NBER WORKING PAPER SERIES

MODELS FOR THE ANALYSIS OF LABOR FORCE DYNAMICS

Christopher J. Flinn

James J. Heckman

Working Paper No. 857

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

February 1982

This research was supported by NSF Grant SOC 77-27136 to NBER and 10-P-9074819-01 from the Social Security Administration to NBER. We are grateful to Takeshi Amemiya, Marjorie McElroy, George Neumann, Guilherme Sedlacek, Burton Singer and especially Gary Chamberlain for comments. George Yates supplied extremely competent programming assistance. The first draft of this paper was presented at the Summer meetings of the Econometric Society, Montreal, June, 1979, and at a NSF-NBER conference on Discrete Choice in Cambridge, Massachusetts, in June, 1980. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #857
February 1982

Models for the Analysis of Labor Force Dynamics

ABSTRACT

This paper presents new econometric methods for the empirical analysis of individual labor market histories. The techniques developed here extend previous work on continuous time models in four ways: (1) A structural economic interpretation of these models is presented. (2) Time varying explanatory variables are introduced into the analysis in a general way. (3) Unobserved heterogeneity components are permitted to be correlated across spells. (4) A flexible model of duration dependence is presented that accommodates many previous models as a special case and that permits tests among competing specifications within a unified framework.

We contrast our methods with more conventional discrete time and regression procedures. The parameters of continuous time models are invariant to the sampling time unit used to record observations. Problems plague the regression approach to analyzing duration data which do not plague the likelihood approach advocated in this paper. The regression approach cannot be readily adopted to accommodate time varying explanatory variables. The functional forms of regression functions depend on the time paths of the explanatory variables. Ad hoc solutions to this problem can make exogenous variables endogenous to the model and so can induce simultaneous equations bias.

Two sets of empirical results are presented. A major conclusion of the first analysis is that the discrete time Markov model widely used in labor market analysis is inconsistent with the data. The second set of empirical results is a test of the hypothesis that "unemployment" and "out of the labor force" are behaviorally different labor market states. Contrary to recent claims, we find that they are separate states for our sample of young men.

Christopher J. Flinn
University of Wisconsin

James J. Heckman
Professor of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637

(312) 753-4533 or -1106

This paper presents new econometric methods for the empirical analysis of individual labor market histories. Jovanovic's equilibrium model of worker turnover, the McCall-Mortensen search model, and Holt's model of labor force dynamics can be estimated, and crucial assumptions tested, with our techniques.

Our point of departure is the continuous time Markov model widely utilized in sociology (Coleman; Singer and Spilerman; Tuma, Hannan and Groneveld). The methodology developed here extends this work in four ways. (1) A structural economic interpretation of these models is provided. It is demonstrated that continuous time models naturally arise from optimal stopping rules that are the essence of a variety of economic problems. (See Brock *et al*). (2) The methods developed below admit the introduction of time varying explanatory variables into the analysis in a general way. Previous work either ignores such variables or utilizes special procedures for selected variables. (e.g., Tuma, Hannan and Groneveld). Regression procedures for introducing time varying variables require special assumptions that are unlikely to be realized in empirical work with labor force data.

We present a flexible empirical procedure which can be used to estimate duration models with time varying variables and demonstrate, both theoretically and empirically, the importance of being careful about the way time varying variables are introduced into duration models. (3) We extend previous work by introducing unobserved components ("heterogeneity") that are correlated across spells. Previous work assumes that unobserved components are independently distributed across spells - a strong - and as we demonstrate below for one data set - a counter factual assumption. (4) We produce a flexible econometric model with general types of "state dependence." Many models commonly used in the analysis of continuous time data can be written as a special case of our model. Our framework can be used to test among competing model specifications.

We present empirical estimates of a two-state model of employment and nonemployment[1] in the youth labor market. We then proceed to test a critical assumption often used in labor market analysis: that "unemployment" and "out of the labor force" are legitimately separate labor market states. We find that this is so and that the behavioral equations that generate movement into and out of these states are fundamentally different.

The structure of the paper is as follows. In Section 1 we present a continuous time model of worker turnover. We demonstrate that this model can be used as a framework within which it is possible to estimate Jovanovic's model and many other models as well. The model is extended to allow for heterogeneity, time varying variables and general types of dependence of labor market transition rates on previous labor force states. The likelihood function for a two-state model is presented and solutions to the problem of correct treatment of the initial conditions of the process (sometimes called the "left censoring" problem) are offered.

---

[1] The nonemployment state is composed of the states "unemployment" and "out of the labor force."

Section II presents a discussion of the advantages of continuous time models over discrete time models. Certain limitations of continuous time models are discussed as well.

Section III discusses pitfalls that arise in using regression methods to analyze duration data. Two problems plague the regression approach. First, standard regression estimators are ill-equipped to deal with censored spells of events that arise in short panels. Failure to account for censored spells leads to biased estimates of the parameters of population regression functions. Second, the regression approach cannot be readily adapted to accommodate time varying explanatory variables. The functional forms of population regression functions depend on the time paths of the explanatory variables. *Ad hoc* solutions to this problem can make exogenous variables endogenous to the model and so can induce simultaneous equations bias.

Section IV presents two sets of empirical results. The first set is an analysis of employment and nonemployment data using both regression and maximum likelihood procedures. The second set is a test of the hypothesis that "unemployment" and "out of the labor force" are behaviorally different labor market states.

Appendix A presents a general multiple state multiple spell likelihood function for a continuous time model. Appendix B presents the Weibull regression model used in some of our empirical work. Appendix C presents a simple economic model in which non-Markovian, long-term dependence between labor market outcomes is generated.

I. CONTINUOUS TIME ECONOMETRIC MODELS OF TURNOVER AND UNEMPLOYMENT

(a) First Passage Time Distributions as an Economic Construct. We start with the influential model of Jovanovic. A worker and firm together constitute a match. At the start of the match, both are uncertain about the productivity of the match but both learn about this productivity through a Bayesian learning algorithm. Both worker and firm start with the same prior about the productivity of the match. The prior does not depend on the previous labor market history of the worker because each match is unique: the productivity is not inherent in either the worker or the firm alone. The prior is then updated as the two partners continue their relationship.

Jovanic demonstrates that a competitive equilibrium wage policy of the firm which is also a socially optimal wage policy is to pay all employees their expected marginal productivity where the expectation is computed with respect to the updated prior on the productivity of the worker - firm match. Workers, however, have alternatives. These alternatives include employment in other firms, participation in social transfer programs which subsidize unemployment, or nonmarket activity.

Jovanovic demonstrates that the worker and firm continue their match until the time when the perceived productivity of the match - $\hat{W}(t)$ - falls below the reservation wage - $Q(t)$ - defined as the monetary value of the best alternative to the current employment match. From his assumption of a Wiener wage growth process $\hat{W}(t)$ is normally distributed. The length of the match is the first passage time to the event $\hat{W}(t) < Q(t)$. The first time that this occurs is denoted t^*.

Another way to formulate this model is in terms of the index function model widely used in labor economics. Define

$$I(t) = \hat{W}(t) - Q(t).$$

When $I(t)<0$, the worker leaves the firm (for a match to exist at all the initial value of the index function - $I(0)$ - must be non-negative). The first time in the match that $I(t)$ becomes negative - T^* - is termed the first passage time. (See Jovanovic, page 981, for the explicit formula for T^* in his model).

Denote the distribution of T^* by $F(t^*)$ with density $f(t^*)$. For the moment, we ignore any dependence of this distribution on observed or unobserved components. The Jovanovic model implies a special functional form for this distribution. Moreover, it implies that the same distribution characterizes all spells of a worker's employment with firms, and that the outcomes of previous matches do not determine the distribution of the first passage time or exit time from a current match.

At this point it is useful to rewrite the Jovanovic model in a more convenient form. To do so requires the introduction of the hazard function, $h(t^*)$. The hazard function is a conditional density of first passage or exit time from a spell given the length of time spent in the spell. For expositional convenience assume that the distribution function is continuous and differentiable.

Let $g(T^*|T^* \geq t^*)$ be the conditional density of the first passage time T^* given that T^* is greater than or equal to t^*. From the definition, $1-F(t^*)$ is the probability that the first passage occurs after t^*. Thus

$$h(t^*) = g(T^*|T^* \geq t^*) = \frac{f(t^*)}{1-F(t^*)} .$$

A given F implies a given h. Conversely, given h, and our assumptions, F is uniquely defined because

$$h(t^*)\,dt^* = \frac{f(t^*)\,dt^*}{1-F(t^*)}$$

so

$$F(t^*) = 1 - \exp\left(-\int_0^{t^*} h(u)du\right).$$

and also

$$\text{(0)} \qquad f(t^*) = h(t^*) \exp\left(-\int_0^{t^*} h(u)du\right).$$

Duration Dependence is said to exist if $\frac{\partial h(t^*)}{\partial t^*} \neq 0$. The only (continuous) density of exit times with no duration dependence is the exponential density. Thus if

$$f(t^*) = a \exp(-at^*),\ a > 0$$

$$h(t^*) = a$$

and

$$\frac{\partial h(t^*)}{\partial t^*} = 0.$$

If $\frac{\partial h(t^*)}{\partial t^*} > 0$ there is positive duration dependence. In this case, the longer a worker has been in a job, the more likely he is to escape it in the next "interval" of time $(t^*, t^* + dt^*)$. If $\frac{\partial h(t^*)}{\partial t^*} < 0$, there is negative duration dependence, and the longer a worker has been in a job the less likely he is to exit it in the next "small interval" of time. Positive duration dependence during unemployment is associated with a declining reservation wage in search theory (e.g., Lippman and McCall). Negative duration dependence during employment is associated with firm specific capital in the theory of turnover.

The hazard function arises as a simple and readily interpreted representation of the structural conditional distribution of first passage times in many models of labor market turnover. The Mortensen, Lippman and McCall search theory generates a structural distribution of first passage times. In the infinite horizon case with a stationary economic environment, the optimal search strategy is stationary so that there is no duration dependence in unemployment spells. A shrinking horizon (Gronau), systematic job search over different wage distributions (Salop), or declining assets in a utility maximizing job search model (with constant relative risk aversion, see Danforth, or Hall, Lippman and McCall), all generate optimal stopping time distributions with positive duration dependence. The precise functional form of the hazard function is determined by the distribution of the random shocks facing the agents. Typically, these specific distributional assumptions are imposed as a matter of mathematical convenience in formulating a theory and are not, themselves, justified by an appeal to theory. For this reason, it is important to develop a flexible approach to estimation that does not require special functional forms to secure estimates. The approach to empirical model building which is developed below permits the analyst to explore the sensitivity of his estimates to special assumptions about functional forms.

Holt's model of labor market dynamics, while not derived from an explicit optimizing model (but see Toikka), offers another example of a continuous time labor market model. Holt works with the probability that an individual will be in one of a set of labor market states at a point in time. For convenience of exposition, we consider only a two state model and we designate those states by "e" and "u" - shorthand notation for employment and unemployment. We assume all workers participate in the labor force at each point in time.

The first passage time or exit time distribution <u>from</u> the employment state is $f_e(t_e)$. The first passage time <u>from</u> the unemployment state is $f_u(t_u)$. The labor market history of an individual is governed by these two distributions. Given the initial state, labor market histories are generated by realizations of these first passage distributions. By assuming no duration dependence (or constant hazard) in either labor market state, Holt specializes these distributions to

$$f_e(t_e) = a_e \exp(- a_e t_e)$$

$$f_u(t_u) = a_u \exp(- a_u t_u).$$

Holt does not work with these duration distributions directly. Rather, he works with the probabilities that a person will be in each state at a point in time, $P_e(t)$ and $P_u(t)$ (= $1 - P_e(t)$). These probabilities can be derived from the densities of exit times by the following argument.

Suppose that a person is in state e at time t. This probability is $P_e(t)$. The conditional probability of exit from the state in time interval $(t + \Delta t)$ is simply the hazard, $(a_e \Delta t)$. Thus the probability of exit from the employment state to the unemployment state is $(a_e \Delta t)$ and by a parallel argument the probability of exit from the unemployment state to the employment state is ($a_u \Delta t$). The conditional probability of remaining in the unemployment state is $1-(a_u \Delta t)$. As $\Delta t \to 0$, the probability of remaining in the state becomes unity - a result consistent with fixed costs of changing state. Assuming that a_e and a_u are bounded positive numbers, the probability that a person is employed at time $t + \Delta t$ is

$$P_e(t + \Delta t) = (1 - a_e \Delta t)P_e(t) + (a_u \Delta t)\ P_u(t)$$

<u>i.e.</u> a person is employed at $t + \Delta t$ either by remaining employed (with probability $(1 - a_e \Delta t)$) or becoming employed from the unemployment state (which occurs with probability $a_u \Delta t$).

Rearranging terms

$$\frac{P_e(t + \Delta t) - P_e(t)}{\Delta t} = - a_e P_e(t) + a_u P_u(t).$$

Passing to the limit as $\Delta t \to o$,

$$\dot{P}_e(t) = - a_e P_e(t) + a_u P_u(t).$$

By a parallel argument

$$\dot{P}_u(t) = a_e P_e(t) - a_u P_u(t).$$

This system of equations generates a continuous time Markov process. Given the probability of being in each initial state, these equations can be solved to yield

$$P_e(t) = \frac{a_u}{a_e + a_u} + \left\{P_e(o) - \frac{a_u}{a_e + a_u}\right\} \exp - (a_e + a_u)t \quad \text{and}$$

$$P_u(t) = \frac{a_e}{a_e + a_u} + \left\{P_u(o) - \frac{a_e}{a_e + a_u}\right\} \exp - (a_e + a_u)t.$$

As $t \to \infty$, these probabilities converge to constants irrespective of initial conditions. If the process starts in equilibrium (so $P_e(o) = \frac{a_u}{a_e + a_u}$ and $P_u(o) = \frac{a_e}{a_e + a_u}$), convergence is immediate.

The equilibrium probabilities have strong intuitive appeal. The larger the exit rate (or hazard) a_u from the unemployment state relative to the exit rate from the employment state, a_e, the more likely is the person to be found in the employment state at a point in time. In equilibrium, the odds of finding someone in the employment state are a_u/a_e.

Burdett and Mortensen present a model of search and labor supply that is developed in terms of state probabilities. It is Markovian conditional on market wages and reservation wages. Knowledge of the hazard function and the initial state of the process is sufficient information to calculate the state probabilities. Thus the methods presented in this paper can be used with some modification (by incorporating wages) to estimate their model.

(b) Introducing Heterogeneity Into the Model

Heterogeneity is defined as unmeasured and measured exogenous variables that differ among individuals and that may differ over time for the same individual. The term is usually reserved for unobservables as perceived by the data analyst. This paper considers both types of variables. Uncorrected heterogeneity leads to biased estimates of duration dependence. If individual exit time distributions are exponential but individuals have different exponential parameters, estimated hazard functions exhibit negative duration dependence.

To see this, let the exit time density be $a[\exp(-at)]$. The density of a in the population is $g(a)$. Assuming an ideal data set in which all spells are completed, for a large random sample of individuals the estimated empirical distribution function of exit times $\hat{K}(t)$ converges to the population distribution $K(t)$ defined as

$$K(t) = 1 - \int e^{-at} g(a)\,da.$$

The empirical hazard function converges to

$$h(t) = \frac{k(t)}{1 - K(t)} = \frac{\int a\, e^{-at} g(a)\,da}{\int e^{-at} g(a)\,da}$$

where

$$\frac{\partial h(t)}{\partial t} < 0$$

by the Cauchy-Schwartz inequality. [1] Intuitively, high a individuals are the first to exit the state leaving behind the low a individuals. This shows up as negative duration dependence in the fitted distribution.

By a theorem of Barlow and Proschan (p. 37) if each individual exit time distribution exhibits negative duration dependence over the entire range of values of exit time the fitted hazard function also exhibits negative duration dependence. The only way for the fitted hazard to exhibit positive duration dependence is for some (but not necessarily all) individual exit time distributions to have positive duration dependence over at least some portion of the domain of the distribution of exit times.

This paper controls for heterogeneity in observed and unobserved variables by parameterizing the hazard function in a general way. The strategy adopted here is to write

$$(1) \qquad h(t) = \exp\{Z(t+\tau)\beta + \gamma_1 \frac{t^{\lambda_1}-1}{\lambda_1} + \gamma_2 \frac{t^{\lambda_2}-1}{\lambda_2} + V(t+\tau)\}$$

$$\lambda_2 > \lambda_1 \geq 0$$

1

$$\frac{\partial h(t)}{\partial t} = \frac{\left[\int a\, e^{-at} g(a)da\right]^2 - \left[\int a^2 e^{-at} g(a)da\right] \cdot \left[\int e^{-at} g(a)da\right]}{\left[\int e^{-at} g(a)da\right]^2}$$

The numerator of the expression on the right hand side is nonpositive by the Cauchy-Schwartz inequality for integrals [Buck, p. 123]. It is strictly negative so long as g(a) is nondegenerate.

where $Z(t+\tau)$ is a $1 \times K$ vector of exogenous variables as of calendar time $t+\tau$. β is a $K \times 1$ vector of coefficients. τ is the calendar time at which the spell commences. Duration dependence is captured by the two terms $\frac{t^{\lambda_1}-1}{\lambda_1}$ and $\frac{t^{\lambda_2}-1}{\lambda_2}$. This treatment of the duration terms is clearly analogous to the Box-Cox transformation used in regression analysis. Unobserved variables $V(t+\tau)$ are permitted to be functions of time $(t+\tau)$. By exponentiating the term in brackets, we ensure that $h(t)$ is positive as required since $h(t)$ is a conditional density function.

This formulation of the hazard is more general than any we have seen. It contains, as special cases, virtually all of the commonly utilized hazard functions. For example, setting all elements of β equal to zero except for the intercept term β_o, and assuming $V(t+\tau) = 0$ for all t and τ, a variety of interesting special cases widely used in the literature on reliability theory can be generated. If we set $\lambda_1 = 0$ and $\gamma_2 = 0$, we obtain the Weibull hazard rate

(2) $$h(t) = e^{\beta_0} t^{\gamma_1} \quad \text{(Weibull)}.$$

Duration dependence is monotone in this model and its sign is the same as the sign of γ_1. If we set $\lambda_1 = 1$ and $\gamma_2 = 0$, we obtain the Gompertz hazard

(3) $$h(t) = e^{(\beta_0 - \gamma_1)} e^{\gamma_1 t} \quad \text{(Gompertz)},$$

By specifying alternative values for λ_1, γ_1, λ_2 and γ_2, a variety of models of duration dependence can be generated. In particular, the essential features of Jovanovic's turnover model can be captured by choosing $\lambda_1 = 1$ and $\lambda_2 = 2$. Jovanovic predicts that $\gamma_1 > 0$ and $\gamma_2 < 0$ so that initial positive duration dependence is eventually followed by negative duration dependence. We demonstrate below that γ_1, γ_2, λ_1 and λ_2 can be estimated and classical hypothesis testing procedures can be used to test among competing models of duration dependence.

Our model (1) extends previous work by permitting the exogenous variables to vary freely within spells.[1] Although time varying variables are a computational nuisance they are a fact of life. In the empirical work reported below we demonstrate that this extension makes an important difference in our estimates of the impact of key economic variables on turnover probabilities and enables us to generate sensible parameter estimates.

Finally, our treatment of heterogeneity generalizes previous work by permitting unobserved components to be correlated across spells. Work by Tuma, Hannan and Groeneveld assumes that unobserved components are uncorrelated across spells.[2] However, in this paper, our treatment of heterogeneity components is somewhat restrictive. We assume that within each spell $V(t+\tau) = V$, i.e. that heterogeneity components are constant within spells. Heterogeneity components are permitted to vary across spells. There is no particular reason to assume that unobserved components behave the way we assume they do unless unobserved heterogeneity components are immutable person-specific effects. It is likely that unobserved components change within spells. This assumption is made solely to simplify the computational procedure discussed below. Its relaxation is a major goal of our future research.

In order to simplify the exposition we have thus far confined our attention to the formulation of the hazard function for a single spell of an event. The procedure outlined above can be extended to multiple episodes of the event. Let j index the episode number. The hazard for the jth episode may be written as

[1] Tuma, Hannan, and Groneveld permit one variable in a set of variables to change within spells.

[2] Heckman and Borjas introduce heterogeneity correlated across spells.

$$(4) \qquad h_j(t_j) = \exp\{Z(t_j + \tau_j)\,\beta_j + \gamma_{1j}\,\frac{(t^{\lambda_{1j}}-1)}{\lambda_{1j}} + \gamma_{2j}\,\frac{(t^{\lambda_{2j}}-1)}{\lambda_{2j}} + V_j\}$$

$$\lambda_{1j} < \lambda_{2j}$$

where τ_j is the date of the onset of the jth spell, β_j, λ_{1j}, γ_{1j}, γ_{2j} and λ_{2j} are coefficients for the jth spell and V_j is an unobserved heterogeneity component for the jth spell. This general parameterization permits behavioral coefficients to differ depending on the serial order of the spell. Such shifts in coefficients has been termed "occurrence dependence" by Heckman and Borjas. A simple economic model of "occurrence dependence" or stigma is presented in Appendix C. To simplify the computations we restrict unobserved heterogeneity across spells to a one factor error specification where $V_j = C_j V$ and C_j is a parameter of the model.

The extension to a two state multiepisode model is immediate. Let "e" and "u" denote the two states. The hazard function for the jth episode of state ℓ (ℓ = e, u) may be written as

$$(5) \qquad h_{j\ell}(t_{j\ell}) = \exp\{Z(t_{j\ell}+\tau_{j\ell})\,\beta_{j\ell} + \gamma_{1j\ell}\,\frac{(t^{\lambda_{1j\ell}}-1)}{\lambda_{1j\ell}} + \gamma_{2j\ell}\,\frac{(t^{\lambda_{2j\ell}}-1)}{\lambda_{2j\ell}} + V_{j\ell}\}$$

where $\lambda_{1j\ell} < \lambda_{2j\ell}$. Again, to simplify the calculations, we restrict heterogeneity to a one factor specification

$$(6) \qquad V_{j\ell} = C_{j\ell}\,V.$$

C. THE LIKELIHOOD FUNCTION FOR A TWO-STATE MODEL

This section presents the likelihood function for a two state version of our model. The general likelihood function is presented in Appendix A.

Individuals are observed for a time interval of length T. At the start of the interval, the individual is in one of two states "e" or "u." To commence the analysis we adopt the simplifying assumption that the beginning of the observation interval is also the initial entry date of the individual into the work force. This assumption permits us to postpone discussion of the problem of correct treatment of initial conditions for the process until the next section. The data analyzed below are consistent with this assumption.

To simplify the notation, write hazard function (5) with heterogeneity specification (6) as

$$h_{j\ell}(t_{j\ell}|V)$$

where conditioning on the exogenous Z variables and the date of the onset of the spell is left implicit. Recall that ℓ = e or u and that j denotes the serial order of the spell. Subscript 3e thus denotes the third spell of employment which starts at calendar time τ_{3e}. "$q(V|\theta)$" is the density of heterogeneity components in the population. "θ" is a parameter vector that characterizes the distribution. The V component is the same for each individual across spells but is independently distributed across people.

The density function of exit times $t_{j\ell}$ for an individual who has K_e completed spells of employment and K_u completed spells of unemployment and who at the end of interval T is in an uncompleted spell of event ℓ_T of length $\bar{t}_{K_T,\ell_T}$ (where K_T is either K_e+1 or K_u+1 and ℓ_T is either e or u) is, using equation [illegible].

$$(7)\quad \int \{ \prod_{j=1}^{K_e} h_{je}(t_{je}|V) \exp - \int_0^{t_{je}} h_{je}(\eta|V)d\eta\} \{ \prod_{j=1}^{K_u} h_{ju}(t_{ju}|V) \exp - \int_0^{t_{ju}} h_{ju}(\eta|V)d\eta\} \cdot$$

$$\{\exp - \int_0^{\bar{t}_{K_T,\ell_T}} h_{K_T,\ell_T}(\eta|V)d\eta\}\ q(V)dV ,$$

where $\sum_{j=1}^{K_e} t_{je} + \sum_{j=1}^{K_u} t_{ju} + \bar{t}_{K_T,\ell_T} = T$. K_e and K_u differ by at most one in absolute value because the individual is always in one of the two states.

The first term in brackets is the product of the K_e densities of exit time from employment. Each of these densities is conditioned on V, the heterogeneity component. The second term in brackets is the product of K_u conditional (on V) densities of exit times from unemployment. The third term is the probability that the K_T^{th} spell of event ℓ_T lasts at least $\bar{t}_{K_T,\ell_T}$. This probability is also conditioned on V, the unobserved heterogeneity component. Integration with respect to V eliminates the conditioning. Integrating out V is formally equivalent to integrating out nuisance parameters. An alternative approach to estimation would be to treat V as a fixed effect for each individual. Except for some very special cases, the latter approach leads to a serious incidental parameters problem (see, e.g., Heckman), and results in inconsistent parameter estimates in short panels. For this reason we adopt a "random effect" approach in our empirical analysis.

Under the assumption that individual event histories are obtained from random samples of individuals, the appropriate log likelihood is the sum of the log of density (7) for each individual in the sample. Maximizing this function with respect to the parameters β, $\gamma_{1j\ell}$, $\gamma_{2j\ell}$, $\lambda_{1j\ell}$, $\lambda_{2j\ell}$ and $C_{j\ell}$, for ℓ = "e" or "u," and θ, the parameters generating the density $q(V|\theta)$, produces maximum likelihood estimators which can be shown to be both consistent and asymptotically normally distributed as the number of event histories becomes large. Valid large sample test statistics for parameter vectors can be based on the estimated information matrix. To implement the model, it is necessary to make some assumption about the functional form of $q(V|\theta)$. In this paper, we assume that it is a standard normal density. Since the parameters C_{je} and C_{ju} can be freely chosen, this specification does not restrict the values of variances of $V_{j\ell}$ across spells. In later work, we plan to experiment with a variety of densities for V in order to check the sensitivity of the estimates to alternative specifications of the density of heterogeneity components.

The likelihood function presented in this section generalizes previous work by permitting (a) introduction of time varying explanatory variables, (b) correction for heterogeneity components correlated across spells, (c) estimation of general forms of duration dependence. By permitting structural coefficients to change across different spells of the event, we can estimate a model of "stigma" or "occurrence dependence" of the sort discussed in Heckman and Borjas and derived in Appendix C. Lagged values of lengths of previous spells can also be introduced as explanatory variables in the model to capture the notion of "lagged duration dependence" advanced by Heckman and Borjas. Since the likelihood function accounts for each unit of time spent in the sampling interval 0 - T, it naturally corrects for incomplete or censored spells of events that are a consequence of the sampling scheme.

Testing for the presence of heterogeneity raises certain delicate statistical problems that arise in testing for values of parameters at the boundaries of parameter spaces (Moran, 1973). A straightforward test for the presence of serially correlated unobserved components can be constructed that avoids these problems.

The following test procedure is proposed. Under the null hypothesis that there is no serial correlation in unobservables. Future values of duration variables should not be statistically significant determinants of current duration distributions if heterogeneity is ignored in estimating the parameters of current duration distributions. Standard (asymptotic) significance tests on the estimated coefficients of future duration variables estimated without correcting for heterogeneity can be used to test the hypothesis of no serial correlation in unobservables for any two spells of the event.[1] This test is not informative on the presence or absence of heterogeneity components distributed independently across spells.

It is possible to estimate both λ_{1j} and λ_{2j} (subject to the restriction that $\lambda_{1j} < \lambda_{2j}$) and the associated coefficients γ_{1j} and γ_{2j}. Using the estimated information matrix one can construct a joint confidence interval for these coefficients and determine whether or not certain restricted models lie within the confidence interval. If they do, the data are consistent with the restricted models. Thus, for example, if the estimated confidence interval includes $\gamma_{2j} = 0$ and $\gamma_{1j} = 1$, a Gompertz hazard is consistent with the data for spell j. If the confidence interval includes $\gamma_{2j} = 0$ and $\gamma_{1j} = 0$, the data are consistent with a Weibull hazard. By examination of the confidence interval for the general model it may thus be possible to select a more parsimonious model.

---

[1]We advocate use of future values rather than lagged values in constructing the test because of the possibility that lagged duration variables may be present in the structural model. See Heckman and Borjas (1980).

(d) Initializing the Process[1]

It is a rare data set for which the beginning of the sample observation interval is also the beginning of the individual's entry into the workforce - the assumption made in the last section. More typically, a sample begins with individuals caught midstream in an employment or unemployment spell. It is commonly the case that we know which state an individual occupies at the beginning of the sample, and possibly the serial order of the spell but we do not know the length of time spent in the spell before the individual is observed. This section presents methods for adjusting the likelihood function presented in the preceding section to account for this problem.

In order to focus on the essential aspects of the problem, we commence the analysis under the assumption of a strictly stationary economic environment, and under the further assumption that the process is in equilibrium. People can be in only one of two states: employed or unemployed. The density of exit time from employment, $f_e(t_e)$, is the same across people and across time for the same person. The same is true of the density of exit time from unemployment, $f_u(t_u)$. The process is assumed to have been in operation for a long time for each person. Let μ_e and μ_u be the mean time in employment and unemployment, respectively. These means are assumed to be finite. We abstract from occurrence and lagged duration dependence, but permit duration dependence.

Suppose that we first observe the process for each individual at calendar time τ_0. The probability that a randomly selected person will be found to be employed at τ_0 is

$$\pi_e = \frac{\mu_e}{\mu_e + \mu_u} .$$

A derivation for this expression can be found in Cox (p. 36). Intuitively, each complete employment and unemployment episode lasts $\mu_e + \mu_u$ on average

[1] Discussions with Marjorie McElroy and Burton Singer clarified our thinking on the problems discussed in this section.

and each complete employment spell lasts μ_e. Hence π_e is the proportion of time we expect to find an individual employed over a long observation period.

Given that a person is employed at time τ_o, what is the distribution of a completed employment spell sampled at τ_o? The completed spell length includes the portion observed after the start of the observation period at τ_o and the unobserved portion completed prior to τ_o. The distribution of exit times from the sampled spell is not the distribution $f_e(t_e)$. This is so because the completed lengths of spells sampled at τ_o are on average longer than the typical spell. Longer spells are more likely to be sampled than shorter spells. Amplifying an heuristic argument due to Cox and Lewis (pp. 61-63), the density of the sampled completed spell, denoted t_e^o, can be derived in the following way.[1]

Condition on the event that the individual is employed at the time he is sampled. The number of employment spells of length x that occur in J episodes of total time spent in employment $\sum_{j=1}^{J} (t_e)_j$ is $n^{(J)}(x)$. For a random sampling scheme, random across time for a single person or across people at a point in time, the probability of sampling a spell of length x is

$$\frac{x n^{(J)}(x)}{\sum_{j=1}^{J} (t_e)_j} .$$

This expression is the ratio of the total length of spells of length x to the total length of the employment process. On average, one will tend to oversample longer spells by the random sampling process. Divide the numerator and denominator by J, the number of spells of employment, and let J grow large, so that total time spent in employment gets large. Then

$$\lim_{J \to \infty} \frac{\sum_{j=1}^{J} (t_e)_j}{J} = \mu_e$$

[1]For discussion of this point in the analysis of unemployment spells see Salant, p. 56.

by the strong law of large numbers and

$$\lim_{J \to \infty} \frac{n^{(J)}(x)}{J} = f_e(x).$$

Then the density of sampled employment spells of length x is

$$\lim_{J \to \infty} \frac{x\,n^{(J)}(x)}{\sum_{j=1}^{J} (t_e)_j} = \frac{x\,f_e(x)}{\mu_e}.$$

The mean of the sampled employment distribution exceeds the mean of employment μ_e so long as the variance of t_e (σ_e^2) is positive. The mean of the sampled distribution is $\frac{\sigma_e^2}{\mu_e} + \mu_e$ which is clearly greater than μ_e since the mean is always positive. For exponential exit time distributions this mean is twice μ_e, the mean value of an employment interval. By randomly sampling across time for a single person or across people, we overestimate the population mean length of employment (and unemployment) duration.

Collecting results, the density of the completed employment spell sampled at time period τ_o is

$$g\,(t_e^o) = \frac{t_e^o\, f_e(t_e^o)}{\mu_e} \quad (8)$$

We do not observe t_e^o. Instead we observe $\bar{t}_e^o$, the time from the origin of the sample to the completion of the spell.

For any length of completed spell t_e^o, as a consequence of stationarity any value of $\bar{t}_e^o$ is equally likely as long as $0 \leq \bar{t}_e^o \leq t_e^o$. Thus the conditional density of $\bar{t}_e^o$ given t_e^o is

$$(9) \qquad k(\bar{t}_e^o \mid t_e^o) = \frac{1}{t_e^o} \quad , \; 0 \leq \bar{t}_e^o \leq t_e^o .$$

The joint density of $\bar{t}_e^o$ and t_e^o is the product of (8) and (9),

$$m(\bar{t}_e^o, t_e^o) = \frac{1}{t_e^o} \quad \frac{t_e^o \, f_e(t_e^o)}{\mu_e} \quad .$$

Thus the marginal density of $\bar{t}_e^o$ is

$$\psi e(\bar{t}_e^o) = \int_{\bar{t}_e^o}^{\infty} \frac{1}{\mu_e} f_e(t_e) dt_e$$

$$= \frac{1-F_e(\bar{t}_e^o)}{\mu_e} \quad . \qquad (1)$$

Recall that these derivations depend on the assumption that the sampled spell is an employment spell. The probability that the sampled spell is an employment spell is $\pi_e = \frac{\mu_e}{\mu_e + \mu_u}$. Thus the

unconditional density of an observed first spell of employment is

$$\psi_e^*(\bar{t}_e^o) = \psi_e(\bar{t}_e^o)\pi_e = \frac{1-F_e(\bar{t}_e^o)}{\mu_e + \mu_u} \quad .$$

---

[1] This density is also the density for time in the spell prior to τ_o, Δ_e^o. See Cox, p. 61. Replacing Δ_e^o for $\bar{t}_e^o$ in $\psi_e(\bar{t}_e^o)$ produces the density for Δ_e^0.

By an entirely parallel argument the unconditional density of an observed first spell of unemployment is

$$\psi_u^* (\bar{t}_u^o) = \psi_u (\bar{t}_u^o) \pi_u = \frac{1 - F_u(\bar{t}_u^o)}{\mu_e + \mu_u}$$

where the new symbols used in this expression are defined in the obvious way replacing "e" with "u."

It is straightforward to modify the analysis by adding exogenous heterogeneity components provided that these components are identical across spells. They may differ across people. Thus, in place of $f_\ell(t_\ell)$ (ℓ = e,u) we may write $f_\ell(t_\ell \mid Z,V)$ and the derivation may be repeated as before with $\psi_\ell^* (\bar{t}_{\ell_o}^o \mid Z,V)$, ℓ = e,u, written in place of $\psi_\ell^* (\bar{t}_{\ell_o}^o)$.

With this modification, the density function for the exit times $\bar{t}_{\ell_o}^o$ (where the subscript denotes the state at sampling time period "o") and $t_{j\ell}$ for an individual who has K_e completed spells of employment and K_u completed spells of unemployment and who at the end of interval T is in an incomplete spell of event ℓ_T of length $\bar{t}_{K_T \ell_T}$ (where K_T is either $K_e + 1$ or $K_u + 1$ and ℓ_T is either e or u) is

$$(10) \qquad \int \psi_{\ell_o}^*(\bar{t}_{\ell_o}^o \mid Z, V)\{ \prod_{j=1}^{K_e} h_e(t_{je} \mid V) \exp - \int_0^{t_{je}} h_e(n \mid V)dn \} \{ \prod_{j=1}^{K_u} h_u(t_{ju} \mid V) \cdot$$

$$\exp - \int_0^{t_{ju}} h_u(n \mid V)dn \} \cdot \{\exp - \int_0^{\bar{t}_{K_T, \ell_T}} h_{\ell_T}(n \mid V)dn\} g(V)dV \ .$$

Under the assumption that individual event histories are obtained from random samples of individuals, the appropriate log likelihood for this sampling scheme is the sum of the log of density (10) for each individual in the sample. Maximizing this function with respect to the parameters of the model produces consistent and asymptotically normally distributed estimators as the number of event histories becomes large.

This procedure produces an exact solution to the problem of initializing the likelihood in a stationary environment if the process is in stationary equilibrium. It is not a general solution. In many cases the two key assumptions - equilibrium and a stationary environment - are unlikely to be even approximately correct. In this case the procedure just presented is not valid.

We postpone a general discussion of this problem to a later occasion. Here we sketch two solutions, both of which are patterned after a discussion by one of us in a previous paper (Heckman). In a general nonstationary environment, in order to correctly initialize the process, we require the probability distribution of the first spell in our sample. Its derivation depends on the rule used to select the sample being analyzed, the probability that the individual is in a given state at the time he is sampled, and the distribution of the length of time spent in the first spell in the observed sample period given the state that the individual is in when the sampling begins. The density of the first spell duration $\bar{t}^o_{\ell_o}$ may be written as

$$\lambda(\bar{t}^o_{\ell_o} \mid V, \{Z(r)\}_{-\infty}^{\bar{t}^o_{\ell_o}}) \qquad \ell_o = e, u \quad .$$

This density does not, in general, have the same functional form as f_e or f_u. It depends on presample values of the exogenous variables as well as within sample values. In general, we do not know its exact functional form. However, we can approximate it using the flexible functional forms for the hazards given in equations (5) and (6). One strategy for empirical work is to parameterize the first spell in the sample differently from that of subsequent spells and utilize density λ in place of $\psi^*_{\ell_o}$ in equation (10).

A practical difficulty with this solution to the problem of initial conditions is that presample values of the exogenous variables are required to form the density of the exit time from the first spell in the sample. While such data may be available for some samples, this is unlikely in most cases.

There are two approaches to this problem. Denote the density function for the exogenous variables by $k(Z|\chi)$ where χ is a vector of parameters of the density generating the data. The first approach estimates this density from data. The exogenous variables within the sample period might be used to estimate the density or else auxiliary data sources might be utilized. Given consistent estimates of the density function of the missing data, denoted $\hat{k}(Z|\chi)$, one can form $\hat{\lambda}^*_{\ell_o}$ defined as

(11) $$\lambda^*_{\ell_o} = \int \lambda_{\ell_o}(\bar{t}^o_{\ell_o} \mid \{Z(\tau)\}^{\bar{t}^o_{\ell_o}}_{-\infty}, V)\ \hat{k}(Z|\chi)dZ .$$

Inserting $\lambda^*_{\ell_o}$ in place of λ_{ℓ} in density (10) is equivalent to integrating out the missing data, and forming an estimated likelihood function. From the exogeneity of Z, and the assumed convergence of $\hat{k}(Z|\chi)$ to $k(Z|\chi)$ the estimated likelihood converges to the true likelihood and maximum likelihood estimators based on the estimated likelihood converge to the true maximum likelihood estimators.[1]

The second approach is to jointly estimate the data density and structural parameters of the model. Given the exogeneity of Z, no advantage accrues to this computationally more demanding approach.

A second solution to the problem of initial conditions proceeds conditionally on V. Substituting λ for ψ^* in density (10) and given V = v, the probability density for the sample exit times is

$$(12) \qquad \lambda(\bar{t}^{o}_{\ell_o} \mid v, \{Z(r)\}_{-\infty}^{\bar{t}^{o}_{o}}) \cdot$$

$$\left\{\prod_{j=1}^{K_e} h_e(t_{je} \mid v) \exp - \int_0^{t_{je}} h_e(\eta \mid v)\, d\eta\right\}\left\{\prod_{j=1}^{K_u} h_u(t_{ju} \mid v) \exp - \int_0^{t_{ju}} h_u(\eta \mid v) d\eta\right\} \cdot$$

$$\left\{\exp - \int_0^{t_{K_T}, \ell_T} h_{\ell_T}(\eta \mid v)\, d\eta\right\} .$$

"v" is a parameter for each individual. One may further condition on $\bar{t}^{o}_{\ell_o}$ to write the density of the sample exit times <u>given v and $\bar{t}^{o}_{\ell_o}$</u>. This conditional density is

$$(13) \quad \left\{\prod_{j=1}^{K_e} h_e(t_{je} \mid v) \exp - \int_0^{t_{je}} h_e(\eta \mid v)\, d\eta\right\}\left\{\prod_{j=1}^{K_u} h_u(t_{ju} \mid v) \exp - \int_0^{t_{ju}} h_u(\eta \mid v) d\eta\right\} \cdot$$

$$\left\{\exp - \int_0^{t_{K_T}, \ell_T} h_{\ell_T}(\eta \mid v)\, d\eta\right\} .$$

---

[1]Due to the nonlinearity of the model, integrating out the missing data with respect to its distribution is a <u>not</u>, in general, equivalent to replacing the missing values with estimated mean values as is customary in standard linear regression models.

and does not depend on the parameters of the initial exit time distribution. Maximizing the conditional likelihood function with respect to the parameters of the model (including a value of v for each person in the sample) generates consistent parameter estimates as number of spells per person becomes large ($K_e + K_u \rightarrow \infty$). However, in the case of short panels, it is in general not possible to consistently estimate v, and due to the nonlinearity of the model estimated parameters will not be consistent because the maximum likelihood estimator will involve joint estimation of v and the structural parameters. The inconsistency in the estimator of v will be transmitted to the estimator of the structural parameters (Heckman).

II CONTINUOUS TIME vs. DISCRETE TIME MODELS

The principal advantage of continuous time models for discrete panel data over more conventional discrete time models for discrete panel data is that the parameters generating the continuous time model are invariant to the time unit used in empirical work. Commonly utilized discrete time models such as logit and probit lack this invariance property. A probit model for the occurrence of an event in a time interval of a specified length does not imply a probit model for the probability of events in time intervals of different length. Dependence of the parameters and functional form of the model on the data used to estimate it is an undesirable feature of discrete time time models that is avoided by use of the continuous time approach. Put differently, a continuous time model can always be used to generate a discrete time model while a discrete time model is critically dependent for its parameterization and interpretation on the particular time interval on which it is estimated.

To demonstrate this point, we consider the first passage time associated with a single spell of an event. We do not observe the process continuously but we know in which of a series of discrete, equispaced intervals the first passage occurs. To simplify the exposition, we ignore heterogeneity in unobserved components and initially assume that measured variables stay constant within each spell. The stochastic process starts at time "0". We observe the occurrence of the event only in equispaced time intervals of length Δ.

For a continuous time model with hazard function $h(u)$[1], the probability that a first passage occurs in the ith interval is

[1]For simplicity we do not explicitly note the dependence of the hazard function on exogenous variables.

(14) $$\exp -\left(\int_0^{(j-1)\Delta} h(u)du\right) - \exp -\left(\int_0^{j\Delta} h(u)du\right).$$

This is simply the probability that the first passage occurs sometime between $(j-1)\Delta$ and $j\Delta$. $(= F(j\Delta)-F((j-1)\Delta))$. No matter how wide or small the intervals are defined and irrespective of whether or not successive intervals are of equal length, the first passage time probability is generated by the same structural hazard function $h(u)$. Thus if Δ_ℓ is the width of the ℓth interval, the probability of a first passage in the jth interval is simply

$$\exp -\left(\int_0^{\sum_{\ell=1}^{j-1}\Delta_\ell} h(u)du\right) - \exp -\left(\int_0^{\sum_{\ell=1}^{j}\Delta_\ell} h(u)du\right).$$

The conventional discrete time model (e.g., Kiefer and Neumann or Heckman and Willis) assumes that there is a "true" time interval of length Δ. P_Δ is the probability that an event does not occur in the interval. For example, if the probit model is adopted it is assumed that

$$P_\Delta = \Phi(\beta Z)$$

where Φ is the cumulative distribution function of the normal distribution, and Z is a vector of explanatory variables with associated coefficient vector β. For this model, the probability of a first passage in the jth interval is

$$P_\Delta^{j-1}(1-P_\Delta) = (\Phi(\beta Z))^{j-1}(1-\Phi(\beta Z)).$$

Suppose that in another data set the interval widths are longer (e.g., quarters and not months). For the continuous time model this enlargement of the interval creates no problem and the probability of first passage can be expressed as a function of the underlying h(u) function by a straightforward modification of expression (14).

For the discrete time model to be defined at all on the new time scale, that scale must be a positive integer multiple of the original scale. For the discrete time model, unlike the continuous time model, fractional intervals have no meaning because the discrete time model is silent on the behavior of the process within any interval. If in fact the new time scale is m times the old scale (with m a positive integer), the probability of a first passage in the jth <u>new interval</u> is

$$(15) \qquad (P_\Delta)^{(j-1)m}\left(1-\sum_{\ell=1}^{m}\binom{m}{\ell}(P_\Delta)^{m-\ell}(1-P_\Delta)^{\ell}\right) = (P_\Delta)^{(j-1)m}(1-(P_\Delta)^{m}).$$

The second term in the expression on the left hand side is the probability that at least one event occurs in the jth time interval measured on the new scale. Neither the probability of occurrence of at least one event in the new interval $(1-(P_\Delta)^m)$ nor the probability of occurrence of exactly one event in the new interval $(1-\binom{m}{1}(P_\Delta)^{m-1}(1-P_\Delta))$ are probit functions if the probability of occurrence of the event in the old interval is probit. Of course, in this simple example, it is possible to write these probabilities in terms of the underlying probit model for the old intervals. But note that it will be necessary to modify a likelihood function used to compute parameter estimates to account for the modification of the interval width. And, as noted above, there is no legitimate procedure available to modify the discrete time model to account for non-integer expansion or shrinkage of the original interval length. It is necessary to postulate a new model--which might be assumed to be probit in the new intervals--with new parameters that cannot be derived from the original parameters of the model defined on the original interval.

Continuous time models can be used to generate discrete time models. Indeed, in light of the discussion in this section, the models consisting of equations (5), (6) and (7) can be interpreted as providing a general algorithm for producing a class of discrete time models for the analysis of discrete panel data. In contrast with conventional approaches in econometrics, the approach offered in this paper provides a parameterization of discrete time models that is independent of the time scale in which the occurrence of discrete events is measured.

We do not want to overstate the case for continuous time models. As noted by Singer and Spilerman, Phillips and others, aggregation of continuous time data into interval data of the sort described in this section may lead to non-identification of the hazard function h(u). This is clearly the case if the hazard is arbitrarily specified. If sufficient smoothness is imposed on the hazard, as is done for the hazards utilized in this paper (see equations (1), (4) and (5)), this identification problem does not arise. Nonetheless, it is important to note that without imposing information of some sort, time aggregated data may not always be used to recover the underlying hazard. An infinity of hazard functions defined over time intervals within observed sampling periods can produce the same time aggregated data. Unless a smoothness assumption is imposed, it is not possible to utilize the time aggregated data to recover the underlying hazard h(u). Without such an assumption, there is no necessary advantage in using continuous time models in place of a conventional discrete time approach. Put differently, in the absence of such identifying assumptions, and in the presence of time aggregated data, the relative merits of the continuous time approach fade, and the models advanced in this paper must be interpreted as just one of a variety of discrete time models that might be used to analyze discrete panel data.

III PITFALLS IN USING REGRESSION METHODS TO ANALYZE DURATION DATA

Appendix B presents a derivation of the properties of a Weibull regression model that can be used to analyze duration data and estimate duration dependence parameters. This section considers two important problems that arise in using standard regression analysis to analyze duration data. The first problem is one of sample selection bias.[1] The second problem is the difficulty that arises from introducing time varying explanatory variables into regression models. The first problem arises because most panel samples are short. In the course of a panel, some individuals never complete a single spell of an event while others will have multiple spells and even those individuals will usually have one unfinished spell in the course of the panel. Commonly used procedures, such as utilizing only completed spells of events for regression analysis, impose a sample selection criterion on the sample used to execute the empirical work. Failure to

[1] This problem is discussed in Tuma and Hannan.

account for such conditioning results in biased parameter estimates. The resulting bias depends on the particular rule chosen and on the length of the panel. The dependence of the bias on the length of the panel makes regression estimates from panel samples of different lengths noncomparable even when the same selection rule is employed for generating data to estimate models.

The second problem--introducing time varying exogenous variables into a regression analysis for duration data--arises because the appropriate functional form of the regression equation depends on the time profile of the exogenous variables. Conditioning duration regressions on "exogenous" variables measured after a spell begins may induce simultaneous equations bias in the regression estimators. Ad hoc solutions to these problems convert truly exogenous variables into endogenous variables and hence result in biased estimates. This section considers both of these problems starting with the first one.

To focus on essential ideas, consider a regression analysis of duration data for a particular type of event--e.g., the lengths of time spent in consecutive jobs. To simplify the analysis we assume that no time elapses between consecutive jobs. The density of duration in a given job for an individual with fixed characteristics Z is

$$f(t|Z).$$

Unobserved heterogeneity components are assumed to be absent from the model. The expected length of t given Z is

$$E(t|Z) = \int_0^\infty t f(t|Z)dt = g(Z). \tag{17}$$

From a regression analysis we seek to estimate the parameters of $g(Z)$. For example, if $f(t|Z) = \theta(Z) \exp-(\theta(Z)t)$, $\theta(Z) > 0$,

$$(18) \qquad E(t|Z) = \frac{1}{\theta(Z)} .$$

Defining $\theta(Z) = (\beta Z)^{-1}$,

$$(19) \qquad E(t|Z) = \beta Z.$$

Under ideal conditions, a regression of t on Z will estimate β. We now specify those conditions.

Suppose that the data at our disposal come from a panel data set of length T. To avoid inessential detail suppose that at the origin of the sample — "0" — everyone begins a spell of the event. This assumption enables us to ignore problems with initial conditions.

We would like to use this data to estimate $E(t|Z)$. But in our panel sample the expected value of the length of the first spell is <u>not</u> $E(t|Z)$ but is rather

$$(20) \qquad E(t|Z,T) = \int_0^T tf(t|Z)dt + T \int_T^\infty f(t|Z)dt \leq E(t|Z).$$

Thus, in the exponential example

$$(21) \qquad E(t|Z,T) = \beta Z \left[1 - e^{-T/\beta Z} \right] .$$

Clearly, a least squares regression of t on Z will not estimate β. As $T\to\infty$, the bias disappears. In the exponential example, as T becomes big relative to the mean duration, $1/\theta(Z)$, the bias becomes small.

One widely used method utilizes only completed first spells. This results in another type of selection bias. The expected value of t given that $t < T$ is

$$(22) \qquad E(t|Z, T, t<T) = \frac{\int_0^T tf(t|Z)dt}{\int_0^T f(t|Z)dt} .$$

In our exponential example

$$(23) \qquad E(t|Z, T, t<T) = \beta Z \frac{\left(1-e^{-(T/\beta Z)}(T/\beta Z+1)\right)}{\left(1-e^{-T/\beta Z}\right)} =$$

$$\beta Z \left(1- \frac{(T/\beta Z)e^{-(T/\beta Z)}}{(1-e^{-T/\beta Z})}\right) .$$

Again, a least squares regression of t on Z does not estimate β for this sample. As $T\to\infty$, the bias disappears.

Clearly there is also selection bias when we analyze the expected duration of a completed second spell of the event. Denote the length of spell i by t_i. The expected length of the second spell is

$$(24) \qquad E(t_2|Z, T, t_1 + t_2<T) = \frac{\int_0^T \int_0^{T-t_2} t_2 f(t_2|Z)f(t_1|Z)dt_1 dt_2}{\int_0^T \int_0^{T-t_2} f(t_2|Z)f(t_1|Z)dt_1 dt_2} .$$

Because t_1 and t_2 are independent, and hence the subscripts "$_1$" and "$_2$" can be interchanged without affecting the validity of the expression, this is also the conditional expection of the length of the first spell.[1] For a sample of individuals with at least two completed spells of the event

$$(25)\qquad E(t_2|Z,\ T,\ t_1+t_2<T) = \beta Z\left[\frac{1-e^{-T/\beta Z}(1+T/\beta Z)-\frac{(\beta Z)}{2}\ e^{-T/\beta Z}}{1-e^{-T/\beta Z}-(T/\beta Z)e^{-T/\beta Z}}\right].$$

Clearly, $E(t_1|Z,\ T,\ t_1+t_2<T) \neq E(t_1|Z,\ T,\ t_1<T)$.

The key point to extract from this discussion is that for short panels in which T is "small", regression estimators do not estimate the parameters of regression function (18). Least squares estimators are critically dependent on both the sample selection rule and the length of the panel. Two studies that utilize the same sample selection rule for generating "usable" observations will produce different regression coefficients if the studies are based on panels of different length.

It is possible to estimate the true structural parameters of the model. The maximum likelihood estimator discussed in section (I) automatically corrects for the panel length bias. Moreover, it is obvious from our exponential examples that by use of non-linear regression it is possible to retrieve the structural parameters of interest.

The same main conclusions can be obtained if we relax the simplying assumption that the process starts up at the origin date of the panel. In addition to sample selection bias and panel length bias, in models with

---

[1]Thus in testing for "mean occurrence dependence" (i.e., different regression functions for consecutive spells), the sample selection bias discussed in the text does not bias the test based on regression methods. This result is critically dependent on the assumption that the Z variables remain constant both within and across spells, and of course, that the same rule is used to generate first and second spells.

duration dependence regression estimators that do not correctly account for the length of time spent in a spell prior to the time the panel begins are subject to a further source of bias. Except for exponential duration distributions, the distribution of a left censored first spell is not the same as the distribution of an uncensored spell. (See equation (11)). The conditional expectation of a left censored distribution is not the same as the conditional expectation of the uncensored distribution. As noted in section (I.d), the maximum likelihood estimator can be used to correct for this source of bias as well.

Regression methods for the analysis of duration data also break down in the presence of time varying explanatory variables. The functional forms of the regression equations depend on the time paths of the explanatory variables. Even if the lengths of consecutive spells of an event are generated by hazard functions of identical functional form, the regression functions for durations of consecutive events will have different functional forms. In the presence of unmeasured heterogeneity components, regression equations fit conditional on the "exogenous variables" measured starting at the onset of a spell may generate biased coefficients. This is so because the particular values assumed by the explanatory variables selected in this fashion may depend on the lengths of the preceding spells. This is certainly the case for time trended explanatory variables like age, and in the empirical work we report below, it is also the case for our national unemployment rate variable. Assuming that unmeasured heterogeneity components are correlated across spells, the initial values of time trended variables in a spell will be correlated with the regression error term, and will become endogenous variables. For the usual reasons, least squares estimators will be biased.

To focus on the essential aspects of the problem, suppose that we have access to a very long panel data set ($T \to \infty$) so that we can safely ignore the length of panel bias previously considered. As before, there is only one type of event and we assume that the stochastic process begins at the beginning of our sample. The duration time distribution depends on time varying variables $Z(\tau)$ where τ denotes calendar time measured from the origin of our sample.

A common functional form for the hazard function, h(.), is assumed for all spells. "V" is a heterogeneity component common across all spells. The density of duration time in the first spell, t_1, is

<u>First spell</u>: $$h(t_1, Z(t_1), V) \exp - \int_0^{t_1} h(u, Z(u), V)du.$$

The density for the duration time in the second spell t_2 <u>given that the first spell ends at calendar time t_1</u> is

<u>Conditional Second Spell</u>: $$h(t_2, Z(t_2 + t_1), V)\exp - \int_0^{t_2} h(u, Z(u + t_1), V)\, du.$$

The marginal second spell density is obtained by integrating out t_1. Thus

$$f^*(t_2, Z, V) =$$

$$\int_0^{\infty} [h(t_2, Z(t_2 + t_1), V) \exp - \int_0^{t_2} h(u, Z(u + t_1), V)du] \cdot$$

$$[h(t_1, Z(t_1), V) \exp - \int_0^{t_1} h(u, Z(u), V)du]dt_1.$$

In the case in which the distribution of $Z(t)$ does not depend on time (i.e., time stationarity in the exogenous variables),

$$f^*(t_2, Z, V) = h(t_2, Z(t_2 + t_1), V) \exp - \int_0^{t_2} h(u, Z(u + t_1), V)\, du.$$

Otherwise the marginal second spell density will be of a different functional form than the marginal first spell density, and the regression function for the second spell will have a functional form different from that of the first spell regression function.

A simple example may serve to clarify the main points. We first demonstrate that the functional form of the regression will depend on the time path of the exogenous variables that drive the model. Consider the following exponential model for the first spell of an event:

$$f(t_1|Z, V) = \theta(Z,V)e^{-\theta(Z,V)t_1}, \quad 0 < t_1 < \infty$$

where $\theta(Z,V) = \frac{1}{\beta Z + V}$, and Z remains constant over the entire spell. The regression function for duration in the first spell is

$$E(t_1|Z, V) = \frac{1}{\theta(Z,V)} = \beta Z + V. \tag{26}$$

Suppose we consider another individual who is subject to a different value of Z before and after calendar time τ_1. The density of t_1 for this person is derived most simply from the conditional density before and after τ_1, i.e.

$$f(t_1|Z, V, t_1 < \tau_1) = \frac{\theta(Z_1, V)\, e^{-\theta(Z_1, V)t_1}}{(1 - e^{-\theta(Z_1,V)\tau_1})}, \quad 0 < t_1 < \tau_1$$

and

$$f(t_1|Z, V, t_1 > \tau_1) = \frac{\theta(Z_2, V)\, e^{-\theta(Z_2, V)t_1}}{e^{-\theta(Z_2, V)\tau_1}} \quad t_1 \geq \tau_1$$

The conditional expectation of duration in the first spell is

$$(27) \quad E(t_1 | Z_1, Z_2, \tau_1, V) = \frac{1}{\theta(Z_1, V)} + e^{-\theta(Z_1, V)\tau_1}\left(\frac{1}{\theta(Z_2, V)} - \frac{1}{\theta(Z_1, V)}\right).$$

The functional form of the regression equation (27) differs dramatically from that of (26), and the problems of inference and parameter estimation differ greatly between the two equations. In general, different time paths for the exogenous variables will result in different functional forms for the regression equations.

A commonly used regression procedure for the analysis of duration data computes regression equations conditional on the values of exogenous variables that occur at or after the start date of a spell. The intuition that underlies this procedure is that it is only the variables that occur during a spell that can explain the duration of a spell. In the presence of unobserved heterogeneity components correlated across spells this procedure is inherently dangerous. In the presence of heterogeneity, selection of explanatory variables in this fashion converts exogenous variables into endogenous ones, and guarantees simultaneous equations bias in least squares estimators.

To show this, assume the same functional form for the hazard function in all spells of the event. The conditional expectation of duration in the second spell given values of the exogenous variables that confront the individual <u>after the end of the first spell</u> is, for a case of no time varying variables

$$(28) \qquad E(t_2 | Z, V) = 1/\theta(Z, V).$$

For the case of time varying variables, the conditional expectation depends on whether or not $t_1 > \tau_1$. If $t_1 \leq \tau_1$, the conditional expectation is

$$(29) \qquad E(t_2 | Z_1, Z_2, \tau_1, V, t_1 < \tau_1) = \frac{1}{\theta(Z_1,V)}$$

$$+ e^{-\theta(Z_1,V)(\tau_1-t_1)}\left(\frac{1}{\theta(Z_2,V)} - \frac{1}{\theta(Z_1,V)}\right), t_1 \leq \tau_1 \quad .$$

For $t_1 > \tau_1$, the conditional expectation is

$$(30) \quad E(t_2|Z_1, Z_2, t_1, V, t_1 > \tau_1) = \frac{1}{\theta(Z_2, V)}, \; t_1 > \tau_1.$$

Although equations (29) and (27) are of the same functional form, there is one important difference: in equation (29) t_1 is an explanatory variable. Since unobserved heterogeneity component V is correlated across spells, t_1 is an endogenous variable in a regression model that treats V as a component of the error term of the model (i.e., a model that is not computed conditional on V). Partitioning the data on the basis of $t_1 < \tau_1$ raises further problems. By Bayes theorem, the conditional mean of V in equations (29) and (30) depends on t_1 and the explanatory variables so that the error term (inclusive of V) associated with regression specifications for equations (29) or (30) does not in general have a zero mean. A standard least squares assumption is violated and least squares estimators of duration equations will be biased and inconsistent.

The main point is quite general: whenever there are time trended or nonstationary explanatory variables in the model, conditioning the durations of subsequent spells on explanatory variables measured from the onset of those spells induces correlation between the explanatory variables and the heterogeneity component in the model.[1]

One solution to these problems is to use the marginal second spell density and compute the conditional expectation of t_2 with respect to it. For the case of no time varying variables, and in the more general case of time stationary exogenous variables, the marginal and conditional densities coincide so that the right hand side of equation (27) is the conditional expectation of t_2 with respect to the marginal second spell density. In the presence of non-stationary explanatory variables the two distributions differ.

[1] Gary Chamberlain (1981) has also discussed this problem.

In our example, the conditional expectation of t_2 computed with respect to the marginal distribution of t_2 is

$$(31)\quad E(t_2 \mid Z_1, Z_2, \tau_1, V) = \frac{1}{\theta(Z_1,V)} + e^{-\theta(Z_1,V)\tau_1} \cdot \left(\frac{\theta(Z_1,V) - \theta(Z_2,V)}{\theta(Z_2,V)}\right)\left(\frac{1}{\theta(Z_1,V)} + \tau_1\right).^{(1)}$$

In long panels, estimation of this equation avoids the endogeneity problem induced by selecting explanatory variables on the basis of past realizations of the process. Note however from inspection of equations (27) and (31), that successive conditional expectations of duration times taken with respect to the successive marginal distributions have different functional forms. This is so even though the functional form of the hazard function is invariant across spells. This means that in the presence of time varying variables it is not possible to simply pool data across successive spells of the event to estimate the common parameters of a regression function. Moreover, due to the nonstationarity of the exogenous variables, simple regression tests of "occurrence dependence" will tend not to reject that hypothesis (comparing differences in regression coefficients across consecutive spells of an event).

Two *ad hoc* procedures for coping with time trended or general non-stationary variables in a regression format are readily discussed and disposed of. The first uses average values of the regression variables

---

[1] This expectation is computed from the joint density of t_1, t_2 which is

$$\theta(Z_1,V)e^{-\theta(Z_1,V)t_1}\,\theta(Z_1,V)e^{-\theta(Z_1,V)t_2} \qquad \begin{matrix} 0<t_2<\tau_1-t_1 \\ 0<t_1<\tau_1 \end{matrix}$$

$$\theta(Z_2,V)e^{-\theta(Z_2,V)t_2}\,\theta(Z_1,V)e^{-\theta(Z_1,V)t_1}\; e^{(\theta(Z_2,V)-\theta(Z_1,V))(\tau_1-t_1)} \qquad \begin{matrix} 0<t_1<\tau_1 \\ \tau_1-t_1<t_2 \end{matrix}$$

$$\theta(Z_2,V)e^{-\theta(Z_2,V)t_2}\,\theta(Z_1,V)e^{-\theta(Z_1,V)t_1}\; e^{(\theta(Z_2,V)-\theta(Z_1,V))\tau_1} \qquad \begin{matrix} 0<t_2 \\ \tau_1<t_1 \end{matrix}$$

within a spell in a standard linear regression format.[1] This procedure ignores the change in functional form that results from different paths for regressors. It fails to capture the essential dependence of first passage time densities on the *entire* sequence of exogenous variables, not just those realized in a spell. Exogenous variables that are selected in this fashion become endogenous variables. To demonstrate this final point by way of an example consider a strictly positively trended explanatory variable. The average value of the variable within a spell is simply the multiple of the length of the spell, and is clearly endogenous.

A second *ad hoc* procedure for introducing time varying variables into a regression format is to use exogenous variables measured at a start of a spell. Again, this procedure results in a misspecification of the true conditional expectation function and for nonstationary explanatory variables manufactures simultaneous equations bias for models with heterogeneity components by selecting "explanatory" variables on the basis of prior realizations of the process.

The standard regression approach to the analysis of duration data is thus seen to be a rather fragile empirical procedure. Only in the case of long panels in stationary economic environments does it produce valid parameter estimates. These conditions are unlikely to be realized in the analysis of microeconomic labor market data. The likelihood approach corrects for length of panel bias and suitably modified corrects for other sampling rules. Time varying explanatory variables can readily be accommodated in the likelihood approach. For both reasons, we strongly prefer the likelihood approach to the regression approach in the analysis of labor market duration data.

---

[1]Cox and Lewis suggest this approach but only for explanatory variables that are not strongly time trended.

IV

EMPIRICAL ANALYSIS

This section reports the results of two empirical analyses. Subsection A compares empirical results obtained from regression and maximum likelihood procedures used to analyze spells of employment and nonemployment for a sample of young men. Our results from this analysis are of considerable substantive interest, and also serve to illustrate the biases inherent in using regression techniques to analyze duration data. Subsection B presents tests of the proposition that it is legitimate to aggregate the "unemployment" and "out of labor force" states into one state called nonemployment. This proposition is of interest given recent claims that the distinction between the two nonemployment states is artificial (see, e.g., Clark and Summers). We find that the states are empirically distinct.

The sample used to perform all the empirical work reported here is selected from the National Longitudinal Survey of Young Men, and is the same as that previously employed by Heckman and Borjas. We follow 122 young men for thiry consecutive months, from the time they graduate from high school. The small size of our sample is due to the stringent selection criteria imposed. To be included in the sample an individual must (1) be white; (2) have received a high school diploma in the spring or early summer of 1969; and (3) not have returned to school in the period beginning in the fall of 1969 and ending in December of 1971.

The sample was selected in this manner in an attempt to minimize the initial conditions problem discussed in part d of Section I. By using individuals who have recently completed schooling, we have selected individuals with little or no previous labor force experience. The vast majority of individuals in our sample have not worked in full time jobs during high school. We can safely ignore the initial conditions problem in deriving the maximum likelihood estimates presented here.

A. Comparison of Regression and Maximum Likelihood Estimates of Employment and Nonemployment Spell Duration

The regression specification employed here is similar to the Weibull model used by Heckman and Borjas . A more complete discussion of the Weibull regression model and its statistical properties is presented in Appendix B. We assume that duration times conditional on heterogeneity components are Weibull distributed, and consider the first two completed spells of employment and nonemployment. In terms of hazard (5), the specification estimated is obtained by setting $\lambda_{1j\ell} = 0$, $\gamma_{2j\ell} = 0$, and $Z(u + \tau_{j\ell}) = Z_{j\ell}$ for $0 \leq u < t_{j\ell}$, $j = 1, 2$, where 1 = employment "e" and 2 = nonemployment "n." (Recall that $\tau_{j\ell}$ is the date in calendar time at which the j^{th} spell of event ℓ begins.) The last condition imposes the requirement that the values of the exogenous variables are fixed over the duration of a spell. In our empirical analysis the values of the exogenous variables are fixed in two different ways. In the first way, we fix the exogenous variables at their beginning of spell values, so that

$$Z_{j\ell}^{(1)} = Z(\tau_{j\ell})$$

The second way uses <u>within spell averages</u> of the Z,

$$Z_{j\ell}^{(2)} = \frac{\int_0^{t_{j\ell}} Z(\tau_{j\ell} + k)\, dk}{t_{j\ell}}$$

to approximate the "true" value of $Z_{j\ell}$ within the jth spell of event ℓ.

We adopt the following one factor Weibull regression model specification:

$$\ell n\ t_{j\ell} = Z_{j\ell}\ \tilde{\beta}_{j\ell} + \tilde{V}_{j\ell} + W_{j\ell}$$

where $\tilde{\beta}_{j\ell} = \frac{-\beta_{j\ell}}{1+\gamma_\ell}$ (equation B-5)

$$\tilde{V}_{j\ell} = \frac{C_\ell \phi}{1+\gamma_\ell}$$

C_ℓ is a factor loading for event ℓ on heterogeneity component ϕ, and

$$E(W_{j\ell}) = 0,\ E(\tilde{V}_{j\ell}) = 0,\ E(\tilde{V}_{j\ell}W_{j\ell}) = 0,$$

and $W_{j\ell}$ is iid across spells and people. The variance of $W_{j\ell}$ is $\frac{\pi^2}{6}/(\gamma_{j\ell} + 1)^2$, a result derived in Appendix B.

Differencing successive completed log durations in state ℓ eliminates $V_{j\ell}$ from the regression. Focusing on the first and second completed spells of event ℓ, we may write

$$\ell n\ t_{1\ell} - \ln t_{2\ell} = (Z_{2\ell} - Z_{1\ell})\tilde{\beta}_{2\ell} + Z_{1\ell}(\tilde{\beta}_{2\ell} - \tilde{\beta}_{1\ell}) + W_{2\ell} - W_{1\ell}.$$

The residual variance for this equation has mean zero and variance $\frac{\pi^2}{3(\gamma_\ell+1)^2}$.

From the estimated residual variance it is thus possible to consistently estimate the duration dependence parameter γ_ℓ. As noted in Appendix B this estimate is derived under the assumption that there is no error in measuring completed durations. If there is, the estimated duration dependence parameter is downward biased.

As noted in Section III, restricting attention in the empirical analysis to completed spells of events generates sample selection bias. In our empirical work, we lose approximately two-thirds of our sample by imposing this sample selection requirement.

We include only two regressors in the empirical results presented here; the first is a dummy variable set to one if the individual is married with spouse present (MSP) and zero otherwise and the second is the national unemployment rate for prime age white males. The second variable is a proxy for aggregate demand. In addition to the first-differenced specification described above, we also present estimates of a two equation system of first and second spell completed log duration times using generalized least squares.

A topic of considerable interest in estimating the regression equations is the determination of the sensitivity of parameter estimates to the method used to select spell constant values for the time varying exogenous variables. For the regressors used here, the national unemployment rate varies by month, while marital status is observed yearly. In our sample period (1969-1971) the within spell variation of the unemployment variable is likely to be far greater than that of marital status. We thus expect to find that the parameters associated with unemployment will show more sensitivity to the method of selecting spell constant exogenous variables.

Since the regression approach has been shown to yield biased parameter estimates (see Section III) it is difficult to isolate the effect of fixing the exogenous variables over spells using the two regression methods examined here. To provide a benchmark we present maximum likelihood estimates of the Weibull duration model using both methods of fixing the exogenous variables. We then incorporate time varying variables. Differences among the estimates are considerable.

Regression results for log employment durations are presented in Table 1. The top panel contains the estimates of the two spell GLS model, while the lower panel contains estimates from the first difference specification. On the left side of the table the values of the regressors is the average value over the spell ($Z_j^{(2)}$) and on the right side of the values are those observed in the first month of the spell ($Z_j^{(1)}$). As discussed in Section III the second method of fixing the regressors may also produce biased coefficients because the values of the "exogenous variables" at the start of the second spell of an event will depend on the lengths of previous spells of events if the true exogenous variables are time trended. In our sample of thirty months the unemployment measure has an extreme positive time trend.

The two methods for fixing the regressors produce large differences in parameter estimates. Although most parameters are not significantly different from zero by conventional standards, the coefficient of the unemployment rate in the first spell GLS equations using $Z_{je}^{(2)}$ is three times as large as the coefficient based on $Z_{je}^{(1)}$. The sign of estimated second spell coefficient unemployment rate differs depending on the method used to fix time varying variables. The coefficients associated with marital status are statistically insignificant and are much less sensitive to the method of selecting explanatory variables.

The two regression specifications yield very different parameter estimates. For the model using $Z_{je}^{(1)}$ the estimate of the coefficient of unemployment for second spell duration is .531 in the GLS system as opposed to .858 from the first difference specification. The coefficients of the unemployment rate on first spell employment duration are 9.31 in the first difference specification and -6.523 in the GLS system. These disparities and the generally poor fits demonstrate the difficulties inherent in the regression approach.

TABLE 1.—Ln Employment Durations (Based on Two Completed Spells)[1]

	Within Spell Averages of Exogenous Variables				Start of Spell Exogenous Variables			
	Spell One	t	Spell Two	t	Spell One	t	Spell Two	t
Intercept	1.232	(2.16)	-1.052	(1.25)	7.094	(2.1)	-2.048	(3.2)
Marital Status (1=if married)	.624	(1.43)	-.449	(.72)	-.409	(.81)	-.490	(.75)
National Unemployment	-2.020	(5.40)	-.975	(.21)	-6.523	(2.5)	.531	(1.88)
			Difference Specifications					
Intercept	2502	(1.95)			12.423	(2.5)		
Marital Status	-.138	(.17)			0.112	(.01)		
Unemployment	.185	(.56)			.858	(.337)		
Marital Status (First Spell)	-.570	(.49)			-.496	(1.11)		
National Unemployment (First Spell)	1.946	(2.41)			8.951	(2.43)		
	γ_e = 1.08				γ_e = .65			

[1]t statistics are reported in parentheses.

The estimate of the Weibull parameter γ_e is substantially greater than zero, using either of the two methods for computing $Z_{j\ell}$. A value of γ_e greater than zero indicates positive duration dependence, i.e., the longer as individual is in a state the more likely he is to leave it. The same results using a different list of variables were also found by Heckman and Borjas.

Regression results for the nonemployment equations are presented in Table 2. Once again estimates based on $Z_{jn}^{(1)}$ and $Z_{jn}^{(2)}$ are very different. When within spell averages are used ($Z_{jn}^{(2)}$), the unemployment measure has a negative

TABLE 2.—Ln Nonemployment Durations (Based on Two Completed Spells)[1]

	Within Spell Averages of Exogenous Variables				Start of Spell Exogenous Variables			
	Spell One	t	Spell Two	t	Spell One	t	Spell Two	t
Intercept	-.398	(.97)	.44	(.65)	-1.051	(2.28)	-1.011	(1.61)
Marital Status (=1 if married spouse present)	-.117	(.24)	.16	(.44)	-.375	(.764)	-.093	(.25)
National Unemployment	-.335	(1.46)	-.53	(1.74)	.057	(.21)	.151	(.51)
		Difference Specifications						
Intercept		-.0746	(.084)			.950	(1.1)	
Δ Marital Status		-.342	(.645)			-.656	(1.31)	
Δ Unemployment		-.402	(.85)			.544	(1.32)	
Marital Status (In First Spell)		.091	(-.13)			.280	(.52)	
Unemployment (In First Spell)		-.252	(.60)			-.653	(1.42)	
		$\gamma_n = 1.23$				$\gamma_n = .74$		

[1] t statistics are reported in parentheses.

estimated effect on employment duration. Since the prime age male unemployment rate is time trended in our sample, using the average value of the employment rate as a regressor causes longer duration times to be associated with higher within spell unemployment rates for purely mechanical reasons. A strong empirical relationship is found not because of any causal link, but only because the unemployment measure is a transformation of the dependent variable.

We again find evidence of positive duration dependence in the unemployment equations. "γ_n" is approximately equal to one. The coefficients associated with the unemployment measures switch signs in both the top and bottom panels according to whether $Z_{jn}^{(1)}$ or $Z_{jn}^{(2)}$ is used.

We now discuss estimates obtained from maximum likelihood estimation of the parameters of the hazard function (5). In order to compare these estimates to those from the regressions, we initially adopt a Weibull specification. The coefficient associated with the log of the duration in the state is γ_{ℓ}. Later on in this section we present results from a more general parameterization of duration dependence and compare results from this model with those from the Weibull.

The maximum likelihood parameter estimates for the employment-nonemployment model are presented in Table 3. Panel A contains estimates from the model using $Z_{je}^{(2)}$, in panel B are the estimates using regressors $Z_{je}^{(1)}$, and in panel C are the parameter estimates for the case when the regressors are allowed to vary within the spells. On the left side of the page are the coefficients associated with the employment to nonemployment transition, and on the right side are the coefficients associated with the nonemployment to employment transition. First note that, in contrast to the regression results, the estimates of γ_{ℓ} indicate negative duration dependence in all states and for all methods of fixing regressors. The duration dependence parameter is highly significant in virtually all transitions. However, spurious negative duration dependence can be generated by not properly accounting for population heterogeneity. In fact, when heterogeneity is introduced in the model there is no evidence of duration dependence in the employment to nonemployment transition (see Table 4).

The coefficient of MSP is significant in all three panels in explaining employment duration, though there are large differences in the absolute values of the estimates. Allowing the MSP variable to change over the course of the

TABLE 3. -- Maximum Likelihood Estimates -- Weibull Model[1]

Panel A: Regressors Fixed at Average Value Over Spell

	Employment to Nonemployment	Nonemployment to Employment
Intercept	.971 (1.535)	-.093 (.221)
ℓn Duration (γ)	-.137 (1.571)	-.287 (2.976)
MSP	-1.093 (2.679)	.347 (1.134)
Unemployment	-1.800 (6.286)	-.577 (3.119)

£[2] = -711.457

Panel B: Regressors Fixed at Value for First Month of Spell

	Employment to Nonemployment	Nonemployment to Employment
Intercept	-3.743 (12.074)	-1.054 (3.464)
ℓn Duration (γ)	-.230 (2.888)	-.363 (4.049)
MSP	-.921 (2.310)	.297 (.902)
Unemployment	.569 (3.951)	-.130 (.900)

£ = -740.998

Panel C: Regressors Free to Vary Over the Spell

	Employment to Nonemployment	Nonemployment to Employment
Intercept	-3.078 (8.670)	-.899 (2.742)
ℓn Duration (γ)	-.341 (3.941)	-.316 (3.279)
MSP	-.610 (1.971)	.362 (1.131)
Unemployment	.209 (1.194)	-.204 (1.321)

£ = -746.515

[1] Absolute value of asymptotic normal statistics in parentheses.

[2] £ denotes the value of the log likelihood function.

spell results in a parameter estimate almost one half as large in absolute value as the coefficient value that results from using the average spell value in the employment to nonemployment transition. In the nonemployment-employment equations the method of introducing exogenous variables has little effect on the value of the MSP parameter, though in all cases the coefficient is approximately equal to its standard error. In all three panels the interpretation of the effect of marital status on exit times is the same: individuals married with spouse present have lower rates of transition out of employment and higher rates of transition from nonemployment than others.

The differences in parameter estimates are much more extreme for the coefficients of the prime age male unemployment variable. Recall that this variable is strongly time trended. The coefficients of the unemployment variable in both panels A and B are highly significant for the employment-nonemployment transition, but are of opposite sign. Allowing the variables to change over the spell results in an estimate which is positive but not statistically significant. Use of spell constant regressors leads to dramatically different interpretations depending on the technique employed to arrive at a spell constant value. This same remark is true with regard to the unemployment rate coefficients in the nonemployment-employment transition. Taken as a whole, these results demonstrate the seriousness of the biases introduced into the parameter estimates by restricting the variability of exogenous variables. These findings cast considerable doubt on the value of regression methods for estimating duration models.

We next turn to an investigation of the effects of not controlling for heterogeneity on estimates of duration dependence. The specification of the one factor scheme we adopt is given in equation (6) in Section I. The same set of regressors is used as was used in the estimation of the model in Table 3, Panel C, so that all variables are allowed to vary within spells. Table 4 presents the estimates.

TABLE 4. — Maximum Likelihood Estimates with Time Varying Variables and Heterogeneity[1]

	Employment to Nonemployment	Nonemployment to Employment
Intercept	-3.600	-.879
	(8.395)	(2.525)
ℓn Duration	.015	-.312
	(.121)	(3.170)
MSP	-.498	.320
	(1.384)	(.961)
Unemployment	-.017	-.172
	(.101)	(1.056)
C_{ij}	1.196	-.133
	(4.651)	(.756)

$\mathcal{L} = -740.126$

[1]Absolute value of asymptotic normal statistics in parentheses.

It is interesting to note that unobservable heterogeneity is an important determinant of the rate of transition out of the employment state, but has little effect on the rate of leaving nonemployment. There has been much attention given to separating the effect of heterogeneity and state dependence in the length of nonemployment spells, but our estimates suggest that for young men estimates of duration dependence are not affected by the inclusion of heterogeneity into the model. However, the duration dependence effect vanishes with the introduction of heterogeneity in the employment-nonemployment transition. The introduction of heterogeneity reduces the magnitude of all the parameters (with the exception of the constant) in the employment equation, while the coefficients in the nonemployment equation are barely affected.

These findings indicate that employment-nonemployment transition probabilities are nonMarkovian and call into question the standard discrete time Markov assumption widely used in labor market analysis (see, e.g., Marston, 1976). The source of departure from the Markov model differs between the employment to nonemployment transition and the nonemployment to employment transition. The former transition is non-Markovian because of uncontrolled heterogeneity which gives rise to the classical mover-stayer problem. The latter transition is non-Markovian because of structural duration dependence.

All the estimation done up to this point has been predicated on the assumption that the exit time distributions are Weibull. We now relax this assumption. We estimate a model which allows for general forms of duration dependence. This specification is obtained by letting $\lambda_{1j\ell} = 1$ and $\lambda_{2j\ell} = 2$ for all values of j and ℓ (see equation 5). Empirical results are presented in Table 5 for models estimated with and without heterogeneity corrections. The models estimated without heterogeneity show some evidence of linear duration dependence in the E to N transition, while the hazard associated with the N to E transition appears to be nonmonotonic in duration. Except for the unemployment rate coefficient in the E to N transition, the estimates of the coefficients of the explanatory variables do not differ much from the estimates obtained in the Weibull model. When the heterogeneity correction is added, there is no evidence of duration dependence in the rate of leaving employment. Recall that a similar result was found in the Weibull case. For the nonemployment to employment transition, the exit rate to employment also appears to be a linear function of duration. In our data we do not reject the null hypothesis that the squared duration term is insignificant in each of the transition densities. Assuming that this is so, it then become possible to test between the Weibull ($\lambda_{1j\ell} = 0$) and Gompertz ($\lambda_{1j\ell} = 1$) specifications. The difference in log likelihoods between the two models is negligible (-739.2 vs -740.6). The evidence suggests a slight preference for the Weibull specification. The estimates of the coefficients of the explanatory variables do not change much between those specifications, and for the sake of brevity we do not report the detailed empirical results from this model.

TABLE 5. -- Maximum Likelihood Estimates with Time Varying Variables, Heterogeneity, and General Duration Dependence[1]

	Without Heterogeneity		With Heterogeneity	
	E → N	N → E	E → N	N → E
Const.	-3.271 (8.901)	-.762 (2.425)	-3.565 (8.537)	-.748 (2.247)
Tenure/10	-.806 (1.858)	-1.714 (2.731)	.045 (.085)	-1.704 (2.685)
$Tenure^2/100$	.028 (.145)	.602 (1.673)	-.120 (.607)	.603 (1.666)
MSP	-.568 (1.731)	.349 (1.098)	-.490 (1.353)	.313 (.956)
Unemployment	.329 (1.865)	-.192 (1.271)	.075 (.392)	-.164 (1.042)
C_{ij}			1.008 (3.572)	-.118 (.650)
£	-742.334		-739.177	

[1]Absolute value of asymptotic normal statistics in parentheses.

B. Tests of a Three State vs. a Two State Model and Tests of Whether or not Unemployment and Out of the Labor Force are Behaviorally Different States[1]

The preceding empirical work lumps the "unemployed" and those "out of the labor force" into a common "nonemployed" category. There is considerable controversy in the literature over the issue of whether or not the categories "unemployed" and "out of the labor force" are behaviorally distinct labor force states (Lucas). This issue is particularly relevant in the study of the labor market dynamics of youth. Given the range of nonmarket options available to many youths, and given practices of many state unemployment compensation agencies which effectively limit the eligibility for unemployment compensation of many youths, it seems especially likely that there is no distinction between "unemployment" and "out of the labor force" status for young people. Recent

[1]Comments by Gary Chamberlain have greatly improved this section of the paper, and eliminated several errors that appeared in previous drafts.

papers by Clark and Summers and Ellwood have made this claim. In this section of the paper we present a test of this proposition and reject it. We find that distinct behavioral equations govern transitions from out of the labor force to employment and from unemployment to employment.

Recent theory suggests that being "unemployed" and "out of the labor force" describes different behavior. For example, in search theory (e.g., Burdett and Mortensen) a key difference between unemployed individuals and those out of the labor force is that the former are at an interior point with respect to the optimal amount of time they devote to search while the latter are at a corner and spend no time searching. Separate behavioral equations generate observations in these two states.

Even though these theoretical distinctions are widely accepted, many economists claim that the empirical distinction between reported "unemployment" and reported "out of the labor force" is so arbitrary that it is of little or no analytical value. This point would seem to have some merit after examining the official current population survey definition of unemployment, which defines those individuals as unemployed "who had no employment during the survey week, were available for work, and (1) had engaged in any specific job seeking activity within the past four weeks, (2) were waiting to be called back to a job from which they had been laid off, or (3) were waiting to report to a new wage or salary job scheduled to start within the following 30 days." Because there is no stipulation as to the quality or quantity of searches made within the month, the unemployment-out of the labor force distinction may be of little value in predicting employment probabilities for the nonemployed.

In this section we present a test to determine whether or not the classifications "u" (unemployed) and "o" (out of the labor force) are behaviorally meaningless distinctions. The idea underlying the test is as follows: controlling for heterogeneity if the hazard rate for exit to employment from unemployment (h_{ue}) is the same as the hazard rate for exit to employment out of the labor force (h_{oe}), the origin state ("o" or "u") is irrelevant in determining the rate at which individuals leave nonemployment to go to employment. In a simple 3 state Markov model, this test is equivalent to testing the proposition that the two nonemployment states can be aggregated into a single state and a properly specified two state Markov model can be defined for employment and nonemployment. To simplify the exposition we assume that there is no heterogeneity in observed or unobserved characteristics. This assumption is not essential and is not used in performing the empirical work reported below.

To motivate the test, we consider two cases. The first case assumes that individuals exit employment at a rate governed by density $f_e(t_e)$. The probability that a person terminating employment classifies himself as a "u" or "o" is determined by tossing a coin that comes up "u" fraction π of the time and "o" fraction $1-\pi$ of the time. Once acquired the person keeps these labels as long as he is nonemployed so there is no switching between "o" and "u" states (a patently counterfactual case). The density of duration in the nonemployment state is governed by density $f_n(t_n)$. The associated hazard h_n is

$$h_n = \frac{f_n(t_n)}{1-F_n(t_n)} .$$

The joint probability density that an individual is classified as unemployed and leaves nonemployment at t_n is

$$\pi f_n(t_n)$$

with associated hazard

$$h_{ue} = h_n .$$

The joint probability that an individual is classified as out of the labor force and leaves nonemployment at t_n is

$$(1-\pi)f_n(t_n)$$

with associated hazard

$$h_{oe} = h_n .$$

The hazard rate for entry to employment will be the same whether or not the nonemployed individual is classified as an "o" or an "u".[1]

In the second case considered here, individuals are allowed to switch their reported nonemployment status "randomly". By this we mean that initial nonemployment classification is random (governed as before by a toss of the coin) and that individuals switch randomly between "o" and "u". The continuous time analogue of discrete time independent Bernoulli trials is an exponential waiting time model (Cox, 1962). Write the hazard for durations from "o" to "u" as h_{ou} and the hazard from "u" to "o" as h_{uo}, the density of time spent going from o to e (t_{oe}) is

$$h_{oe} \exp - \left\{ (h_{oe} + h_{ou}) t_{oe} \right\}$$

while the density of time spent going from u to e (t_{ue}) is

$$h_{ue} \exp - \left\{ (h_{ue} + h_{uo}) t_{ue} \right\} .$$

Individuals may change among reported nonemployment states for whatever reason. All that is required for the origin state (o or u) to be irrelevant for characterizing transitions from nonemployment to employment is for $h_{oe} = h_{ue}$.

---

[1] The proof is trivial. Assume $f_n(t_n)$ is not defective so that $\int_0^\infty f_n(t_n)dt_n = 1$. The hazard rate for exit from unemployment to employment is

$$h_{ue} = \frac{\pi f_n(t_n)}{\pi - \pi F_n(t_n)} = h_n .$$

The term in the denominator is the probability that the exit occurs from u to e after time t_n. A parallel argument demonstrates that $h_{oe} = h_n$.

The condition $h_{oe} = h_{ue}$ is also the requirement that must be satisfied in a Markov model to aggregate "o" and "u" into a single state n, and for the resulting two state model for e and n to be a properly defined Markov model. To demonstrate this it is most convenient to work with the state probability representation of the three state Markov model. Define $P_j(t)$ as the probability that state j is occupied at time t and $\dot{P}_j(t)$ as the instantaneous rate of change of this probability. The three state generalization of the two state model presented in Section I.b is

$$\begin{bmatrix} \dot{P}_e(t) \\ \dot{P}_o(t) \\ \dot{P}_u(t) \end{bmatrix} = \begin{bmatrix} -(h_{eu} + h_{eo}) & h_{oe} & h_{ue} \\ h_{eo} & -(h_{oe} + h_{ou}) & h_{uo} \\ h_{eu} & h_{ou} & -(h_{ue} + h_{uo}) \end{bmatrix} \begin{bmatrix} P_e(t) \\ P_o(t) \\ P_u(t) \end{bmatrix}$$

or

$$\dot{P}^{(3)}(t) = AP^{(3)}(t)$$

using matrix notation. Note that the rank of A is at most 2.

In order to aggregate "o" and "u" into a two state model defined in terms of n, we require that we be able to collapse the three state system into

$$\begin{bmatrix} \dot{P}_e(t) \\ \dot{P}_n(t) \end{bmatrix} = \begin{bmatrix} -h_{en} & h_{ne} \\ h_{en} & -h_{ne} \end{bmatrix} \begin{bmatrix} P_e(t) \\ P_n(t) \end{bmatrix}$$

where $P_n(t) = P_o(t) + P_u(t)$. In matrix notation $\dot{P}^{(2)}(t) = B\,P^{(2)}(t)$. The rank of B is 1. For this to be an equivalent representation of the three state model, a necessary condition is that rank (A) = rank (B) = 1. A necessary and sufficient condition is that $h_{oe} = h_{ue} = h_{ne}$. Sufficiency may be checked by direct substitution into A.

This interpretation of the test is also informative in that it makes precise the sense in which o and u are "irrelevant". Aggregating o and u into a single state for the purpose of statistical analysis does not alter the Markov property of the model. The rate at which individuals leave nonemployment to enter employment does not depend on which nonemployment state individuals are in.

It is tempting to extend this type of reasoning to consider transitions from employment to the two nonemployment states. Thus it might be argued that if u and o are "irrelevant" distinctions, the rate of transition from e to u (h_{eu}) whould be the same as the rate of transition from e to o (h_{eo}). This argument is correct only if $\pi = 1 - \pi = 1/2$. If $f_e(t_e)$ is the density of employment length durations with hazard rate $h_e(t_e)$ the hazard rate for transitions from e to u is

$$h_{eu} = \pi h_e$$

while the hazard rate for transitions from e to o is

$$h_{eo} = (1 - \pi) h_e .$$

Obviously $h_{eu} + h_{eo} = h_e$, as is required by the law of conditional probability. But unless $\pi = 1 - \pi = 1/2$, $h_{eu} \neq h_{eo}$. We have no theory of π. Even reporting oneself as unemployed is strictly a matter of tossing a coin, nothing requires $\pi = 1/2$.

Table 6 presents estimates of the three state model estimated with heterogeneity. The fact that the standard errors are so large relative to the magnitude of the parameters is to be expected given that we are attempting to estimate twenty parameters with so few degrees of freedom. The parameter signs are generally consistent with our earlier results from the employment-nonemployment model. Only the constant terms and the factor loading of the employment to unemployment transition are greater than twice their standard errors.

TABLE 6 -- Parameter Estimates from the Three State Unrestricted Model

	From Employment to:		To Employment from:	
	Unemployment	OLF	Unemployment	OLF
Constant	-3.822 (9.778)[1]	-7.193 (2.768)	-.698 (3.782)	-2.384 (2.078)
Tenure/10	.482 (.846)	.700 (.379)	-1.253 (1.530)	1.441 (.365)
$Tenure^2/100$	-.240 (1.004)	-.019 (.030)	.481 (.547)	.208 (.084)
MSP	-.355 (.837)	.086 (.068)	-.065 (.193)	1.154 (.400)
C_{ij}	1.396 (3.336)	2.788 (1.025)	-.342 (1.633)	-1.866 (1.081)

$\mathcal{L} = -784.33$

[1]Absolute value of asymptotic normal statistics in parentheses.

The estimates from the restricted three state model are given in Table 7. Let $\theta_{ij} \equiv (\beta_{ij} \;\; c_{ij})$. The restrictions imposed are $\theta_{oe} = \theta_{ue}$, which forces all parameters in the "unemployment" to employment and "out of the labor force" to employment transitions to equality. There are a total of five restrictions. Performing the likelihood ration test on the restricted versus the unrestricted model, the value of the test statistic is 28.72 which is distributed $\chi^2(5)$. The critical value for a 5 percent significance level is

11.07. We are able to reject the null hypothesis of the equality of the parameters governing the two nonemployment states. These empirical results suggest that "out of the labor force" and "unemployment" are not artificial distinctions for this sample of young men.

TABLE 7 — Parameter Estimates from the Three State Restricted Model

	From Employment to: Unemployment	From Employment to: OLF	Nonemployment to Employment
Constant	-3.735 (9.934)[1]	-7.718 (2.596)	-.857 (4.756)
Tenure/10	.400 (.706)	.782 (.528)	-1.460 (1.790)
$Tenure^2/100$	-.220 (.940)	-.004 (.007)	.683 (1.116)
MSP	-.397 (.966)	.160 (.148)	.202 (.577)
C_{ij}	1.327 (4.195)	3.102 (1.078)	-.421 (1.894)

$\mathcal{L}$ = -798.69

[1]Absolute value of asymptotic normal statistics in parentheses.

SUMMARY AND CONCLUSIONS

This paper presents new econometric methods for the empirical analysis of individual labor market histories. The techniques developed here extend previous work on continuous time models in four ways: (1) A structural economic interpretation of these models is presented. (2) Time varying explanatory variables are introduced into the analysis in a general way. (3) Unobserved heterogeneity components are permitted to be correlated across spells. (4) A flexible model of duration dependence is presented that accommodates many previous models as a special case and that permits tests among competing specifications within a unified framework. In addition, longer range types of state dependence can be introduced into the model and their empirical importance tested with our model.

We contrast our methods with more conventional discrete time and regression procedures. The parameters of continuous time models are invariant to the sampling time unit used to record observations. Parameters of discrete time models defined for one time unit are not in general comparable to parameters of discrete time models defined for other time units. Two problems plague the regression approach to analyzing duration data which do not plague the likelihood approach advocated in this paper. The first problem is that standard regression estimators are ill equipped to deal with censored spells of events that arise in short panels. The second problem is that the regression approach cannot be readily adopted to accommodate time varying explanatory variables. The functional forms of regression functions depend on the time paths of the explanatory variables. *Ad hoc* solutions to this problem can make exogenous variables endogenous to the model and so can induce simultaneous equations bias. The likelihood approach advocated in this paper can readily accommodate time varying explanatory variables.

Two sets of empirical results are presented. The first set is an analysis of employment and nonemployment data using both regression and maximum likelihood procedures. Standard regression methods are shown to perform rather poorly, and to produce estimates wildly at variance with the estimates from our maximum likelihood procedure. The maximum likelihood estimates are more in accord with *a priori* theoretical notions. A major conclusion of this analysis is that the discrete time Markov model widely used in labor market analysis is inconsistent with the data.

The second set of empirical results is a test of the hypothesis that "unemployment" and "out of the labor force" are behaviorally different labor market states. Contrary to recent claims, we find that they are separate states for our sample of young men.

REFERENCES

Anderson, T. W. "Some Scaling Models and Estimation Procedures in the Latent Class Model." In Probability and Statistics. Edited by U. Grenander. Uppsala, 1959.

Barlow, R. E. and Proschan F. Mathematical Theory of Reliability. New York: John Wiley and Sons, 1965.

Becker, G. "Imperfect Information, Marriage and Divorce." In "A Treatise on the Family." Chapter 12. Unpublished manuscript. University of Chicago, 1980.

Becker G. and Stigler, G. "Law Enforcement, Malfeasance and Compensation of Enforcers." Journal of Legal Studies 3, no. 1 (1974): 1-18.

Box, G. and Cox, D. R. "The Analysis of Transformations." Journal of the Royal Statistical Society, Series B 26 (1964): 211-52.

Brock, W., Rothschild, M. and Stiglitz, J. "Notes on Stochastic Capital Theory." University of Wisconsin, 1979.

Buck, G. Advanced Calculus. 2d ed. New York: McMillan, 1968.

Burdett, K. and Mortensen, D. "Labor Supply Under Uncertainty." Discussion paper 297, Northwestern University, August, 1977.

Chamberlain, G. "On the Use of Panel Data." In Heckman, J. and B. Singer, The Analysis of Longitudinal Labor Market Data. Academic Press, 1981.

Clark, K. and Summers, L. "The Dynamics of Youth Unemployment." N.B.E.R. Working Paper No. 274, 1978.

Coleman, J. S. Introduction to Mathematical Sociology. Glencoe, Ill.: Free Press of Glencoe, 1964.

Cox, D. R. Renewal Theory. London: Butler and Tanner, 1962.

Cox, D. R. and Lewis, P. A. W. The Statistical Analysis of Series of Events. London: Metheun, 1966.

Danforth, J. "On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search." In Studies in the Economics of Search. Edited by S. Lippman and J. McCall. Amsterdam, North Holland: 1979, pp. 109-32.

Ellwood, D. "Teenage Unemployment: Temporary Scar or Permanent Blemish," N.B.E.R. Working Paper No. 399, October, 1979.

Gronau, R. "Information and Frictional Unemployment." American Economic Review 60 (1971): 290-301.

Hall, J., Lippman, S., and McCall, J. "Expected Utility Maximizing Job Search." In Studies in the Economics of Search. Edited by S. Lippman and J. McCall. Amsterdam, North Holland: 1979. Pp. 133-56.

Heckman, J. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating Discrete Time--Discrete Data Stochastic Process and Some Monte Carlo Evidence." In Structural Analysis of Discrete Data with Econometric Applications. Edited by C. Manski and D. McFadden. Cambridge: MIT Press, 1980.

Heckman, J. and Borjas, G. "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." Forthcoming in a special issue on the economics of unemployment, Economica, August, 1980.

Heckman, J. and Willis, R. "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In Household Production and Consumption. Edited by N. Terleckyj. Stanford, Calif.: National Bureau of Economic Research, 1975.

Holt, C. "Job Search, Phillips Wage Resolutions and Union Influence: Theory and Evidence." In Microeconomic Foundations. Edited by E. Phelps et al. New York: Norton, 1970. Pp. 53-123.

Jovanovic, B. "Job Matching and the Theory of Turnover." Journal of Political Economy 87, no. 5 (October 1979): 972-90.

Kiefer, N. and Neumann, G. "An Empirical Job Search Model with a Test of the Constant Resolution Wage Hypothesis." Journal of Political Economy 87, no. 1 (February 1979): 89-108.

Lippman, S. and McCall, J. J. "Search Unemployment: Mismatches, Layoffs and Unemployment Insurance." Western Management Science Institute, UCLA, September, 1979.

________. "The Economics of Job Search: A Survey." Part I. Economic Inquiry 14 (1976): 155-89.

Lucas, Robert E., Jr. "Unemployment Policy." American Economic Review 68 (1978): 353-57.

Marston, S. "Employment Instability and High Unemployment Rates." Brookings Papers on Economic Activity 7 (1976): 169-203.

Moran, P. A. P. "Maximum Likelihood Estimation in Non Standard Conditions." Proceedings Cambridge Philosophical Society 70 (1971): 441-50.

Mortensen, D. "Job Search, The Duration of Unemployment and the Phillips Curve." American Economic Review 60 (1970): 847-62.

Phillips, P. C. B. "The Problem of Identification in Finite Parameter Continuous Time Models." In Statistical Inference in Continuous Time Economic Models. Edited by A. R. Bergstrom. Amsterdam, North Holland: 1976. Pp. 123-34.

Salant, S. "Search Theory and Duration Data," The Quarterly Journal of Economics 91, No. 1: 39-57.

Salop, S. "Systematic Job Search and Unemployment." Review of Economic Studies 40, no. 1 (1973): 191-201.

Salop, S. and Salop, J. "Self Selection and Turnover in the Labor Market." Quarterly Journal of Economics 90, no. 4 (November 1976): 629-49.

Singer, B. and Spilerman, S. "Some Methodological Issues in the Analysis of Longitudinal Surveys." Annals of Economic and Social Measurement 4 (Fall 1976): 447-74.

Toikka, R. S. "A Markovian Model of Labor Market Decisions by Workers. American Economic Review 66 (December 1976): 821-34.

Tuma, N., Hannan, M., and Groeneveld, L. "Dynamic Analysis of Event Histories." American Journal of Sociology 84, no. 4 (January 1979): 820-54.

Tuma, N. and Hannan, M. "Approaches to the Censoring Problem in Analysis of Event Histories." In Sociological Methodology. Edited by K. Schuessler. San Francisco: Jossey Bass, 1979. Pp. 209-40.

APPENDIX A

The Likelihood Function for a Multistate-Multiepisode model.

In this appendix we discuss general issues in the estimation of continuous time probability models and the particular features of the likelihood function we employ in the empirical work presented in the text. Since our data start at the beginning of the labor market history of individuals, we can safely ignore the initial conditions problem discussed in the text. We defer a general discussion of this problem to a later paper.

Let there be N states the individual can occupy at any moment of time. If the individual begins "life" in state i there are N-1 "latent times" with densities

$$\text{(A-1)} \qquad f_{ij}(t_{ij}) = h_{ij}(t_{ij}) \exp - \{\int_0^{t_{ij}} h_{ij}(u)du\}$$

$$(j = 1,\ldots,N, j \neq i)$$

where $f_{ij}(.)$ is the density function of exit times from state i into state j, and $h_{ij}(.)$ is the associated hazard function. The joint density of the N-1 latent exit times is given by

$$\text{(A-2)} \qquad \prod_{\substack{j=1 \\ j\neq i}}^{N} h_{ij}(t_{ij}) \exp - \{\int_0^{t_{ij}} h_{ij}(u)du\} \cdot$$

An individual exits from state i to state j′ if the j′-th first passage time is the smallest of the N-1 potential first passage times i.e. if

$$t_{ij'} < t_{ij}\ , \{j = 1,\ldots,N; j \neq j';\ j,j' \neq i\}.$$

Let the probability that the individual leaves state i and enters state j′ be denoted $p_{ij'}$. Then

$$\text{(A-3)}\quad P_{ij'} = \int_0^\infty \left[\int_{t_{ij'}}^\infty \cdots \int_{t_{ij'}}^\infty \left[\prod_{\substack{j=1 \\ j\neq i \\ j\neq j'}}^{N} h_{ij}(t_{ij}) \exp - \{\int_0^{t_{ij}} h_{ij}(u)du\} \, dt_{ij} \right] \right.$$

$$\left. \times \ (h_{ij'}(t_{ij'}) \exp - \{\int_0^{t_{ij'}} h_{ij'}(u)du\}) \right] dt_{ij'}$$

$$= \int_0^\infty h_{ij'}(t_{ij'}) \exp - \{ \int_0^{t_{ij'}} \left(\sum_{\substack{k=1 \\ k\neq i}}^{N} h_{ik}(u) \right) du\} \, dt_{ij'} .$$

The conditional density of exit times from state i into state j$'$ given that $t_{ij'} < t_{ij}$, $(\forall j;\ j\neq j';\ j,j'\neq i)$ is

$$\text{(A-4)}\quad g(t_{ij'} \mid t_{ij'} < t_{ij})(\forall j; j\neq j'; j,j'\neq i) = \frac{h_{ij'}(t_{ij'}) \exp - \{\int_0^{t_{ij'}} \left(\sum_{\substack{k=1 \\ k\neq i}}^{N} h_{ik}(u) \right) du\}}{P_{ij'}} .$$

It follows that the density of exit times from state i into any other state can be written

$$\text{(A-5)}\quad f_{i.}(t_{i.}) = \sum_{\substack{j'=1 \\ j'\neq i}}^{N} P_{ij'} \, g(t_{ij'} \mid t_{ij'} < t_{ij})(\forall j; j\neq j'; j,j'\neq i)$$

$$= \left(\sum_{\substack{k=1 \\ k \neq i}}^{N} h_{ik}(t_{i.}) \right) \exp - \{ \int_{o}^{t_{i.}} \left(\sum_{\substack{k=1 \\ k \neq i}}^{N} h_{ik}(u) \right) du \} \cdot$$

The probability that the spell is uncompleted by time T is simply

$$\text{(A-6)} \qquad \text{Prob } (t_{i.} > T) = \int_{T}^{\infty} f_i(t)\, dt$$

$$= \exp - \{ \int_{o}^{T} \left(\sum_{\substack{k=1 \\ k \neq i}}^{N} h_{ik}(u) \right) du \} \quad .$$

This term enters the likelihood function for spells uncompleted as of the end of sample period T. In this manner all spells, not only completed ones, are used in the estimation of the parameters of the hazard function. This is not the case in regression analyses of durations in a state (or some transformation of duration) on exogenous variables, where only completed spells can be used in a straightforward fashion.

We now describe in some detail the specific form of the likelihood function used in the analysis performed here. Let $Z_{rm}(u + \tau_{rm})$ be a $K \times 1$ vector of explanatory variables of the r^{th} individual in his m^{th} spell at time $(u + \tau_{rm})$, where u is the duration of time spent in the current spell, and τ_{rm} is the date in calendar time at which the individual began his m^{th} spell. Included among the explanatory variables are functions of the spell duration variables. In particular, the form of $Z'_{rm}(u + \tau_{rm})$ is

$$[1 \quad Z_{1rm}(u + \tau_{rm}) \ \cdots \ Z_{(K-2)rm}(u + \tau_{rm}) \ V_r]$$

The last element of Z is an unobserved heterogeneity term, invariant over time for the individual, which is assumed to have a standard normal distribution, i.e.

$$V_r \sim N(0, 1) \quad \forall r \quad .$$

Parameter vectors are indexed by transition. "β_{ij}" is $K \times 1$ vector of coefficients of explanatory variables in the hazard function.

To be specific

$$\beta_{ij} = [\beta_{oij} \ \beta_{1ij} \ \cdots \ \beta_{(K-2)ij} \ C_{ij}]$$

As discussed in the text (eq. 6), we impose a one factor specification, so that C_{ij} is the factor loading associated with the i to j transition.

Now we write the hazard function for the m^{th} spell and r^{th} individual as

$$h_{i_m j_m}(t_{r_m}) = \exp\left(Z'_{r_m}(t_{r_m} + \tau_{r_m})\,\beta_{i_m j_m}\right)$$

$$\text{(A-8)} \qquad = \frac{f_{i_m j_m}(t_{r_m})}{1 - F_{i_m j_m}(t_{r_m})}$$

Then the cumulative density for i to j exit times is

$$\text{(A-9)} \qquad F_{i_m j_m}(t_{r_m}) = 1 - \exp\left\{ - \int_0^{t_{r_m}} \left[h_{i_m j_m}(x)\right] dx\right\}$$

and the probability density is

$$\text{(A-10)} \qquad f_{i_m j_m}(t_{r_m}) = h_{i_m j_m}(t_{r_m}) \exp\left(- \int_0^{t_{r_m}} \left[h_{i_m j_m}(x)\right] dx\right) .$$

Now consider one individual's contribution to the likelihood function (we henceforth suppress the individual subscript in the density and distribution function)

$$\text{(A-11)} \qquad L_r(\beta, V) = \left[\prod_{m=1}^{M-1} f_{i_m j_m}(t_m)\right] \left[\prod_{\substack{k=1 \\ k \neq i_M}}^{N} \left(1 - F_{i_M k}(t_M)\right)\right]$$

where t_M is the time spent in the M^{th} spell, which is censored. Because heterogeneity is modeled as a random effect, it is necessary to obtain the expected value of L_r with respect to the assumed distribution of V. Denote the conditional likelihood by $\bar{L}_r(\beta)$, where

(A-12)

$$\bar{L}_r(\beta) = \int_{-\infty}^{\infty} L_r(\beta, V) \exp(-V^2/2) \frac{dV}{\sqrt{2\pi}} .$$

Now define

$$\pounds_r(\beta) \equiv \ln(\bar{L}_r(\beta)) .$$

In taking first partials note that

$$\text{(A-13)} \qquad \frac{\partial \pounds_r(\beta)}{\partial \beta_{gij}} = \frac{1}{\bar{L}_r(\beta)} \frac{\partial \bar{L}_r(\beta)}{\partial \beta_{gij}}$$

$$= \frac{1}{\bar{L}_r(\beta)} \int_{-\infty}^{\infty} \frac{\partial L_r(\beta, V)}{\partial \beta_{gij}} e^{-V^2/2} \frac{dV}{\sqrt{2\pi}}$$

where g denotes an element of the parameter vector.

For notational convenience define the set

$$S_{ij} = \{ m \mid i_m = i, j_m = j \}$$

which consists of all spells that begin in state i and end in state j for the individual. Then

$$\text{(A-14)} \qquad \frac{\partial L_r(\beta, V)}{\partial \beta_{gij}} = \left\{ \sum_{\substack{m \in S_{ij} \\ m < M}} \left[\left. \frac{\partial f_{i_m j_m}}{\partial \beta_{gij}} \right|_{t_m} \cdot \prod_{\substack{b \neq m \\ b < M}} f_{i_b j_b}(t_b) \right] \right\} .$$

$$\cdot \prod_{\substack{h \neq i_M \\ h=1}} \left((1 - F_{i_m h} \quad (t_M) \right)$$

$$+ \left[\prod_{m=1}^{M-1} f_{i_m j_m} (t_m) \right] \left[- \frac{\partial F_{i_M j}}{\partial \beta_{gij}} \Bigg|_{t_M} \right] \left[\prod_{\substack{h \neq i_m, j \\ h=1}} (1 - F_{i_M h}(t_M)) \right] .$$

Now we evaluate the partial derivative of the density with respect to the parameters,

$$\text{(A-15)} \quad \frac{\partial f_{i_m j_m}}{\partial \beta_{gij}} \Bigg|_{t_m} = \left[\frac{\partial h_{i_m j_m}(t_m)}{\partial \beta_{gij}} - h_{i_m j_m} (t_m) \int_0^{t_m} \frac{\partial h_{i_m j_m}(u)}{\partial \beta_{gij}} du \right]$$

$$\cdot \exp \left(- \int_0^{t_m} h_{i_m j_m} (u) \, du \right)$$

But note that

$$\frac{\partial h_{i_m j_m}(u)}{\partial \beta_{gij}} = Z_g(u+\tau_m) h_{i_m j_m}(u) \quad .$$

After making the appropriate substitutions, we have

$$\text{(A-16)} \quad \frac{\partial f_{i_m j_m}}{\partial \beta_{gij}} \Bigg|_{t_m} = \left[Z_g(t_m + \tau_m) - \int_0^{t_m} Z_g(u+\tau_m) \, h_{i_m j_m}(u) \, du \right] f_{i_m j_m}(t_m)$$

and

(A-17)

$$\left.\frac{\partial F_{i_M j}}{\partial \beta_{gij}}\right|_{t_m} = \int_0^{t_M} \frac{\partial h_{ij}(u)}{\partial \beta_{gij}} du \quad . \quad \exp\{-\int_0^{t_M} h_{ij}(u)\, du\} \quad \text{for } i = i_m$$

$$= \int_0^{t_M} Z_g(u+\tau_M)\, h_{ij}(u)\, du\, (1 - F_{ij}(t_M)) \qquad \text{for } i = i_m.$$

Upon making a last round of substitutions, we get that

(A-18)

$$\frac{\partial L_r(\beta,V)}{\partial \beta_{gij}} = \{ \sum_{\substack{m \in S_{ij} \\ m<M}} \left[Z_g(t_m+\tau_m) - \int_0^{t_m} Z_g(u+\tau_m)\, h_{i_m j_m}(u)\, du \right] \}$$

$$. \ L_r(\beta,V) - \left[\int_0^{t_m} Z_g(u+\tau_m)\, h_{ij}(u)\, du \right] L_r(\beta,V)$$

$$= L_r(\beta,V) \left\{ \sum_{\substack{m \in S_{ij} \\ m<M}} \left[Z_g(t_m+\tau_m) - \int_0^{t_m} Z_g(u+\tau_m)\, h_{i_m j_m}(u)\, du \right] \right.$$

$$\left. - \int_0^{t_M} Z_g(u+\tau_m)\, h_{i_m j}(u)\, du \right\}$$

$$= A_{gij}\, L_r(\beta,V)$$

where A_{gij} is defined to be the expression in braces in the second to last line above.

(A-19)
$$\frac{\partial^2 \ell_r(\beta)}{\partial\beta_{gij}\partial\beta_{g'i'j'}} = -\frac{1}{[\bar{L}_r(\beta)]^2}\frac{\partial\bar{L}_r(\beta)}{\partial\beta_{gij}}\frac{\partial\bar{L}_r(\beta)}{\partial\beta_{g'i'j'}}$$

$$+\frac{1}{\bar{L}_r(\beta)}\frac{\partial^2\bar{L}_r(\beta)}{\partial\beta_{gij}\partial\beta_{g'i'j'}}$$

(A-20)
$$\frac{\partial^2\bar{L}_r(\beta)}{\partial\beta_{gij}\partial\beta_{g'i'j'}} = \int_{-\infty}^{\infty}\frac{\partial^2 L_r(\beta,V)}{\partial\beta_{gij}\partial\beta_{g'i'j'}}\, e^{-V^2/2}\frac{dV}{\sqrt{2\pi}} .$$

If $i \neq i'$ or $j \neq j'$ then this last term is

(A-21)
$$\int_{-\infty}^{\infty} A_{g'i'j'}A_{gij}\, L_r(\beta,V)\, e^{-V^2/2}\frac{dV}{\sqrt{2\pi}}$$

and if $i = i'$ and $j = j'$ then it is

$$\left\{ A_{g'i'j'}A_{gij} - \sum_{\substack{m\epsilon S_{ij}\\ m<M}}\left[\int_0^{t_m} Z_g(u+\tau_m)Z_{g'}(u+\tau_m)h_{i_mj_m}(u)\, du\right]\right. .$$

$$\left. -\int_0^{t_M} Z_g(u+\tau_M)Z_{g'}(u+\tau_M)\, h_{i_Mj}(u)\, du\right\} L_r(\beta,V)$$

where the last term in the braces enters only if $i_M = i = i'$.

To derive the sample likelihood, simply sum the individual contributions, i.e.,

$$\mathcal{L}(\beta) = \sum_{r=1}^{R} \mathcal{L}_r(\beta)$$

where R is the number of individuals in the sample. First and second partials are similarly computed for the sample by the summation of each individual's contribution. In computing the matrix of second partials we can employ the well known approximation based on the summed outer produce of the vector of first partials for each observation based on a suggestion of T. W. Anderson (1959) or the exact second partials presented above.

APPENDIX B

The Weibull Regression Model

This appendix presents a brief derivation and discussion of the Weibull regression model utilized by Heckman and Borjas and employed in this paper as well. We derive the conditional regression function of the log of duration in the jth spell of state ℓ, $\ell n\ t_{j\ell}$, discuss the distribution of the regression errors and demonstrate how it is possible to estimate duration dependence parameters using regression analysis. The efficiency of least squares estimators relative to maximum likelihood is derived for a special case.

The regression model derived here is the conditional regression function computed with respect to the density generated by hazard function (5) in the text with exogenous variables assumed to be fixed within each spell at value $Z_{j\ell}$ and with $\lambda_{1j\ell} = 0$ and $\gamma_{2j\ell} = 0$. To simplify notation we define $\gamma_{j\ell}$ in place of $\gamma_{1j\ell}$ in the more general model. The hazard function that generates the density of duration times is

$$\text{(B-1)} \qquad h_{j\ell}(t_{j\ell}) = \exp\{Z_{j\ell}\beta_{j\ell} + (\ell n\ t_{j\ell})\gamma_{j\ell} + V_{j\ell}\} .$$

Define $M_{j\ell} = \exp\{Z_{j\ell}\beta_{j\ell} + V_{j\ell}\}$. The expectation of $\ell n\, t_{j\ell}$ given $Z_{j\ell}$ and $V_{j\ell}$ and assuming a long panel (so there is no panel length bias as discussed in the text) is

$$E(\ell n\, t_{j\ell} | Z_{j\ell}, V_{j\ell}) = \int_0^\infty (\ell n\ t_{j\ell}) M_{j\ell}(t_{j\ell})^{\gamma_{j\ell}} \exp - \left(\frac{M_{j\ell}}{\gamma_{j\ell}+1}\right) t_{j\ell}^{\gamma_{j\ell}+1}\, dt_{j\ell} .$$

Define $\phi_{j\ell} = \left(\frac{M_{j\ell}}{\gamma_{j\ell}+1}\right) t_{j\ell}^{\gamma_{j\ell}+1}$. Then $d\phi_{j\ell} = (M_{j\ell}) t_{j\ell}^{\gamma_{j\ell}}\, dt_{j\ell}$. Substituting $\phi_{j\ell}$ for $t_{j\ell}$ and using standard LaPlace transforms

(B-2) $$E(\ln t_{j\ell}|Z_{j\ell},V_{j\ell}) = \int_0^\infty(\ln\phi_{j\ell} - \ln(\frac{M_{j\ell}}{\gamma_{j\ell}+1}))\frac{1}{\gamma_{j\ell}+1}(\exp-\phi_{j\ell})d\phi_{j\ell}$$

$$= (\frac{1}{\gamma_{j\ell}+1})\left[\Gamma'(1) + \ln(\gamma_{j\ell}+1) - Z_{j\ell}\beta_{j\ell} - V_{j\ell}\right],$$

where $\Gamma'(1)$ is the derivative of the gamma function evaluated at one ($=-.5772$).

The full regression model may be written as

$$\ln t_{j\ell} = E(\ln t_{j\ell}|Z_{j\ell},V_{j\ell}) + W_{j\ell}$$

where

(B-3) $$E(W_{j\ell}) = 0 \qquad E(W^2_{j\ell}) = \frac{B}{(\gamma_{j\ell}+1)^2}$$

where $B = \frac{d\ln\Gamma(\theta)}{d\theta}$ evaluated at $\theta = 1$ ($B=\pi^2/6$).

Note that the variance of error term $W_{j\ell}$ does <u>not</u> depend on $\beta_{j\ell}$.

To prove these results most directly it is helpful to derive the characteristic function of $W_{j\ell}$. Let $E(\ln t_{j\ell}) = \mu_{j\ell}$ so $W_{j\ell} = \ln t_{j\ell}-\mu_{j\ell}$.

Then

(B-4) $$E(\exp(i\theta W_{j\ell})) = E(\exp(i\theta(\ln t_{j\ell}-\mu_{j\ell})))$$

$$= \exp-(i\theta\mu_{j\ell})\int_0^\infty M_{j\ell}t^{i\theta}t^{\gamma_{j\ell}}\exp-\frac{M_{j\ell}}{(\gamma_{j\ell}+1)}t^{\gamma_{j\ell}+1}dt_{j\ell}.$$

$$= \exp-(i\theta\mu_{j\ell})(M_{j\ell})^{-(\frac{i\theta}{\gamma_{j\ell}+1})}\Gamma(\frac{i\theta}{\gamma_{j\ell}+1}+1)(\gamma_{j\ell}+1)^{\frac{i\theta}{\gamma_{j\ell}+1}}.$$

Since $\mu_{j\ell} = \frac{\Gamma'(1)-\ln M_{j\ell}+\ln(\gamma_{j\ell}+1)}{\gamma_{j\ell}+1}$,

$$E(\exp(i\theta W_{j\ell})) = \Gamma(\frac{i\theta}{\gamma_{j\ell}+1}+1)\exp-(\frac{i\theta\Gamma'(1)}{\gamma_{j\ell}+1}).$$

Thus the characteristic function does not depend on $\beta_{j\ell}$ as asserted and the moments of $W_{j\ell}$ are not functions of $Z_{j\ell}$, $\beta_{j\ell}$, or $V_{j\ell}$. Differentiating the characteristic function produces the moments stated above.

The density of $W_{j\ell}$ can be obtained by a direct application of the inversion theorem or by direct substitution. In our notation the density of $W_{j\ell}$ is

(B-5) $$k(W_{j\ell}) = \frac{M_{j\ell}\exp\{(\gamma_{j\ell}+1)(W_{j\ell}+\mu_{j\ell})\}}{(1+\gamma_{j\ell})}\exp-\left\{\frac{M_{j\ell}}{1+\gamma_{j\ell}}\exp\{(1+\gamma_{j\ell})(\mu_{j\ell}+W_{j\ell})\}\right\}$$

If there is no duration dependence ($\gamma_{j\ell} = 0$), the density simplifies to

$$k(W_{j\ell}) = M_{j\ell}\exp\{W_{j\ell}+\mu_{j\ell}\}\exp-M_{j\ell}\{\exp\{W_{j\ell}+\mu_{j\ell}\}\}.$$

Because of the conditional independence of the $t_{j\ell}$, $j=1,\ldots,J$, $\ell=1,\ldots,L$ (given $Z_{j\ell}$ and $V_{j\ell}$), the $W_{j\ell}$ terms are independently distributed, and $E(W_{j\ell}W_{j'\ell'}) = 0$ for $j \neq j'$ or $\ell \neq \ell'$ or both.

Collecting results and redefining the intercept term in $\beta_{j\ell}$ to include $\Gamma'(1) + \ln(\gamma_{j\ell}+1)$, the new coefficients may be defined as $\vec{\beta}_{j\ell}$. Collecting the unobservables into a composite error term $\frac{-1}{\gamma_{j\ell}+1} V_{j\ell} + W_{j\ell}$, we may write the conditional expectation of $\ln t_{j\ell}$ given $Z_{j\ell}$ as

(B-6) $$E(\ln t_{j\ell}|Z_{j\ell}) = \frac{-1}{\gamma_{j\ell}+1} Z_{j\ell}\vec{\beta}_{j\ell}$$

with associated disturbance term $U_{j\ell}$ defined by

(B-7) $$U_{j\ell} = \frac{-1}{\gamma_{j\ell}+1} V_{j\ell} + W_{j\ell}$$

with zero mean and variance $(\frac{1}{\gamma_{j\ell}+1})^2\ [\sigma^2_{j\ell} + B]$

where $E(V^2_{j\ell}) = \sigma^2_{j\ell}$. Heterogeneity components are independent of the $W_{j\ell}$

(see equation B-4) but are freely correlated across states and spells. Thus $E(V_{j\ell}V_{j'\ell'}) = \sigma_{j\ell,j'\ell'} \neq 0$ in general for $j \neq j'$ and/or $\ell \neq \ell'$.

Estimation and Identification in the Weibull Model

Under standard assumptions, least squares estimators of $\frac{\beta_{j\ell}}{1+\gamma_{j\ell}}$ derived from samples of durations of the jth spell of event ℓ are unbiased, consistent but not efficient. Consistency is achieved in the usual cross-sectional sense: by letting the number of individuals in the sample become large. The lack of efficiency is due to the highly nonnormal skewed distribution of $W_{j\ell}$ and is proved in the next section for an instructive special case.

Provided that the intercept term in $\beta_{j\ell}$ is zero, it is possible under certain conditions to estimate $1+\gamma_{j\ell}$. The intercept term in the reparameterized model is

$$\frac{\Gamma'(1) + \ell n(\gamma_{j\ell}+1)}{\gamma_{j\ell}+1}.$$

If the estimated intercept is negative, it is possible to solve for $\gamma_{j\ell}$ uniquely from the estimated intercept in the reparameterized model. If the estimated intercept is positive, and less than or equal to $\exp\{\Gamma'(1)-1\}$, a solution exists but is not unique unless it equals $\exp\{\Gamma'(1)-1\}$.

In the general case, it is not possible to consistently estimate $\beta_{j\ell}$ or $\gamma_{j\ell}$ from the estimated regression coefficients. Provided that further structure is imposed on the distribution of the heterogeneity components $V_{j\ell}$, it is possible to consistently estimate $\beta_{j\ell}$ and $\gamma_{j\ell}$ provided that we have access to panel data in which more than one spell of an event is observed.

Consider the following one-factor structure

(B-8) $$V_{j\ell} = C_\ell \phi$$

where $E(\phi) = 0$, $E(\phi^2) = \sigma_\phi^2$, and the C_ℓ are parameters $\ell=1,\ldots,L$. As is customary in factor analysis, we normalize $C_1 = 1$. $E(V_{j\ell}V_{j'\ell'}) = C_\ell C_{\ell'}\sigma_\phi^2$.

Provided that there are data on two or more spells, we can consistently estimate $\gamma_{j\ell}$, $\beta_{j\ell}$, and C_ℓ. To see how, note that the residual variance of the regression equation for log duration of the jth spell of event ℓ is

(B-9) $$\frac{C_\ell^2\ \sigma_\phi^2}{(\gamma_{j\ell}+1)^2} + \frac{B}{(\gamma_{j\ell}+1)^2}$$

The covariance between the residual in the jth duration equation and that in the j'th duration equation for event ℓ is

(B-10) $$\frac{C_\ell^2\ \sigma_\phi^2}{(\gamma_{j\ell}+1)(\gamma_{j'\ell}+1)}.$$

From the two residual variances (for spells j and j') and the covariance, it is possible to solve for $\gamma_{j\ell}$ and $\gamma_{j'\ell}$. Replacing population moments by estimated sample moments, we derive unique consistent estimators for these parameters.

Define S_{jj} as the estimated residual variance from the jth duration interval. $S_{j'j'}$ is defined in a similar fashion. $S_{jj'}$ is the estimated interspell residual covariance. The interspell residual correlation is $r_{jj'}$. From these sample moments, which are consistently estimated, it is possible to estimate $C_\ell^2\sigma_\phi^2$, $\gamma_{j'\ell}$, and $\gamma_{j\ell}$. Let "^" denote estimate. Then

$$\text{(B-11)} \qquad \hat{C}_{\ell}^{2}\sigma_{\phi}^{2} = \frac{B r_{jj'}}{1-r_{jj'}}$$

$$\hat{\gamma}_{j\ell} = \frac{B}{S_{jj}} \left(\frac{1}{1-r_{jj'}}\right)^{1/2} - 1$$

$$\hat{\gamma}_{j'\ell} = \frac{B}{S_{jj'}} \left(\frac{1}{1-r_{jj'}}\right)^{1/2} - 1$$

where positive values of square roots are used to evaluate these expressions. From consistency of the sample moments we have consistency of the estimators. Note that a negative $r_{jj'}$ is evidence against the one-factor structure. Given consistent estimators of these parameters, it is clearly possible to estimate $\beta_{j\ell}$ and $\beta_{j'\ell}$.

These estimators of the duration dependence parameters are sensitive to measurement error in the dependent variable that is independently distributed across spells. Such measurement error generates downward biased estimators of duration dependence parameters. Permanent measurement error components are absorbed in ϕ leading to an upward bias in the estimate of $C_{\ell}^{2}\sigma_{\phi}^{2}$ but no bias in the other coefficients.

Other covariance restrictions could be imposed to secure estimates of model parameters but we do not pursue the matter further here. For additional discussion of the regression approach, see Heckman and Borjas (1980).

The Relative Inefficiency of the Least Squares Estimator[1]

We consider only a single episode of an event. Further assume that the heterogeneity component is zero for everyone in our sample, $V_{j\ell} = 0$. To simplify the notation, we suppress all subscripts for events and spells.

[1]Takeshi Amemiya suggested the line of proof used in this section.

The panel is assumed to be of sufficient length that there is no censoring. We further assume that γ is known, in order to simplify the analysis.

The regression model may be written for individual i as

$$\ln t_i = \left(\frac{1}{\gamma+1}\right)(\Gamma'(1) + \ln(\gamma+1) - Z_i\beta) + W_i, \quad i=1,\ldots,I.$$

$E(W_i) = 0$, $E(W_i^2) = (\pi^2/6)\left(\frac{1}{\gamma+1}\right)^2$.

Given γ, the least squares estimators of $\hat{\beta}, \beta$ has the sampling variance

$$\text{VAR}(\hat{\beta}) = \frac{\pi^2}{6}(\Sigma Z_i'Z_i)^{-1}. \tag{B-12}$$

The likelihood for the sample is

$$\mathcal{L} = \prod_{i=1}^{I} t_i^{\gamma} \exp(Z_i\beta) \exp -[\{\exp(Z_i\beta)\}\frac{1}{\gamma+1} t_i^{\gamma+1}].$$

The log likelihood is

$$\ln\mathcal{L} = \gamma \sum_{i=1}^{I} \ln t_i + \sum_{i=1}^{I} Z_i\beta - \frac{1}{\gamma+1} \sum_{i=1}^{I} [\exp(Z_i\beta)] t_i^{\gamma+1}.$$

Given γ, the maximum likelihood estimator of β is obtained by solving

$$\frac{\partial \ln\mathcal{L}}{\partial\beta} = \sum_{i=1}^{I} Z_i - \sum_{i=1}^{I} [\exp(Z_i\beta)] \frac{t_i^{\gamma+1}}{\gamma+1} Z_i = 0.$$

Note further that

$$\frac{\partial^2 \ln\mathcal{L}}{\partial\beta'\partial\beta} = -\sum_{i=1}^{I} \exp(Z_i\beta) \frac{t_i^{\gamma+1}}{\gamma+1} Z_i'Z_i$$

and

$$E\left(-\frac{\partial^2 \ln\mathcal{L}}{\partial\beta'\partial\beta}\right) = \sum_{i=1}^{I} Z_i'Z_i.$$

Therefore the asymptotic variance-covariance matrix of the maximum likelihood estimator is

$$(\Sigma Z_i'Z_i)^{-1}. \tag{B-13}$$

Thus the asymptotic relative efficiency of the maximum likelihood estimator compared to the least squares estimator is $\pi^2/6$.

APPENDIX C

A Model of "Stigma" or "Occurrence Dependence"

Workers differ in turnover propensities and productivity characteristics. These attributes cannot be directly observed. Because of fixed costs of hiring, firm specific capital investments or costs of monitoring worker output, some information - albeit imperfect - about unobserved attributes may be valued by firms in making wage offers to potential employees.

One source of information about a worker is his employment record. Information about the number of previous jobs held, their duration, circumstances under which these jobs were terminated and what the worker did after his termination is useful in estimating the productivity of a potential match. It is not obvious that this information is of any market value. This is so because (1) workers have an incentive to misrepresent their work history or, more generally, accurate work histories are hard to come by. (2) Contingent contracts (c.f. Becker and Stigler, and Salop and Salop) might be written that select workers out by their productivity characteristics and turnover propensities, so the past record of an employee is irrelevant.

Here we sketch an idealized model of "stigma" and focus only on one piece of information: the <u>number</u> of jobs held by a worker of a given age. We assume that each job terminates with a spell of unemployment and that there is no recall. (Becker, 1980, considers a model of stigma in the marriage market).

Jobs terminate for many reasons. We assume that termination probabilities are determined in part by worker "quit" or "mismatch" characteristics as well as by micro demand shocks experienced by firms. To capture this notion most simply we suppose that $1-P$ is the per period probability of termination of a match. This probability is exogenously determined. P varies across workers but we assume it stays constant across unemployment spells for the worker. In a more general model P would be influenced by firm wage policy.

Firms know the number of jobs held by workers. They adjust to this information by offering them wages contingent on their age and on their job history. For the moment, abstract further and suppose that the wage offer distributions are indexed only by the number of employment spells. Suppose that after n jobs, the worker falls into a terminal state and is "marked" as a loser for life. We assume n is finite but this assumption is inessential.

Let $F^{(j)}(X)$ denote the wage offer distribution available to a worker with a history of j jobs. $F^{(j)}$ is stochastically dominated by $F^{(j-1)}$ and dominates $F^{(j+1)}$. For a risk neutral worker with n jobs, the optimal job search strategy for an unemployed worker is trivial to establish. Let $W_n(X)$ be the value function for a worker who receives a wage offer of X from distribution $F^{(n)}$. The job terminates with probability 1-P and the discount factor is β. W_n is defined as the expected value of $W_n(X)$ with respect to $F^{(n)}$. There is a fixed cost of search C with one offer per "period".[1]

Thus

$$W_n(X) = \text{Max}\ \left\{\frac{X}{1-P\beta} + \frac{\beta(1-P)}{1-P\beta} W_n \ ; \ W_n\right\}$$

and

$$W_n = -C + \beta \underset{X}{E} \max \left\{\frac{X}{1-P\beta} + \frac{(1-P)\beta}{1-P\beta} W_n \ ; \ W_n\right\}$$

The reservation wage, ε_n, is $(1-\beta)W_n$.

[1] Our analysis of the terminal state is similar to that of Lippman and McCall (1979). They do not, however, discuss "stigma" or "occurrence dependence."

It is obvious that the greater the values of β and P, the greater the reservation wage and the longer the expected length of search. It is also well known (see <u>e.g.</u> Kiefer and Neumann, 1979) that if the mean of the wage offer distribution increases (a pure translation of the distribution), the reservation wage increases by less than the increase in the mean so that the expected length of time in search decreases with increases in the mean of the distribution. Defining θ_n as a translation parameter,

$$\frac{\partial \varepsilon_n}{\partial \theta_n} = \frac{\frac{\beta}{1-\beta P}(1-F(\varepsilon_n))}{1+\frac{\beta}{1-\beta P}(1-F(\varepsilon_n))} \leq 1. \quad (=1 \text{ if } P=\beta=1).$$

We next consider a worker who has n-1 spells of employment. The value of a wage offer of X obtained from distribution $F^{(n-1)}$ is

$$W_{n-1}(X) = \max\left\{\frac{X}{1-P\beta} + \frac{(1-P)\beta W_n}{1-P\beta} \; ; \; W_{n-1}\right\} .$$

The first term in the braces is the expected value of accepting a wage offer of X. It is the sum of two terms: the discounted value of the wage offer (inclusive of the termination probability) and the discounted expected value of search from distribution $F^{(n)}$. The second term in braces is the discounted expected value of search from distribution $F^{(n-1)}$.

It is obvious that $W_n < W_{n-1}$. Further, it is easy to see that the reservation wage for the worker is

$$\varepsilon_{n-1} = (1-P\beta)W_{n-1} - (1-P)\beta W_n$$

Alternatively

$$\varepsilon_{n-1} - \varepsilon_n = (1-\beta P)\,[W_{n-1} - W_n]$$

so that the reservation wage for a worker with n-1 spells of employment exceeds that for a worker with n spells. This does not necessarily imply that workers with n-1 jobs search longer on an average than workers with n jobs. Although the mean of distribution $F^{(n-1)}$ necessarily exceeds the mean of distribution $F^{(n)}$ by the assumed stochastic dominance relations, the variance of $F^{(n-1)}$ may or may not be smaller hence the expected search time in spell n-1 may be shorter or longer than for spell n. However, if it is assumed that the distributions differ only in the mean, then workers with n-1 jobs necessarily search less on the average than workers with n or more jobs.

To show this in a direct way note that the equation for W_{n-1} may be written as

$$W_{n-1}(1-P\beta)(1-\beta) = -C(1-P\beta) + \beta \int_{\varepsilon_{n-1}}^{\infty} (X-\varepsilon_{n-1})\, dF^{(n-1)}(X-\theta_{n-1})$$

where $\theta_{(n-1)}$ is a translation parameter for distribution n-1. If $\theta_{n-1} = \theta_n = 0$, $F^{(n)} = F^{(n-1)}$, $W_n = W_{n-1}$, and $\varepsilon_n = \varepsilon_{n-1}$. Note that $\frac{\partial \varepsilon_{n-1}}{\partial \theta_{n-1}} = (1-\beta P)\frac{\partial W_{n-1}}{\partial \theta_{n-1}}$.

In a neighborhood of $\theta_{(n-1)} = \theta_{(n)} = 0$

$$\frac{\partial \varepsilon_{n-1}}{\partial \theta_{n-1}} = \frac{\beta(1-F)}{1-\beta F} \leq 1 \ . \quad (=1 \text{ for } \beta = 1) \ .$$

Thus, for the case of a negative translation in wage distributions across successive spells, on average individuals with n-1 spells of employment will spend less time in search than individuals with n spells of employment. By recursion, this argument can be extended to demonstrate that geometric exit time distributions have successively greater means for individuals who have held more jobs. Thus the exit time distribution for the jth spell of unemployment is stochastically dominated by the exit time distribution for spell j+1 and stochastically dominates the exit time distribution for spell j-1. This generates structural occurrence dependence.

This model can be extended to account for age although it is not especially illuminating to do so. However, it is obvious that employers would utilize age to estimate mismatch and turnover propensities. Holding P fixed, older workers will hold more jobs than younger workers. Moreover, it is also clear that the length of previous employment spells may also provide information about expected worker productivity. This gives rise to lagged duration dependence as defined by Heckman and Borjas (1980).