Structural Analysis of Labor Market Transitions Using Indirect Inference [1]

Mark Yuying An and Ming Liu [2]

Duke University and *Chinese University of Hong Kong*

July, 1996

[1] Presented at the Optimization Days Conference, Montreal, Quebec, May 13-15, 1996. We thank conference participants, especially Jacque Raynauld, Eric Renault and Paul Ruud for helpful comments. Guido Imbens kindly provided us with the Dutch ORIN dataset used in the paper. Partial financial support is received from Duke University. An updated electronic version of the paper in postscript format can be retrieved using ftp as "home/man/papers/indinf.ps" on "lewis.econ.duke.edu".

[2] Direct correspondence to Mark An, Department of Economics, Duke University, Durham, NC 27708-0097. Telephone: (919) 660-1809. Fax (919) 684-8974. Email: man@econ.duke.edu, ming@econ.duke.edu.

Abstract

In the econometric analysis of labor market transitions, the data generating process is often specified as a continuous-time semi-Markovian process with a finite state space. With typically short panel data, analysts have long been concerned with the *initial conditions problem*— a complication associated with the very first spells observed in the data which are typically left-hand censored or interrupted. For practical convenience, one may want to discard the left-hand censored spells altogether. When there is uncontrolled heterogeneity, this passive approach results in inconsistent parameter estimates for the structural model. It has been well documented that consistent estimation, based on the specification of the correct likelihood function, requires explicit functional forms for the density function of the left-hand censored spells. Such a requirement can not be met, except in very special cases. In this paper we investigate an estimation procedure using indirect inference (II). Our procedure consists of two easy steps. In the first step, a pseudo likelihood function is maximized. In the second, simulations are employed to eliminate the discrepancy of such a pseudo maximum likelihood estimator. We discuss the consistency and asymptotic normality of the II estimator. We describe in detail the pseudo maximum likelihood estimation, the auxiliary parameter estimates and the simulation algorithm. We also implement the procedure using a Dutch data set on labor market histories.

Key Words: Labor Market Histories, Initial Conditions Problem, Indirect Inference.

JEL Classification: C41, C23.

1 Introduction

During the last two decades, a sizable body of work has accumulated in the empirical analysis of labor market dynamics. Devine and Kiefer (1991), for example, survey more than 500 such studies. These studies have been possible mainly due to the recently collected longitudinal surveys on labor market histories. In a typical survey, a sample of individuals are followed for a considerable period of time. Their experience in the labor market in terms of changes in labor market status is recorded. To analyze this type of data, labor economists have adopted, for good reasons, statistical duration models for single-spell data or semi-Markovian process models for multiple-spell data.

Unlike data used in medical studies where the duration and transition models are initiated, economic longitudinal surveys are usually non-experimental. Two important issues have to be taken care of when applying the standard duration or Markovian chain models to the analysis of labor market transitions. The first is the necessity of controlling for unobserved as well as observed heterogeneities. The second issue is associated with the various ways the labor market histories are sampled and recorded. Rarely is there a duration data set for which the start of the survey coincides with the origin date of all sampled individuals. Sampling from interrupted spells is common.

Unobserved heterogeneity and interrupted spells together create a major complication called the *initial conditions problem.* [3] Under the maintained model of semi-Markovian process, if there is no unobserved heterogeneity, interrupted spells can be handled easily. At least they can be discarded to safeguard consistency at the expense of efficiency. Vice versa, if there are no interrupted spells, unobserved heterogeneity can be also handled at ease. The unobserved heterogeneity can be integrated out at the expense of computing time. However when both are present, the initial conditions problem quickly becomes intractable.

For practical convenience, there are ad hoc procedures to deal with the interrupted spells. The first one ignores the left censoring and treats the interrupted spells the same way as a completed spell observed within the sampling frame. The second

[3] For a lucid introduction of the initial conditions problem in duration analysis, we refer the reader to Heckman and Singer (1986). For a more recent account, see Ham and LaLonde (1996).

approach simply discard the interrupted spells altogether. Neither approach provides consistent parameter estimates in the presence of unobserved heterogeneity.

This paper demonstrates why the initial conditions problem is inherently difficult. On the one hand, the interrupted spells contribute to the sample likelihood in a way too complicated to specify. On the other, they cannot be discarded due to the existence of unobserved heterogeneity. Effort continues in searching for solutions. Nickell (1979), Heckman and Singer (1984), and Ondrich (1985) remain the only three known solutions. Both Nickell and Ondrich deal only with highly specialized cases. Neither of the two generalizes beyond the proposed special setting. Heckman and Singer's approach introduces an extra set of nuisance parameters to the model. The efficiency loss could be severe. [4] To our knowledge, systematic solutions to the initial conditions problem for general semi-Markovian processes have not been previously proposed.

This paper investigates the solution to the initial conditions problem using indirect inference (Gallant and Tauchen, 1995, Gourieroux *et al*, 1993, Smith, 1993). This procedure consists of two easy steps. In the first step, a seemingly incorrect criterion is used to obtain consistent estimates for some auxiliary parameter(s). In the second, simulations of the structural model are employed to recover the structural parameters through repetitive usage of the auxiliary model. The idea of applying indirect inference to the initial conditions problem was first outlined in An (1996). Another application of indirect inference to labor market history analysis is reported in Magnac *et al* (1995), where data information is incomplete in that not all the transitions in labor market status are reported. In this paper, we (1) document fully the need for an indirect inference procedure, (2) discuss in greater detail the pseudo likelihood function used to generate auxiliary parameter estimates and the simulation algorithm used to recover the structural parameter estimates, (3) provide argument for the indirect inference estimator to be consistent and asymptotically normal, and (4) implement the procedure using a Dutch data set on labor market histories. Our empirical results demonstrate clearly the severe bias using ad hoc treatments of the

[4] More on this in section 2.

interrupted spells. The indirect inference procedure proposed here provides a consistent and asymptotically normal estimator of the structural parameters.

The remainder of the paper is structured as follows. In the next section we briefly introduce the semi-Markovian framework for the analysis of labor market histories and pin down the initial conditions problem. Examples are used to demonstrate why it is extremely difficult to specify the full likelihood function for direct inference. In section 3 we describe the indirect inference procedure as a solution to the problem. After a brief introduction of the idea of indirect inference, we lay out our choices of the pseudo likelihood function and the simulation algorithms. Asymptotic properties of the estimation procedure and specification tests using indirect inference are then discussed. Section 4 reports the empirical results using a field data set. Section 5 provides some concluding remarks.

2 The Initial Conditions Problem

In this section we first offer a brief account of the dynamic decision theory that underlies the semi-Markovian models widely adopted in the analysis of labor market history data. A discussion of the common features found in most longitudinal surveys reveals the initial conditions problem. Previous solutions to the initial conditions problem are critically reviewed.

2.1 A Prototypical Model of Labor Market History

In the analysis of labor market transitions, the maintained hypothesis has been that each individual, facing an uncertain and dynamic environment, makes sequential discrete choices. In other words, the feasible choices set at each point of time can be characterized as a finite set of alternatives. [5]

[5] The choice set can be {"full-time work," "part-time work," "non-employment"} for one example, and {"employment," "unemployment," "non-participation in labor force"} for another. Other applications of dynamic discrete choices in labor market dynamics include occupational choices (McCall, 1990) and welfare program participation (Keane and Moffitt, 1995).

The prototypical model of labor market histories is the following dynamic optimization framework. Flinn and Heckman (1982) and Burdett *et al* (1984) were among the first to use such a formalization. The discussion here borrows from An (1995) where the general framework of controlled jump processes is first introduced. [6] Let the choice set be denoted by $A = \{1, 2, \cdots, M\}$. A typical individual chooses one and only one alternative from A at any given time. The current reward, $u(w, a)$, depends on the current action $a \in A$, and on the outstanding environment, $w = (\mathbf{x}, \epsilon)$, characterized by a vector of observed variables $\mathbf{x}$ and a vector of unobserved variables ϵ. In the language of optimal control, A is called the action space, whereas $W = W_x \times W_\epsilon \subset \mathbf{R}^q$, the set of all possible values of the environment w, is called the state space. The evolution of the environment follows a piece-wise constant jump process characterized by the following ,

$$p^a(w'; t|w, s) = g^a(t - s, s) f^a(w'|w), \qquad a \in A, \tag{1}$$

which is the conditional probability distribution of the new state w' at a future time $t > s$, given that the process entered into the current state w at time s. Notice that the *law of motion* specified in (1) separates into two parts. The first part governs the conditional distribution of the waiting time until a "jump" in w. The second part governs the conditional distribution of the new environment. The hazard rate, $g^a(t - s, s)$, of the waiting time depends on the elapsed duration since the last change in w. The conditional distribution, $f^a(w'|w)$, of the new state given the current environment does not depend on the elapsed duration. Both g^a and f^a depend on the current action a.

The individual sequentially chooses actions $a_1, a_2, \cdots, a_k, \cdots$ at calendar time points $y_0 = 0, y_1, \cdots, y_{k-1}, \cdots$ where $0 < y_1 < y_2 < \cdots < y_k < \cdots$. For each k, $a_k = f_k(.)$ is a measurable function of the history $H(y_{k-1})$ up to time y_{k-1} to the action space A. For all $k = 1, 2, \cdots$, define $t_k = y_k - y_{k-1}$. We call a typical sequence

[6] Econometric analysis of sequential discrete choices in a discrete-time framework, although quite advanced and fast growing (see Eckstein and Wolpin (1989) and Rust (1994) for surveys), does not seem to apply well in labor market transitions where workers' changes in actions are "random event driven" rather than "clock driven." See, for example, Heckman and Singer (1986).

of such choices $\pi = (a_1, t_1, a_2, t_2, ...)$ a policy. Corresponding to each policy π and each initial state $w = (\mathbf{x}, \epsilon)$, there exist three sequences of random variables, [7]

$$\{W_k\}, \{A_{k+1}\}, \{T_{k+1}\}, \qquad k = 0, 1, 2, \cdots. \tag{2}$$

By optimal choice of policy π the individual maximizes a discounted expected reward of the form,

$$V^{\pi}(w) = E_w^{\pi} \sum_{k=0}^{\infty} \int_{Y_k}^{Y_{k+1}} e^{-\rho t} u(W_k, A_{k+1}) dt \tag{3}$$

where $Y_0 = 0$, $Y_k = \sum_{j=1}^{k} T_j$, and ρ is the time discounting parameter.

Under further regularity conditions which essentially guarantee that the dynamic decision problem is well defined, one can prove, [8]

Lemma 1 *For the above mentioned dynamic optimization problem,*

(a) there exists an optimal policy $\pi^* = (f_1^*, f_2^*, \cdots)$ which maximizes the objective defined in (3) for every initial state $w \in W$;

(b) according to the optimal policy π^*, the point of time t_k at which a new action a_k is taken is always a point at which there is a new realization of the state;

(c) the optimal policy is not only Markovian but also stationary, i.e., $f_k^*(H(t_{k-1})) = f^*(w_{k-1}|t-s)$ for all k; and

(d) the optimal policy is of the acceptance region type*, i.e., there is a partition of the state space W into $\{W_a(t-s), a \in A\}$ such that*

$$f^*(w_{k-1}|t-s) = \sum_{a \in A} 1_{W_a(t-s)}. \tag{4}$$

If the analyst had observations on the three sequences of random variables, the construction of the sample likelihood function for statistical inference would have been a simple matter, save the problem of numerically solving the optimal policy π^*

[7] In the sequel we adopt the convention to use capital letters for the random variables and their small-case counterparts for generic realizations.

[8] For a list of regularity conditions needed and a proof of the result, see An (1995).

for each individual in the sample. The main problem has to do with the unobservability of ϵ and possible partial information on $\mathbf{x}$ in the data. For the purpose of this paper, we assume the extreme case where the analyst observes only the realization of $\{T_k, A_k\}$, $k = 1, 2, \cdots$ for each individual in the sample together with a vector of exogenous covariates. The following result characterizing the property of the observable is instrumental for statistical inference.

Proposition 1 *The observed labor market history is a continuous-time, semi-Markovian process with state space A, whose transition intensity from state $a \in A$ to state $b \in A$ is*[9]

$$h_{ab}(s,t) = g^{\,a}(t-s|s)\theta^{\,*}_{ab}(w) \tag{5}$$

where θ^*_{ab} is the conditional probability that action b will be chosen, according to the optimal policy π^*, given there is a new draw w' of the state and given the current action is a, that is,

$$\theta^*_{ab} = \int_{W_b(t-s)} f^a(w'|w)dw', \tag{6}$$

where $W_b(t-s)$ is the subset of the state space W defined by the optimal policy π^*.

Proof. The result follows easily from Lemma 1.

Without the observability of the $\mathbf{x}$ sequence, it is impossible to separately identify the two structural parts in (5). It is in general difficult even with the observation of $\mathbf{x}$ (Rust, 1992, An 1995). To focus on the main theme, i.e. the initial condition problem, later in this sequel we will adopt a reduced-form parametrization of (5),

$$h_{ab}(t-s|z(t), v; \beta), \tag{7}$$

which depends on a vector of strongly exogenous, and possibly time-varying covariates $Z(t) = z$, on a scalar variable of uncontrolled heterogeneity $V = v$, and on a vector of parameters β which is the subject of inference.

[9]Earlier the set A was called the action space in the optimal decision problem from the stand point of the economic agent, in contrast to the state space W. In the semi-Markovian process observed by the analyst, A is now called the state space.

2.2 Data Structure

Consider a data set that contains a collection of N independent observations. For each individual n in the sample we observe a labor market history H^n which consists of K_n spells of durations,

$$H^n = (A_1^n, T_1^n; A_2^n, T_2^n; \cdots; A_{K_n}^n, T_{K_n}^n), \qquad n = 1, 2, \cdots N, \tag{8}$$

where $A_k^n \in A$, $T_k^n > 0$ are the state indicator and the duration of the k-th spell. Figure 1 depicts a typical history. Also in the data is a vector of weakly exogenous covariates Z^n for individual n.[10] Denote all the observed histories by $\mathbf{H}^N = \{H^n\}_{n=1}^N$.

There are no standard ways in which longitudinal surveys are conducted. However since economic surveys are both non-experimental and costly to maintain, they share the following common features.

(a) K_n is small relative to N. Attrition is abundant in follow-up surveys. Weak memory restricts how far a retrospective survey can go.

(b) The last spells ($k = K_n$) are typically *right-hand censored*. By the time the survey ends, it is expected that the last spell has not ended with a known destination state.

(c) The first spells ($k = 1$) are interrupted or *left censored*. For interrupted spells, one of the three duration times may be recorded: (i) time in the state preceding the sampling date, (ii) time in the state after the sampling date and (iii) total time in the state of the interrupted spell. Later we will demonstrate that none of the three duration times contributes to the sample likelihood the same way as a non-interrupted spell.

(d) There is *unmeasured* individual-specific heterogeneity beyond the observed $Z(t)$ vector. The heterogeneity is characterized by a scalar random variable V.

The last two features create what is now called the initial conditions problem.

[10] Typically many covariates in Z^n are time-varying. To simplify the notation we will, for the main part of paper, assume time-invariant Z^n. See section 5 for a discussion of possible extensions.

2.3 Sample Likelihood Functions

In constructing the sample likelihood function for a semi-Markovian process with data (8), there are two basic building blocks. The first is the probability density $f(t,b|a,z,v)$ of the duration $T_k^n (= t)$ and the destination state $A_{k+1}^n (= b)$, conditional on the current state $A_k^n = a$, on the observed covariates $Z_k^n = z$, and on the unobserved heterogeneity $V = v$. The second is the corresponding survivor function $S(t|a,z,v)$. For $k = 2, \cdots$, both $f()$ and $S()$ are simple one-to-one functional transformations of the primitive transition intensities $h()$,

$$f(t,b|a,z,v;\beta) = h_{ab}(t|z,v;\beta)e^{-\int_0^t h_a(u,z,v;\beta)du} \tag{9}$$

and

$$S(t|a,z,v;\beta) = e^{-\int_0^t h_a(u|z,v;\beta)du} \tag{10}$$

where

$$h_a(t|z,v;\beta) = \sum_{b \neq a} h_{ab}(t|z,v;\beta).$$

As mentioned earlier, the contribution to the likelihood function of a first spell is in general different from that of the rest. To emphasize the difference, denote by $g(t_1^n, a_1^n|z^n, v)$ the joint density of (T_1, A_1), conditional on $Z = z$ and $V = v$. To ease notational burden, denote by

$$L_n(\beta) = \prod_{k=2}^{K_n - 1} f\left(t_k^n, a_{k+1}^n|a_k^n, z^n, v;\beta\right) \cdot S(t_{K_n}^n|a_{K_n}^n, z^n, v;\beta) \tag{11}$$

the joint density of the $(A_2, T_2, \cdots, T_K)$ conditional on the initial conditions and the heterogeneity. In (11), we have implicitly taken that the last spell of each history is right-hand censored.

Suppose also that the uncontrolled heterogeneity has a distribution function $Q(v;\gamma)$ whose support is $B \subset \mathbf{R}_+$. Three log likelihood functions can be specified. The full log likelihood function, in terms of $\theta = (\beta, \gamma)$ is

$$l(\theta;\mathbf{H}^N) = N^{-1} \sum_{n=1}^{N} log\left\{\int_B g(t_1^n, a_1^n|z^n, v)L_n(\beta)dQ(v;\gamma)\right\}. \tag{12}$$

The first alternative to (12) is the pseudo log likelihood function,

$$l^p(\theta; \mathbf{H}^N) = N^{-1} \sum_{n=1}^{N} log \left\{ \int f(t_1^n | a_1^n, a_1^n, z^n, v) L_n(\beta) dQ(v;\gamma) \right\} \tag{13}$$

which simply treats the interrupted spells the same way as the rest of a completed spell. The second alternative is to discard the first spells altogether so that the log likelihood function is simply,

$$l^0(\theta; \mathbf{H}^N) = N^{-1} \sum_{n=1}^{N} \log \left\{ \int_B L_n(\beta) dQ(v;\gamma) \right\}. \tag{14}$$

Let $\hat{\theta}$, $\hat{\theta}^p = \hat{\mu}$, $\hat{\theta}^0$, respectively, be the corresponding maximizers of the three log likelihood functions. Consider the three maximizers as alternative estimators for the structural parameters θ. $\hat{\mu}$ is clearly inconsistent since l^p is a wrong likelihood function. $\hat{\theta}^0$ is obviously inefficient, since the first spells usually constitute a non-trivial proportion of the observed spells. More importantly it has been well documented that maximization of (14) would in general result in inconsistent estimates for β even under the assumption that V is independent of Z, unless the distribution $Q(v;\gamma)$ of V is degenerate. (Heckman and Singer, 1986, Lancaster, 1990). The only valid choice seems to be $\hat{\theta}$ that maximizes (12). Unfortunately, except in extremely restricted environments, the first component $g(t_1^n, a_1^n | z^n, v)$ is not analytically obtainable. We turn now to demonstrate such a difficulty.

2.4 A Difficulty in Direct Inference

Direct inference requires specification of the distribution of the interrupt spells. To ease exposition, we now examine a semi-Markovian process with two generic states $A = \{1, 2\}$. Conditional on observed and unobservable heterogeneities $(Z, V) = (z, v)$, the state-specific transition (into the alternative state) intensities are

$$h_a(t|z, v; \beta) = h_{ab}(t|z, v; \beta), \qquad a, b \in A \text{ and } a \neq b.$$

Suppose the sampling period starts at calendar time 0. We observe an individual residing in state a. Suppose this already-in-progress spell started at calendar time

$-S$ (entering from state b) and ends at a future time T (transiting into state b). Let $W = S + T$. The contribution of such a spell to the likelihood depends on whether S or T or both are observable. Let $\lambda_a(-u|z, v)$ be the conditional intake rate at which the process enters state a from state b. The proportion of people with the (z,v) characteristics who are found in state a at time 0 is obtained by integrating over the survivors of each cohort of entrants, that is,

$$p_a(z, v) = \int_0^\infty \left[\lambda_a(-s|z, v) \exp\left\{-v \int_0^s h_a(u|z, v)du\right\}\right] ds. \tag{15}$$

The following result is well-known. (See for example Heckman and Singer, 1986.)

Proposition 2 *Conditional on $Z = z$, $V = v$ the density functions for S, T and W are respectively,*

$$g_S(s|z, v) = \frac{\lambda_a(-s|z, v) \exp\{- \int_0^s h_a(u|z, v)du\}}{p_a(z, v)}, \tag{16}$$

$$g_T(t|z, v) = \frac{\int_0^\infty \left[\lambda_a(-s|z, v)h_a(s + t|z, v) \exp\{- \int_0^{s+t} h_a(u|z, v)du\}\right] ds}{p_a(z, v)}, \tag{17}$$

and

$$g_W(w|z, v) = \frac{\int_0^\infty [\lambda_a(-s|z, v)]\, ds\, [h_a(w|z, v) \exp\{- \int_0^w h_a(u|z, v)du\}]}{p_a(z, v)}. \tag{18}$$

Proof. See Appendix A.1. □

The basic message of Proposition 2 is that, in order to estimate the parameters of $h_{ab}(t|z, v; \beta)$, as long as v appears in the conditional hazard and λ_a is time-varying or Z is time-varying, λ must be specified along with $Q(v; \gamma)$ and $h_{ab}(t|x, v; \beta)$. This rate, $\lambda_a(u)$, is not free. On the contrary, it is determined by the primitive quantities of the process, namely, the transition intensities h. For maximum likelihood estimation of the structural parameters to be possible, the exact relationship between λ and h has to be specified. So far there seems to be no break-through that would enable the analyst to characterize such a relationship in general.

2.5 Previous Solutions

In the previous literature, three approaches were proposed as solutions to the initial conditions problem.

(a) Nickell's (1979) special case. In Nickell's setting, suppose the number, $N_a(-u)$, of the people in the population entering the state a in question at time $-u$ are known as far back in the past as empirically relevant. He assumes [11]

$$\lambda_a(-u|z,v) = N_a(-u) \cdot c(z,v) \qquad \text{for all } (z,v). \tag{19}$$

Under this assumption $c(z,v)$ would disappear from all three equations (16), (17) and (18). This ingenious specification to get around the initial conditions problem relies on the strong assumption (19) and requires additional data on the aggregate pre-sampling transition rates ($N_a(-u)$). It is not clear how this approach generalizes to be useful.

(b) The Bayesian method of Ondrich (1985). In a two-state Markovian setting, Ondrich first uses the interrupted spells to obtain the *posterior distribution* of the unmeasured heterogeneity

$$dQ(v|a_1^n, t_1^n) \quad \propto \quad dQ(v;\gamma) \cdot g(t_1^n, a_1^n|a_2^n, Z^n, v)$$

where $g(t_1^n, a_1^n|a_2^n, Z^n, v)$ is the conditional density for the interrupted spells. The posterior is then used as the mixing distribution in the log likelihood function (14). The fact that this approach needs as a prerequisite the functional form for $g(t_1^n, a_1^n|a_2^n, Z^n, v)$ shows that it should not be thought of as a solution to the initial conditions problem. It seems that this approach only applies to the two-state stationary Markovian framework, for which the initial conditions problem is analytically solvable.

[11] In Nickell's original setting there is no unobserved heterogeneity. Since there is only one spell observed for each individual, discarding the interrupted spells would amount to throwing away virtually all the information in the data.

(c) The approach of Heckman and Singer (1984). Heckman and Singer suggest using a different functional form, hence a set of extra parameters associated with it, for the conditional density $g(t\ _1^n, a_1^n | a_2^n, Z^n, v)$ as well as a separate heterogeneity term for the interrupted spells. In a recent application, Ham and LaLonde (1996) adopt this approach. The assessment of the statistical merit of this approach is still welcome. Intuitively, the interrupted spells are now mainly used to recover the essentially nuisance and add-on parameters. The efficiency loss is expected to be sizable.

3 The Indirect Inference Procedure

In this section we describe our solution to the initial conditions problem using indirect inference. The procedure exploits the facts that the structural model can be easily simulated and that an easily computed auxiliary model is readily available.

3.1 Basic Idea

Sometimes in econometrics, because the maintained model involves a complicated structure, direct inference is intractable. Provided the structural model is easily simulated for any fixed parameter value in the parameter space, indirect inference is the most appropriate and natural procedure. According to indirect inference, estimation of the structural parameters consists of two steps. In the first, an easy-to-compute auxiliary model is used to achieve consistent and asymptotically normal estimates for some auxiliary parameters (pseudo true value). In the second step, simulations are used to correct the discrepancy of the auxiliary parameters from the structural ones.

There are three ingredients in an indirect inference procedure: the structural model whose parameters θ are of primary interest, an auxiliary model which provides consistent estimates for a vector of auxiliary parameters μ, and a binding function $\mu = \mu(\theta)$. For identification reasons, the dimension of μ is at least as large as that of θ. In general the binding function is not known. Under mild conditions, simulations of the structural model with fine tuning of the structural parameters will lead to indirect inference of the structural parameters. The basic idea is that if the estimated structural parameters are close to the truth, then the auxiliary parameter estimates using the simulated data should be close — in some metric — to those using the actual data at hand. Depending on the metric used in measuring this closeness, there have been two approaches of indirect inference. One approach, pioneered by Gallant and Tauchen (1995), measures the geometric length of the expectation under the structural model of the score vector from the auxiliary model. [12] Another approach, advocated by Gourieroux, Monfort and Renault (1993, GMR), measures the distance between the auxiliary parameter estimates using the simulated data and those using actual data.

We propose to use indirect inference to solve the initial conditions problem described in Section 2. Here the maintained model does not involve a conceptually complicated structure. Rather the data at hand are collected in a way that make the sample likelihood function intractable. In the current context, we adopt the approach of Gourieroux *et al.* The remainder of this section describes the technical details of the auxiliary models, the simulation algorithm and the asymptotic properties of the indirect inference estimators so constructed.

3.2 The Auxiliary Model

The first step of indirect inference is to devise an auxiliary model which involves a set of auxiliary parameters. There are two minimal requirements for a model to be considered as a candidate for the auxiliary model. First, for identification reasons, the dimension of the auxiliary parameters must be at least as large as that of the structural parameters. Second, for feasibility reasons, an auxiliary model has to be computationally easy to generate consistent and asymptotically normal estimates for some auxiliary parameters. In general, auxiliary models that are closer to the maintained model are better on the grounds of efficiency. As a matter of fact, if the auxiliary model encompasses the structural model then full efficiency, as compared to

[12] Gallant and Tauchen's work is motivated in proposing a systematic way to generate moment conditions for the generalized method of moments (GMM) estimator (Hansen, 1982).

direct inference, can be achieved. For general guidance on how to choose an auxiliary model on efficiency grounds and under the possibility that the maintained structural model is itself mis-specified, see Tauchen (1995).

In the current context, we choose quasi maximum likelihood estimation as the auxiliary model. Instead of maximizing the full likelihood (12) we maximize the pseudo likelihood (13), which treats the interrupted spells as completed. It follows from standard conditions that the objective function converges almost surely to a non-stochastic limit and that the limit has a unique maximum at μ^0. And the extremum estimator,

$$\hat{\mu} = \arg\max_{\mu \in \Gamma} l^p(\mu; \mathbf{H}^N) \tag{20}$$

converges in probability to μ^0 which can be expressed as $\mu^0 = \mu(\theta^0)$, for some unknown function $\mu(.)$. The conditions under which a general quasi maximum likelihood is consistent and asymptotically normal can be found in, for example, White (1994).

3.3 Structural Parameter Estimates

In this subsection we assume that we can easily simulate the structural model. [13] Our indirect inference procedure proceeds as follows. First, for each trial value of the structural parameters θ, we simulate all the labor market histories S times over. For each of the simulated histories, we throw away the pre-sampling episodes according to the actual data censoring mechanism that had generated the real data. Let the resulting simulated histories be $\tilde{\mathbf{H}}^N(\theta; s) = \{\tilde{H}^n(\theta; s)\}_{n=1}^N$ for $s = 1, 2, \cdots, S$. Define

$$\tilde{\mu}(\theta; s) = \arg\max_{\mu \in \Gamma} l^p(\mu; \tilde{\mathbf{H}}^N(\theta; s)),$$

which maximizes the pseudo likelihood function with data $\tilde{\mathbf{H}}^N(\theta; s)$.

Since in the current context,

$C.1$ $$\dim(\theta) = \dim(\mu)$$

[13] Given the semi-Markovian structure, this is indeed a trivial task. See the Appendix for computational details.

one has the option to fine tune θ until $\hat{\theta}$ is found such that $\tilde{\mu}^{\,n}(\hat{\theta}) \approx \hat{\mu}$. The search mechanism defines the new trial value of θ as

$$\theta^{j+1} = \theta^j + \tilde{\mu}(\theta^j) - \hat{\mu}. \tag{21}$$

While this procedure is intuitive and has been adopted in Magnac *et al* (1995), we found that it is extremely slow in our setting. One possible reason is that the parameter θ and μ have different relative perturbation scales in the Θ and Γ spaces. A better algorithm should replace (21) by

$$\theta^{j+1} = \theta^j + P\left[\tilde{\mu}(\theta^j) - \hat{\mu}\right],$$

with careful choice of the matrix P. This leads to the second way to recover the structural parameter in which the indirect inference estimator $\hat{\theta}$ is defined as

$$\hat{\theta} = \arg\min_{\theta\in\Theta} (S^{-1}\sum_{s=1}^{S} \tilde{\mu}(\theta; s) - \hat{\mu})'\Omega(S^{-1}\sum_{s=1}^{S} \tilde{\mu}(\theta; s) - \hat{\mu}), \tag{22}$$

where Ω is a positive-definite weighting matrix in defining the distance between the two auxiliary parameter estimates. To obtain $\hat{\theta}$ we embed the simulation algorithm and the pseudo maximum likelihood procedure as the two subroutines of the main minimization algorithm which minimizes (22). See Appendix A.2 for more details.

GMR show that under condition C.1 the indirect inference estimation is independent of the choice of Ω in (22) and that the objective function in (22) is equal to zero at $\hat{\theta}$. It then follows that the fine tuning algorithm (21), at its convergence, will result in estimates $\hat{\theta}$ equivalent to those using algorithm (22).

3.4 Asymptotics

As demonstrated by GMR, under the usual regularity conditions, the indirect estimator $\hat{\theta}$ is consistent and asymptotically normal, when the number S of simulations is fixed and the sample size N goes to infinity. Namely, if we denote $\mu(\theta)$ as the true binding function, $\mathbf{H}^N$ as $(H^1, H^2, \cdots, H^N)'$, $\mathbf{z}$ as $(z^1, z^2, \cdots, z^N)'$ and $\mathbf{v}$ as

$(v^1, v^2, \cdots, v^N)'$, we will have

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, W(S, \Omega)) \tag{23}$$

where

$$\begin{aligned} W(S,\Omega) &= \frac{S+1}{S}\left(\frac{\partial \mu'(\theta_0)}{\partial \theta}\Omega\frac{\partial \mu(\theta_0)}{\partial \theta'}\right)^{-1} \\ &\cdot \frac{\partial \mu'(\theta_0)}{\partial \theta}\Omega J_0^{-1}(I_0 - K_0)J_0^{-1}\Omega\frac{\partial \mu(\theta_0)}{\partial \theta'} \\ &\cdot \left(\frac{\partial \mu'(\theta_0)}{\partial \theta}\Omega\frac{\partial \mu(\theta_0)}{\partial \theta'}\right)^{-1}, \end{aligned}$$

where

$$I_0 = \lim_{N\to\infty} \mathrm{V}(\sqrt{N}\frac{\partial \mathbf{L}_N}{\partial \mu'}(\mathbf{H}^N, \mathbf{z}, \mathbf{v}; \theta_0)],$$

$$K_0 = \lim_{N\to\infty} \mathrm{Cov}(\sqrt{N}\frac{\partial l_N^p}{\partial \mu'}(\mathbf{H}^N, \mathbf{z}, \mathbf{v}; \theta_0)), \sqrt{N}\frac{\partial l_N^p}{\partial \mu'}(\mathbf{H}^N, \mathbf{z}, \mathbf{v}; \theta_0)))$$

and

$$J_0 = \mathrm{plim}_{N\to\infty} - \frac{\partial^2 l_N^p}{\partial \mu \partial \mu'}[\mathbf{H}^N, \mathbf{z}, \mathbf{v}].$$

When the auxiliary model is good enough, K_0 should be close to 0 and $J_0 = I_0$. As in the case of GMM of Hansen (1982), the optimal choice of the Ω matrix is: $\Omega^* = J_0(I_0 - K_0)^{-1}J_0$ and a two step procedure similar to Hansen's two step procedure could be used to do the estimation. With this optimal choice of weighting matrix, the covariance matrix of the indirect estimator is

$$W_S^* = W(S, \Omega^*) = (1 + \frac{1}{S})\left[\frac{\partial \mu'}{\partial \theta}(\theta_0)J_0(I_0 - K_0)^{-1}J_0\frac{\partial \mu}{\partial \theta'}(\theta_0)\right]^{-1}. \tag{24}$$

Given that the binding function $\mu(\theta)$ is not known, we could use

$$\frac{\partial \mu}{\partial \theta'}(\theta_0) = J_0^{-1}\frac{\partial^2 l_\infty^p}{\partial \mu \partial \theta'}(\theta_0, \mu_0),$$

and W_S^* could be written as

$$W_S^* = (1 + \frac{1}{S})\left[\frac{\partial^2 l_\infty^p}{\partial \mu \partial \theta'}(\theta_0, \mu_0)(I_0 - K_0)^{-1}\frac{\partial^2 l_\infty^p}{\partial \mu \partial \theta'}(\theta_0, \mu_0)\right]^{-1}. \tag{25}$$

4 An Empirical Application

In this section the indirect inference discussed in Section 3 will be applied to a Dutch data set on labor market transitions.

4.1 The Data

The data set used in this paper was taken from the ORIN (Onderzoek naar Ralatievormen in Nederland) surveys. It was set up by NIDI (Nederlands Interuniversitair Demografisch Instituut) in cooperation with the Universities of Tilburg, Amsterdam, and Wageningen. [14] The original data contains 1601 individuals randomly selected in 1984. Each respondent was asked to reconstruct his or her labor market history going back to the last change in labor market status before January 1977. Like Imbens (1994), we only use the data on men aged between 23 and 50 in January, 1977. This leaves us with 360 constructed histories. To serve the purpose of this paper we will treat the data set as if it were constructed by taking a random sample in January, 1977 from the population of males between 23 and 50 years of age.

For each labor market history, there is information on all the spells of durations within the 7 year period. The data set identifies whether each spell is a non-employment, part-time employment or full-time employment and records the entrance date, and the ending date, unless it is the last spell. For males within this age group, part-time employment is rare. We do not distinguish it from full-time employment. Also the data do not distinguish unemployment from non-labor force participation. This leaves us with a two-status framework. However an employment spell can end with another employment spell as people move from one job to another without an intervening spell of unemployment. Therefore we encounter the problem of modeling a two- status process with three possible types of transitions, $U \to E$, $E \to E$, and $E \to U$. A standard way to accommodate this is to invent a second employment state and call the employment states E_1 and E_2 (Lancaster 1990: p114.) We would then

[14] Professor Guido Imbens kindly provided us with this ideal data set for which we are deeply indebted.

allow the movement from U to E_1, from both E_1 and E_2 to U at the same intensity rate, and from E_1 to E_2 and vice versa,

	E_1	E_2	U
E_1	—	h_{EE}	h_{EU}
E_2	h_{EE}	—	h_{EU}
U	h_{UE}	0	—

There are many social demographic variables in the data. To concentrate on the main theme of the current paper we take only two as our covariates —Education Attainment and Age at January, 1977 – both of which are time invariant.

Table 1 reports the sample summary statistics at the individual level. For the 360 men observed in the data, the average age at the beginning of the sampling period is 34.6 years. Their average education is close to level 3, about the level of high school in the U.S. They on average experience 1.57 number of spells during the 7 year period. Table 2 reports the basic sample information at the spell level. For the 566 observed spells, the average duration is 54.7 months. On average an $U \to E$ transition takes the least time of 23 months compared with an $E \to E$ transition in 29 months and an $E \to U$ transition in 41 months. Of those 566 spells, 54% of them are right-censored employment and 10% are right-censored unemployment spells. Among the 164 employment spells with known destination, exactly half are followed by another employment spell. Table 3 records the numbers of individuals with certain number of transitions by type. Among the 360 men, 241 have no transition whatsoever, and 68 have one transition of some kind. Only 8 experience more than 3 transitions. This confirms that the Dutch labor market is not as mobile as the labor market in the U.S., at least for this age group.

4.2 Parameterization

As discussed before, we need to model a two-status process with three types of transitions. In the sequel, use subscripts $(1, 2, 3) = (E \to E, E \to U, U \to E)$. We adopt the following parametrization for the three conditional hazard functions,

$$h_j(t|z, v; \beta) = \beta_{0j} \exp\{\beta_j z\} v t^{\beta_{0j}-1}, \qquad j \in \{1, 2, 3\} = \{EE, EU, UE\}. \tag{26}$$

For the mixing distribution of the unobserved heterogeneity V, two alternatives are used. Both are very simple. The first is the unit-mean Gamma distribution $G(\gamma, \gamma)$. This makes the model a Weibull-Gamma mixture that has been widely used in duration models for computational simplicity (Lancaster 1990). The second alternative is the discrete distribution with a small number ($K = 3$, say) of supports, $P(V = v_i) = \pi_i$, for $i = 1, 2, 3$. For the discrete distribution, $\gamma = (v_1, v_2, v_3, \pi_1, \pi_2)$. In the current context both the Weibull-Gamma mixture and the discrete factor loading continue to produce convenient functional forms for the quasi maximum likelihood functions as demonstrated in Appendix A.3.

4.3 Results

Parameter estimates are reported in Table 4. Panel (1) of the table reports the parameter estimates from the quasi maximum likelihood estimation defined in (20). They are consistent estimates for the pseudo true values μ^0. Since in the current context, the auxiliary model can also be viewed as the maximum likelihood estimation of the true structural model with an ad hoc treatment of the interrupted spells, it is interesting to compare these estimates with the consistent estimates from the indirect inference (reported in Panel (2) of the table). Such a comparison reveals the severe bias of the former.

The indirect inference estimates of the structural parameter are much more sensible. We plan to conduct Monte Carlo simulations in the future to investigate the accuracy of the first-order asymptotic normality approximation of the distribution of the indirect inference estimator.

5 Discussions

The idea of indirect inference is still in its infancy. So far only its theoretical properties have been investigated. In this paper we demonstrate that indirect inference can be applied to solve the otherwise intractable initial conditions problem in the analysis of labor market histories. This procedure provides a consistent and asymptotically normal estimator of the structural parameters of the semi-Markovian model. The simulation required in the second step is easy since it does not depend on the observed histories at hand. The pseudo likelihood used in the first step is natural and is likely to capture most features of the data.

To focus on the main theme, we have throughout the paper assumed that the covariates are all time-invariant. This is of course an unrealistic assumption for any longitudinal survey. The methodology proposed in this paper applies to cases with time-varying covariates as well. First of all, having time-varying covariates in the picture makes direct inference even harder. For our indirect inference, we need to be able to simulate the pre-sampling labor market histories. For the latter we need to construct every individual's covariates going back to the origination of the process (fresh out of school). For a reasonable set of covariates, this does not seem to be a big obstacle. Our auxiliary model, the quasi maximum likelihood estimation framework, can handle easily time-varying covariates.

For convenience, we have adopted the popular Weibull-Gamma mixture and Weibull-discrete parametrizations for our empirical example. Other mixtures can be employed as well. The difficulty with other mixtures stems from the computing time constraints in implementing the corresponding quasi maximum likelihood estimation. The difficulty is not at the conceptual level. As a matter of fact, used as the auxiliary model for the indirect inference, in principle, the quasi maximum likelihood estimation procedure does not have to be specified in accordance with the structural model. Provided the structural model, be it the Weibull-Gamma mixture or not, can be easily simulated, one still has the option to use the Weibull-Gamma mixture for the auxiliary model. Without the initial conditions problem, this mismatch between the structural model and the auxiliary model is essentially the essence and the heart of the indirect inference. The statistical properties of this mismatch in the current context should be investigated. We leave this for possible future research.

Mathematical Appendices

In this appendix we gather mathematical details used in the text.

A.1 Proof of Proportion 2

To simplify the notation, suppress the dependence of λ_a, h_a, etc. on (z, v). First derive the density function for S. By Equation (15), p_a is the probability that an individual be seen at 0 in state a. Among these people, the subgroup who have been in the state for exactly s period are those who entered state a at -s and have not left yet. The size of this subgroup is exactly the integrand in calculating p_a. Equation (16) follows immediately.

To prove the other two equations, we first characterize the joint density of (S, T),

$$\begin{aligned}
g_{(S,T)}(s,t) &= g_S(s) g_{(T|S)}(t,s) \\
&= g_S(s) \left[-\frac{d}{dt} (Pr(T > t|S = s) \right] \\
&= g_S(s) \left[-\frac{d}{dt} (Pr(S + T > s + t|S = s) \right] \\
&= g_S(s) \left[-\frac{d}{dt} \exp\left\{ -\int_s^{s+t} h_a(u)du \right\} \right] \\
&= \frac{\lambda_a(-s) \exp\{-\int_0^s h_a(u)du\}}{p_a} \exp\left\{ -\int_s^{s+t} h_a(u)du \right\} h(s+t) \\
&= \frac{\lambda_a(-s) h_a(s+t) \exp\left\{ -\int_0^{s+t} h_a(u)du \right\}}{p_a}
\end{aligned}$$

Integrating out s from $g_{(S,T)}$ results in the density of T which is given in (17). Do transformation of variables $u = s$ and $w = s + t$, the joint density of (U,W) is

$$g_{(U,W)}(u,w) = \frac{\lambda_a(-u) h_a(w) \exp\{-\int_0^w h_a(t)dt\}}{p_a}.$$

Further integrating out w results in the density of W as in (18).

A.2 Computational Details Using Indirect Inference

The empirical application of the paper uses FORTRAN programs. For the pseudo likelihood maximization, we use the DFP routine of GQOPT by Goldfeld and Quandt. For the calibration minimization, we use the NMSIMP simplex routine in GQOPT. The complete set of codes can be downloaded using anonymous ftp from "lewis.econ.duke.edu" in /pub/man/program/indinf directory.

To simulate a typical observation of labor histories $\tilde{\mathbf{H}}^N(\beta,\gamma)$, we first simulate N Gamma variates, $(v^1, v^2, ..., v^N)$ from $G(\gamma,\gamma)$. The observed covariates $(z^1, z^2, ..., z^N)$ and β determine the parameters of the Weibull distributions for each of the three transition types. We then generate labor histories for all the individual in the sample. For each spell, we generate the duration by inverting the survival function. For the state of employment, we simulate two durations associated respectively with h_{EU} and h_{EE} and keep the shorter one as the realized duration. To create left-censored spells in accordance with the actual data, we simulate the pre-sample labor history for each individual starting from when he enters the labor market. We assume that those individuals without college education enter the labor market at age 18, and those with college education enter at age 22 years old and so forth. We then cut the period for each individual according to the sampling period. Noted here, because there is a distributional parameter in θ which means the gamma variate has to be updated across different parameter vectors, we are no longer able to keep the variate the same across simulations associated with different parameter vectors. This has given us grave numerical instability in our implementation and further exploration on how to get around this problem could be very helpful.

Also we use formula (25) to estimate the covariance matrix of the indirect inference estimator. We estimate the part $\partial^2 l^p_\infty/\partial\mu\partial\theta'$, using the finite difference at the point $\hat{\theta}$ and $\hat{\mu}$. For the part $(I_0 - K_0)$, notice that it is simply the covariance of $\sqrt{N}\partial l^p_N/\partial\mu$ at the point of θ_0 given the covariates. So the following formula as suggested in GMR is used,

$$\frac{N}{S}\sum_{s=1}^{S}(W_s - \bar{W})(W_s - \bar{W})',$$

with $W_s = \partial l^p_N(\hat{\mu})/\partial\mu$ and $\bar{W} = S^{-1}\sum_{s=1}^{S} W_s$.

A.3 The Pseudo Log likelihood Function under the Weibull-Gamma Mixture

Let $\mu = \{\beta_{01}, \beta_{02}, \beta_{03}, \beta_1, \beta_2, \beta_3, \gamma\}'$. Here β_j is a row vector consisting of the coefficients associated with the covariates. The log quasi likelihood of the model is

$$l_N^p(\mu) = N^{-1} \sum_{n=1}^{N} \log L_n^p(\mu)$$

with

$$L_n^p(\mu) = \int_0^\infty \prod_{k=1}^{K_n-1} f(t_k^n, a_{k+1}^n | a_k^n, z^n, v; \mu) S(t_{K_n}^n \mid a_{K_n}^n, z^n, v; \mu) dG(v; \gamma)$$

For each spell (k, n) with $k < K_n$, define two binary variables indicating the type of transition, $D_{1kn} = \chi\left(a_k^n = E, a_{k+1}^n = E\right)$ and $D_{2kn} = \chi\left(a_k^n = E, a_{k+1}^n = U\right)$. Also for each $1 \leq k \leq K_n$, set $D_{3kn} = \chi(a_k = U)$. Let

$$\log(I_1^n) = \sum_{j=1}^{3} \sum_{k=1}^{K_n-1} \{D_{jkn} \left[\log(\beta_{0j}) + (\beta_{0j} - 1) \log(t_k^n) + \beta_j(z^n)\right]\}$$

and

$$\begin{aligned} I_2^n &= \sum_{k=1}^{K_n} \Big\{(1 - D_{3jn}) \left[\exp\{\beta_1 z^n\}(t_k^n)^{\beta_{01}} + \exp\{\beta_2 z^n\}(t_k^n)^{\beta_{02}}\right] \\ &\quad + D_{3kn} \left[\exp\{\beta_3 z^n\}(t_k^n)^{\beta_{03}}\right]\Big\} \end{aligned}$$

Using these definitions,

$$\begin{aligned} L_n^p(\mu) &= \int_0^\infty \left\{ \prod_{k=1}^{K_n-1} \prod_{j=1}^{3} [h_j(t_k^n | z_n, v; \beta)]^{D_{jkn}} \prod_{k=1}^{K_n} \exp\left\{-v e^{\beta_3 z_n} (t_k^n)^{\beta_{03}} D_{3kn}\right\} \right. \\ &\quad \left. \prod_{k=1}^{K_n} \exp\left\{-v[e^{\beta_1 z_n}(t_k^n)^{\beta_{01}} + e^{\beta_2 z_n}(t_k^n)^{\beta_{02}}](1 - D_{3kn})\right\} \right\} dG(v; \gamma) \\ &= \int_0^\infty v^{K_n-1} I_1^n \cdot \exp\{-I_2^n v\} dG(v, \gamma) \end{aligned}$$

For the Gamma mixing distribution,,

$$L_n^p(\mu) = \frac{\gamma^\gamma}{(\gamma + I_2^n)^{\gamma + K_n - 1}} \cdot \frac{\Gamma(\gamma + K_n - 1)}{\Gamma(\gamma)} I_1^n,$$

and for the discrete mixing distribution

$$L_n^p(\mu) = \sum_{I=1}^{3} \pi_i \cdot \left[(v_i)^{K_n-1} I_1^n \cdot \exp\{-I_2^n v_i\}\right].$$

References

An, Mark Y. (1995): "Econometric Analysis of Sequential Discrete Choice models," Paper presented at the 7th World Congress of the Econometric Society, Tokyo, Japan.

An, Mark Y. (1996): "Initial Conditions Problem in Event History Analysis: An Indirect Inference Procedure," in Jewell, N.P. *et al* eds. *Life Time Data Models in Reliability and Survival Analysis*, New York: Kluwer Academic Publisher, 7-11.

Burdett, Kenneth, Nicholas M. Kiefer, Dale Mortensen and George R. Neumann (1984): "Earning, Unemployment, and Allocation of Time over Time," *Review of Economic Studies*, **LI**, 559-578.

Hansen, Lars P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, **50**, 1029-1054.

Devine, Teresa and Nicholas M. Kiefer (1991): *Empirical Labor Economics: The Search Approach*, New York: Oxford University Press.

Eckstein, Zvi and Kenneth Wolpin (1989): "The Specification and Estimation of Dynamic Stochastic Choice Models: A Survey," *Journal of Human Resources*, **23**, 562-598.

Flinn, Christopher J. and James J. Heckman (1982), "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, **18**, pp.115-168.

Gallant, A. Ronald and George E. Tauchen (1995): "Which Moment to Match?" forthcoming *Econometric Theory.*

Gourieroux, Christian, Alan Monfort and Eric Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, **8**, S85-S118.

Ham, John C. and Robert J. LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, **64**, 175-205.

Heckman, James J. and Burton Singer (1984): "A Method Minimizing the Impact of Distribution Assumptions in Econometric Models for Duration Data," *Econometrica*, **47**, 247-283.

Heckman, James J. and Burton Singer (1986): "Econometric Analysis of Longitudinal Data," in Z. Griliches and M.D. Intriligator (eds.) *Handbook of Econometrics Vol. III*, 1917-1978.

Imbens, Guido (1994): "Transition Models in a Non-Stationary Environment," *Review of Economics and Statistics*, 703-720.

Keane, Michael and Robert Moffitt (1995): "A Structural Model of Multiple Welfare Program Participation and Labor Supply," Working Paper, Brown University.

Lancaster, Tony (1990): *The Econometric Analysis of Transition Data*, New York: Cambridge University Press.

Magnac, Thierry, J.-M. Robin, and M. Visser (1995): "Analyzing Incomplete Individual Employment Histories Using Indirect Inference," *Journal of Applied Econometrics*, **?.** S??-??.

McCall, Brian (1990): "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, **98**, 45-69.

Nickell, Stephen (1979): "Estimating the Probability of Leaving Unemployment," *Econometrica*, **47**, 1249-1266.

Ondrich, Jan I. (1985): "The Initial Conditions Problem in Work History Data," *Review of Economics and Statistics*, **67**, 411-421.

Rust, John (1992): "Do People Behave According to Bellman's Principle of Optimality?" Hoover Institution Working Paper E-92-10.

Rust, John (1994): "Estimation of Dynamic Structural Models: Problems and Prospects Part I: Discrete Decision Process," forthcoming in *Proceedings of the 6th World Congress of the Econometric Society*, Cambridge University Press.

Smith, Anthony (1993): "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, **8**, S63-84.

Tauchen, George E. (1995): "New Minimum Chi-square Methods in Financial Economics," Invited Symposiums at the 7th World Congress of the Econometric Society, Tokyo, Japan.

White, Halbert (1994): *Estimation, Inference, and Specification Analysis*, Cambridge University Press.

Table 1. Summary Statistics for Observed Individuals

Variable	Mean	S.D.	Min.	Max
Age	34.63	6.92	23	47
Education	2.90	1.17	1	5
Number of Spells	1.57	1.05	1	8
No. of Individuals	360			

Table 2. Summary Statistics for Observed Spells

Type of Transition	Length of Spell (Months)		Number of Spells	
	Mean	S.D.	No.	Percentage
$E \to E$	28.56	19.75	82	14.5
$E \to U$	40.84	25.32	82	14.5
$U \to E$	22.88	21.09	42	7.4
Censored E	73.99	25.71	305	53.9
Censored U	31.65	24.36	55	9.7
Total spells	54.69	23.01	566	100

Table 3. Distribution by Number of Spells

Number of Transitions	Number of Men	Percentage
0	241	66.9
1	68	18.9
2	32	8.9
3	11	3.0
≥ 4	8	2.2

Table 4. Parameter Estimates

	Auxiliary Parameter		Structural Parameter	
	$\hat{\theta}^p = \hat{\mu}$	s.d.	$\hat{\theta}$	s.d.
E $\rightarrow$ **E**:				
Shape Parameter β_{01}	**1.663**	(0.147)	**1.074**	(0.131)
Intercept β_{11}	**-10.60**	(0.961)	**-18.148**	(8.466)
Age β_{21}	**0.022**	(0.00138)	**-0.005**	(0.017)
Education β_{31}	**-1.168**	(1.667)	**4.401**	(0.886)
E $\rightarrow$ **U**:				
Shape Parameter β_{02}	**1.381**	(0.196)	**1.357**	(0.142)
Intercept β_{12}	**-12.385**	(1.063)	**-19.927**	(8.414)
Age β_{22}	**0.223**	(0.0138)	**-0.009**	(0.017)
Education β_{32}	**-1.262**	(1.669)	**4.308**	(0.910)
U $\rightarrow$ **E**:				
Shape Parameter β_{03}	**2.313**	(0.182)	**2.395**	(0.138)
Intercept β_{13}	**-9.154**	(1.307)	**-18.697**	(7.336)
Age β_{23}	**0.221**	(0.0138)	**-0.038**	(0.017)
Education β_{33}	**0.890**	(1.672)	**4.681**	(0.801)
Gamma Mixing Parameter γ		( )	**0.665**	(3.829)

Figure 1 A Typical Labor Market History