A Locally Stationary Markov Chain
Model for Labor Dynamics

Enrique E. Alvarez[1], Francisco J. Ciocchini[2] and Kishori Konwar[3]
[1] *Universidad Nacional de La Plata y CONICET*, [2] *Universidad Católica Argentina, and* [3] *University of Connecticut*

Abstract: Labor market surveys usually partition individuals into three states: employed, unemployed, and out of the labor force. In particular, the Argentine " Encuesta Permanente de Hogares (EPH)" follows a rotating scheme so that each selected household is interviewed four times within two years. Each time, the current labor state of individuals is recorded, together with extensive demographic information. We model those labor paths as consecutive observations from independent Markov chains, were transition matrixes are related to covariates through a multivariate logistic link.

Because the EPH is severely affected by attrition, a significant fraction of the surveyed paths contain just one single point. Instead of discarding those observations, we opt to base estimation on the full data by (*i*) assuming the Markov chains are stationary and (*ii*) incorporating the chronological time of the first interview as an additional covariate for each individual. This novel treatment represents a convenient approximation, which we illustrate with data from Argentina in the period 1995-2002 via maximum likelihood estimation. Several interesting labor market indexes, which are functionally related to the transition matrixes, are also presented in the last portion of the paper and illustrated with real data.

Key words: Markov Ccain, sationarity, unemployment.

1. Introduction

In April 1991 the Argentine government launched a major stabilization program, commonly known as the *Convertibility Plan*. At that time, the unemployment rate was around 7%, a figure similar to the average of the previous decade. By May 1995 the unemployment rate had reached 18.4 %, in spite of the rapid growth of the previous four years. After a deep recession in 1995, growth resumed and the Gross Domestic Product (GDP) reached its peak in the second quarter of 1998, while the unemployment rate was down to 12.4% by October of that year. From that point on, the economy entered a protracted recession

that triggered the end of the convertibility program in January of 2002, with unemployment reaching 21.5% in May. By the second quarter of 2003, GDP was still 13 percentage points below its peak and the unemployment rate was 15.6%. Several papers analyze unemployment in Argentina; some take a macroeconomic perspective (*e.g.*, Frenkel and Ros 2004, Llach and Llach 1998), others take a regional approach (*e.g.*, Lamarche, Porto, and Sosa-Escudero 1998), while others use individual data to estimate either conditional models (Servy, Hachuel, Boggio, and Cuesta 2003) or duration models (Galiani and Hopenhayn 2003, and Hopenhayn 2004).

In this paper we take a novel approach and study the labor dynamics of the population by fitting a locally stationary Markov Chain to the Argentine official labor survey. We focus on how individual characteristics affect people's performance in the labor market. Data are taken from the Argentine *Encuesta Permanente de Hogares* (EPH), which follows a *rotating* scheme: data are collected at two *waves* each year in the months of May and October. Once selected, each household is interviewed at four consecutive waves (*i.e.*, 2 years) and in each new wave a fourth of the sampled is renewed. Within a household, extensive demographic information is collected on all its members. Labor information is collected in a *current-status-censored* fashion; *i.e.* it is only known with certainty the employment status of each participating adult precisely at the interview times. Individuals are partitioned into three possible states: employed (E), unemployed (U), and out of the labor force (O). This entails for each person in the survey a labor path, *i.e.* a 4-tuple of labor states, which we view as a realization of a time-homogeneous Markov chain. Its directing transition matrix is functionally related to the individual´s characteristics. In that set-up, all *labor dynamics* information is contained in the transition matrixes. They determine the *long-run unemployment rate*, the *employment availability or reliability*, the *labor inertia*, or other labor indexes that we define in Section 4.

As for the need of stationarity, it is paramount to observe that the EPH is severely affected by *attrition*. As we will see in Section 3, almost 50% of the labor paths in the dataset consisted of just one time point. Thus, if we were to discard those observations we would be losing about half the sample. In a addition to a severe sample size reduction, this entails the danger of seriously biasing the results. We deemed this aspect of the dataset was crucial and needed special care. We opted to base estimation on the full data by assuming the Markov chains were stationary, but also incorporating the chronological time of the first interview as an additional covariate for each individual. This treatment is novel; it is equivalent to assuming labor paths are stationary throughout the two years people remain in the sample, but it allows simultaneously that the parameters may change in time. Also interestingly from the economic point of

view, this formulation enables for testing for a change on the *labor dynamics* in time, through the inclusion of appropriate interaction terms.

The application of Markovian-type processes to model the labor market has, by now, a fairly vast background in the Economics literature. Classical references are, among others, Flinn and Heckman (1982), Burdett, Kiefer, Mortensen and Neumann (1984) and Burdett, Kiefer and Sharma (1985). An excellent review until the late 80's can be found in Lancaster (1990). An up-to-date survey of estimation using search-theoretic structural models is Eckstein and van den Berg (2005). Estimation of reduced-form transition models is reviewed in Fougère and Kamionka (2005).

This paper is organized as follows. In Section 2 we present the model and we discuss issues on estimation. In Section 3 we fit the model to the Argentine data and we present some interpretations. Finally, in Section 4 we construct some indexes from the labor transition matrixes and we illustrate with an example how they exhibit different aspects of the dynamics of the population in the Labor Market.

2. Stationary Markov Chain Labor Model (SMCM)

In this Section we present a statistical model able to (i) estimate the typical path for people with given covariates; (ii) asses how a person's labor market performance is affected by his/her characteristics. Specifically, we view each person's labor path $\mathbf{w} := (w_1, w_2, w_3, w_4)$ as generated by a time-homogeneous Markov chain in the state space $\mathcal{K} := \{\mathsf{E}, \mathsf{U}, \mathsf{O}\}$ with transition probability matrix

$$
\boldsymbol{\theta} = \begin{pmatrix} \theta_{\mathsf{EE}} & \theta_{\mathsf{EU}} & \theta_{\mathsf{EO}} \\ \theta_{\mathsf{UE}} & \theta_{\mathsf{UU}} & \theta_{\mathsf{UO}} \\ \theta_{\mathsf{OE}} & \theta_{\mathsf{OU}} & \theta_{\mathsf{OO}} \end{pmatrix}. \tag{2.1}
$$

Each chain is related to the individual's characteristics, as measured by a vector of *covariates* or *independent variables* $\mathbf{y}$ in the following manner. First we propose a *multivariate logistic link* $\boldsymbol{\eta} := \boldsymbol{\eta}(\boldsymbol{\theta})$ given by

$$
\eta_{\mathsf{EU}} := \log\left(\frac{\theta_{\mathsf{EU}}}{\theta_{\mathsf{EE}}}\right) \qquad \eta_{\mathsf{EO}} := \log\left(\frac{\theta_{\mathsf{EO}}}{\theta_{\mathsf{EE}}}\right)
$$

$$
\eta_{\mathsf{UE}} := \log\left(\frac{\theta_{\mathsf{UE}}}{\theta_{\mathsf{UU}}}\right) \qquad \eta_{\mathsf{UO}} := \log\left(\frac{\theta_{\mathsf{UO}}}{\theta_{\mathsf{UU}}}\right) \tag{2.2}
$$

$$
\eta_{\mathsf{OE}} := \log\left(\frac{\theta_{\mathsf{OE}}}{\theta_{\mathsf{OO}}}\right) \qquad \eta_{\mathsf{OU}} := \log\left(\frac{\theta_{\mathsf{OU}}}{\theta_{\mathsf{OO}}}\right).
$$

With this link $\boldsymbol{\eta} := (\eta_{\mathsf{EU}}, \eta_{\mathsf{EO}}, \eta_{\mathsf{UE}}, \eta_{\mathsf{UO}}, \eta_{\mathsf{OE}}, \eta_{\mathsf{OU}}) \in I\!\!R^6$ and each transition probability is strictly between 0 and 1, so that the chain is irreducible and aperiodic. These parameters are related to the independent variables, which in addition to the chronological time ($\mathtt{t}$) a household was first interviewed, include age ($\mathtt{X}$), gender ($\mathtt{G}$), and education ($\mathtt{M,H}$). Their values are all fixed at the time of the first interview so that we have *time-independent* covariates in the terminology of Survival Analysis. Specifically, we propose

$$
\begin{aligned}
\eta_{\mathsf{EU}} &:= \beta_{\mathsf{EU}}^{(0)} + \beta_{\mathsf{EU}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{EU}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{EU}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{EU}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{EU}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{EU}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{EU}}^{(\mathsf{H})}\ \mathsf{H} \\
\eta_{\mathsf{EO}} &:= \beta_{\mathsf{EO}}^{(0)} + \beta_{\mathsf{EO}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{EO}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{EO}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{EO}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{EO}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{EO}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{EO}}^{(\mathsf{H})}\ \mathsf{H} \\
\eta_{\mathsf{UE}} &:= \beta_{\mathsf{UE}}^{(0)} + \beta_{\mathsf{UE}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{UE}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{UE}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{UE}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{UE}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{UE}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{UE}}^{(\mathsf{H})}\ \mathsf{H} \\
\eta_{\mathsf{UO}} &:= \beta_{\mathsf{UO}}^{(0)} + \beta_{\mathsf{UO}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{UO}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{UO}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{UO}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{UO}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{UO}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{UO}}^{(\mathsf{H})}\ \mathsf{H} \\
\eta_{\mathsf{OE}} &:= \beta_{\mathsf{OE}}^{(0)} + \beta_{\mathsf{OE}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{OE}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{OE}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{OE}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{OE}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{OE}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{OE}}^{(\mathsf{H})}\ \mathsf{H} \\
\eta_{\mathsf{OU}} &:= \beta_{\mathsf{OU}}^{(0)} + \beta_{\mathsf{OU}}^{(\mathtt{t})}\ \mathtt{t} + \beta_{\mathsf{OU}}^{(\mathtt{t2})}\ \mathtt{t}^2 + \beta_{\mathsf{OU}}^{(\mathtt{x})}\ \mathtt{x} + \beta_{\mathsf{OU}}^{(\mathtt{x2})}\ \mathtt{x}^2 + \beta_{\mathsf{OU}}^{(\mathsf{G})}\ \mathsf{G} + \beta_{\mathsf{OU}}^{(\mathsf{M})}\ \mathsf{M} + \beta_{\mathsf{OU}}^{(\mathsf{H})}\ \mathsf{H},
\end{aligned}
$$

where G is 0 for males or 1 for females, M is 1 for individuals with *medium* education or 0 for (*low* or *high*), and H is 1 for individuals with *high* education or 0 otherwise. With the above specification, the model accounts for possible quadratic time trends as well as main effects in sex and education, and a quadratic term in age.

It is noteworthy that the functional relation $\boldsymbol{\theta} \mapsto \boldsymbol{\eta}$ expressed in Equation (2.2) is one-to-one. To reconstruct the transition probability matrix from the η's we use the inverse mapping:

$$
\begin{aligned}
&\theta_{\mathsf{EE}} = \frac{1}{1 + e^{\eta_{\mathsf{EU}}} + e^{\eta_{\mathsf{EO}}}} &&\theta_{\mathsf{EU}} = \frac{e^{\eta_{\mathsf{EU}}}}{1 + e^{\eta_{\mathsf{EU}}} + e^{\eta_{\mathsf{EO}}}} &&\theta_{\mathsf{EO}} = \frac{e^{\eta_{\mathsf{EO}}}}{1 + e^{\eta_{\mathsf{EU}}} + e^{\eta_{\mathsf{EO}}}} \\[2mm]
&\theta_{\mathsf{UE}} = \frac{e^{\eta_{\mathsf{UE}}}}{1 + e^{\eta_{\mathsf{UE}}} + e^{\eta_{\mathsf{UO}}}} &&\theta_{\mathsf{UU}} = \frac{1}{1 + e^{\eta_{\mathsf{UE}}} + e^{\eta_{\mathsf{UO}}}} &&\theta_{\mathsf{UO}} = \frac{e^{\eta_{\mathsf{UO}}}}{1 + e^{\eta_{\mathsf{UE}}} + e^{\eta_{\mathsf{UO}}}} \quad (2.3) \\[2mm]
&\theta_{\mathsf{OE}} = \frac{e^{\eta_{\mathsf{OE}}}}{1 + e^{\eta_{\mathsf{OE}}} + e^{\eta_{\mathsf{OU}}}} &&\theta_{\mathsf{OU}} = \frac{e^{\eta_{\mathsf{OU}}}}{1 + e^{\eta_{\mathsf{OE}}} + e^{\eta_{\mathsf{OU}}}} &&\theta_{\mathsf{OO}} = \frac{1}{1 + e^{\eta_{\mathsf{OE}}} + e^{\eta_{\mathsf{OU}}}}.
\end{aligned}
$$

We have exploited this inverse mapping in the numerical algorithms described in the Appendix.

2.1 Stationarity

Because the observed labor chains are short and, worse, a substantial portion of the surveyed paths contain just one time point, it makes sense to try to use the initial states of each history to aid in the estimation. We achieve this by assuming stationary Markov chains. In the context of chains of length 2 it was

shown in Alvarez (2006) that this method can achieve large gains in asymptotic efficiency comparative to a method that discards or " conditions away" the initial states from each path.

Naturally, we realize that the labor market is by no means stationary and our treatment accommodates this, since calendar time t is among the independent variables in the model. We claim that because we only get to observe each individual's labor history for at most two years stationarity is a very reasonable approximation. Thus each person's labor history is assumed to be a realization of a stationary Markov chain over 4 consecutive states; but people that incorporate to the survey at different times may well face different parameters. On a more pedantic note, observe that assuming that individual labor paths are independent is also questionable. Given the macroeconomic conditions, the number of jobs available for a person with a given set of skills is finite. Thus, finding that a given person is employed reduces the probability that others with the same characteristics are also employed. However, because the population is large relative to the sample, we consider independence as a very reasonable approximation for our dataset. Hence, even though they may not hold strictly, both independence and stationarity are very convenient and reasonable approximations for the EPH.

The stationary vector $\boldsymbol{\pi}$ satisfies the usual conditions $\boldsymbol{\theta}^T \boldsymbol{\pi} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}^T \mathbf{1} = 1$; explicit expressions for $\boldsymbol{\pi}$ are given in the Appendix. Here we note that since transition matrixes are irreducible, unique stationary distributions exist with state probabilities that are strictly positive. Furthermore, those stationary probabilities are a continuous function of the transition matrix (e.g., Chung 1967). Therefore with stationarity, the initial labor state observed for each person provides information on the θ's and thus the β's.

2.2 Likelihood maximization

The contribution of a person with labor path (w_1, w_2, w_3, w_4) to the sample likelihood is a product $L = \pi_{w_1} \; \theta_{w_1 w_2} \; \theta_{w_2 w_3} \; \theta_{w_3 w_4}$, where each of the factors is a function of the global parameter $\boldsymbol{\beta}$ and the individual's covariate values $\mathbf{y}_i$. The sample log-likelihood is, by independence, $\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta})$ where

$$l_i(\boldsymbol{\beta}) = \log \pi_{w_0^i} \left(\boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{y}_i) \right) + \log \theta_{w_0^i w_1^i}(\boldsymbol{\beta}, \mathbf{y}_i)$$
$$+ \log \theta_{w_1^i w_2^i}(\boldsymbol{\beta}, \mathbf{y}_i) + \log \theta_{w_2^i w_3^i}(\boldsymbol{\beta}, \mathbf{y}_i). \quad (2.4)$$

The first term, namely $\theta_{w_0^i w_1^i}(\boldsymbol{\beta}, \mathbf{y}_i)$, contains the information provided by the initial state of an individual labor path. For those individuals observed only at one time point, this would be the sole term in the individual's contribution to the likelihood. It is important to notice that, without stationarity, those individuals would provide no information about the transition probabilities (the θ's) nor their

directing parameters (the β's), simply because they show no transition. On the other hand, assuming stationarity makes this first term a function of the β's and hence contribute to its estimation. This is the indirect way in which assuming stationarity makes "usable" even those labor histories observed at just one point. For a longer discussion of the implications of stationarity in the context of 0-1 Markov chains, we refer the reader to Alvarez(2003).

On the computational side, it is a nuisance that formulae for the maximum likelihood estimators of $\boldsymbol{\beta}$ are not available in closed form. In order to maximize the log-likelihood we developed a numerical algorithm in C++, which is available from the authors upon request. It involves two choices: (i) a selection of initial estimators $\dot{\boldsymbol{\beta}}$ and (ii) a rule to update each iteration towards the maximization of the log-likelihood.

Initialization We estimate the overall transition matrix non-parametrically, ignoring all covariate information. For example, in the case of the transitions from E to U, we calculate the ratio of all such transitions relative to all transitions that start in E. That gives $\dot{\theta}_{\mathsf{EU}}^{(0)} := \#(\mathsf{E} \to \mathsf{U})/\#\mathsf{E}\cdot$. Next, we set the initial estimator for the intercept to be $\dot{\beta}_{\mathsf{EU}}^{0} = \log(\dot{\theta}_{\mathsf{EU}}^{(0)}/\dot{\theta}_{\mathsf{EE}}^{(0)})$ and the remaining coefficients are set to zero, *i.e.* $\dot{\beta}_{\mathsf{EU}} = 0$.

Updates We use a modified quasi-Newton method, called the *Broyden-Fletcher-Goldfarb-Shanno* algorithm, which is described in detail in Press, Flannery, Teukolsky, and Vetterlin (2002).

With actual data, estimation of the SMCM model may well be impossible as log-likelihood in Equation (2.4) is not necessarily convex for small samples. This phenomenon also occurs for processes without explanatory variables. It has been analyzed in the context of renewal processes by Vardi (1982) and in the context of discrete and continuous Alternative Renewal Processes by Alvarez (2006); those papers give formal conditions for the existence and uniqueness of the MLEs in finite samples, in their respective contexts. We do not attempt to characterize the irregular data configurations in this article. Instead, we conclude by mentioning that degenerate situations only occur when the number of histories is small, or the number of covariates is large. They are usually detected when a Hessian matrix for the sample log likelihood that is not negative definite or when the maximization algorithm fails to converge in a reasonable number of iterations. We recommend to the practitioner in those situations the usual remedies: to enlarge the sample size, drop variables, and/or collapse categorical variables into fewer levels.

3. Argentine Unemployment

In this section we fit the stationary Markov chain model to the labor market in Argentina during the period Oct. 1995 through May 2003. As mentioned in the Introduction, the dataset is taken from the *Argentine Permanent Household Survey* (EPH) and is available online after free registration from the website "http://www.indec.gov.ar/". Even though retrospective continuous information is collected by the EPH on employment tenures and unemployment durations, that information is disregarded as unreliable by official statisticians. We have preferred, instead, to model current labor state information at the interview times. Those records are considered very accurate by EPH personnel, as they go through a set of consistency checks and data-quality-controls.

It is also noteworthy that in EPH is a sample of *households*, not individuals. This entails that there could exist a "household effect" in the labor process. Two examples of how this effect may operate in practice are the following: (*i*) for an unemployed person who has many of his/her housemates unemployed there may be more motivation for finding employment due to the need to support the family, thus increasing the chance of moving to E; or (*ii*) having many housemates unemployed may be the consequence of the household having the wrong skills or some characteristic that potential employers find unnatractive, making it more difficult for the person to find employment. In addition, because individuals are sampled within households, individual labor paths are not truly independent, as they share a household-level kind of correlation. Having acknowledged these effects, we have opted to ignore the household hierarchal structure in our model. We believe that this omission, which represents an important gain in model parsimony, cannot affect the estimates by much. This is because the number of households is very large relative to the number of individuals within a household. Notwithstandingly, the authors are presently investigating the incorporation of "frailty effects" to the model. This extension would allow for the treatment of household effects as well as other types of latent factors.

Table 1: Length of the observed labor histories

	$k=1$	$k=2$	$k=3$	$k=4$	Total
Number of Histories	20,386	6,510	5,598	13,045	45539
Percentage	44.77%	14.30%	12.30%	28.65%	100

We include all labor histories for people of age 18–75 in the Greater Buenos Aires region, which includes the country's capital and its 24 neighboring districts. This area represents 33% of the country's population. The time variable $\mathtt{t}$ is set to represent $\mathtt{t=12}$ for October 1995, $\mathtt{t=13}$ for May 1996, and so forth until $\mathtt{t=27}$

for May 2003. Data contains labor histories for 45,539 people, of which only 28.8% are complete paths exhibiting all four labor states. The distribution of the length of the labor histories is presented next in Table 1.

In the sample the composition by gender was Males 47% and females 53%, and by educational levels : *Low* (54%) *Medium* (23%) and *High* (23%). The median age was 39 years and the first and third quartiles were respectively 26 and 53. Further, the aggregate distribution of employment states throughout the was E:= 55.68%, U := 10.71% and O:=33.61%. It implies an overall unemployment rate of 16.13% throughout the period.

Table 2: Estimated coefficients for Argentine data

	EU	EO	UE	UO	OE	OU
$\beta^{(0)}$	2.8612	-1.2748	-1.3362	2.7874	-8.5305	-5.8052
	0.3792	*0.4309*	*0.5919*	*0.6337*	*0.3831*	*0.4823*
$\beta^{(t)}$	-0.1762	-0.0487*	-0.0510*	0.0638*	0.1734	-0.1121*
	0.0395	*0.0442*	*0.0628*	*0.0680*	*0.0413*	*0.0529*
$\beta^{(t2)}$	0.0053	0.0013*	0.0008*	-0.0012*	-0.0036	0.0029*
	0.0010	*0.0012*	*0.0016*	*0.0018*	*0.0011*	*0.0014*
$\beta^{(x)}$	-0.1732	-0.1204	0.1504	-0.2881	0.2527	0.3032
	0.0050	*0.0066*	*0.0091*	*0.0091*	*0.0040*	*0.0061*
$\beta^{(x2)}$	0.0018	0.0016	-0.0021	0.0036	-0.0029	-0.0039
	0.0001	*0.0001*	*0.0001*	*0.0001*	*0.0000*	*0.0001*
$\beta^{(G)}$	-0.2617	2.1373	-0.3050	0.9568	-0.4216	-0.7552
	0.0353	*0.0418*	*0.0530*	*0.0576*	*0.0378*	*0.0521*
$\beta^{(M)}$	-0.3016	-0.2877	-0.2128	0.0335*	0.1335	0.2721
	0.0382	*0.0408*	*0.0548*	*0.0632*	*0.0401*	*0.0476*
$\beta^{(H)}$	-0.7738	-1.0825	-0.3767	0.5034	0.6546	0.4931
	0.0389	*0.0493*	*0.0643*	*0.0705*	*0.0448*	*0.0589*

In Table 2 we present the MLE's and their standard errors for the Argentine labor data. In *italics* underneath each number we present the approximate standard errors, estimated by observed Fisher information. It is noteworthy that chronological time is significant in the model. It represents the influence of exogenous macroeconomic or country-specific conditions in the individual labor dynamics. Interestingly, the time variable t is only relevant for the transitions EU and OE and in both cases the effect is quadratic in the 'log-odds' scale. In Figure 1 we exhibit the transition probability estimates against chronological time for a representative highly educated 40-year old male. Its shows a monotonically increasing trends for both type of transitions EU and OE. The remaining trends are not-significant.
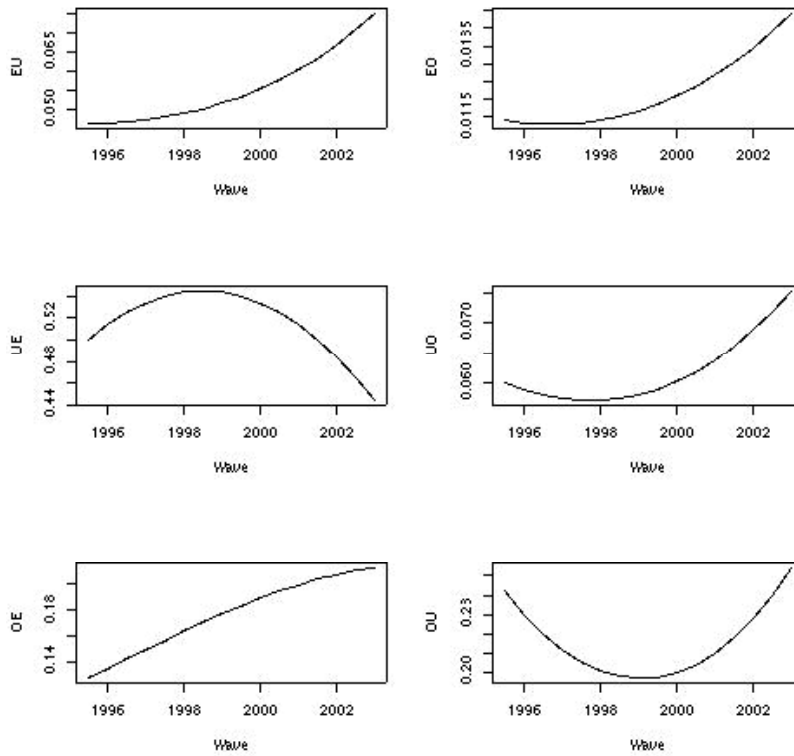
Figure 1: Transition probability trends for highly educated 40-year old males

All other coefficients are significant, with the exception of medium education for transitions UO. This means that people with medium education are indistinguishable from people with low education in terms of their odds of moving from U to O, relative to staying in U. The Education variable as a whole remains significant. There is a quadratic effect of age in all the transitions. Its maxima and minima are given by Table 3.

Table 3: Maxima and minima for the effect of Age

	EU	EO	UE	UO	OE	OU
Age	48	37	35	40	44	39
	min	min	max	min	max	max

As for the coefficient of Gender, it is interesting that while women are less likely than men to lose their jobs $\hat{\beta}_{EU}^{(G)} = -0.2617$, they are also less likely to

find a position when they are unemployed $\hat{\beta}_{\mathsf{UE}}^{(\mathsf{G})} = -0.3050$. Also, females are more prone that men to move out of the labor market, either when employed or unemployed and they are more reluctant to come back to the labor market (either to get a job or to look for one) than males; this is because $\hat{\beta}_{\mathsf{OE}}^{(\mathsf{G})}$ and , $\hat{\beta}_{\mathsf{OE}}^{(\mathsf{G})}$ are both negative.

Finally, in order to assess the predictive power of the SMC model we conclude calculate, in analogy with the linear models, a *pseudo-R^2* statistic, *i.e.*,

$$R^2 = 1 - [l_\Omega(\tilde{\boldsymbol{\beta}})/l_\Gamma(\tilde{\boldsymbol{\beta}})] = 0.1360,$$

where l_Ω is the log-likelihood of our model and l_Γ is the log-likelihood of the null model (*i.e.* a model which only includes intercepts). Observe that the R^2 coefficient is rather small but, at the same time, the chosen covariates appear very significant (as their standard errors place almost all of them away from zero in Table 2). This apparent paradox has an interesting interpretation, *i.e.*, that the proposed independent variables are indeed influential in people's labor dynamics, but they are not, as a whole, sufficient to predict typical labor paths. The latter may be either because more independent variables are needed, or because there is an inherent variability in labor experiences even after including all relevant covariate information. We plan to examine these aspects, including the addition of independent variables, interactions and frailty effects in a future paper, specially tailored to Economic Data-Analysis.

4. Labor Dynamics Functionals

Individuals' experience in the labor market is a complex phenomenon which reaches far beyond the unemployment rate. For illustration we consider the predicted labor matrix for men and women aged 55 with high education:

$$\boldsymbol{\theta}_M = \begin{pmatrix} 0.9573 & 0.0321 & 0.0106 \\ 0.3191 & 0.4765 & 0.2045 \\ 0.1798 & 0.0873 & 0.7329 \end{pmatrix}, \quad \boldsymbol{\theta}_F = \begin{pmatrix} 0.9023 & 0.0263 & 0.0714 \\ 0.1571 & 0.3487 & 0.4942 \\ 0.1233 & 0.0374 & 0.8393 \end{pmatrix}.$$

The stationary vector is given respectively by $\boldsymbol{\pi}_M = (0.8550, 0.0662, 0.0843)$ with male unemployment $U_M = 0.0722$ and $\boldsymbol{\pi}_F = (0.5649, 0.0452, 0.3990)$ with female unemployment $U_F = 0.0740$. Even though unemployment rates are almost identical, a more careful look unveils important differences in the labor market dynamics. To facilitate the description of the labor market performance we have opted to define the following indexes

Activity rate $a(t) := \pi_{\mathsf{E}}(t) + \pi_{\mathsf{U}}(t)$ measures the proportion of people in the labor market either employed or seeking work.

Unemployment Rate $u(t) := \pi_U(t)/(\pi_U(t) + \pi_E(t))$ measures the proportion of those individuals in the labor market who do not have a job.

Inertia $\mathcal{I}(t) := 3^{-1}\text{trace}[\theta(t)]$ it measures the proclivity of the current employment state to perpetuate in time.

Give-up rate $g(t) := \theta_{UO}(t)/(\theta_{UU}(t) + \theta_{UO}(t))$ is the probability, for the unemployed, that a transition to O occurs before a transition to E.

Net Outflow $n(t) := \pi_E(t)\theta_{EO}(t) - \pi_O(t)\theta_{OE}(t)$ is the difference between the percentage of people who leave employment in order to join the idle and those who transit from being out of the labor force to a job tenure.

Average Durations Employed, Unemployed and out of the Labor Force The "number of trials" until transitions are geometric random variables with expectations $\bar{E} = (1 - \theta_{EE}(t))^{-1}$, $\bar{U} = (1 - \theta_{UU}(t))^{-1}$ and $\bar{O} = (1 - \theta_{OO}(t))^{-1}$.

Reliability Its the probability the person retires from the labor force before becoming unemployed $\rho(t) := \theta_{EO}(t)/(1 - \theta_{EE})(t)$

Availability Its the probability the person leaves inactivity in order to find employment before becoming unemployed $\alpha(t) := \theta_{OE}(t)/(1 - \theta_{OO}(t))$

For illustration, in Table 4 we exhibit those labor indexes for males and females of high education, 55 years of age in May 1999. For both groups the unemployment rate is about the same 7% and the average unemployment spell last for about 1.5 to 2 waves (*i.e.* 9 to 12 months); further, their inertia is very similar as well as their job availability. However those two groups behave very differently in the labor market. First of all we notice that while 92% of the males are active (*i.e.* either employed or seeking work) the corresponding figure for females is only 61%. Secondly while the *net outward flow* for males is positive, women are opting to retire from the labor force and; additionally the *give up* rate for females is twice that of males. Finally, while the job reliability for men is about three quarters, the corresponding figure for females is just one quarter.

Table 4: Labor market functionals by sex for highly educated 55 year old subjects in May 1999

Sex	a	u	$\mathcal{I}$	g	n	$\bar{E}$	$\bar{U}$	$\bar{O}$	ρ	α
M	0.92	0.0722	0.7222	0.3003	0.0431	23.43	1.91	3.74	0.2477	0.67
F	0.61	0.0740	0.6968	0.5863	-0.0070	10.24	1.54	6.22	0.7311	0.77

Next, in Figure 2 we exhibit how labor market functionals depend on the individual's age for males and females of low education.
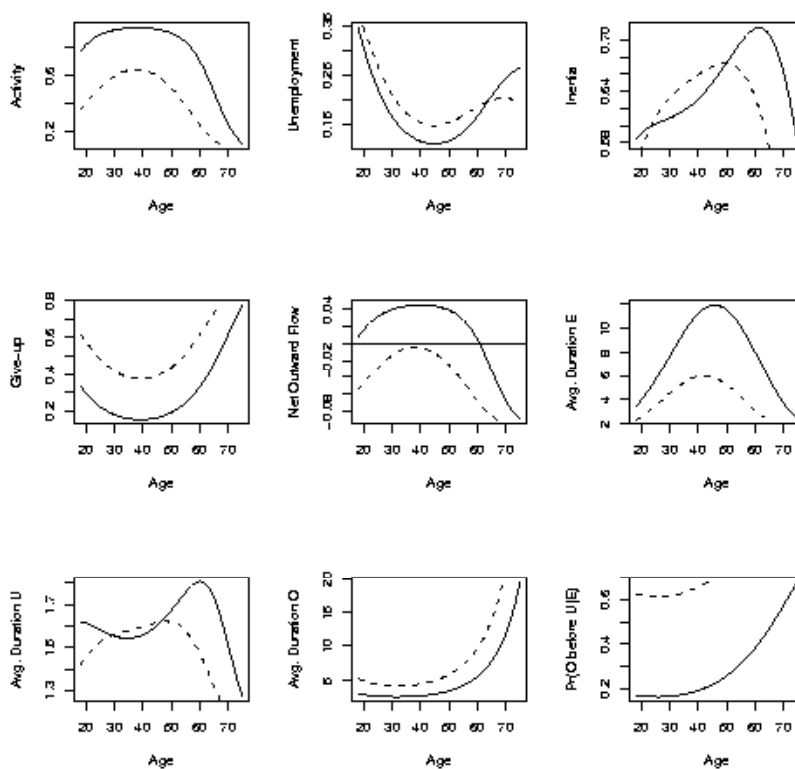
Figure 2: Low Education. Solid=Males, Dashed=Females.

First we notice,as expected, that the activity rate a is larger for males of all ages with peaks at about 40 years of age. The unemployment for both males and females has a draught at about 45 years of age and is larger for females than for men at ages 20-60, after which the direction is reversed. The inertia reaches peaks at 50 and 65 for females and males respectively. Interestingly the *net outward flow* is negative for females of all ages, and is positive for males up to the age of 60.

There are wide differences in the average lengths of employment spells for males and less so for women, with peaks at ages 45-50. The average length of an unemployment spell also varies with age. For women they achieve a peak at age 40 and for men the behavior is apparently bimodal with peaks at ages 20 and 60 and a drought at about 38. The average duration of a spell out of the labor force is monotonically increasing for males and females with high slopes after the age of about 55. The reliability of employment is much larger for men than for females of all ages.

The interpretation of the give up coefficient requires care; this is because in practice it is often times vague to classify a person as unemployed or out of the labor force, due to the technical definitions of those concepts. Exactly because of this arbitrariness in the definition of unemployment, Clark and Summers (1979) define the probability of transiting from unemployment to employment for what they call " indomitable job seekers" as $\theta^*_{\mathsf{UE}} := \theta_{\mathsf{UE}}/(\theta_{\mathsf{UE}} + \theta_{\mathsf{UU}})$ (the probability of finding a job, conditional on not dropping out of the labor force). Using this, we can calculate the average duration of an unemployment spell for an indomitable job seeker as $\overline{\mathsf{U}}^* := (1/\theta^*_{\mathsf{UE}})$. For illustration, we continue looking at the dynamics of males and females of age 55 with high education to get:

	θ^*_{UE}	$\overline{\mathsf{U}}^*$
Males	0.401	2.49
Females	0.311	3.22

Notice that, while $\overline{\mathsf{U}}$ is higher for males (1.91 vs. 1.54, see Table 4), $\overline{\mathsf{U}}^*$ is lower (2.49 vs. 3.22). This reflects the fact that the estimated probability of transiting from unemployment to inactivity (θ_{UO}) is much higher for women (0.49 vs. 0.20).

Appendix A. Stationarity

The vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\theta}^T \boldsymbol{\pi} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}^T \mathbf{1} = 1$ and is specifically given by

$$\pi_\mathsf{E}(\boldsymbol{\theta}) = \frac{\mathsf{Num}_\mathsf{E}(\boldsymbol{\theta})}{\mathsf{Den}(\boldsymbol{\theta})}, \quad \pi_\mathsf{U}(\boldsymbol{\theta}) = \frac{\mathsf{Num}_\mathsf{U}(\boldsymbol{\theta})}{\mathsf{Den}(\boldsymbol{\theta})}, \quad \text{and} \quad \pi_\mathsf{O}(\boldsymbol{\theta}) = \frac{\mathsf{Num}_\mathsf{O}(\boldsymbol{\theta})}{\mathsf{Den}(\boldsymbol{\theta})},$$

where the numerators above are $\mathsf{Num}_\mathsf{E}(\boldsymbol{\theta}) = \theta_{\mathsf{OE}} \theta_{\mathsf{UE}} + \theta_{\mathsf{UE}} \theta_{\mathsf{OU}} + \theta_{\mathsf{OE}} \theta_{\mathsf{UO}}$, $\mathsf{Num}_\mathsf{U}(\boldsymbol{\theta}) = \theta_{\mathsf{OE}} \theta_{\mathsf{EU}} + \theta_{\mathsf{EU}} \theta_{\mathsf{OU}} + \theta_{\mathsf{EO}} \theta_{\mathsf{OU}}$, and $\mathsf{Num}_\mathsf{O}(\boldsymbol{\theta}) = \theta_{\mathsf{EU}} \theta_{\mathsf{UO}} + \theta_{\mathsf{EO}} \theta_{\mathsf{UE}} + \theta_{\mathsf{EO}} \theta_{\mathsf{UO}}$. The common denominator equals

$$\mathsf{Den}(\boldsymbol{\theta}) = \mathsf{Num}_\mathsf{E}(\boldsymbol{\theta}) + \mathsf{Num}_\mathsf{U}(\boldsymbol{\theta}) + \mathsf{Num}_\mathsf{O}(\boldsymbol{\theta}) = \theta_{\mathsf{EO}} \theta_{\mathsf{OU}} + \theta_{\mathsf{UE}} \theta_{\mathsf{OU}} + \theta_{\mathsf{EU}} \theta_{\mathsf{OU}}$$
$$+ \theta_{\mathsf{OE}} \theta_{\mathsf{UO}} + \theta_{\mathsf{EU}} \theta_{\mathsf{UO}} + \theta_{\mathsf{OE}} \theta_{\mathsf{EU}} + \theta_{\mathsf{OE}} \theta_{\mathsf{UE}} + \theta_{\mathsf{EO}} \theta_{\mathsf{UE}} + \theta_{\mathsf{EO}} \theta_{\mathsf{UO}}.$$

Appendix B. Log-likelihood evaluation

Equation (2.4) apart from its interest in its own right suggest an algorithm for calculation. The method proceeds by scrolling down through the labor histories and for each person perform the following tasks:

1. Calculate the individual's transition matrix $\boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{y}_i)$ using the overall vector $\boldsymbol{\beta}$ and the values of the person's covariates $\mathbf{y}_i$ in two steps:

    (a) Calculate $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{y}_i)$

    (b) Calculate $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\beta}, \mathbf{y}_i)$ from the inverse mapping of Equation (2.2)

2. Calculate the persons stationary vector $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta}, \mathbf{y}_i)$.

3. Calculate the person's contribution to the log-likelihood

$$l_i(\boldsymbol{\beta}, \mathbf{y}_i) = \log \pi_{i_0} + \log \theta_{i_0 i_1} + \log \theta_{i_1 i_2} + \log \theta_{i_2 i_3}.$$

4. Sum up over all people in the sample, *i.e.*, $1 \leq i \leq n$.

Appendix C. Gradient and Hessian

Calculation of the derivative of the log-likelihood with respect to $\boldsymbol{\beta}$ is accomplished by using the chain rule in Equation (2.4) to obtain

$$\frac{\partial \ell_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}_i(\boldsymbol{\eta})^T}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{y}_i)^T}{\partial \boldsymbol{\beta}}. \tag{C.1}$$

The last two matrixes in each summand are readily available:

$$\underbrace{\frac{\partial \boldsymbol{\theta}(\boldsymbol{\eta})^T}{\partial \boldsymbol{\eta}}}_{(6 \times 6)} = \begin{pmatrix} \theta_{\mathsf{EU}}(1 - \theta_{\mathsf{EU}}) & -\theta_{\mathsf{EU}}\theta_{\mathsf{EO}} & 0 & 0 & 0 & 0 \\ -\theta_{\mathsf{EU}}\theta_{\mathsf{EO}} & \theta_{\mathsf{EO}}(1 - \theta_{\mathsf{EO}}) & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{\mathsf{UE}}(1 - \theta_{\mathsf{UE}}) & -\theta_{\mathsf{UE}}\theta_{\mathsf{UO}} & 0 & 0 \\ 0 & 0 & -\theta_{\mathsf{UE}}\theta_{\mathsf{UO}} & \theta_{\mathsf{UO}}(1 - \theta_{\mathsf{UO}}) & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{\mathsf{UE}}(1 - \theta_{\mathsf{UE}}) & -\theta_{\mathsf{UE}}\theta_{\mathsf{UO}} \\ 0 & 0 & 0 & 0 & -\theta_{\mathsf{UE}}\theta_{\mathsf{UO}} & \theta_{\mathsf{UO}}(1 - \theta_{\mathsf{UO}}) \end{pmatrix}.$$

and

$$\underbrace{\frac{\partial \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{y}_i)^T}{\partial \boldsymbol{\beta}}}_{(6 \times 48)} = \begin{pmatrix} \mathbf{v}_i & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_i & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{v}_i \end{pmatrix}, \tag{C.2}$$

where $\mathbf{v}_i = (1, \mathbf{y}_i)$ is a row vector with components 1 (for the intercepts) and the values of the individual's covariates $\mathbf{y}_i$, and $\mathbf{0}$ is a vector of zeroes.

Explicit calculation of the Hessian matrix is accomplished by using the chain rule in Equation (2.4) to obtain

$$\frac{\partial^2 \ell_n}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} = \sum_{i=1}^n \left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right) \left(\frac{\partial \boldsymbol{\theta}_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \right) \left\{ \left(\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right) \left(\frac{\partial \boldsymbol{\theta}_i(\boldsymbol{\eta})^T}{\partial \boldsymbol{\eta}} \right) \right.$$

$$\left. + \left[\left(\frac{\partial l_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \otimes \mathbf{1}_{(6 \times 6)} \right] \left(\frac{\partial^2 \boldsymbol{\theta}_i(\boldsymbol{\eta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\eta}} \right) \right\} \left(\frac{\partial \boldsymbol{\eta}_i(\boldsymbol{\beta})^T}{\partial \boldsymbol{\beta}} \right). \tag{C.3}$$

Formulae in this appendix are part of an algorithm implemented in C++ and and available from the authors upon request.

References

Alvarez, E. E. (2003). *Likelihood Based Estimation of Stationary Semi-Markov Processes under Window Censoring.* Ph.D. Thesis. The University of Michigan, Ann Arbor.

Alvarez, E. E. (2006). Maximum likelihood estimation in alternating renewal processes under window censoring. *Stochastic Models* **22** 55-76.

Burdett, K., Kiefer, N., Mortensen, D., and Neumann, G. (1984). Earnings, unemployment, and the allocation of time over time, *The Review of Economic Studies* **51**, 559-578.

Burdett, K., Kiefer, N. and Sharma, S. (1985). Layoffs and duration dependence in a model of turnover. *Journal of Econometrics* **28**, 51-69.

Clark, K. and Summers, L. (1979). Labor market dynamics and unemployment: A reconsideration, *Brookings Papers on Economic Activity*, **1** 13-60.

Chung, K. L. (1967). *Markov chains with stationary transition probabilities*, Springer.

Eckstein, Z. and van den Berg, G. (2007). Empirical labor search: A survey, *Journal of Econometrics* **136**, 531-564.

Flinn, C. and Heckman, J. (1982). New methods for analyzing structural models of labor force dynamics, *Journal of Econometrics*, **18**, 115-168.

Fougere, D. and Kamionka, T. (2005). Econometrics of individual labor market transitions, IZA Discussion Papers 1850, Institute for the Study of Labor (IZA).

Frenkel, R. and Ros, J. (2004). Unemployment, macroeconomic policy and labor market flexibility: Argentina and Mexico in the 1990s, Kellogg Institute, University of Notre Dame, Working Paper #309 (February).

Galiani, S. and Hopenhayn, H. (2003). Duration and risk of unemployment in Argentina, *Journal of Development Economics*, **1**, 199-212.

Hopenhayn, H. (2004). *Labor market policies and employment duration: The effects of labor market reform in Argentina.* In *Law and Employment: Lessons from Latin America and the Caribbean* (Editeb by James J. Heckman and Carmen Pagés, National Bureau of Economic Research Conference Report), Chapter 9. The University of Chicago Press.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press.

Lamarche, C., Porto, A., and Sosa Escudero, W. (1998). *Aspectos regionales del Ddsempleo en la Argentina*, Universidad Nacional de La Plata, Departamento de Economía, Documento de Trabajo No. 8.

Llach, J. and Llach, L. (1998)*Cancelando la Hipoteca. Hiperinflación, Reforma de la Economía, Empelo y Desempleo en la Argentina de los 90*, Inter-American Development Bank, Office of the Chief Economist, Working Paper No.385.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterlin, W. T. (2002). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press.

Servy, E., Hachuel, L., Boggio, G., Cuesta, C. (2003). Un Modelo Dinámico Para el Estudio de la Desocupación. *Revista de la Sociedad Argentina de Estadística. SAE.* **7**, pp. 1-19.

Vardi, Y. (1982). Nonparametric estimation in renewal processes. *Ann. Statist.* **10** 772-785.

Enrique E. Alvarez
Department of Mathematics
Universidad Nacional de La Plata
Edificio de Matemática - Calle 50 e/115
(1900) La Plata - Argentina
ealvarez@mate.unlp.edu.ar

Francisco J. Ciocchini
Departamento de Economía
Universidad Católica Argentina
Buenos Aires, Argentina
f_ciocchini@uca.edu.ar

Kishori Konwar
Departament of Computer Science
University of Connecticut
Storrs, USA
kishori@cse.uconn.edu