,j}(0,T)} \frac{\partial p_{i,j}(0,T)}{\partial q_{k\ell}} \frac{\partial p_{i,j}(0,T)}{\partial q_{k'\ell'}}\right\} \quad (84)$$

(see Kalbfleisch and Lawless, 1985, p. 864), is estimated by:

$$M(\theta) = \left\{\sum_{\ell=1}^{L}\sum_{i,j=1}^{K} \frac{n_i(\ell-1)}{p_{i,j}(0,T)} \frac{\partial p_{i,j}(0,T)}{\partial q_{k\ell}} \frac{\partial p_{i,j}(0,T)}{\partial q_{k'\ell'}}\right\} \quad (85)$$

The iterative formula for the scoring algorithm being:

$$\theta_{n+1} = \theta_n + M(\theta_n)^{-1} S(\theta_n)$$

where $n \geq 0$ and an initial value $\theta_0 = \theta(Q_0)$ is still to be chosen. Two cases must be considered (the case with multiple solutions in Q is excluded):

• equation (77) admits only one solution for $\widehat{Q}$ and this solution belongs to the set $\mathbb{Q}$ of transition intensity matrices: $\hat{Q}$ is the MLE of the transition matrix Q of the time-homogeneous markovian process, and the matrix $M(\theta(\widehat{Q}))^{-1}$ gives a consistent estimate of the covariance matrix of $\hat{\theta} = \theta(\widehat{Q})$;

• the unique solution $Q_0 = \widehat{Q}$ to equation (77) doesn't belong to the set $\mathbb{Q}$; however, it may exist matrices $\tilde{P}(0,T) = \exp(\tilde{Q}T)$ "close" to $\widehat{P}(0,T)$ and which are embeddable, i.e. such that $\tilde{Q} \in \mathbb{Q}$; in this case, the scoring algorithm of Kalbfleisch and Lawless (1985) can be applied to the maximization of the likelihood (81) subject to the constraint $Q \in \mathbb{Q}$; this constraint can be directly introduced into the iterative procedure by setting

$$q_{i,j} = \begin{cases} \exp(a_{i,j}), \ a_{i,j} \in \mathbb{R}, j \neq i, (i,j) \in E \times E \\ q_{ii} = -\sum_{\substack{k=1 \\ k\neq i}}^{K} q_{ik}, i = j, \ i \in E \end{cases} \quad (86)$$

and the initial value Q_0 can be chosen to verify:

$$Q_0 = \underset{Q\in \mathbb{Q}}{argmin} \parallel Q_0 - \widehat{Q} \parallel \quad (87)$$

where $\widehat{Q} = \frac{1}{T}\log \widehat{P}(0,T)$.

3.1.5 Bayesian inference

Geweke, Marshall and Zarkin (1986a) have developed a bayesian approach for statistical inference on Q (and functions of Q) by using a diffuse prior on the set of stochastic matrices. This approach can be justified with two arguments:

• when the MLE of Q is on the parameter set boundary, standard asymptotic theory cannot be applied any more; bayesian inference overcomes this difficulty: the posterior confidence interval for Q can be viewed as its asymptotic approximation;

• moreover, bayesian inference allows incorporating into the choice of the prior distribution some information external to the sample (for example, the distribution of sojourn durations in each state).

Let us denote P_K the set of $(K \times K)$ stochastic matrices, i.e. $P_k = \{P \in M_{K,K} : \forall i,j \in E, p_{i,j} \geq 0 \text{ and } \sum_{j=1}^{K} p_{i,j} = 1\}, P_K^*$ the set of $(K \times K)$ embeddable stochastic matrices, i.e. $P_K^* = \{P \in M_{K,K} : P \in P_K \text{ and } \exists Q \in \mathbb{Q}, P(O,T) = \exp(QT), T > 0\}$. For any $P \in P_K^*, k^*(P)$ denotes the number of combinations of the form (78) belonging to $\mathbb{Q}$ and verifying equation (77). Now let us consider a prior distribution on $P \in P_K$, denoted $\mu(P)$, a prior distribution on Q, denoted $h_k(P)$ and verifying $\sum_{k=1}^{k^*(P)} h_k(P) = 1$ for $P \in P_K$, and a $\mathbb{R}$-valued function of interest denoted $g(Q)$. If the posterior embeddability probability of P is defined as:

$$\Pr(P \in P_K^* \mid N) = \frac{\int_{P_K^*} L(P;N)\mu(P)dP}{\int_{P_K} L(P;N)\mu(P)dP} > 0 \tag{88}$$

then the expectation of $g(Q)$ is equal to

$$E[g(Q) \mid N, P \in P_K^*] = \frac{\int_{P_K^*} \sum_{k=1}^{k^*(P)} h_k(P) g[Q_k(P)] L(P;N)\mu(P)dP}{\int_{P_K^*} L(P;N)\mu(P)dP} \tag{89}$$

where the entry (i,j) of the matrix N is $\sum_{\ell=1}^{L} n_{i,j}(\ell), L(P;N)$ is the likelihood function and $Q_k(P)$ is the transition intensity matrix corresponding to the k-th combination of logarithms of the eigenvalues of matrix P. The function of interest $g(Q)$ can be, for example, $g(Q) = q_{i,j}, (i,j) \in E \times E$, or:

$$g(Q) = E\left\{(q_{i,j} - E(q_{i,j} \mid N; P \in P_K^*))^2 \mid N; P \in P_K^*\right\}$$

which is equivalent to:

$$g(Q) = E\left\{q_{i,j}^2 \mid N; P \in P_K^*\right\} - E^2\left\{q_{i,j} \mid N; P \in P_K^*\right\}$$

The embeddability probability for P and the first moment of $g(Q)$ may be computed using Monte-Carlo integration. This involves the choice of an importance function from which a sequence of matrices $\{P_i\} \in P_K$ can be easily generated (see Geweke et al., 1986a, for such a function). Now let us consider a function $J(P_i)$ such that $J(P_i) = 1$ if $P_i \in P_K^*$ and $J(P_i) = 0$ otherwise. If $\mu(P_i)$ is bounded above, then:

$$\lim_{I\to\infty} \frac{\sum_{i=1}^{I} J(P_i)L(P_i;N)\mu(P_i)/I(P_i)}{\sum_{i=1}^{I} L(P_i;N)\mu(P_i)/I(P_i)} = \Pr[P \in P_K^* \mid N] \qquad \text{a.s.} \tag{90}$$

Moreover, if $H_k(P)$ is a multinomial random variable such that $\Pr[H_k(P) = 1] = h_k(P)$, and if $g(Q)$ is bounded above, then

$$\lim_{I\to\infty} \frac{\sum_{i=1}^{I}\sum_{k=1}^{k^*(P_i)} H_k(P_i)g[Q_k(P_i)]J(P_i)L(P_i;N)\mu(P_i)/I(P_i)}{\sum_{i=1}^{I} J(P_i)L(P_i;N)\mu(P_i)/I(P_i)} = E[g(Q) \mid N; P \in P_K^*] \qquad \text{a.s.} \tag{91}$$

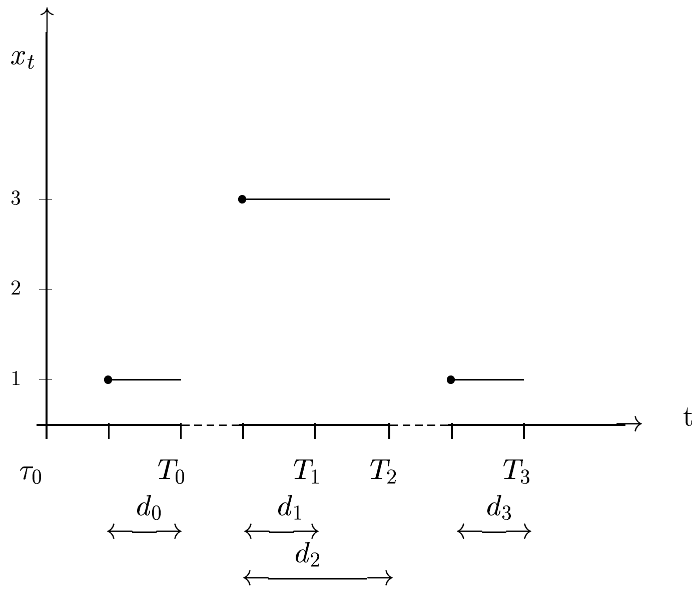
(see Geweke et al., 1986a, p. 658).

3.1.6 Tenure records

Up to now we concentrated on the statistical analysis of discrete-time observations of an underlying continuous-time Markov process. The available information is sometimes richer than the one brought by discrete-time data, but not as complete as the one contained in continuous-time data. Indeed it can consist, for a given individual, in the joint sequence $\{(x_{T_\ell}, d_{T_\ell})\}_{\ell=0,\cdots,L}$ of occupied states $\{x_{T_\ell}\}_{\ell=0,\cdots,L}$ and of times $\{d_{T_\ell}\}_{\ell=0,\cdots,L}$ already spent in these states at distant observation times $\{T_\ell\}_{\ell=0,\cdots L}$. Such data have been studied in the continuous-time markovian framework by Magnac and Robin (1994), who proposed to call this kind of observations "tenure records". Figure 3 gives an example of a tenure record.

In this example, T_0, T_1, T_2 and T_3 are the exogenous survey dates. The process $\{X_t\}_{t\geq 0}$ is first observed to be in state $x_{T_0} = 1$ at time T_0: it

Figure 3



occupies this state from date $(T_0 - d_0)$ on. It is then observed to be in state 3 at successive times T_1 and T_2. This state was entered at time $(T_1 - d_1) = (T_2 - d_2)$. Finally, the process is at time T_3 in state $x_{T_3} = 1$ from date $(T_3 - d_3)$ on. Indeed it is possible that a spell covers two survey dates, as it is the case for the second observed spell in the preceding example: obviously, the information collected in T_1 is redundant.

Let us remark that in tenure records data sets, any sojourn duration is right-censored with probability one. Typically, a tenure record consists of a sequence $\{x_{T_\ell}, d_\ell, t_\ell\}_{\ell=0,\cdots,L}$ with the convention $t_L = \infty$. The process $\{X_t\}_{t\geq 0}$ enters state x_{T_ℓ} at time $(T_\ell - d_\ell)$ and is observed to stay in this state for a duration greater than d_ℓ. Then the process is not observed (i.e. is truncated) during a period of length $t_\ell = (T_{\ell+1} - d_{\ell+1}) - T_\ell$. Let $h_{ij}(s,t)$ be the probability that the process $\{X_t\}$ enters state j at time t given that it was in state i at time $s(s < t)$. If $\{X_t\}$ is time-homogeneous markovian, then $h_{ij}(0, t-s) \equiv h_{ij}(t-s), s < t$. In this case, $h_{ij}(t)$ is equal to:

$$h_{ij}(t) = \sum_{\substack{k=1 \\ k\neq j}}^{K} p_{ik}(t)\ q_{kj},\ \ (i,j) \in E \times E \tag{92}$$

Consequently, the likelihood function for a tenure record $\{x_{T_\ell}, d_\ell, t_\ell\}_{\ell=0,\cdots,L}$

is the following:

$$L = \left\{ \prod_{\ell=0}^{L-1} S(d_\ell \mid x_{T_\ell})\ h_{x_{T_\ell}, x_{T_{\ell+1}}}(t_\ell) \right\} S(d_L \mid x_{T_L})$$
$$= \exp(-\lambda_{x_{T_L}} d_L) \prod_{\ell=0}^{L-1} \left\{ \exp(-\lambda_{x_{T_\ell}} d_\ell) \sum_{\substack{k=1 \\ k \neq x_{T_{\ell+1}}}}^{K} \{\exp(Q\, t_\ell)\}_{(x_{T_\ell}, k)} \cdot q_{k, x_{T_{\ell+1}}} \right\} \tag{93}$$

where $S(u \mid x_{T_\ell})$ is the survivor function of the sojourn duration in state x_{T_ℓ} and Q is the transition intensity matrix with entries:

$$Q(i,j) = \begin{cases} -\lambda_i = -\sum_{\substack{k=1 \\ k \neq i}}^{K} q_{ik}, \text{ if } j = i \\ \\ q_{ij}, \text{ if } j \neq i \end{cases}$$

Magnac and Robin (1994) show that tenure records allow to identify the intensity of transition from one state to the same state (for example, employment) when within-state mobility is allowed (i.e. when a worker can directly move from one job to another). Discrete-time observations do not present this advantage.

For a treatment of incomplete records, particularly in presence of unobserved heterogeneity see, for instance, Kamionka (1998). Magnac et al. (1995) propose to use indirect inference to estimate the parameters of a transition model under a semi-Markov assumption in the context of a censoring mechanism.

3.2 The Mover-Stayer model

3.2.1 MLE for the discrete-time mover-stayer model

The mover-stayer model has been introduced by Blumen et al. (1955) for studying the mobility of workers in the labor market. Subsequently, Goodman (1961), Spilerman (1972) and Singer and Spilerman (1976a) have developed the statistical analysis of this model, essentially on the discrete-time axis. The mover-stayer model in discrete time is a stochastic process $\{X_\ell,\ \ell \in I\!N\}$, defined on a discrete state-space $E = \{1, \ldots, K\}$, $K \in I\!N$, and resulting from the mixture of two independent Markov chains; the first of these two chains, denoted $\{X_\ell^1, \ell \in I\!N\}$ is degenerate, i.e. its transition probability matrix is the identity matrix, denoted I. The other chain, denoted $\{X^2_\ell\ ,\ \ell \in I\!N\}$ is characterized by a non-degenerate transition matrix $M(s,u) = \parallel m_{i,j}(s,u) \parallel$, $i,j = 1,\ldots,K$, $0 \le s \le u$, where:

$$m_{i,j}(s,u) = \Pr\{X^2_u = j \mid X^2_s = i\},\ i,j \in E,\ s,u \in I\!N,\ s \le u \tag{94}$$

and $\displaystyle\sum_{j=1}^{K} m_{i,j}(s,u) = 1.$

Moreover, the Markov chain $\{X^2_\ell, \ell \in I\!N\}$ is assumed to be time homogeneous, i.e.:

$$m_{i,j}(s,u) = m_{i,j}(0,u-s) \equiv m_{i,j}(u-s),\ \ 0 \le s \le u \tag{95}$$

which is equivalent to:

$$M(s,u) = M(0,u-s) \equiv M(u-s),\ \ 0 \le s \le u \tag{96}$$

Now let us assume that the mixed process $\{X_\ell, \ell \in I\!N\}$ is observed at fixed and equally distant times: $0, T, 2T, \ldots, LT$, with $T > 0$ and $L \in I\!N$ $(L \ge 1)$. Transition probabilities for this process are given by the formulas:

$$p_{i,j}(0,kT) = \Pr[X_{kT} = j \mid X_0 = i],\ i,j \in E,\ k = 1,\ldots,L \tag{97}$$

$$= \begin{cases} (1-s_i)[m_{i,j}(T)]^{(k)} & if\ \ j \ne i \\ s_i + (1-s_i)[m_{i,i}(T)]^{(k)} & if\ \ j = i \end{cases}$$

where $[m_{i,j}(T)]^{(k)}$ is the entry (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 $M(s,u), s \le u$. The proportion of "stayers" in state i $(i \in E)$ is equal to s_i.

The estimation of the transition matrix $M(0,T)$ and of the mixing measure s from a sample of N independent realizations of the process $\{X_\ell, \ell \in I\!N\}$, has been extensively treated by Frydman (1984) and then carried out by Sampson (1990). The method developed by Frydman relies on a simple recursive procedure, which will be rapidly surveyed. Formally, the form of the sample is:

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

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, [8] i.e. $i_0 = i_T = \ldots = i_{LT} = i$, $i \in E$, $n_{ij} = \sum_{k=1}^{L} n_{ij}(kT)$ the total number of observed transitions from state i to state j, $n_i^* = \sum_{k=0}^{L-1} n_i(kT)$ the total number of visits to state i before time (LT), $\eta_i \geq 0$ 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):

$$L = \prod_{i=1}^{K} \eta_i^{n_i(0)} \prod_{i=1}^{K} L_i \tag{98}$$

where:

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

In this last expression, $n_i(0)$ is the number of individuals in state i at time 0, n_i is the number of individuals permanently observed in state i, $(n_i(0) - n_i)$ is the number of individuals initially in state i who experience at least one transition in the L following periods, n_{ik} is the total number of transitions from state i to state k. Maximizing the function (98) with respect to M and s subject to the constraints $s_i \geq 0$, $i \in E$, is equivalent to maximize the K expressions:

$$\mathcal{L}_i = Log\ L_i + \lambda_i s_i,\ i = 1, \ldots, K \tag{99}$$

for which the first-order derivatives relatively to s_i are:

$$\frac{\partial \mathcal{L}_i}{\partial s_i} = \frac{n_i\{1 - [m_{ii}(0,T)]^L\}}{s_i + (1 - s_i)[m_{ii}(0,T)]^L} - \frac{n_i(0) - n_i}{1 - s_i} + \lambda_i = 0 \tag{100}$$

[8] 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 $\{m_{ii}(0,T)\}^L$.

Two situations should be considered:
First case: If $s_i > 0$, then $\lambda_i = 0$ and:

$$s_i = \frac{n_i - n_i(0)[m_{ii}(0,T)]^L}{n_i(0)\{1 - [m_{ii}(0,T)]^L\}} \tag{101}$$

As shown by Frydman (1984, p. 634-635), the ML estimators of transition probabilities m_{ij} (with fixed i, and j varying from 1 to K) are given by the recursive equation:

$$\widehat{m}_{ij}(0,T) = n_{ij}\{1 - \widehat{m}_{ii}(0,T) - \sum_{\substack{k=1 \\ k \neq i}}^{j-1} \widehat{m}_{ik}(0,T)\} / \sum_{\substack{k=j \\ k \neq i}}^{K} n_{ik}, \; j \neq i, \; i,j \in E \tag{102}$$

To solve equation (102), it is necessary to begin by setting $j = 1$ if $i \neq 1$ and $j = 2$ if $i = 1$. Furthermore, $\widehat{m}_{ii}(0,T)$ is the solution, belonging to the interval $[0,1]$, to the equation:

$$\begin{array}{c} [n_i^* - Ln_i(0)][m_{ii}(0,T)]^{L+1} + [Ln_i(0) - n_{ii}][m_{ii}(0,T)]^L \\ +[Ln_i - n_i^*]m_{ii}(0,T) + (n_{ii} - Ln_i) \;=\; 0 \end{array} \tag{103}$$

Frydman (1984) doesn't notice that $s_i \leq 0$ whenever $(\frac{n_i}{n_i(0)}) \leq [m_{ii}(0,T)]^L$, where $(n_i/n_i(0))$ is the proportion of individuals permanently observed in state i. In that case, the initial assumption $s_i > 0$ is violated, and it is necessary to consider the case where $s_i = 0$.
Second case: If $s_i = 0$, then:

$$\widehat{m}_{ij}(0,T) = n_{ij}/n_i^*, \; \forall \; i,j = 1,\ldots,K \tag{104}$$

This is the usual ML estimator for the probability of transition from i to j for a first-order Markov chain in discrete time (for example, see Anderson and Goodman, 1957, or Billingsley, 1961). A remark, which is not contained in the paper by Frydman (1984), must be made. It may appear that $Ln_i = n_{ii}$ (with $n_{ii} \neq 0$), which means that no transition from state i to any other distinct state is observed. This case arises when the number n_i of individuals permanently observed in state i is equal to the number $n_i(0)$ of individuals initially present in state i (if $n_i(0) \neq 0$). Then the estimation problem has two solutions:

- s_i=1 and m_{ii} is non-identifiable (see equations (101) and (103)),
- $s_i = 0$ and $m_{ii} = 1$.

The first solution corresponds to a pure model of "stayers" in state i, the second to a time-homogeneous Markov chain in which state i is absorbing. The mover-stayer model, as a mixture of two Markov chains, is not appropriate any more for state i. When this case appears in the applied work, we propose to choose the solution $s_i = 0$ and $m_{ii} = 1$, especially for computing the estimated marginal probabilities of the form $\Pr[X_{kT} = i]$, $k = 0, \ldots, L$, $i = 1, \ldots, K$. The analytical expression of the estimated asymptotic covariance matrix for ML estimators $\widehat{M}$ and $\widehat{s}$ can be calculated using second derivatives of expression (99).

3.2.2 Bayesian inference for the continuous-time mover-stayer model

The mover-stayer model in continuous-time is a mixture of two independent Markov chains; the first one denoted $\{X_t^1,\ t \in \mathbb{R}^+\}$ has a degenerate transition matrix equal to the identity matrix I; the second one denoted $\{X_t^2,\ t \in \mathbb{R}^+\}$ has a non-degenerate transition matrix $M(s,t)$, $0 \leq s \leq t$, verifying over any interval of length T:

$$M(0,T) = \exp(QT), \quad T \in \mathbb{R}^+ \tag{105}$$

Setting $M(0,kT) = \|m_{i,j}(0,kT)\|$, we get:

$$P(0,kT) = diag(s) + diag(\mathbb{1}_K - s)\{\exp(QT)\}^K,\ T \geq 0,\ k = 1, \ldots, L \tag{106}$$

where $s = (s_1, \ldots, s_K)'$, $(\mathbb{1}_K - s) = (1 - s_1, \ldots, 1 - s_K)'$, and $diag(x)$ is a diagonal matrix with vector x on the main diagonal. From the discrete-time ML estimators of stayers' proportions s and of the transition probability matrix $M(0,T)$, it is then possible to obtain the ML estimator of the intensity matrix Q by resolving equation (105) (see subsection 2.1 above). But, due to the possible problem of non-embeddability of the matrix $M(0,T)$, it could be better to adopt a bayesian approach, as the one proposed by Fougère and Kamionka (2003). This approach is summarized below.

3.2.2.1 Definitions To write the likelihood-function and the expected value under the posterior of some function of parameters, additional notation is needed. Let M_K be the space of $K \times K$ stochastic matrices:

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

Clearly, the transition probability matrix $M(0,T)$ belongs to M_K. Let $\mu(M,s)$ be a prior mapping $M_K \times [0,1]$ into $I\!R$ (the uniform prior will be used in the application). $\mu(M,s)$ is defined for $M \in M_K$ and for a vector of mixing measures $s = \{s_i \ , \ i \ \in \ E\} \ \in \ [0,1]^K$. $[0,1]^K$ denotes the cartesian product of K copies of $[0,1]$. Let us denote $\mathbb{Q}$ the space of intensity matrices:

$$\mathbb{Q} = \{Q = \| \ q_{ij} \ \| \ \ : \ q_{ij} \geq 0, \ i,j \in E, \ i \neq j \text{ and } q_{ii} \leq 0, \forall i \in E\}.$$

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

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

If $D_K = M_K \times [0,1]^K$ represents the parameters space for the model, then the space $D_K^* = M_K^* \times [0,1]^K$ denotes the set of embeddable parameters and $D_K^* \ \subset \ D_K$. As it was shown in subsection 2.1, the solution to $M(0,T) = \exp(QT)$ may not be unique: this is the **aliasing problem**.

Let us consider now the set of matrices $Q^{(k)} \in \mathbb{Q}$, solutions of the equation $Q^{(k)} = \log(M(0,T))/T$, for $k = 1 \ ,\ldots, \ B(M)$. $B(M)$ 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 $B(M) \in I\!N$ and $B(M) = 0$ if $M \notin M_K^*$. Denote $Q^{(k)}(M)$ the intensity matrix that corresponds to the $k{-}th$ solution of log(M), $k = 1,\ldots,B(M)$. $Q^{(k)}(M), \ 1 \leq k \leq B(M)$, is a function defined for $M \in M_K^*$, $Q^{(k)}(M) \ \in \ Q$. Let $h^{(k)}(M)$ be a probability density function induced by a prior probability distribution on the $k{-}th$ solution of the equation $M(0,T) = \exp(QT)$ when $M \in M_K^*$. By definition, $h^{(k)}(M)$ verifies $\sum_{k=1}^{B(M)} h^{(k)}(M) = 1$.

Let $g(Q,s)$ be a function defined for $(Q,s) \in \mathbb{Q} \times [0,1]^K$. 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. Thus, the posterior probability that the transition probability matrix M is embeddable has the form:

$$\Pr[(M,s) \in D_K^* \ \mid \ (N,n)] = \frac{\displaystyle\int_{D_K^*} L(M,s;N,n)\mu(M,s) \ d(M,s)}{\displaystyle\int_{D_K} L(M,s;N,n)\mu(M,s) \ d(M,s)} \qquad (107)$$

3.2.2.2 Likelihood and importance functions The likelihood function $L \equiv L(M,s;N,n)$ up to the initial distribution of the process $\{X(t), \ t \geq 0\}$

is

$$L \propto \prod_{i=1}^{K} L_i \tag{108}$$

where:

$$\begin{aligned} L_i = & \ [\ s_i + (1-s_i) \times \{\exp(QT)\}_{ii}^{L}\ ]^{n_i} \times (1-s_i)^{n_i(0)-n_i} \\ & \times \{\exp(QT)\}_{ii}^{n_{ii}-Ln_i} \prod_{k \neq i, k=1}^{K} \{\exp(QT)\}_{ik}^{n_{ik}}, \end{aligned} \tag{109}$$

$\{\exp(Q\,T)\}_{i,k}$ denoting the entry (i,k) of the $K \times K$ matrix $\exp(Q\,T)$. If $\Pr[M \in M_K^* \mid N, n] > 0$, then

$$E[g(Q,s) \mid (N,n); (M,s) \in D_K^*] \tag{110}$$

$$= \frac{\int_{D_K^*} \sum_{k=1}^{B(M)} h^{(k)}(M)\ g(Q^{(k)}(M), s)\ L(M,s;N,n)\ \mu(M,s)\ d(M,s)}{\int_{D_K} L(M,s;N,n)\ \mu(M,s)\ d(M,s)}$$

In order to evaluate the integrals inside expressions (107) and (110), an adaptation of the Monte-Carlo method may be used because an analytical expression for $Q^{(k)}(M)$ or $B(M)$ when $K \geq 3$ has not been found yet. Let $I(M,s)$ be a probability density function defined for $(M,s) \in D_K$. $I(M,s)$ is 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.

Let $J(M)$ a function defined for $M \in M_K$:

$$J(M) = \begin{cases} 1 \text{ if } M \in M_K^* \\ 0 \text{ otherwise} \end{cases}$$

Then

$$\lim_{I \to +\infty} \frac{\sum_{i=1}^{I} J(M_i)\ L(M_i, s_i; N, n)\ \mu(M_i, s_i)/I(M_i, s_i)}{\sum_{i=1}^{I} L(M_i, s_i; N, n)\ \mu(M_i, s_i)/I(M_i, s_i)} \tag{111}$$

$$\overset{a.s}{=} \Pr[(M,s) \in D_K^* \mid N, n]$$

and

$$E[g(Q,s) \mid N,n;(M,s) \in D_K^*] \stackrel{a.s}{=}$$

$$\lim_{I \to +\infty} \frac{\displaystyle\sum_{i=1}^{I} \sum_{k=1}^{B(M)} \frac{h^{(k)}(M_i)\ g[Q^{(k)}(M_i), s_i] J(M_i)\ L(M_i, s_i; N, n)\ \mu(M_i, s_i)}{I(M_i, s_i)}}{\displaystyle\sum_{i=1}^{I} J(M_i)\ L(M_i, s_i; N, n)\ \mu(M_i, s_i)/I(M_i, s_i)} \tag{112}$$

where $\Pr[(M,s) \in D_K^* \mid N,n]$ is the probability under the posterior that the discrete-time Mover-Stayer model is embeddable with the continuous-time one, and $E[g(Q,s) \mid N,n;(M,s) \in D_K^*]$ defines the posterior moments of the parameters' function of interest.

For a better convergence of estimators (111) and (112), $I(M,s)$ should be concentrated on the part of D_K where $L(M,s;N,n)$ is nonnegligible. For that purpose, if $\mu(M,s)$ is not concentrated on some part of the set D_K (that's the case when μ is uniform), $I(M,s)$ can be taken proportional to the likelihood $L(M,s;N,n)$. Because drawing (M,s) from $L(M,s;N,n)$ is difficult, Fougère and Kamionka (2003) choose a normal expansion for $L(M,s;N,n)$ with mean the ML estimator $(\widehat{M},\widehat{s})$ and with covariance matrix the inverse of the information matrix estimated at $(\widehat{M},\widehat{s})$.

When $g(Q,s)$ and $\mu(M,s)$ are bounded above, the convergence of the estimator (112) is obtained almost surely. When the function $g(Q,s)$ does not verify this property (for instance, if we are interested in the estimation of q_{ij}), the convergence of the expression (112) relies on the existence of the posterior mean: $E[g(Q,s) \mid (M,s) \in D_K^*; N,n]$.

The covariance matrix V associated to $L(M,s;N,n)$ is block diagonal with blocks consisting of matrices V_i, $i = 1,\ldots,K$, defined as:

$$V_i(M,s) = -E\left[\frac{\partial^2 Log(L_i(M,s;N,n))}{\partial\theta_k \partial\theta_l}\right]^{-1} = R_i(M,s)^{-1} \tag{113}$$

with $\theta_k, \theta_l = \begin{cases} m_{i,j}\ ,\ i,j \in E \\ s_i\ ,\ i \in E \end{cases}$ where $R_i(M,s)$ is the $i-th$ diagonal block of the information matrix $R(M,s)$ associated to $L(M,s;N,n)$. Then a sequence of draws $\{(M_k,s_k)\}_{k=1,\ldots,I}$ can be generated according to the density of a multivariate normal distribution with mean (M,s) and covariance matrix $V(M,s) = R(M,s)^{-1}$. If we suppose that $V_{i-} = P_i P_i'$ is the Choleski's decomposition of the matrix V_{i-} obtained by dropping the last row and column of matrix V_i , and if $y^k \sim N(0_K, I_K)$, then

$$z^k = P_i\, y^k + \begin{pmatrix} s_i \\ m_{i1} \\ \vdots \\ m_{iK-1} \end{pmatrix} \sim N(\begin{pmatrix} s_i \\ m_{i1} \\ \vdots \\ m_{iK-1} \end{pmatrix}, V_{i-}) \tag{114}$$

Finally, we can obtain m_{iK} by setting $m_{iK} = 1 - \sum_{j=1}^{K-1} m_{i,j}$. Inside the procedure, s_i, $(m_{i,1}, \ldots, m_{i,K})$, and V_i are estimated by their MLE, respectively $\hat{s}_i$, $(\hat{m}_{i,1}, \ldots, \hat{m}_{i,K})$, and $\hat{V}_i$. For more details, see Fougère and Kamionka (2003).

3.2.2.3 - Limiting probability distribution and mobility indices
The mobility of movers can be appreciated by examination of the mobility indices for continuous-time Markov processes proposed by Geweke et al. (1986b). For the movers process with intensity matrix Q, four indices of mobility can be considered:

$$\begin{aligned} M_1(Q) &= -\log[\det(M(0,T))]/K = -tr(Q)/K \\ M_2(Q) &= \sum_{i=1}^{K} \Pi_i^{(m)} \sum_{j=1}^{K} q_{ij} \mid i-j \mid \\ M_3(Q) &= -\sum_{j=1}^{K} \Pi_j^{(m)} q_{ij} \\ M_4(Q) &= -\Re e[\log(\lambda_2)] \end{aligned} \tag{115}$$

where:

• $\Pi_i^{(m)}$ is the equilibrium probability in state i for the movers, given by equation $Q'\pi_i^{(m)} = 0$, with $\Sigma_{i=1}^{K} \pi_i^{(m)} = 1$,

• the eigenvalues of the matrix $M(0,T)$ denoted by $\lambda_1, \ldots, \lambda_K$, are ordered so that $\mid \lambda_1 \mid \geq \ldots \geq \mid \lambda_K \mid$,

• $\Re e$ denotes the real part of the logarithm of the eigenvalue λ_2.

We can also define the equilibrium (or limiting) probability distribution for the mixed "mover-stayer" process $\{X_t, t \in I\!R^+\}$. For state i, the limiting probability, denoted π_i, is given by:

$$\pi_i = s_i \eta_i + \pi_i^{(m)} \sum_{j=1}^{K} (1 - s_j)\eta_j, \; i \in E \tag{116}$$

where:

• $\eta = \{\eta_i, i \in E\}$ is the initial probability distribution (i.e. at the date 0) for the process $\{X_t, t \in I\!R^+\}$,

• and $\pi_i^{(m)}$ is the limiting probability of "movers" in state i.

It is easily verified that, for a purely markovian process (one for which $s_i = 0,\ \forall i \in E$), the formula (116) becomes $\pi_i = \pi_i^{(m)}$. The mobility indices (115) and the limiting distribution (116) can be estimated using formula (112) and taking respectively $g(Q,s) = M_k(Q)$ $(1 \leq k \leq 4)$, or $g(Q,s) = \pi$.

3.2.2.4 Bayesian inference using Gibbs sampling The likelihood function of the sample X can be written

$$L(X|s,M,X_0)=\prod_{n=1}^{N}\sum_{k=1}^{2}\mathcal{L}(X_{(n)}|s,M,X_{o(n)},z_n{=}k)\ \Pr[z_n{=}k|s,M,X_{o(n)}]$$

where $\mathcal{L}$ is the conditional contribution of the individual n given the initial state $X_{o(n)}$ and the unobserved heterogeneity type z_n. z_n is an unobserved indicator taking the value 1 if the individual is a stayer or the value 2 if the individual is a mover.

The prior density on the parameter $\theta = (s,M)$ is assumed to be the product of the conjugate densities $\mu_1(s)$ and $\mu_2(M)$, where

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

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

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

is the matrix beta distribution with parameter $\alpha_{ij} > 0,\ i,j = 1,\ldots,K$.

The conditional distribution of the unobserved type z_n is thus

$$z_n \mid \theta, X_{(n)} \sim B(1; p(X_{(n)};\theta)) \tag{117}$$

where

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

Combining the prior and the sample informations we obtain that

$$s_j \mid X, Z \sim \text{ Dirichlet}\left(a_j + \sum_{n=1}^{N} i_j^{(n)}(2 - z_n), b_j + \sum_{n=1}^{N} i_j^{(n)}(z_n - 1)\right) \quad (118)$$

$$M \mid X, Z \sim \text{ Matrix beta}\left(\alpha_{ik} + \sum_{n=1}^{N}(z_n - 1)N_{ik}^{(n)}; i, k = 1, \ldots, K\right) \quad (119)$$

The Gibbs sampling algorithm runs like this:
Initialization: Fix an initial value $\theta^{(0)} = (s^{(0)}, M^{(0)})$.

Update from $\theta^{(m)}$ to $\theta^{(m+1)}$ by doing :

1 - Generate $Z^{(m)}$ according to the conditional distribution (117), given $\theta = \theta^{(m)}$ and X;
2 - Generate $\theta^{(m+1)} = (s^{(m+1)}, M^{(m+1)})$ using the conditional distribution (118) and (119), given $Z = Z^{(m)}$ and X.

Under general regularity conditions and for m large enough, the resulting random variable $\theta^{(m)}$ is distributed according to the stationary posterior distribution $\mu(\theta \mid X)$. Draws from the stationary posterior distribution $\mu(\theta \mid X)$ may be used to obtain posterior estimates of θ using an expression similar to the one given by equation (112) (see Fougère and Kamionka, 2003). Step one of the algorithm corresponds to a data augmentation step (see, Robert and Casella, 2002).

4 Concluding remarks

This chapter has introduced reduced-form models and statistical methods allowing to analyse longitudinal panel data on individual labor market transitions. The first section gave a very general presentation of methods concerning continuous-time observations, while the second section focused on the treatment of discrete-time observations for continuous-time discrete-state processes.

Obviously, our survey did not intend to cover exhaustively a continuously and rapidly growing literature. Among subjects treated in this field of research, two topics seem to be especially important. The first one is the treatment of endogenous selection bias in dynamic populations (see Lancaster and Imbens, 1990, 1995, Lancaster, 1990b, Ham and Lalonde, 1996, and Fougère, Kamionka and Prieto, 2005). Indeed, some sampling schemes for continuous-time discrete state space processes are such that the probability of being in the sample depends on the endogenous variable, i.e. being in a given state (for example, unemployment) at some date. Consequently inference from these endogenous samples requires specific statistical methods which have begun to be elaborated (see the papers quoted above). Another research area is the evaluation of the effect of public interventions such as employment and training programs. Here the main problem is knowing if these programs have a joint positive effect on earnings and employment rates of beneficiaries (see, for example, papers by Card and Sullivan, 1988, Ham and Lalonde, 1990, Heckman, 1990, Eberwein, Ham and Lalonde, 1997, Bonnal, Fougère and Sérandon, 1997, Heckman, Lalonde and Smith, 1999). In order to avoid misleading results, this evaluation must take into account the selection biases induced simultaneously by the process of eligibility to the program and by the sampling scheme. Thus these two fields of research are very closely connected.

5 References

Aalen O.O., 1978, "Non parametric inference for a family of counting processes", The Annals of Statistics, Vol. 6, 701-726.

Amemiya T., 2001, "Endogenous Sampling in Duration Models", Monetary and Economic Studies, Vol. 19, $n°3$, 77-96.

Anderson T.W. and Goodman L.A., 1957, "Statistical inference about Markov chains", Annals of Mathematical Statistics, Vol. 28, 89-110.

Baker M. and Melino A., 2000, "Duration dependence and nonparametric heterogeneity: a Monte Carlo study", Journal of Econometrics, Vol. 96, $n°2$, 357-393.

Billingsley P., 1961, Statistical inference for Markov processes. The University of Chicago Press.

Bonnal L., Fougère D. and A. Sérandon, 1997, "Evaluating the impact of french employment policies on individual labour market histories", Review of Economic Studies, Vol. 64, $n°4$, 683-713.

BURDETT K., KIEFER N., MORTENSEN D.T. AND NEUMAN G., 1984a, "Earnings, unemployment and the allocation of time over time" , Review of Economic Studies, Vol. 51, $n°4$, 559-578.

BURDETT K., KIEFER N., MORTENSEN D.T. AND NEUMAN G., 1984b, "Steady states as natural rates in a dynamic discrete choice model of labor supply", in: Studies in labor market dynamics, G.R. Neuman and N.C. Westergard-Nielsen (eds.), Springer-Verlag, Berlin, 74-97.

BURDETT K., KIEFER N. AND SHARMA S., 1985, "Layoffs and duration dependence in a model of turnover", Journal of Econometrics, Vol. 28, $n°1$, 51-69.

BUTLER J.S., ANDERSON K.H. AND BURKHAUSER R.V., 1986, "Testing the relationship between work and health: a bivariate hazard model", Economics Letters, Vol. 20, 383-386.

BUTLER J.S., ANDERSON K.H. AND BURKHAUSER R.V., 1989, "Work and health after retirement: a competing risks model with semiparametric unobserved heterogeneity", Review of Economics and Statistics, Vol. 70, $n°1$, 46-53.

CARD D. AND SULLIVAN D., 1988, "Measuring the effect of subsidized training programs on movements in and out of employment", Econometrica, Vol. 56, 497-530.

CHESHER A., LANCASTER T., 1983, "The estimation of models of labour market behavior", Review of Economic Studies, Vol. 50, $n°4$, 609-624.

CHUNG K.L., 1967, Markov chains with stationary transition probabilities. Berlin: Springer - Verlag.

DOOB J.L., 1953, Stochastic processes. New-York: Wiley.

EBERWEIN C., HAM J.C. AND R.J. LALONDE, 1997, "The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: evidence from experimental data", The Review of Economic Studies, Vol. 64, n○4, 655-682.

EBERWEIN C., HAM J.C. AND R.J. LALONDE, 2002, "Alternative methods of estimating program effects in event history mdoels", Labour Economics, Vol. 9, 249-278.

ECKSTEIN Z. AND WOLPIN K.I., 1989, "The specification and estimation of dynamic stochastic discrete choice models", Journal of Human Resources, Vol. 24, n○4, 562-598.

ELFVING G., 1937, "Zur Theorie der Markoffschen Ketten", Acta Social Science Fenicae, series A2, n°8, 1-17.

FLINN C.J. AND HECKMAN J.J., 1982a, "New methods for analyzing structural models of labor force dynamics", Journal of econometrics, Vol. 18, 115-168.

FLINN C.J. AND HECKMAN J.J., 1982b, "Models for the analysis of labor force dynamics", in: Advances in Econometrics, Vol. 1, R. Basmann and G. Rhodes (eds.), Greenwich, Conn.: JAI Press, 35-95.

FLINN C.J. AND HECKMAN J.J., 1983a, "The likelihood function for the multi-state-multiepisode model", in: Advances in econometrics, Vol. 2, R. Basmann and G. Rhodes (eds.), Greenwich, Conn.: JAI Press, 225-231.

FLINN C.J. AND HECKMAN J.J., 1983b, "Are unemployment and out of the labor force behaviorally distinct labor force states ?", Journal of labor economics, Vol. 1, n°1, 28-42.

FOUGÈRE D. AND KAMIONKA T., 2003, "Bayesian inference for the mover-stayer model in continuous-time with an application to labour market transition data", Journal of Applied Econometrics, 18(6), 697-723.

FOUGÈRE D., KAMIONKA T. AND PRIETO A., 2005, "Stock-sampling bias in unemployment competing-risks duration models", mimeo, CREST-INSEE, Paris.

FRYDMAN H., 1984, "Maximum likelihood estimation in the mover-stayer model", Journal of the American Statistical Association, Vol. 79, 632-638.

GEWEKE J., MARSHALL R.C. AND ZARKIN G.A., 1986a, "Exact inference for continuous time Markov chains". Review of Economic Studies, Vol. 53, 653-669.

GEWEKE J., MARSHALL R.C. AND ZARKIN G.A., 1986b, "Mobility indices in continuous-time Markov chains", Econometrica, Vol. 54, n°6, 1407-1423.

GILBERT L., KAMIONKA T. AND LACROIX G., 2001, "The Impact of Government Sponsored Training Programs on the Labor Market Transitions of Disadvantaged Men", Crest Working Paper 2001-15, Paris.

GOURIÉROUX C. AND MONFORT A., 1997, Simulation-based econometric methods, Oxford: oxford University Press.

GRITZ R.M., 1993, "The impact of training on the frequency and the duration of employment", Journal of Econometrics, Vol. 57, 21-51.

Ham J.C. and Lalonde R.J., 1996, “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training”, Econometrica, Vol. 64, n°1, 175-205.

Heckman J., 1990, “Alternative approaches to the evaluation of social programs ; Econometric and experimental methods”, Barcelona Lecture, 6th World Congress of the Econometric Society, Barcelona, Spain.

Heckman J.J. and Borjas G.J., 1980, “Does unemployment cause future unemployment ? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence”, Economica, Vol.47, 247-283.

Heckman J.J., Lalonde R. and Smith J., 1999, “The economics and econometrics of active labor market policies”, in: The Handbook of Labor Economics, Vol. 3A, O. Ashenfelter and D. Card (eds.), Amsterdam, North-Holland, 1865-2097.

Heckman J.J., and Singer B., 1984, “Econometric duration analysis”, Journal of Econometrics, vol. 24, 63-132.

Kalbfleisch J.D. and Lawless J.F., 1985, “The analysis of panel data under a Markov assumption”, Journal of the American Statistical Association, Vol. 80, n°392, 863-871.

Kamionka T., 1998, “Simulated maximum likelihood estimation in transition models”, Econometrics Journal, Vol. 1, C129–153.

Kamionka T. and G. Lacroix, 2003, "Assessing the Impact of Non-Response on the Treatment Effect in the Canadian Self-Sufficiency Experiment", Crest working paper 2003-37.

Kingman J.F.C., 1962, “The imbedding problem for finite Markov chains”, Zeitschrift fur Wahrscheinlichkeitstheorie, n°1, 14-24.

Klein J.P. and Moeschberger M.L., 2003, “Survival analysis : techniques for censored and truncated data”, Second Edition, Springer.

Lancaster T., 1990a, The econometric analysis of transition data. Cambridge University Press, New-York, USA, 336p.

Lancaster T., 1990b, “A paradox in choice-based sampling”, Working Paper 9017, Department of Economics, Brown University.

Lancaster T. and Imbens G.W., 1990, “Choice-based sampling of dynamic populations”, in: Panel data and labor market studies, J. Hartog, G. Ridder and J. Theeuwes (eds.), Amsterdam, North-Holland, 21-44.

LANCASTER T. AND IMBENS G.W., 1995, "Optimal stock/flow panels", Journal of Econometrics, Vol. 66, n°1-2, 325-348.

LINDEBOOM M. AND VAN DEN BERG G., 1994, "Heterogeneity in models for bivariate survival: the importance of the mixing distribution", Journal of the Royal Statistical Society, Series B, 56, 49-60.

MAGNAC T. AND ROBIN J.M., 1994, "An econometric analysis of labour market transitions using discrete and tenure data", Labour Economics, vol. 1, 327-346.

MAGNAC T., ROBIN J.M. AND VISSER M., 1995, "Analysing Incomplete Individual Employment Histories Using Indirect Inference", Journal of Applied Econometrics, vol. 10, S153-S169.

MEALLI F. AND PUDNEY S., 1996, "Occupational pensions and job mobility in Britain: estimation of a random-effects competing risks model", Journal of Applied Econometrics, vol. 11, 293-320.

MEALLI F. AND PUDNEY S., 2003, "Applying heterogeneous transition models in labor economics: the role of youth training in labour market transitions", in: Analysis of Survey Data, R. L. Chambers and C. J. Skinner (eds.), New-York, Wiley, 245-274.

MORTENSEN D., 1986, "Job search and labor market analysis", in: Handbook of Labor Economics, Vol. 2, O. Ashenfelter and R. Layard (eds), Amsterdam, North-Holland, 849-919.

MORTENSEN D.T. AND NEUMAN G.R., 1984, "Choice or chance ? A structural interpretation of individual labor market histories", in: Studies in labor market dynamics, edited by G.R. Neuman and N.C. Westergard-Nielsen, Springer-Verlag, Berlin, 98-131.

OLSEN R., SMITH D. AND FARKAS G., 1986, "Structural and reduced-form models of choice among alternatives in continuous time: youth employment under a guaranted jobs program", Econometrica, Vol. 54, 375-394.

ONDRICH J., 1985, "The Initial Conditions Problem in Work History Data", Review of Economics and Statistics, Vol. 67, $n°3$, 411-421.

PHILLIPS P. C. B., 1973, "The problem of identification in finite parameter continuous time models", Journal of Econometrics, Vol. 1, $n°4$, 351-362.

RIDDER G., 1986, "An event history approach to the evaluation of training, recruitment and employment programmes", Journal of applied econometrics, Vol. 1, n°2, 109-126.

ROBERT CH. P. AND G. CASELLA, 2002, Monte Carlo Statistical Methods. Springer.

RUNNENBERG J.TH., 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, ser. A, Mathematical Sciences, Vol. 65, n°5, 536-41.

RUST J., 1994, "Dynamic structural models ; Problems and prospects: discrete decision processes", in C. Sims (ed): Advances in Econometrics, Sixth World Congress, vol. II, Econometric Society Monographs, Cambridge University Press, Cambridge, 119-170.

SAMPSON M., 1990, "A Markov chain model for unskilled workers and the highly mobile", Journal of the American Statistical Association, Vol. 85, n°409, 177- 180.

SINGER B., 1981, "Estimation of nonstationary Markov chains from panel data", Sociological Methodology, Samuel Leinhart-Jossey-Bass Publishers, San Francisco.

SINGER B., 1982, "Aspects of non-stationarity", Journal of Econometrics, Vol. 18, 169-190.

SINGER B. AND SPILERMAN S., 1976a, "The representation of social processes by Markov models", American Journal of Sociology, Vol. 82, n°1, 1-54.

SINGER B. AND SPILERMAN S., 1976b, "Some methodological issues in the analysis of longitudinal surveys", Annals of Economic and Social Measurement, Vol. 5, n° 4, 447-474.

TRIVEDI P.K. AND ALEXANDER J.N., 1989, "Re-employment probability and multiple spells: a partial-likelihood approach", Journal of business and economic statistics, Vol. 7, n°3, 395-401.

TUMA N.B. AND ROBINS P.K., 1980, "A dynamic model of employment behavior: an application to the Seattle and Denver Income Maintenance Experiments", Econometrica, vol. 48, n°4, 1031-1052.

VAN DEN BERG G.J., 1997, "Association measures for durations in bivariate hazard rate models", Journal of Econometrics, Vol. 79, n°2, 221-245.