# Multistate Duration Models

James J. Heckman
The University of Chicago

Econ 312, Spring 2021

# Single Spell Models

A nonnegative random variable $T$ with absolutely continuous distribution function $G(t)$ and density $g(t)$ may be uniquely characterized by its hazard function. The hazard for $T$ is the conditional density of $T$ given $T > t \geq 0$ *i.e.*

$$h(t) = f(t \mid T > t) = \frac{g(t)}{1 - G(t)} \geq 0. \tag{1}$$

Knowledge of $G$ determines $h$.

Conversely, knowledge of $h$ determines $G$ because by integration of (1)

$$\int_0^t h(u)du = -\ln(1 - G(x))\,|_0^t + c, \tag{2}$$

$$G(t) = 1 - \exp\left[-\int_0^t h(u)\,du\right]; \tag{3}$$

$c = 0$ since $G(0) = 0$.

The density of $T$ is

$$g(t) = h(t) \exp\left[-\int_0^t h(u)\, du\right]. \tag{4}$$

$$\begin{aligned} h(t) &= \lim_{\Delta \to 0} \Pr(t < T < t + \Delta \mid T > t) \qquad (5) \\ & \lim_{\Delta \to 0} \left[\frac{G(t+\Delta) - G(t)}{\Delta}\right] \frac{1}{(1 - G(t))} \\ &= \frac{g(t)}{1 - G(t)}. \end{aligned}$$

The survivor function is the probability that a duration exceeds $t$. Thus

$$S(t) = P(T > 1) = 1 - G(t) = \exp\left[-\int_0^t h(u)\,du\right]. \qquad (6)$$

In terms of the survivor function we may write the density $g(t)$ as

$$g(t) = h(t)S(t).$$

Note that there is no requirement that

$$\lim_{t\to\infty} \int_0^t h(u)du \to \infty \tag{7}$$

or equivalently that $S(\infty) = 0$. Duration dependence is said to exist if

$$\frac{dh(t)}{dt} \neq 0.$$

If $\dfrac{dh(t)}{dt} > 0$, at $t = t_0$, there is said to be *positive duration dependence at* $t_0$. If $\dfrac{dh(t)}{dt} < 0$, at $t = t_0$, there is said to be *negative duration dependence at* $t_0$. In job search models of unemployment, positive duration dependence arises in the case of a "declining reservation wage" (see, e.g., Lippman and McCall, 1976). We define the conditional hazard as

$$h(t \mid \underset{\sim}{x}(t), \underset{\sim}{\theta}(t)) = \lim_{\Delta \to 0} \frac{\Pr(t < T < t + \Delta \mid T > t, \underset{\sim}{x}(t), \underset{\sim}{\theta}(t))}{\Delta}, \quad (8)$$

The dating on regressor $\underset{\sim}{x}(t)$ is an innocuous convention, $\underset{\sim}{x}(t)$ may include functions of the entire past or future or the entire paths of some variables *e.g.*

$$\begin{aligned} x_1(t) &= \int_t^{\infty} k_1(z_1(u))du \\ x_2(t) &= \int_{-\infty}^{t} k_2(z_2(u))du \\ x_3(t) &= \int_{-\infty}^{t} k_3(z_3(u), t)du \end{aligned}$$

where the $z_i(u)$ are underlying time-dated regressor variables.

(A-1) $\underset{\sim}{\theta}(t)$ is distributed independently of $x(\text{t}')$ for all $t$, $t'$. The distribution of $\underset{\sim}{\theta}$ is $\mu(\underset{\sim}{\theta})$. The distribution of $\underset{\sim}{x}$ is $D(\underset{\sim}{x})$.

(A-2) There is no functional restrictions connecting the conditional distribution of $T$ given $\underset{\sim}{\theta}$ and $\underset{\sim}{x}$ and the marginal distributions of $\underset{\sim}{\theta}$ and $\underset{\sim}{x}$.

By analogy with the definitions presented for the raw duration models, we may integrate (8) to produce the conditional duration distribution

$$G(t \mid \underset{\sim}{\theta}, \underset{\sim}{x}) = 1 - \exp\left[ - \int_0^t h(u \mid \underset{\sim}{x}(u), \underset{\sim}{\theta}(u))\, du \right] \tag{9}$$

the conditional survivor function

$$S(t \mid \underset{\sim}{\theta}, \underset{\sim}{x}) = P(T > t \mid \underset{\sim}{\theta}, \underset{\sim}{x}) = \exp\left[-\int_0^t h(u \mid \underset{\sim}{x}(u), \underset{\sim}{\theta}(u))\, du\right]$$

and the conditional density

$$g(t \mid \underset{\sim}{\theta}, \underset{\sim}{x}) = h(t \mid \underset{\sim}{x}(t), \underset{\sim}{\theta}(t)) S(t \mid \underset{\sim}{\theta}, \underset{\sim}{x}). \tag{10}$$

One specification of conditional hazard (8) that has received much attention in the literature is the *proportional hazard specification*

$$h(t \mid \underset{\sim}{x}(t), \underset{\sim}{\theta}(t)) = \psi(t)\varphi(\underset{\sim}{x}(t))\eta(\underset{\sim}{\theta}(t)) \tag{11}$$

which postulates that the log of the conditional hazard is linear in functions of $t$, $\underset{\sim}{x}$ and $\underset{\sim}{\theta}$ and that

$$\psi(t) \geqslant 0, \eta(\underset{\sim}{\theta}(t)) > 0, \varphi(\underset{\sim}{x}(t)) \geqslant 0 \text{ for all } t.$$

# Multiple Spell Models

Let $\{Y(\tau), \tau > 0\}$, $Y(\tau) \in \bar{N}$ where $\bar{N} = \{1, ..., C\}$, $C < \infty$, be a finite state continuous time stochastic process. We define random variable $R(j)$, $j \in \{1, ..., \infty\}$ as the value assumed by $Y$ at the $j^{th}$ transition time. $Y(\tau)$ or $R(j)$ is generated by the following sequence.

THE UNIVERSITY OF CHICAGO

(i) An individual begins his evolution in a state $Y(0) = R(0) = r(0)$ and waits there for a random length of time $T_1$ governed by a conditional survivor function

$$P(T_1 > t_1 \mid r(0)) = \exp\left( - \int_0^{t_1} h(u \mid \underset{\sim}{x}(u), r(0))\, du \right).$$

As before $h(u \mid \underset{\sim}{x}(u), r(0))$ is a calendar time (or age) dependent function and we now make explicit the origin state of the process.

THE UNIVERSITY OF CHICAGO

(ii) At time $\Upsilon(1) = \tau(1)$, the individual moves to a new state $R(1) = r(1)$ governed by a conditional probability law

$$P(R(1) = r(1) \mid \tau(1), r(0))$$

which may also be age dependent.

(iii) The individual waits in state $R(1)$ for a random length of time $\Upsilon_2$ governed by

$$P(T_2 > t_2 \mid \tau(1), r(1), r(0)) = \exp\left( - \int_0^{t_2} h(u \mid \underset{\sim}{x}(u + \tau(1)), r(1), r(0))\, du \right).$$

Note that one coordinate of $\underset{\sim}{x}(u)$ may be $u + \tau(1)$, and that $\Upsilon(2) - \Upsilon(1) = T_2$. At the transition time $\Upsilon(2) = \tau(2)$ he switches to a new state $R(2) = r(2)$ where the transition probability

$$P(R(2) = r(2) \mid \tau(1), \tau(2), r(1), r(0))$$

may be calendar time dependent.

Continuing this sequence of waiting times and moves to new states gives rise to a sequence of random variables

$$\begin{aligned} R(0) &= r(0), \Upsilon(1) = \tau(1), \\ R(1) &= r(1), \Upsilon(2) = \tau(2), \\ R(2) &= r(2), \ldots \end{aligned}$$

and suggests the definitions

$$Y(\tau) = R(k) \text{ for } \tau(k) \leq \tau < \tau(k+1)$$

where $R(k)$, $k = 0, 1, 2, ...$ is a discrete time stochastic process governed by the conditional probabilities

$$P(R(k) = r(k) \mid \underset{\sim}{t}_{k}, \underset{\sim}{r}_{k-1})$$

where

$$\underset{\sim}{t}_k = (t_1, \ldots, t_k) \text{ and } \underset{\sim}{r}_{k-1} = (r(0), ..., r(k-1)).$$

$T_k = \Upsilon(k) - \Upsilon(k-1)$ is governed by the conditional survivor function

$$P(T_k \geq t_k \mid \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}) = \exp\left( - \int_0^{t_k} h(u \mid \underset{\sim}{x}(u + \tau(k-1)), \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}) \right.$$

# Specializations of Interest

**Repeated events of the same kind.**

This is a one state process, *e.g.* births in a fertility history. $R(\cdot)$ is a degenerate process and attention focuses on the sequence of waiting times $T_1$, $T_2$,.... .

One example of such a process writes

$$P(T_k > t_k) = \exp\left(-\int_0^{t_k} h_k(u \mid \underset{\sim}{x}(u + \tau(k-1)))\, du\right).$$

The hazard for the $k^{th}$ interval depends on the number of previous spells. This special form of dependence is referred to as *occurrence dependence*. In a study of fertility, $k - 1$ corresponds to birth parity for a woman at risk. Heckman and Borjas (1980) consider such models for the analysis of unemployment.

Another variant writes the hazard of a current spell as a function of the mean duration of previous spells *i.e.* for spell $j > 1$

$$h(u \mid \underset{\sim}{x}(u + \tau(j-1)), \underset{\sim}{t}_{j-1}) = h\left(u \mid \frac{1}{j-1}\sum_{i=1}^{j-1} t_i, \tau(j-1) + u\right)$$

(see, e.g., Braun and Hoem (1978)).

THE UNIVERSITY OF CHICAGO

Yet another version of the general model writes for the $j^{th}$ spell
$$h(u \mid \underset{\sim}{x}(u+\tau(j-1)), \underset{\sim}{t}_{j-1}) = h_j(u \mid \underset{\sim}{x},\ t_1, t_2, ...., t_{j-1}).$$
This is a model with both occurrence dependence and lagged duration dependence, where the latter is defined as dependence on lengths of preceding spells.

A final specification writes

$$h(u \mid \underset{\sim}{x}(u+\tau(j-1)), \underset{\sim}{t}_{j-1}) = h(\underset{\sim}{x}(u+\tau(j-1))).$$

For spell $j$ this is a model for independent non-identically distributed durations; and $Y(\tau)$ is a nonstationary renewal process.

# Multistate Processes

Let

$$P(R(k) = r(k) \mid \underset{\sim}{t}_{k}, \underset{\sim}{r}_{k-1}) = m_{r(k-1),r(k)}$$

where

$$\| \, m_{ij} \, \| \; = M$$

is a finite stochastic matrix with and

$$P(T_k > t_k \mid \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}) = \exp\left(-\lambda_{r(k-1)} t_k\right)$$

where the elements of $\{\lambda_i\}$ are positive constants. Then $Y(\tau)$ is a time homogeneous Markov chain with constant intensity matrix

$$Q = \Lambda(M - I)$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \lambda_C \end{pmatrix}$$

and $C$ is the number of states in the chain.

In a dynamic McFadden model for a stationary environment, $M$ has the special structure $m_{ij} = m_{\ell j} = P_j$ for all $i$ and $\ell$ *i.e.* the origin state is irrelevant in determining the destination state. This restricted model can be tested against a more general specification.

A time inhomogeneous semi-Markov process emerges as a special case of the general model if we let

$$P(R(k) = r(k) \mid \underset{\sim}{t}_k, \underset{\sim}{r}_{k-1}, \tau(k-1)) = \pi_{r(k-1),r(k)}(\tau(k), t_k)$$

where

$$\| \pi_{ij}(\tau, u) \| = \underset{\sim}{\pi}(\tau, u)$$

is a two parameter family of time ($\tau$) and duration ($u$) dependent stochastic matrices with each element a function $\tau$ and $u$ and $m_{ii} = 0$.

We further define

$$\begin{aligned} P(T_k &> t_k \mid \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}, \tau(k-1)) \\ &= \exp\left(-\int_0^{t_k} h(u \mid \underset{\sim}{r}_{(k-1)}, \tau(k-1))\, du\right). \end{aligned}$$

With this restricted form of dependence, $Y(\tau)$ is a time inhomogeneous semi-Markov process. (Hoem, 1972, provides a nice expository discussion of such processes). Moore and Pyke (1968) consider the problem of estimating a time inhomogeneous semi-Markov model without observed or unobserved explanatory variables.

The natural estimator for a model without restrictions connecting the parameters of

$$P(R(k) = r(k) \mid \underset{\sim}{t}_{k}, \underset{\sim}{r}_{k-1}, \tau(k-1))$$

and

$$P(T_k > t_k \mid \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}, \tau(k-1))$$

breaks the estimation into two components.

a. Estimate $\underset{\sim}{\Pi}$ by using data on transitions from $i$ to $j$ for observations with transitions having identical (calendar time $\tau$, duration $u$) pairs. A special case of this procedure for a model with no duration dependence in a time homogeneous environment pools $i$ to $j$ transitions for all spells to estimate the components of $M$ (see also Billingsley, 1061). Another special case for a model with duration dependence in a time homogeneous environment pools $i$ to $j$ transitions for all spells of a given duration.

b. Estimate $P(T_k > t_k \mid \underset{\sim}{t}_{k-1}, \underset{\sim}{r}_{k-1}, \tau(k-1))$ using standard survival methods on times between transitions.

# General Duration Models For The Analysis of Event History Data

An individual event history is assumed to evolve according to the following steps.

(i) At time $\tau = 0$, an individual is in state $i$, there are $N_i \leq C - 1$ possible destinations. The limit (as $\Delta t \to 0$) of the probability that a person who starts in $i$ at calendar time $\tau = 0$ leaves the state in interval $(t_1, t_1 + \Delta t)$ given regressor path $\{\underset{\sim}{x}(u)\}_0^{t_1+\Delta t}$ and unobservable $\theta$ is the conditional hazard or escape rate

$$\lim \frac{P(t_1 < T_1 < t_1 + \Delta t \mid \underset{\sim}{r}_{(0)}{=}(i), \Upsilon(0){=}0, \underset{\sim}{x}(t_1), \theta)}{\Delta t} \tag{12}$$

$$= h(t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0){=}0, \underset{\sim}{x}(t_1), \theta). \tag{13}$$

This limit is assumed to exist. (Assumed to be "regular"). The limit (as $\Delta t \to 0$) of the probability that a person starting in $\underset{\sim}{r}_{(0)} = (i)$ at time $\tau(0)$ leaves to go to $j \neq i$, $j \in N_i$ in interval $(t_1, t_1 + \Delta t)$ given regressor path $\{\underset{\sim}{x}(u)\}_0^{t_1+\Delta t}$ and $\theta$ is

$$\begin{aligned} &\lim_{\Delta t \to 0} \frac{P(t_1 < T_1 < t_1 + \Delta t, R(1) = j \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \underset{\sim}{x}(t_1),}{\Delta t} \\ = \;& h(t_1, j \mid \underset{\sim}{r}_{(0)}{=}(i), \Upsilon(0){=}0, \underset{\sim}{x}(t_1), \theta). \qquad (1 \end{aligned}$$

From the laws of conditional probability

$$\sum_{j=1}^{N_i} h(t_1, j \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \underset{\sim}{x}(t_1), \theta) = h(t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0,$$

(ii) The probability that a person starting in state $i$ at calendar time $\tau = 0$ survives to $T_1 = t_1$ is (from the definition of the survivor function in (8) and from hazard (69))

$$\begin{aligned} P(T_1 &> t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta) \\ &= \exp\left( - \int_0^{t_1} h(u \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \underset{\sim}{x}(u), \theta)\, du \right). \end{aligned}$$

Thus the density of $T_1$ is

$$f(t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta)$$

$$= -\frac{\partial P(T_1 > t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta)}{\partial t_1}$$

$$= h(t_1 \mid \underset{\sim}{r}_{(0)} = i, \Upsilon(0) = 0, x(t_1), \theta) P(T_1 > t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0,$$

The density of the joint event $R(1) = j$ and $T_1 = t_1$ is

$$f(t_1, j \quad | \quad \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta) =$$
$$h(t_1, j \quad | \quad \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \underset{\sim}{x}(t_1), \theta) P(T_1 > t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0$$

The density is sometimes called a subdensity. Note that

$$\sum_{j=1}^{N_i} f(t_1, j \quad | \quad \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta)$$

$$= \quad f(t_1 \mid \underset{\sim}{r}_{(0)} = (i), \Upsilon(0) = 0, \{\underset{\sim}{x}(u)\}_0^{t_1}, \theta).$$

Proceeding in this fashion, one can define densities corresponding to each duration in the individual's event history. Thus, for an individual who starts in state $\underset{\sim}{r}_{(m)}$ after his $m^{th}$ transition, the subdensity for $T_{m+1} = t_{m+1}$ and $R(m+1) = j, j = 1, ..., N_\ell$ is

$$f(t_{m+1}, f \mid \underset{\sim}{r}_{(m)}, \Upsilon(m) = \tau(m), \{\underset{\sim}{x}(u)\}_0^{\tau(m+1)}, \theta)$$

where

$$\tau(m+1) = \sum_{n=1}^{m+1} t_n. \tag{15}$$

The conditional density of completed spells $T_1, ..., T_k$ and right censored spell $T_{k+1}$ given $\{\underset{\sim}{x}(u)\}_0^{\tau(k)+t_{k+1}}$ assuming that $\Upsilon(0) = 0$ is the exogenous start date of the event history (and so corresponds to the origin date of the sample) is, allowing for more general forms of dependence,

$$g(t_1, r(1), t_2, r(2), ..., t_k, r(k), t_{k+1} \mid \{\underset{\sim}{x}(u)\}_0^{\tau(k)+t_{k+1}}) = \tag{16}$$

$$\int \left\{ P(T_{k+1} > t_{k+1} \mid \underset{\sim}{r}_{(k)}, \underset{\sim}{t}_{(k)}, \tau(k), \{\underset{\sim}{x}(u)\}_{\tau(k)}^{\tau(k)+t_k}, \theta \right\} d\mu(\theta).$$

As noted in section 5, it is unlikely that the origin date of the sample coincides with the start date of the event history. Let

$$\varphi(r(0), \Upsilon(0) = 0, r(1), t_{1a}, \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(1)}, \theta)$$

be the probability density for the random variables describing the events that a appropriate $h$ as defined in (8.2). The joint density of $(r(0), t_{1c}, r(1))$ the *completed spell* density sampled at $\Upsilon(0) = 0$ terminating in state $r(1)$ is defined analogously. For such spells we write the density as

$$\varphi(r(0), \Upsilon(0) = 0, t_{1c}, r(1), \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(1)}, \theta).$$

In a multiple spell model setting in which it is plausible that the process has been in operation prior to the origin date of the sample, intake rate $k$ is the density of the random variable $\Upsilon$ describing the event "entered the state $r(o)$ at time $\Upsilon = \tau \leq 0$ and did not leave the state until $\Upsilon > 0$." The expression for $k$ in terms of the exit rate depends on (i) presample values of $\underset{\sim}{x}$ and (ii) the date at which the process began. Thus in principle given (i) and (ii) it is possible to determine the functional form of $k$. In this context it is plausible that $k$ depends on $\theta$.

The joint likelihood for $r(0)$, $t_{1\ell}(\ell = a, c), r(1), t_2, ..., r(k), t_{k+1}$ conditional on $\theta$ and $\{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(k)+t_{k+1}}$ for a right censored $k+1$ st spell is

$$g(r(0), t_{1\ell}, r(1), t_2, r(2), ..., t_k, r(k), t_{k+1} \Big| \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(k)+t_{k+1}}, \theta)$$

$$= \varphi(r(0), \Upsilon(0) = 0, t_{1\ell}, r(1) \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(1)}, \theta)$$

$$\left[\prod_{i=2}^{k} f(t_i, r(i) \mid \underset{\sim}{r}_{(i-1)}, \underset{\sim}{t}_{(k-1)} \tau(i-1), \underset{\sim}{x}(u)\}_{\tau(i-1)}^{\tau(i)}, \theta\right]$$

$$P(T_{k+1} > t_{k+1} \mid \underset{\sim}{r}_{k}, \underset{\sim}{t}_{(k-1)}, \tau(k-1), \{\underset{\sim}{x}(u)\}_{0}^{\tau(k)+t_{k+1}}, \theta). \qquad (17)$$

THE UNIVERSITY OF CHICAGO

The marginal likelihood obtained by integrating out $\theta$ is,

$$g(r(0), t_{1\ell}, r(1), t_2, \ldots, t_k, r(k), t_{k+1} \quad | \quad \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(k)+t_{k+1}}) =$$

$$\int_{\underset{=}{\theta}} g(r(0), t_{1\ell}, r(1), t_2, \ldots, t_k, r(k), t_{k+1} \quad | \quad \{\underset{\sim}{x}(u)\}^{\tau(k)+t_{k+1}}, \theta) du(\theta) \quad (18)$$

Using (16) and conditioning on $T_{1\ell} = t_{1\ell}$ produces conditional likelihood

$$g(r(0), t_{1\ell}, r(1), t_2, ..., t_k, r(k), t_{k+1} \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{\tau(k)+t_{k+1}}, \theta, T_{1\ell} = t_{1\ell})$$

$$= \prod_{i=2}^{k} f(t_i, r(i) \mid \underset{\sim}{r}_{(i-1)}, \underset{\sim}{t}_{(i-1)}, \tau(i-1), \{x(u)_{\tau(i-1)}^{\tau(i)}, \theta) \cdot$$

$$P(T_{k+1} > t_{k+1} \mid \underset{\sim}{r}_{k}, \underset{\sim}{t}_{k}, \tau(k), \{x(u)\}_{\tau(k)}^{\tau(k)+t_{k+1}}, \theta).$$

# Competing Risk Specifications

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

$$f_{ij}(t_{ij}) = h_{ij}(t_{ij}) \exp\left[-\int_0^{t_{ij}} h_{ij}(u)du\right] \quad (j = 1..., N; \, j \neq i) \qquad \text{(A1)}$$

where $f_{ij}(\cdot)$ is the density function of exit times from state $i$ into state $j$, and $h_{ij}(\cdot)$ is the associated hazard function.

The joint density of the $N-1$ latent exit times is given by

$$\prod_{\substack{j=1 \\ j\neq i}}^{N} h_{ij}(t_{ij}) \exp\left[-\int_0^{t_{ij}} h_{ij}(u)du\right]. \tag{A2}$$

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

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

Let the probability that the individual leaves state $i$ and then directly enters state $j'$ be denoted by $P_{ij'}$.
Then

$$P_{ij'} = A3 \tag{19}$$

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

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

$$= \int_0^\infty h_{ij'}(t_{ij'}) \exp\left\{ -\int_0^{t_{ij'}} \left[ \sum_{\substack{k=1 \\ k \neq i}}^N h_{ik}(u) \right] du \right\} dt_{ij'} . A3 \tag{20}$$

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

$$g(t_{ij'} \quad | \quad t_{ij'} < t_{ij})(\forall j : j \neq j'; j, j' \neq i) A4 \tag{21}$$

$$= \frac{h_{ij'}(t_{ij'}) \exp\left\{ -\int_0^{t_{ij'}} \left[ \sum_{\substack{k=1 \\ k \neq i}}^{N} h_{ik}(u) \right] du \right\}}{P_{ij'}} . A4 \tag{22}$$

It follows that the density of exit times from state $i$ into all other states combined can be written

$$
\begin{aligned}
f_{i\cdot}(t_{i\cdot}) &= \sum_{\substack{j'=1 \\ j'\neq i}}^{N} P_{ij'} g(t_{ij'} \mid t_{ij'} < t_{ij})(\forall j : j \neq j';\; j, j' \neq i) \\
&= \left[\sum_{\substack{k=1 \\ k\neq i}}^{N} h_{ik}(t_{i\cdot})\right] \exp\left\{-\int_{0}^{t_{i\cdot}} \left[\sum_{\substack{k=1 \\ k\neq i}}^{N} h_{ik}(\mathrm{u})\right] \mathrm{du}\right\} . A5\,(23)
\end{aligned}
$$

The probability that the spell is not complete by time $C$ is simply

$$\text{Prob}(T > C) = \int_C^\infty f_{i\cdot}(t)dt = \exp\left\{-\int_0^C \left[\sum_{\substack{k=1 \\ k\neq i}}^{N} h_{ik}(u)\right] du\right\}. \quad \text{(A6)}$$

This term enters the likelihood function for incomplete spells of at least $C$ in length. 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. (Obviously a nonlinear regression procedure can account for censoring). $[Z_{1rm}(u + \tau_{rm})...z_{(K-2)rm}(u + \tau_{rm})V_r]$.

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

$$\beta_{ij} = [\beta_{0ij}\beta_{1ij}...\beta_{(K-2)ij}C_{ij}]. \tag{A7}$$

As discussed in (Flinn and Heckman, 1982) we impose a one factor specification, so that $C_{ij}$ is the factor loading associated with the $i$ to $j$ transition. The usual normalizations required in factor analysis in discrete data models are imposed. (See Heckman, 1981, pp. 167-174).

We write the hazard function for the mth spell and rth individual as

$$h_{i_m j_m}(t_{rm}) = \exp\left[Z'_{r_m}(t_{rm} + \tau_{rm})\beta_{i_m j_m}\right] = \frac{f_{i_m j_m}(t_{rm})}{1 - F_{i_m j_m}(t_{rm})} \qquad \text{(A8)}$$

where $f_{i_m j_m}$ is the cdf associated with (A1). From state $i_m$ into state $j_m$ is given by expression (A4). The density of a duration in the mth spell that begins in $i_m$ and terminates in state $j_m$ is the product of these two terms and is written as

$$g_{i_m j_m}(t_{rm}) = h_{i_m j_m}(t_{rm}) \exp\left\{ -\int_0^{t_{rm}} \sum_{\substack{k=1 \\ k \neq i_m}}^{N} h_{i_m k}(u) du \right\}. \qquad \text{(A9)}$$

Densities of this type are called subdensities in the duration analysis literature (see Kalbfleisch and Prentice, 1980, p. 167). It is notationally convenient to write the exp term in (A9) as

$$S_{i_m}(t_{rm}) = \exp\left\{ -\int_0^{t_{rm}} \sum_{\substack{k=1 \\ k \neq i_m}}^{N} h_{i_m k}(u)du \right\} \tag{A10}$$

so

$$g_{i_m j_m}(t_{rm}) = h_{i_m j_m} S_{i_m}(t_{rm}).$$

Note that from (A6) the probability that the mth spell lasts more than $t_{rm}$ is

$$P(T_{rm} > t_{rm}) = S_{i_m}(t_{rm}) \tag{A11}$$

so $S$ has a substantive probabilistic interpretation. It is called the survivor function in the duration analysis literature.

Consider an individual's contribution to the likelihood function (suppressing the individual's subscript for notational convenience)

$$L(\beta, V) = \left[\sum_{i=1}^{M-1} g_{i_m j_m}(t_m)\right] S_{i_M}(t_M) \quad \text{(A12)}$$

where the Mth (and final) censored spell is assumed to last longer than $t_M$. Treating $V$ as a random effect, the integrated likelihood is

$$L(\beta) = \int_{-\infty}^{\infty} L(\beta, V) d\mu(V) \quad \text{(A13)}$$

where $d\mu(V)$ is the density of $V$. Now define

$$\$(\beta) = \ln[\bar{L}(\beta)].$$

Note that

$$\frac{\partial \$(\beta)}{\partial \beta_{gij}} = \frac{1}{\bar{L}(\beta)} \frac{\partial \bar{L}(\beta)}{\partial \beta_{gij}} = \frac{1}{\bar{L}(\beta)} \int_{-\infty}^{\infty} \frac{\partial \bar{L}(\beta, V)}{\partial \beta_{gij}} d\mu(V) \tag{A14}$$

a. It allows for a flexible Box-Cox hazard with scalar heterogeneity

$$h \mid \underset{\sim}{x}, \theta) = \exp\left(\underset{\sim}{x}(t)\beta + \left(\frac{t^{\lambda_1} - 1}{\lambda_1}\right)\gamma_1 + \left(\frac{t^{\lambda_2} - 1}{\lambda_2}\right)\gamma_2 + \right. \tag{13'}$$

where $\beta, \gamma_1, \gamma_2, \lambda_1, \lambda_2$ and $c$ are permitted to depend on the origin state, the destination state and the serial order of the spell. Lagged durations may be included among the $\underset{\sim}{x}$. Using maximum likelihood procedures it is possible to estimate all of these parameters except for one normalization of $c$.

THE UNIVERSITY OF CHICAGO

b. It allows for general time varying variables and right censoring. The regressors may include lagged durations.
c. $\mu(\theta)$ can be specified as either normal, log normal or gamma or the NPMLE procedure can be used.
d. It solves the left censoring or initial conditions problem by assuming that the functional form of the initial duration distribution for each origin state is different from that of the other spells.

# Conventional Reduced Form Models

## Proposition 1

*Uncontrolled unobservables bias estimated hazards towards negative duration dependence.*

The proof is a straightforward application of the Cauchy-Schwartz theorem. Let $h(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})$ be the hazard conditional on $\underset{\sim}{x}, \underset{\sim}{\theta}$ and $h(t \mid \underset{\sim}{x})$ is the hazard conditional only on $\underset{\sim}{x}$. These hazards are associated respectively with conditional distributions $G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})$ and $G(t \mid \underset{\sim}{x})$.

From the definition,

$$h(t \mid \underset{\sim}{x}) = \frac{\int g(t \mid \underset{\sim}{x}, \underset{\sim}{\theta}) d\mu(\underset{\sim}{\theta})}{\int (1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})) d\mu(\underset{\sim}{\theta})}.$$

Thus

$$\frac{\partial h(t \mid \underset{\sim}{x})}{\partial t} = \frac{\int (1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})) \frac{\partial h(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})}{\partial t} d\mu(\underset{\sim}{\theta})}{\int (1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})) d\mu(\underset{\sim}{\theta})}$$

$$+ \frac{\left[\int g(t \mid \underset{\sim}{x}, \underset{\sim}{\theta}) d\mu(\underset{\sim}{\theta})\right]^2 - \int \frac{g^2(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})}{1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})} d\mu(\underset{\sim}{\theta}) \int (1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta}) d}{\left[\int (1 - G(t \mid \underset{\sim}{x}, \underset{\sim}{\theta})) d\mu(\underset{\sim}{\theta})\right]^2}.$$

The second term on the right hand side is always nonpositive as a consequence of the Cauchy-Schwartz theorem.■

Intuitively, more mobility prone persons are the ones likely to leave first.

There are many possible conditional hazard functions (see, e.g., Lawless (1982)). One class of proportional hazard models that nests many previous models as a special case and therefore might be termed "flexible" is the Box-Cox conditional hazard

$$h(t \mid \underset{\sim}{x}, \underset{\sim}{\theta}) = \exp\left\{ \underset{\sim}{x}(t)\beta + \left(\frac{t^{\lambda_1}-1}{\lambda_1}\right) + \left(\frac{t^{\lambda_2}-1}{\lambda_2}\right)\gamma_2 + \underset{\sim}{\theta}(t) \right\} \tag{24}$$

$\gamma_2 = 0$ and $\lambda_1 = 0$; Weibull $\gamma_2 = 0$ and $\lambda_1 = 1$ Gompertz.
The conventional approach does, however, allow for right censored spells assuming independent censoring mechanism. We consider two such schemes.

Let $V(t)$ be the probability that a spell is censored at duration $t$ or later. If

$$\begin{aligned} V(t) &= 0 \qquad t < L \\ V(t) &= 1 \qquad t \geq L \end{aligned} \tag{25}$$

there is censoring at fixed duration $L$. This type of censoring is common in many economic data sets. More generally, for continuous censoring times let $v(t)$ be the density associated with $V(t)$. Let $d = 1$ if a spell is not right censored and $d = 0$ if it is. Let $t$ denote an observed spell length. Then the joint frequency of $(t, d)$ conditional on $\underset{\sim}{x}$ for the case of absolutely continuous distribution $V(t)$ is

$$\begin{aligned} f\left(t, d \,\middle|\, \underset{\sim}{x}\right) &= v(t)^{(1-d)} \int_{\underset{=}{\theta}} [h(t \mid \underset{\sim}{x}(t), \theta)] V(t)]^{d} S(t \mid \underset{\sim}{x}(t), \theta) d\mu(\theta) & (26) \\ &= \{v(t)^{1-d} V(t)^{d}\} \int [h(t \mid x(t), \theta)]^{d} S(t) \mid x(t), \theta) d\mu(\theta) & (27) \end{aligned}$$

For a Dirac censoring distribution, the density of observed durations is

$$f(t, d \mid \underset{\sim}{x}) = \int\limits_{\underset{=}{\theta}} [h(t \mid \underset{\sim}{x}(t), \theta)]^d S(t \mid \underset{\sim}{x}(t), \theta) d\mu(\theta). \tag{28}$$

Except for special time paths of variables the term

$$\int\limits_0^t h(u \mid \underset{\sim}{x}(u), \theta) du$$

which appears the survivor function does not have a closed form expression. To evaluate it requires numerical integration.

To circumvent this difficulty, one of two expedients is often adopted (see, *e.g.* Lundberg, 1981, Cox and Lewis, 1966).

1. 1. Replacing time trended variables with their within spell average

$$\bar{\underset{\sim}{x}}(t) = \frac{1}{t}\int_0^t \underset{\sim}{x}(u)du \qquad t > 0$$

   2. Using beginning of spell values $\underset{\sim}{x}(0)$. Expedient (i) has the undesirable effect of building spurious dependence between duration time $t$ and the manufactured regressor variable. To see this most clearly, suppose that $\underset{\sim}{x}$ is a scalar and $\underset{\sim}{x}(u) = a + bu$. Then clearly

$$\bar{x}(t) = a + \frac{b}{2}t,$$

   and $t$ and $\bar{x}(t)$ are clearly linearly dependent. Expedient (ii) ignores the time inhomogeneity in the environment.

Table 1. Weibull Model-Employment to Nonemployment Transitions
(Absolute Value of Normal Statistics in Parentheses)

| | Regressors Fixed At Average Value Over The Spell (expedient i) | Regressors Fixed At Value As Of tart of Spell (expedient ii) | Regressors Vary Freely |
|---|---|---|---|
| Intercept | .971<br>(1.535) | -3.743<br>(12.074) | -3.078<br>(8.670) |
| In duration ($\gamma_1$) | -.137<br>(1.571) | -.230<br>(2.888) | -.341<br>(3.941) |
| Married with Spouse Present? (=1 if yes; = Otherwise) | -1.093<br>(2.679) | -.921<br>(2.310) | -.610<br>(1.971) |
| National Unemployment Rate | -1.800<br>(6.286) | .569<br>(3.951) | .209<br>(1.194) |

Source: See Flinn and Heckman, 1982b, p. 69.

These empirical results are typical. Introducing time varying variables into single spell duration models is inherently dangerous and *ad hoc* methods for doing so can produce wildly misleading results. More basically, separating the effect of time varying variables from duration dependence is only possible if there is "sufficient" independent variation in $\underset{\sim}{x}(t)$ scalar. Taking logs, we reach

$$\ln(h(t \mid x, \theta)) = x(t)\beta + \left(\frac{t^{\lambda_1} - 1}{\lambda_1}\right)\gamma_1 + \theta(t).$$

# Identification and Estimation Strategies

1. (A) What features, if any, of $h(t \mid \underset{\sim}{x}, \theta)$ and/or $h(t \mid \underset{\sim}{x}, \theta)$ are identified from the "raw data", i.e., $G(t \mid \underset{\sim}{x})$?

   (B) Under what conditions are $h(t \mid \underset{\sim}{x}, \theta)$ and $\mu(\theta)$ identified? i.e., how much *a priori* information has to be imposed on the model before these functions are identified?

   (C) What empirical strategies exist for estimating $h(t \mid \underset{\sim}{x}, \theta)$ and/or $\mu(\theta)$ nonparametrically and what is their performance?

## Nonparametric Procedures to Assess The Structural Hazard $h(t \mid \underset{\sim}{x}, \theta)$

In order to test whether or not an empirical $G(t \mid \underset{\sim}{x})$ exhibits positive duration dependence, it is possible to use the *total time on test statistic* (Barlow *et.al.*, 1972, p. 267). This statistic is briefly described here. For each set of $\underset{\sim}{x}$ values, constituting a simple of $I_{\underset{\sim}{x}}$ durations, order the first $k$ durations starting with the smallest

$$t_1 \leq t_2 \leq t_k, \; 1 \leq k \leq I_{\underset{\sim}{x}}.$$

Let $D_{i:I_{\underset{\sim}{x}}} = [I_{\underset{\sim}{x}} - (i+1)](t_i - t_{i-1})$, where $t_0 \equiv 0$. Define

$$V_k = k^{-1} \sum_{i=1}^{k-1} \left[ \sum_{j=1}^{i} D_{j:I_{\underset{\sim}{x}}} \right] / \; k^{-1} \sum_{i=1}^{k} D_{i:I_{\underset{\sim}{x}}}.$$

$V_k$ is called the cumulative total time on test statistic. If the observations are from a distribution with an increasing hazard rate, $V_k$ tends to be large. Intuitively, if $G(t \mid \underset{\sim}{x})$ is a distribution that exhibits positive duration dependence. $D_{1:I_{\underset{\sim}{x}}}$ stochastically dominates $D_{2:I_{\underset{\sim}{x}}}, D_{2:I_{\underset{\sim}{x}}}$ stochastically dominates $D_{3:I_{\underset{\sim}{x}}}$, and so forth. Critical values for testing the null hypothesis.

Let $G_1 = \{G : -\ln[1 - G(t \mid \underset{\sim}{x})]$ is concave in $t$ holding $\underset{\sim}{x}$ fixed$\}$. Membership in this class can be determined from the total time on test statistic. If $G_1$ is log concave, the $D_{i:I_{\underset{\sim}{x}}}$ defined earlier are stochastically increasing in $i$ for fixed $I_{\underset{\sim}{x}}$ and $\underset{\sim}{x}$. Ordering the observations from the largest to the smallest and changing the subscripts appropriately, we can use $V_k$ to test for log concavity.

Next let $G_2 = \{G : G(t \mid \underset{\sim}{x}) = \int (1 - \exp\left(-t\varphi(\underset{\sim}{x})\eta(\theta)\right)\, d\mu(\theta)$ for some probability measure $\mu$ on $[0,\infty]\}$. It is often erroneously suggested that $G_1 = G_2$ *i.e.* that negative duration dependence by a homogenous population ($G \in G_1$) cannot be distinguished from a pure heterogeneity explanation ($G \in G_2$).

In fact, by virtue of Bernstein's theorem (see, *e.g.* Feller, 1971, p. 439-440) if $G \in G_2$ it is completely monotone *i.e.*

$$(-1)^n \frac{\partial^n}{\partial t^n}(1 - G(t \mid \underset{\sim}{x})) \geq 0 \text{ for } n \geq 1 \text{ and all } t \geq 0 \tag{29}$$

and if $G(t \mid \underset{\sim}{x})$ satisfies (23), $G(t \mid \underset{\sim}{x}) \in G_2$.

Setting $n = 3$, (23) is violated if $(-1)^3 \frac{\partial^3}{\partial t^3}(1 - G(t \mid \underset{\sim}{x})) < 0$ *i.e.* if for some $t = t_0$

$$\left[ -\frac{\partial^2 h(t \mid \underset{\sim}{x})}{\partial t^2} + 3h(t \mid \underset{\sim}{x}) \frac{\partial h(t \mid \underset{\sim}{x})}{\partial t} - h^3(t \mid \underset{\sim}{x}) \right]_{t=t_0} > 0.$$

Formal verification of (23) requires uncensored data sufficiently rich to support numerical differentiation twice. Note that if the data are right censored at $t = t^*$, we may apply (23) over the interval $0 < t \leq t^*$ provided that we define

$$1 - G^*(t \mid \underset{\sim}{x}) = \frac{\int \left(1 - e^{-t\varphi(\underset{\sim}{x})\eta(\theta)}\right) \, d\mu(\theta)}{\int \left(1 - e^{-t^*\varphi(\underset{\sim}{x})\eta(\theta)}\right) \, d\mu(\theta)}$$

and test whether

$$(-1)^n \frac{\partial^n}{\partial t^n}(1 - G^*(t \mid \underset{\sim}{x})) \geq 0 \text{ for } n \geq 1 \text{ and } 0 < t \leq t^*. \qquad (30)$$

The key insight in his test is as follows. For $G \in G_2$, the probability that $T > k$ is the survivor function

By a transformation of variables $z = \exp\left(-\varphi(\underset{\sim}{x})\eta(\theta)\right)$, we may transform (25) for fixed $\underset{\sim}{x}$ to

$$S(k \mid \underset{\sim}{x}) = \int_0^1 z^k d\mu^*(z)$$

*i.e.* as the $k^{th}$ moment of a random variable defined on the unit interval.

From the solution to the classical Hausdorff moment problem (see, *e.g.*, Shohat and Tamarkin, 1943, p. 9) it is known that there exists a $\mu^*(z)$ that satisfies (23) if

$$\Delta^k S(\ell \mid \underset{\sim}{x}) \geq 0 \qquad k, \ell = 0, 1, .., \infty \tag{31}$$

where

$$\Delta^0 S(\ell \mid \underset{\sim}{x}) = S(\ell \mid \underset{\sim}{x})$$

$$\Delta^1 S(\ell \mid \underset{\sim}{x}) = S(\ell \mid \underset{\sim}{x}) - S(\ell + 1 \mid \underset{\sim}{x})$$

$$\Delta^k S(\ell \mid \underset{\sim}{x}) = S(\ell \mid \underset{\sim}{x})$$

$$- \begin{pmatrix} k \\ 1 \end{pmatrix} S(\ell \Big| \underset{\sim}{x}) + \begin{pmatrix} k \\ 2 \end{pmatrix} S(\ell + 2 \mid \underset{\sim}{x}) + \ldots + (-1)^k S(\ell + k \mid \underset{\sim}{x})$$

.

Choosing equispaced intervals $(0, 1, ....[t^*])$ where $[t^*]$ is the nearest whole integer less than $t^*$, form the $S(\ell \mid \underset{\sim}{x})$ functions $\ell = 0, ..., [t^*]$.

Compute the survivor functions so defined and test a subset of the necessary conditions. $(\ell = 1, ..., k)$.

It is important to note that these are rejection criteria. There are other models that may satisfy (23).

For example

$$S(t) = \int_0^\infty e^{-t^{\alpha}\theta} \, d\mu(\theta) \tag{32}$$

for $\alpha < 1$ is completely monotone. By Bernstein's theorem this distribution has one representation in $G_2$ but it is not unique.

# Identifiability

$$h(t \mid \underset{\sim}{x}, \theta) = \psi(t)\varphi(\underset{\sim}{x})\theta. \tag{33}$$

Before stating identifiability conditions, it is useful to define

$$Z(t) = \int_0^t \psi(u)du.$$

Then for the proportional hazard model we have the following proposition due to Elbers and Ridder (1982).

## Proposition 2

*If (i)* $E(\Theta) = 1$, *(ii)* $Z(t)$ *defined on* $[0, \infty)$ *can be written as the integral of a nonnegative integrable function* $\psi(t)$ *defined on* $[0, \infty)$,

$Z(t) = \int_0^t \psi(u)du$, *(iii) the set* $\underset{\sim}{S}$, $\underset{\sim}{x} \in \underset{\sim}{S}$ *is an open set in* $R^k$ *and the function* $\varphi$ *is defined on* $\underset{\sim}{S}$ *and is nonnegative, differentiable and nonconstant on* $S$, *then* $Z, \varphi$, *and* $\mu(\theta)$ *are identified.* ■

A general strategy of proof for this case is as follows (for details see Heckman and Singer (1984a)) Assume that $Z'_{\underset{\sim}{\alpha}}(t)$ is a member of a parametric family of nonnegative functions and that the pair $(\underset{\sim}{\alpha}, \mu)$ is not identified. Assuming that $Z'_{\underset{\sim}{\alpha}}$ is differentiable to order $j$, nonidentifiability implies the identities

$$1 = \frac{g_1(t)}{g_0(t)} = \frac{Z'_{\underset{\sim}{\alpha}_1}(t) \int_0^\infty \theta \; e^{-Z'_{\underset{\sim}{\alpha}_1}(t)\theta} d\mu_1(\theta)}{Z'_{\underset{\sim}{\alpha}_0}(t) \int_0^\infty \theta \; e^{-Z'_{\underset{\sim}{\alpha}_0}(t)\theta} d\mu_0(\theta)}$$

$$\cdots$$

$$1 = \frac{g_l^{(j)}(t)}{g_0^{(j)}(t)}.$$

## Proposition 3

*For the true value of* $\lambda, \lambda_0$*, defined so that* $\lambda_0 \leq 0$*, if* $E(\Theta) < \infty$ *for all admissible* $\mu$*, and for all bounded* $\gamma$*, then the triple* $(\gamma_0, \lambda_0, \mu_0)$ *is uniquely identified.*■ *(For proof, see Heckman and Singer 1984a).*

THE UNIVERSITY OF CHICAGO

## Proposition 4

*For the true value of* $\lambda, \lambda_0$, *such that* $0 < \lambda_0 < 1$, *if all admissible* $\mu$ *are restricted to have a common finite mean that is assumed to be known a priori* ($E(\Theta) = m_1$) *and a bounded (but not necessarily common) second moment* $E(\Theta^2) < \infty$, *and all admissible* $\gamma$ *are bounded, then the triple* $(\gamma_0, \lambda_0, \mu_0)$ *is uniquely identified.*■ *(For proof see Heckman and Singer, 1984a).*

## Proposition 5

*For the true value of $\lambda, \lambda_0$, restricted so that $0 < \lambda_0 < j$, $j$ a positive integer, if all admissible $\mu$ are restricted to have a common finite mean that is assumed to be known a priori $(E(\Theta) = m_1)$ and a bounded (but not necessarily common) $j + 1^{st}$ moment $(E(\Theta^{j+1}) < \infty)$, and all admissible $\gamma$ are bounded, then the triple $(\gamma_0, \lambda_0, \mu_0)$ is uniquely identified.■ (For proof see Heckman and Singer, 1984a).*

THE UNIVERSITY OF CHICAGO

It is interesting that each integer increase in the value of $\lambda_0 > 0$ requires an integer increase in the highest moment that must be assumed finite for all admissible $\mu$.

The general strategy of specifying a flexible functional form for the hazard and placing moment restrictions on the admissible $\mu$ works in other models besides the Box-Cox class of hazards. For example consider a nonmonotonic log logistic model used by Trussell and Richards (1983).

$$Z'(t) = \frac{(\lambda\alpha)(\lambda t)^{\alpha-1}}{1+(\lambda t)^{\alpha}}, \qquad \infty > \lambda, \alpha > 0 \tag{34}$$

## Proposition 6

*For hazards model (4.8), the triple $(\lambda_0, \alpha_0, \mu_0)$ is identified provided that the admissible $\mu$ are restricted to have a common finite mean $E(\Theta) = m_1 < \infty$.■ (For proof, see Heckman and Singer, 1984a).*

## Sampling Plans and Initial Conditions Problems

Begin after the date of the sample. For interrupted spells one of the following duration times may be observed: (1) time in the state up to the sampling date ($T_b$) (2) time in the state after the sampling date ($T_a$) or (3) total time in a completed spell observed at the origin of the sample ($T_c = T_a + T_b$). Durations of spells that begin after the origin date of the sample are denoted $T_d$.

# Time Homogeneous Environments and Models Without Observed and Unobserved Explanatory Variables

Time 0. Looking backward, a spell of length $t_b$ interrupted at 0 began $t_b$ periods ago. Looking forward, the spell lasts $t_a$ periods after the sampling date. The completed spell is $t_c = t_b + t_a$ in length. We ignore right censoring and assume that the underlying distribution is nondefective. (These assumptions are relaxed below.)

Let $k(-t_b)$ be the intake rate *i.e.* $t_b$ periods before the sample begins, $k(-t_b)$ is the proportion of the population that enters the state of interest at time $\tau = -t_b$. The time homogeneity assumption implies that

$$k(-t_b) = k, \forall t_b.$$

Let $g(t) = h(t)\exp\left[-\int_0^t h(u)\,du\right]$ be the density of completed durations *in the population*. The associated survivor function is

$$S(t) = 1 - G(t) = \exp\left[-\int_0^t h(u)du\right].$$

The proportion of the population experiencing a spell at calendar time $\tau = 0, P_0$, is obtained by integrating over the survivors from each cohort, *i.e.*

$$P_0 = \int_0^\infty k(-t_b)(1-G(t_b)dt_b = \int_0^\infty k(-t_b)\exp\left[-\int_0^{t_b} h(u)\,du\right] dt_b. \tag{35}$$

Thus the density of an interrupted spell of length $t_b$ is the ratio of the proportion surviving from those who entered $t_b$ periods ago to the total stock

$$f(t_b) = \frac{k(-t_b)(1 - G(t_b))}{P_0} = \frac{k(-t_b)\exp\left[-\int_0^{t_b} h(u)\,du\right]}{P_0}.$$

This rules out defective distributions. Assuming $m = \int_0^\infty x g(x) dx < \infty$ and integrating the denominator of the preceding expression by parts, we reach the familiar expression (see, *e.g.* Cox and Lewis (1966))

$$f(t_b) = \frac{(1 - G(t_b))}{m} = \frac{S(t_b)}{m} = \frac{1}{m} \exp\left[ - \int_0^{t_b} h(u)\, du \right].$$

The density of sampled interrupted spells is *not* the same as the population density of completed spells.

The density of sampled completed spells is obtained by the following straightforward argument. In the population, the conditional density of $t_c$ given $0 < t_b < t_c$ is

$$g(t_c \mid t_b) = \frac{g(t_c)}{(1 - G(t_b))} = h(t_c) \exp\left[-\int_{t_b}^{t_c} h(u)\, du\right], \;\; t_c > t_b. \tag{36}$$

Using the density of $f(t_b)$, the marginal density of $t_c$ in the sample is

$$f(t_c) = \int_0^{t_c} g(t_c \mid t_b) f(t_b) dt_b = \int_0^{t_c} \frac{g(t_c)}{m}\, dt_b \qquad (37)$$

so

$$f(t_c) = \frac{t_c g(t_{c)}}{m}.$$

The density of the forward time $t_a$ can be derived in a similar fashion.

$$f(t_a) = \int_0^\infty g(t_a + t_b \mid t_b) f(t_b) dt_b = \int_0^\infty \frac{g(t_a + t_b)}{m} dt_b$$

$$\frac{1}{m} \int_{t_a}^\infty g(z) dz = \frac{(1 - G(t_a))}{m} = \frac{S(t_a)}{m} = \frac{\exp\left[-\int_0^{t_a} h(u)\, du\right]}{m} \quad (38)$$

So in a time homogeneous environment the functional form of $f(t_a)$ is identical to $f(t_b)$.

THE UNIVERSITY OF CHICAGO

The following results are well known about the distributions of the random variables $T_a$, $T_b$ and $T_c$.

1. If $g(t)$ is exponential with parameter $\theta$ (*i.e.* $g(t) = \theta e^{-t\theta}$) then so are $f(t_a)$ and $f(t_b)$. The proof is immediate.
2. $E(T_a) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$

   where $\sigma^2 = E(T - m)^2 = \int_0^\infty (t - m)^2 g(t) dt$.
3. $E(T_b) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$
   (since $T_a$ and $T_b$ have the same density)

1. $E(T_c) = m\left(1 + \frac{\sigma^2}{m^2}\right)$
   so $E(T_c) = 2E(T_a) = 2E(T_b)$ and $E(T_c) > m$ unless $\sigma^2 = 0$.
2. If $\frac{-\ln(1-G(t))}{t} \uparrow$ in $t$, $\frac{\sigma^2}{m^2} > 1$. (This condition is implied if $h(t) = \frac{g(t)}{1-G(t)}$ is decreasing in $t$, i.e., $h'(t) < 0$. In this case, $E(T_a) = E(T_b) > m$. (See Barlow and Proschan 1975 for proof.
3. If $\frac{-\ln(1-G(t))}{t} \downarrow t$, $\frac{\sigma^2}{m^2} < 1$. (This condition is implied if $h'(t) > 0$). In this case $E(T_a) = E(T_b) < m$. (See Barlow and Proschan 1975 for proof).

We next present the distribution of $T_d$, the duration time for spells that begin after the origin date of the sample. Let $\Upsilon$ denote the time a spell begins. The density of $\Upsilon$ is $k(\tau)$. Assuming that $\Upsilon$ and $T_d$ are independent the joint probability that a spell begins at $\Upsilon = \tau$ and lasts less than $t_d$ periods is

$$\Pr\{\Upsilon = \tau \text{ and } T_d < t_d\} = k(\tau)G(t_d).$$

Thus the density of $T_d$ in a time homogeneous environment is

$$f(t_d) = g(t_d). \tag{39}$$

It is common to “solve” the left censoring problem by assuming that $G(t)$ is exponential. The bias that results from invoking this assumption when it is false can be severe. As an example suppose that the population distribution of $t$ is Weibull so

$$g(t) = \alpha\varphi \; t^{\alpha-1} e^{-\varphi t^{\alpha}} \quad \varphi > 0, \quad c > 0.$$

For $\alpha = 2$

$$\text{plim } \hat{\lambda} = (\varphi)^{1/2}\Gamma(1/2).$$

As another example, suppose the sample being analyzed consists of complete spells sampled at time zero (*i.e.* $T_c$) generated by an underlying population exponential density

$$g(t) = \lambda e^{-t\lambda}.$$

Then from (32)

$$f(t_c) = \lambda^2 t_c e^{-\lambda t_c}.$$

If it is falsely assumed that $g(t)$ characterizes the duration data and $\theta$ is estimated by maximum likelihood plim $\hat{\lambda} = 2\lambda$.

Continuing this example, suppose instead that a Weibull model is falsely assumed *i.e.*

$$g^*(t) = \alpha \ t^{\alpha-1}\varphi e^{-t^{\alpha}\varphi}$$

and the parameters $\alpha$ and $\varphi$ are estimated by maximum likelihood. The maximum likelihood estimator solves the following equations,

$$\frac{1}{\hat{\varphi}} = \frac{\sum_{i=1}^{I} t_i^{\hat{\alpha}}}{I}$$

$$\frac{1}{\hat{\alpha}} + \frac{\sum_{i=1}^{I} \ln t_i}{I} = \frac{\hat{\varphi}\sum_{i=1}^{I} (\ln t_i)\, t_i^{\hat{\alpha}}}{I}$$

so

$$\frac{1}{\hat{\alpha}} + \frac{\sum_{i=1}^{I} \ln t_i}{I} = \frac{\sum_{i=1}^{I} t_i^{\hat{\alpha}} \ln t_i}{\sum_{i=1}^{I} t_i^{\hat{\alpha}}}. \tag{40}$$

Using the easily verified result that

$$\int_0^\infty t^{P-1} \ln\left(te^{-t\lambda}\right)\, dt = \lambda^{-P}\left\{\frac{\partial\Gamma(P)}{\partial P} - \ln\left(\lambda\Gamma(p)\right)\right\}$$

and that fact that in large samples plim $\hat{\alpha} = \alpha^*$ is the value of $\alpha^*$ that solves (31), $\alpha^*$ is the solution to

$$\frac{1}{\alpha^*} + E(\ln t) = \frac{E(t^{\alpha^*} \ln t)}{E(t^{\alpha^*})}$$

we obtain the equation

$$\frac{1}{\alpha^*} + \left(\frac{\partial\Gamma(P)}{\partial P}\Big|_{P=2} - \ln\lambda\right) = \left(\frac{\partial\ln\Gamma(P)}{\partial P}\Big|_{P=\alpha^*+2} - \ln\lambda\right). \quad (41)$$

Using the fact that

$$\frac{\Gamma^*(P+1)}{\Gamma(P+1)} = \frac{1}{P} + \frac{\Gamma'(P)}{\Gamma(P)}$$

and collecting terms, w may rewrite (32) as

$$\frac{1}{\alpha^*(\alpha^*+1)} + \frac{\partial \Gamma(P)}{\partial P}\,|_{P=2} = \frac{1}{\Gamma(P)}\frac{\partial \Gamma(P)}{\partial P}\,|_{P=\alpha^*+1}\ . \qquad (42)$$

Since $\Gamma(2) = 1$, it is clear that $\alpha^* = 1$ is never a solution of this equation. In fact, since the left hand side is monotone decreasing in $\alpha^*$ and the right hand side is monotone increasing in $\alpha^*$, and since at $\alpha^* = 1$, the left hand side $\alpha^* > 1$.

It can also be shown that

$$\text{plim } \hat{\varphi} = \frac{\lambda^{\alpha^*-1}}{\Gamma(\alpha^* + 2)}.$$

## The Densities of $T_a$, $T_b$, $T_c$ and $T_d$ in Time Inhomogeneous Environments For Models With Observed and Unobserved Explanatory Variables

We define $k(\tau \mid \underset{\sim}{x}(\tau), \theta)$ to be the intake rate into a given state at calendar time $\tau$. We assume that $\theta$ is a scalar heterogeneity component and $\underset{\sim}{x}(\tau)$ is a vector of explanatory variables. It is convenient and correct to think of $k(\tau \mid \underset{\sim}{x}(\tau), \theta)$ as the density associated with the random variable $\Upsilon$ for a person with characteristics $(\underset{\sim}{x}(\tau), \theta)$. We continue the useful convention that spells are sampled at $\Upsilon = 0$. The densities of $T_a$, $T_b.T_c$ and $T_d$ are derived for two cases: (a) conditional on a sample path $\{\underset{\sim}{x}(u)\}_{-\infty}^{t}$ and (b) marginally on the sample path $\{\underset{\sim}{x}(u)\}_{-\infty}^{t}$ (*i.e.* integrating it out). We denote the distribution of $\{\underset{\sim}{x}(u)\}_{-\infty}^{t}$ as $D(\underset{\sim}{x})$ with associated density $dD(\underset{\sim}{x})$.

The derivation of the density of $T_b$ conditional on $\{\underset{\sim}{x}(u)\}_{-\infty}^{0}$ is as follows. The proportion of the population in the state at time $\tau = 0$ is obtained by integrating over the survivors of each cohort of entrants. Thus

$$P_0(\underset{\sim}{x}) \quad =$$

$$\int\limits_0^\infty \int\limits_{\underset{=}{\theta}} k(-t_b \quad | \quad \underset{\sim}{x}(-t_b), \theta) \exp\left( - \int_0^{t_b} h(u \mid \underset{\sim}{x}(u - t_b), \theta)\, du \right) d\mu(\theta) d$$

The proportion of people in the state with sample path $\{\underset{\sim}{x}(u)\}_{-\infty}^{0}$ whose spells are exactly of length $t_b$ is the set of survivors from a spell that initiated at $\tau = -t_b$ or

$$\int_{\underset{=}{\theta}} k(-t_b \mid \underset{\sim}{x}(-t_b), \theta) \exp\left(-\int_0^{t_b} h(u \mid \underset{\sim}{x}(u - t_b), \theta)\, du\right) d\mu(\theta).$$

Thus the density of $T_b$ conditional on $\{\underset{\sim}{x}(u)\}_{-\infty}^{0}$ is

$$f(t_b \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{0} \quad (43$$

$$= \frac{\int\limits_{\underset{=}{\theta}} k(-t_b \mid \underset{\sim}{x}(-t_b), \theta) \exp\left(-\int_0^{t_b} h(u \mid \underset{\sim}{x}(u - t_b), \theta)\, du\right)\, d\mu(\theta}{P_0(\underset{\sim}{x})}$$

The marginal density of $T_b$ (integrating out $\underset{\sim}{x}$) is obtained by an analogous argument: divide the marginal flow rate as of time $\Upsilon = -t_b$ (the integrated flow rate) by the marginal (integrated) proportion of the population in the state at $\tau = 0$.

Thus defining

$$P_0 = \int\limits_{\underset{=}{X}} P_0(\underset{\sim}{x}) dD(\underset{\sim}{x})$$

where $\underset{=}{X}$ is the domain of integration for $\underset{\sim}{x}$ we write

$$f(t_b) = \frac{\int\limits_{\underset{=}{X}} \int\limits_{\underset{=}{\theta}} k(\text{-t}_\text{b}\text{-t}_\text{b}) \mid \underset{\sim}{x}(\text{-t}_\text{b}\text{-t}_\text{b}), \theta) \exp\left(- \int_0^{t_b} h(u \mid \underset{\sim}{x}(u - t_b), \theta)\, du\right.}{P_0}$$

Note that we use a function space integral to integrate out $\{\underset{\sim}{x}(u)\}_{-\infty}^{0}$. (See Kac (1959) for a discussion of such integrals).

Note further that one obtains an incorrect expression for the marginal density of $T_b$ if one integrates (43) against the population density of $\underset{\sim}{x}$ $(dD(\underset{\sim}{x}))$.

The error in this procedure is that the appropriate density for $\underset{\sim}{x}$ against which (43) should be integrated is a density of $\underset{\sim}{x}$ conditional on the event that an observation is in the sample at $\tau = 0$. By Bayes' theorem this density is

$$f(\underset{\sim}{x} \mid T_b > 0) = \left( \int_0^\infty f(t_b \mid \{\underset{\sim}{x}(u))\}_{-\infty}^{0}) dD(\underset{\sim}{x}) dt_b \right) \frac{P_0(\underset{\sim}{x})}{P_0}$$

which is not in general the same as the density $dD(\ \underset{\sim}{x})$. For proper distributions for $T_b$,

$$f(\underset{\sim}{x} \mid T_b > 0) = dD(\underset{\sim}{x}) \frac{P_0(\underset{\sim}{x})}{P_0}.$$

The derivatives of the density of $T_c$, the completed length of a spell sampled at $\Upsilon = 0$ is equally straightforward. For simplicity we ignore right censoring problems so that we assume that the sampling frame is of sufficient length that all spells are not censored and further assume that the underlying duration distribution is not defective. (But see the remarks at the conclusion of this section.)

Conditional on $\{\underset{\sim}{x}(u)\}_{-\infty}^{t}$ and $\theta$ the probability that the spell began at $\tau$ is

$$k(\tau \mid \underset{\sim}{x}(\tau), \theta).$$

The conditional density of a completed spell of length $t$ that begins at $\tau$ is

$$\int h(t \mid \underset{\sim}{x}(\tau + t), \theta) \exp\left(-\int_0^t h(u \mid \underset{\sim}{x}(\tau + u), \theta)\right) d\mu(\theta).$$

For a fixed $\tau \leq 0$, $t_c$ by definition exceeds $-\tau$. Conditional on $\underset{\sim}{x}$, the probability that $T_c$ exceeds $\tau$ is $P_0(\underset{\sim}{x})$.

Thus, integrating out $\tau$, respecting the fact that $t_c > -\tau$

$$f\left(t_c \,\middle|\, \{\underset{\sim}{x}(u)\}_{-\infty}^{t_c}\right) =$$

$$\frac{\int\limits_{-t_c}^{0} \int\limits_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t \mid \underset{\sim}{x}(\tau + t_c), \theta) \exp\left[-\int_0^{t_c} h(u \mid \underset{\sim}{x}(\tau + u), \theta) d\right.}{P_0(\underset{\sim}{x})}$$

The marginal density of $T_c$ is

$$f(t_c) \tag{46}$$

$$= \frac{\int_{-t_c}^{0} \int_{\underset{=}{X}} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_c \mid \underset{\sim}{x}(\tau + t), \theta) \exp\left[- \int_0^{t_c} h(u \mid \underset{\sim}{x}(\tau + u), \theta)\, du\right]}{P_0}$$

Ignoring right censoring, the derivation of the density of $T_a$ proceeds by recognizing that $T_a$ conditional on $\Upsilon \leq 0$ is the right tail portion of random variable $-\Upsilon + T_a$, the duration of a completed spell that begins at $\Upsilon = \tau$. The probability that the spell is sampled is $P_0(\underset{\sim}{x})$.

Thus the conditional density of $T_a = t_a$ given $\{\underset{\sim}{x}(u)\}_{-\infty}^{t_a}$ is obtained by integrating out $\tau$ and correctly conditioning on the event that the spell is sampled *i.e.*

$$f(t_a \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{t_a} \tag{47}$$

$$= \frac{\int\limits_{-\infty}^{0} \int\limits_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_a - \tau \mid \underset{\sim}{x}(t_a - \tau), \theta) \exp\left(-\int_0^{t_a - \tau} h(u \mid \underset{\sim}{x}(u + \tau), \theta\right.}{P_0(\underset{\sim}{x})}$$

and the corresponding marginal density is

$$f(t_a) \tag{48}$$

$$= \frac{\int_{-\infty}^{0} \int_{\underset{=}{X}} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_a - \tau) \mid \underset{\sim}{x}(t_a + \tau), \theta) \exp \left[ - \int_0^{t_a - \tau} h(u \mid \underset{\sim}{x}(u + \tau), \theta)\, du \right.}{P_0} \tag{49}$$

Of special interest is the case $k(\tau \mid x, \theta) = k(\underset{\sim}{x})$ in which the intake rate does not depend on unobservables and is constant for all $\tau$ given $x$, and in which $\underset{\sim}{x}$ is time invariant. Then (43) specializes to

$$f(t_b \mid \underset{\sim}{x}) = \frac{1}{m(x)} \int_{\underset{=}{\theta}} \exp\left[-\int_0^{t_b} h(u \mid \underset{\sim}{x}, \theta)\, du\right] d\mu(\theta) \qquad (50)$$

where

$$m(\underset{\sim}{x}) = \int_0^{\infty} \int_{\theta} \exp\left[-\int_0^{z} h(u \mid \underset{\sim}{x}, \theta)\, du\right] d\mu(\theta)\, dz.$$

This density is essentially of the same functional form as the density after (30). Under the same restrictions on $k$ and $\underset{\sim}{x}$, (43) and (44) specialize respectively to

$$f(t_c \mid \underset{\sim}{x}) = \frac{t_c \int_{\underset{=}{\theta}} h(t_c \mid \underset{\sim}{x}, \theta) \exp\left[-\int_0^{t_c} h(u \mid \underset{\sim}{x}, \theta)\, du\right] d\mu(\theta)}{m(\underset{\sim}{x})} \tag{51}$$

and

$$f(t_a \mid \underset{\sim}{x}) = \frac{\int_{\underset{=}{\theta}} \exp\left[-\int_0^{t_a} h(u \mid \underset{\sim}{x}, \theta)\, du\right] d\mu(\theta)}{m(\underset{\sim}{x})} \tag{52}$$

For this special case all of the results (i)-(vi) stated in subsection $A$ go through with obvious redefinition of the densities to account for observed

It is only for this special case of $k(\tau \mid \underset{\sim}{x}, \theta)$ with time invariant regressors that the densities of $T_a$, $T_b$ and $T_c$ do not depend on the parameters of $k$.

The common expedient for "solving" the initial conditions problem for the density of $T_a$—assuming that $G(t \mid \underset{\sim}{x}, \theta)$ is exponential—does not avoid the dependence of the density of $T_a$ on $k$ even if $k$ does not depend on $\theta$ as long as it depends on $\tau$ or $\underset{\sim}{x}(\tau)$ where $\underset{\sim}{x}(\tau)$ is not time invariant.

Thus in the exponential case in which $h(u \mid \underset{\sim}{x}(u+\tau), \theta) = h(\underset{\sim}{x}(u+\tau), \theta)$, we may write (44) for the case $k = k(\tau \mid \underset{\sim}{x}(\tau))$ as

$$f(t_a \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{t_a})$$

$$= \frac{\int_{-\infty}^{0} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau)) e^{-\int_0^{-\tau} h(\underset{\sim}{x}|(u+\tau),\theta)du} h(\underset{\sim}{x}(t_a), \theta) e^{-\int_0^{t_a} h(\underset{\sim}{x}(u),\theta)\, du} d\mu(\theta) d}{\int_{-\infty}^{0} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau)) e^{-\int_0^{-\tau} h(\underset{\sim}{x}(u+\tau),\theta)du} d\mu(\theta) d\tau}$$

Only if $h(\underset{\sim}{x}(u+\tau),\theta) = h(\underset{\sim}{x}(u+\tau))$, so that unobservables do not enter the model (or equivalently that the distribution of $\Theta$ is degenerate), does $k$ disappear from the expression. In this case the numerator factors into two components, one of which is the denominator of the density. "$k$" also disappears if it is a time invariant constant that is functionally independent of $\theta$.

a. The functional form of $k(\tau \mid \underset{\sim}{x}(\tau), \theta)$ is not in general known. This includes as a special case the possibility that for some known $\tau^* < 0$, $k(\tau \mid \underset{\sim}{x}(\tau), \theta) \equiv 0$ for $\tau < \tau^*$. In addition, the value of $\tau^*$ may vary among individuals so that if it is unknown it must be treated as another unobservable.

b. If $\underset{\sim}{x}(\tau)$ is not time invariant, its value may not be known for $\tau < 0$ so that even if the functional form of $k$ is known, the correct conditional duration densities cannot be constructed.

The initial conditions problem stated in its most general form is intractable. However, various special cases of it can be solved. For example, suppose that the functional form of $k$ is known up to some finite number of parameters, but presample values of $\underset{\sim}{x}(\tau)$ are not.

If the distribution of these presample values is known or can be estimated, one method of solution to the initial conditions problem is to define duration distributions conditional on post sample values of $\underset{\sim}{x}(\tau)$ from the model using the distribution of their values.

This approach suggests using

$$f\left(t_c \mid \{\underset{\sim}{x}(u)\}_0^{t_c}\right)$$

$$= \int_{-t_c}^{0} \int_{\underset{=}{\theta}} \int_{\{\underset{\sim}{x}(\tau):\tau<0\}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_c \mid \underset{\sim}{x}(t+\tau), \theta) \exp\left[-\int_0^{t_c} h(u \mid \underset{\sim}{x}(\tau+u), \theta) du\right]$$

Recall, however, that the distribution of $\underset{\sim}{x}$ within the sample is *not* the distribution of $\underset{\sim}{x}$ in the population, $D(\underset{\sim}{x})$. This is a consequence of the impact of the sample selection rule on the joint distribution of $\underset{\sim}{x}$ and $T$. The distribution of the $\underset{\sim}{x}$ within sample depends on the distribution of $\theta$, and the parameters of $h(t \mid \underset{\sim}{x}, \theta)$ and the presample distribution of $\underset{\sim}{x}$. Thus, for example, the joint density of $T_a$ and $\underset{\sim}{x}$ for $\tau > 0$ is

$$f(t_a, \underset{\sim}{x}(\tau))$$

$$= \frac{dD(\underset{\sim}{x}) \int_{-t_a}^{0} \int_{\underset{=}{\theta}} \int_{\{\underset{\sim}{x}:\tau<0\}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_a + \tau \mid \underset{\sim}{x}(t_a + \tau), \theta) e^{-\int_0^{t_a - \tau} h(u \mid \underset{\sim}{x}(u+\tau), \theta)}}{P_0}$$

so, the density of within sample $\underset{\sim}{x}(\tau)$ is

$$f(\underset{\sim}{x}(\tau) \mid \tau \geq 0) = \int_0^\infty f(t_a, \underset{\sim}{x}(\tau)) dt_a$$

$$= \frac{dD(\underset{\sim}{x})}{P_0} \int_0^\infty \int_{-t_a}^{0} \int_{\underset{=}{\theta}} \int_{\{\underset{\sim}{x}:\tau<0\}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_a + \tau \mid \underset{\sim}{x}(t_a + \tau), \theta) e^{-\int_0^{t_a - \tau} h(u \mid \underset{\sim}{x}(u+\tau), \theta)\, du} dD(\underset{\sim}{x})\, d$$

It is *this* density and not $dD(\underset{\sim}{x})$ that is estimated using within sample data on $\underset{\sim}{x}$.

A partial avenue of escape from the initial conditions problem exploits $T_d$ *i.e.* durations for spells initiated after the origin date of the sample. The density of $T_d$ conditional on $\{\underset{\sim}{x}(u)\}_0^{t_d+\tau_d}$ where $\tau_d > 0$ is the start date of the spell is

$$f(t_d \mid \{\underset{\sim}{x}(u)\}_0^{t_d+\tau_d})$$

$$= \frac{\displaystyle\int_0^\infty \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_d \mid \underset{\sim}{x}(\tau + t_d), \theta) e^{-\int_0^{t_d} h(u|\underset{\sim}{x}(\tau+u),\theta)\, du} \, d\mu(\theta)\, d\tau}{\displaystyle\int_0^\infty \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta)\, d\mu(\theta)\, d\tau}$$

THE UNIVERSITY OF CHICAGO

The denominator is the probability that $\Upsilon \geq 0$. Only if $k$ does not depend on $\theta$ will be the density of $T_d$ not depend on the parameters of $k$. More efficient inference is based on the joint density of $\Upsilon$ and that $t_d$

$$f\left(t_d, \tau \mid \{\underset{\sim}{x}(u)\}_0^{t_d+\tau_d}\right)$$

$$= \frac{\int\limits_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_d \mid \underset{\sim}{x}(\tau + t_d), \theta) \exp\left[-\int\limits_0^{t_d} h(u \mid \underset{\sim}{x}(\tau + u), \theta) du\right]}{\int\limits_0^{\infty} \int\limits_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta)\, d\mu(\theta)\, d\tau}$$

For example, the density of measured completed spells that begin after the start date of the sample incorporates the facts that $0 \leq \Upsilon \leq \tau^*$ and $T_d \leq \tau^* - \Upsilon$ *i.e.* that the onset of the spell occurs after $\tau = 0$ and that all completed spells must be length $\tau^* - \Upsilon$ or less. Thus we write (recalling that $\tau_d$ is the start date of the spell)

$$f(t_d \mid \{\underset{\sim}{x}(u)\}_0^{t_d+\tau_d}, T_d \leq \tau^* - \Upsilon, \Upsilon \geq 0)$$

$$= \frac{\int_0^{\tau^*-t_d} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_d \mid \underset{\sim}{x}(\tau + t_d), \theta) e^{-\int_0^{t_d} h(u|\underset{\sim}{x}(\tau+u),\theta)\, du}\, d\mu(}{\int_0^{\tau^*} \int_0^{\tau^*-t_d} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_d \mid \underset{\sim}{x}(\tau + t_d), \theta) e^{-\int h(u|\underset{\sim}{x}(\tau+u),\theta) du}\, d\mu(\theta}$$

The density of right censored spells that begin after the start date of the sample is simply the joint probability of the events $0 < \Upsilon < \tau^*$ and $T_d > \tau^* - \Upsilon$ *i.e.*

$$P(0 < \Upsilon < \tau^* \wedge T_d > \tau^* - \Upsilon \mid \{\underset{\sim}{x}(u)\}_0^{\tau^*}) =$$

$$= \int_0^{\tau^*} \int_{\tau^*-\tau}^{\infty} \int_{\underline{\underline{\theta}}} k(\tau \mid$$

$$\underset{\sim}{x}(\tau), \theta) \exp\left[-\int_0^{\tau^*-t_d} h(u \mid \underset{\sim}{x}(\tau+u), \theta) du\right] d\mu(\theta) dt_d d\tau.$$

The modification required in the other formulae presented in this subsection to account for the finiteness of the sampling plan are equally straightforward. For spells sampled at $\tau = 0$ for which we observe presample values of the duration and post sample *completed* durations ($T_c$), it must be the case that (a) $\Upsilon \leq 0$ and (b) $\tau^* - \Upsilon > T_c > -\Upsilon$ where $\tau^* > 0$ is the length of the sampling plan. Thus in place of (41) we write

$$f(t_c \mid \{\underset{\sim}{x}(u)\}_{-\infty}^{t_c}, -\Upsilon < T_c \leq \tau^* - \Upsilon, \Upsilon \leq 0)$$

$$= \frac{\int_{-t_c}^{\tau^*-t_c} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_c \mid \underset{\sim}{x}(\tau + t_c), \theta) e^{-\int_0^{t_c} h(u \mid \underset{\sim}{x}(\tau+u), \theta)\, du}\, d\mu(\theta}{\int_{-\infty}^{0} \int_{-\tau}^{\tau^*-t} \int_{\underset{=}{\theta}}^{-\tau} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_c \mid \underset{\sim}{x}(\tau + t_c), \theta) e^{-\int_0^{t_c} h(u \mid \underset{\sim}{x}(\tau+u), \theta)\, du}\, d\mu(\theta}$$

The denominator of this expression is the joint probability of the events that $-\Upsilon < T_c < \tau^* - \Upsilon$ and $\Upsilon \leq 0$. For spells sampled at $\tau = 0$ for which we observe presample values of the duration and post sample *right censored durations*, it must be the case that (a) $\Upsilon < 0$ and (b) $T_c \geq \tau^* - \Upsilon$ so the density for such spell is

$$f\left(t_c \mid \underset{\sim}{x}(u)\right\}_{-\infty}^{t_c}, T_c \geq \tau^* - \Upsilon, \Upsilon \leq 0)$$

$$= \int_{-\infty}^{0} \int_{\tau^*-\tau}^{\infty} \int_{\underset{=}{\theta}} k(\tau \mid \underset{\sim}{x}(\tau), \theta) h(t_c \mid \underset{\sim}{x}(\tau + t_c), \theta) e^{-\int_0^{t_c} h(u|\underset{\sim}{x}(\tau+u),\theta)\, du} d\mu(\theta)$$

## Examples of Duration Models Produced by Economic Theory

## Example A

## A Dynamic Model of Labor Force Participation

The consumer works at age $a$ if the marginal rate of substitution between goods and leisure evaluated at the no work position (also known as the nonmarket wage)

$$M(Y(a)) = \frac{U_2(Y(a), 1)}{U_1(Y(a), 1)} \tag{53}$$

function $I(a)$ written as

$$I(a) = W(a) - M(Y(a)). \tag{54}$$

If $I(a) \geq 0$, the consumer works at age $a$ and we record this event by setting $d(a) = 1$. If $I(a) < 0$, $d(a) = 0$.

For each person successive values of $\varepsilon(a)$ may be correlated but it is assumed that $\varepsilon(a)$ is independent of $Y(a)$ and $W(a)$. We define the index function inclusive of $\varepsilon(a)$ as

$$I^*(a) = W(a) - M(Y(a)) + \varepsilon(a). \tag{55}$$

The probability that an employed person does not leave the employed state is

$$1 - F(\psi) \tag{56}$$

where $\psi = M(Y) - W$. The probability of receiving $j$ new values of $\varepsilon$ in interval $t_e$ is

$$P_j = \binom{t_e}{j} P^j (1-P)^{t_e - j}.$$

The probability that a spell is longer than $t_e$ is the sum over $j$ of the products of the probability of receiving $j$ innovations in $t_e (P^j)$ and the probability that the person does not leave the employed state on each of the $j$ occasions $(1 - F(\psi))^j$. Thus

$$
\begin{aligned}
P(T_e &> t_e) = \sum_{j=0}^{t_e} \binom{t_e}{j} P^j (1-P)^{t_e - j} (1 - F(\psi))^j \\
&= (1 - P(F(\psi))^{t_e}.
\end{aligned}
$$

Thus the probability that an employment spell terminates at $t_e$ is

$$\begin{aligned} P(T_e = t_e) &= P(T_e > t_e - 1) - P(T_e > t_e) \\ &= (1 - PF(\psi))^{t_e - 1}(P(F(\psi)). \end{aligned} \tag{57}$$

THE UNIVERSITY OF CHICAGO

By similar reasoning it can be shown that the probability that a non-employment spell terminates in $t_n$ periods is

$$P(T_n = t_n) = [(1 - P(1 - F(\psi))]^{t_n - 1} P(1 - F(\psi)). \tag{58}$$

From single spell, can only estimate $P(F(\psi)$.

In place of the Bernoulli assumption for the arrival of fresh values of $\varepsilon$, suppose instead that a Poisson process governs the arrival of shocks. As is well known (see, *e.g.* Feller (1970)) the Poisson distribution is the limit of a Bernoulli trial process in which the probability of success in each interval $\eta = \frac{\Delta}{n}, P_\eta$ , goes to zero in such a way that $\lim_{n \to 0} nP_\eta \to \lambda \neq 0$.

For a time homogeneous environment the probability of receiving $j$ offers time period $t_e$ is

$$P(j \mid t_e) = \exp\left(-\lambda t_e\right) \frac{(\lambda t_e)^j}{j!}. \tag{59}$$

Thus for the continuous time model the probability that a person who begins employment at $a = a_1$ will stay in the employed state at least $t_e$ periods is, by reasoning analogous to that use to derive (50),

$$\Pr(T_e > t_e) = \sum_{j=0}^{\infty} \exp\left(-\lambda t_e\right) \frac{(\lambda t_e)^j}{j!} (1 - F(\psi))^j = \exp\left[-\lambda F(\psi) t_e\right] \tag{60}$$

so the density of spell lengths is

$$g(t_e) = \lambda F(\psi) \exp\left[-\lambda F(\psi) t_e\right].$$

A more direct way to derive this notes that from the definition of a Poisson process, the probability of receiving a new value of $\varepsilon$ in interval $(a, a+\Delta)$ is

$$p = \lambda\Delta + o(\Delta)$$

where $\lim_{\Delta\to 0} \frac{o(\Delta)}{\Delta} \to 0$, the probability of exiting the employment state conditional on an arrival of $\varepsilon$ is $F(\psi)$). Hence the exit rate or hazard rate from the employment state is

$$h_e = \lim_{\Delta\to 0} \frac{\lambda\Delta F(\psi)}{\Delta} + o(\Delta) = \lambda F(\psi).$$

Using (4) relating the hazard function and the survivor function we conclude that

$$\Pr(T_e > t_e) = e^{-\int_0^{t_e} h_e(u)\, du} = e^{-\lambda F(\psi) t_e}.$$

By similar reasoning, the probability that a person starting in the nonemployed state will stay on in that state for at least duration $t_n$ is

$$\Pr(T_n > t_n \mid \lambda) = e^{-\lambda(1-F(\psi))t_n}.$$

Analogous to the identification result already presented for the discrete time model, it is impossible using single spell employment or nonemployment data to separate $\lambda$ from $F(\psi)$ or $1 - F(\psi)$ respectively. However, access to data on both employment and nonemployment spells make it possible to identify both $\lambda$ and $F(\psi)$.

The assumption of time homogeneity of the environment is only made to simplify the argument. Suppose that nonmarket time arrives *via* a nonhomogeneous Poisson process so that the probability of receiving one nonmarket draw in interval $(a + a + \Delta)$ is

$$p(a) = \lambda(a)\Delta + o(\Delta). \tag{61}$$

Assuming that $W$ and $Y$ remain constant, the hazard rate for exit from employment at time period $a$ for a spell that begins at $a_1$ is

$$h_e(a \mid a_1) = \lambda(a)F(\psi) \tag{62}$$

so that the survivor function for the spell is

$$P(T_e > t_e \mid a_1) = \exp\left[-F(\psi)\int_{a_1}^{a_1+t_e} \lambda(u)\,du\right]. \tag{63}$$

By similar reasoning

$$P(T_n > t_n \mid a_1) = \exp\left[-(1 - F(\psi))\int_{a_1}^{a_{1+t_n}} \lambda(u)\,du\right].$$

# Example B

# A One State Model of Search Unemployment

$V$ is the value of search. Using Bellman's optimality principle for dynamic programming [see, *e.g.* Ross (1970)]. $V$ may be decomposed into three components plus a negligible component [of order o($\Delta t$)].

$$\begin{aligned} V &= \frac{c\Delta t}{1+r\Delta t} + \frac{(1-\lambda\Delta t)}{1+r\Delta t}V \\ &\quad + \frac{\lambda\Delta t}{1+r\Delta t}E\max[w/r; V] + o(\Delta t), \\ &= 0 \text{ otherwise.} \end{aligned} \tag{64}$$

for $V > 0$.

- $\lambda$ is the rate of arrival of job offeres (externally specified).
- $r$ is discount rate, $C$ is cost of search per instant, $\Delta t$ is time interval.

Collecting terms in (64) and passing to the limit, we reach the familiar formula [Lippman and McCall (1976a)]

$$c + rV = (\lambda/r) \int_{rV}^{\infty} (w - rV) dF(w) \text{ for } V > 0. \tag{65}$$

To calculate the probability that an unemployment spell $T_u$ exceeds $t_u$, we note that the probability of receiving an offer in term interval $(a, a + \Delta)$ is

$$p = \lambda\Delta + o(\Delta) \tag{66}$$

and further note that the probability that an offer is accepted is $(1 - F(rV))$ so

$$h_u = \lambda(1 - F(rV) \tag{67}$$

and

$$P(T_u > t_u) = e^{-\lambda(1-F(rV))t}. \tag{68}$$

For discussion of the economic content of this model, see, *e.g.*, Lippman and McCall (1976) or Flinn and Heckman (1982a). Accepted wages are truncated random variables with $rV$ as the lower point of truncation. The density of accepted wages is

$$g(w \mid w > rV) = \frac{f(w)}{1 - F(rV)}, \; w \geqq rV.$$

From the assumption that wages are distributed independently of wage arrival times, the joint density of duration time $t_u$ and accepted wages $(w)$ is the product of the density of each random variable,

$$\begin{aligned} m(t_u, w) &= \{\lambda(\text{1-F(rV)}) \exp[-\lambda(1 - F(rV))]\} \frac{f(w)}{1 - F(rV)} \\ &= \lambda \exp[-\lambda(1 - F(rV))t_u] f(w), \\ w &\geqq r \end{aligned} \tag{69}$$

For simplicity we assume that a reservation wage property characterizes the optimal policy noting that for general time inhomogeneous models it need not. We denote the reservation wage at time $\tau$ as $rV(\tau)$. The probability that an individual receives a wage offer in time period $(\tau, \tau + \Delta)$ is

$$p(\tau) = \lambda(\tau)\Delta + o(\Delta). \tag{70}$$

The probability that it is accepted is $(1 - F(rV(\tau)))$. Thus the hazard rate at time $\tau$ for exit from an unemployment spell is

$$h(\tau) = \lambda(\tau)(1 - F(rV(\tau)) \tag{71}$$

so that the probability that a spell that began at $\tau_1$ lasts at least $t_u$ is

$$P(T_u > t_u \mid \tau_1) = \exp\left[-\int_{\tau_1}^{\tau_1+t_u} \lambda(z)(1 - F(rV(z)))\,dz\right]. \tag{72}$$

The associated density is

$$g(t_u \mid \tau_1) = \lambda(\tau_1+t_u)(1-F(rV(\tau_1+t_u)))\exp\left[-\int_{\tau_1}^{\tau_1+t_u} \lambda(z)(1 - F(rV($$

THE UNIVERSITY OF CHICAGO

# Example C

# A Dynamic McFadden Model

$$h(j \mid \tau) = \lambda(\tau) P_j(\tau) \tag{73}$$

so that the probability that the next purchase is item $j$ at a time $t = \tau + \tau_1$ or later is

$$P(t, j \mid \tau_1) = \exp\left[- \int_{\tau_1}^{\tau_1 + t} \lambda(u) P_j(u)\, du\right]. \tag{74}$$

The $P_j$ may be specified using one of the many discrete choice models discussed in Amemiya's survey (1981). For the McFadden random utility model with Weibull errors (1973), the $P_j$ are multinomial logit. For the Domencich-McFadden (1975) random coefficients preference model with normal coefficients the $P_j$ are specified by multivariate probit.

Following McFadden (1974), the utility associated with each of $J$ possible choices at time $\tau$ is written as

$$U(\tau) = V(s, \underset{\sim}{x}(\tau)) + \varepsilon(s, \underset{\sim}{x}(\tau))$$

where $s$ denotes vectors of measured attributes of individuals, $\underset{\sim}{x}(\tau)$ represents vectors of attributes of choices, $V$ is a nonstochastic and $\varepsilon(s, \underset{\sim}{x}(\tau))$ are iid Weibull, *i.e.*,

$$P(\varepsilon(s, \underset{\sim}{x}_j(\tau)) \leq \varphi) = e^{-e^{-\varphi}}.$$

Then as demonstrated by McFadden (p. 110),

$$P_j(s, \underset{\sim}{x}_j(\tau)) = \frac{e^{V(s, \underset{\sim}{x}_j(\tau))}}{\sum_{\ell=1}^{J} e^{V(s, \underset{\sim}{x}_\ell(\tau))}}.$$

Adopting a linear specification for $V$ we write

$$V(s, \underset{\sim}{x}(\tau)) = \underset{\sim}{x}'(\tau)\beta(s)$$

so

$$P_j(s, \underset{\sim}{x}_j(\tau)) = \frac{e^{\underset{\sim}{x}'_j(\tau)\beta(s)}}{\sum_{\ell=1}^{J} e^{\underset{\sim}{x}'_\ell(\tau)\beta(s)}}.$$

## New Issues That Arises in Formulating and Estimating Choice Theoretic Duration Models

1. Without data on accepted wages, the models of previous sections are underidentified even if there are no regressors or unobservables in the model.
2. Even with data on accepted wages, the model is not identified unless the distribution of wage offers satisfies a recoverability condition to be defined below.
3. For models without unobserved variables, the asymptotic estimator of the model is non-standard.
4. Allowing for individuals to differ in observed and unobserved variables injects an element of arbitrariness into model specification, creates new identification and computational problems, and virtually guarantees that the hazard is not of the proportional hazards functional form.
5. A new feature of duration models with unobservables produced by optimizing theory is that the support of $\theta$ now depends on parameters of the model.

# Point A

From a random sample of durations of unemployment spells in a model without observed or unobserved explanatory variables, it is possible to estimate $h_u$ via maximum likelihood or Kaplan-Meir procedures (see, *e.g* Kalbfleisch and Prentice, 1980), pp. 10-16). It is obviously not possible using such data alone to separate $\lambda$ from $(1 - F(rV))$ much less to estimate the reservation wage $rV$.

# Point B

Given access to data on accepted wage offers it is possible to estimate the reservation wage $rV$. A strongly consistent estimator of $rV$ is the minimum of the accepted wages observed in the sample

$$\widehat{rV} = \min\{W_i\}_{i=1}^{I}. \tag{75}$$

For proof see Flinn and Heckman (1982a).

Access to accepted wages does not secure identification of $F$. Only the truncated wage offer distribution can be estimated

$$F(w \mid w \geq rV) = \frac{F(w) - F(rV)}{1 - F(rV)}, \qquad w \geq rV.$$

To recover an untruncated distribution from a truncated distribution with a known point of truncation requires further conditions. If $F$ is normal, such recovery is possible. If it is Pareto, it is not. A sufficient condition that

## Point C

Using density (63) in a maximum likelihood procedure creates a non-standard statistical problem. The range of random variable $W$ depends on a parameter of the model ($W \geq rV$). For a model without observed or unobserved explanatory variables, the maximum likelihood estimator of $rV$ is in fact the order statistic estimator (30). The likelihood based on (63) is monotonically increasing in $rV$, so that imposing the restriction that $W \geq rV$ is essential in securing maximum likelihood estimates of the model.

Assuming that the density of $W$ is such that $f(rV) \neq 0$, the consistent maximum likelihood estimator of the remaining parameters of the model can be obtained by inserting $\widehat{rV}$ in place of $rV$ everywhere in (63) and the *sampling distribution of this estimator is the same whether or not rV is known a priori or estimated*. For a proof, see Flinn and Heckman (1982a). In a model with observed explanatory variables but without unobserved explanatory variables, a similar phenomenon occurs. However, at the time of this writing, a rigorous asymptotic distribution theory is only available for models with discrete valued regressor variables which assumes a finite number of values.

i. Economic theory provides no guidance on the functional form of the $c, r, \lambda$ and $F$ functions (other than the restriction given by (59)). Estimates secured from these models are very sensitive to the choice of these functional forms. Model identification is difficult to check and is very functional form dependent.
ii. In order to impose the restrictions produced by economic theory to secure estimates, it is necessary to solve nonlinear equation (59).
iii. Because of the restrictions like (59), proportional hazard specifications are rarely produced by economic models

# Point D

In the search model without observed variables, the restriction that $W \geq rV$ is an essential piece of identifying information. In a model with unobservable $\Theta$ introduced in $c, r, \lambda$ or $F, rV = rV(\theta)$ as a consequence of functional restriction (59). In this model, the restriction that $W \geq rV$ is replaced with an implicit equation restriction on the support of $\Theta$ $i.e.$ for an observation with accepted wage $W$ and reservation wage $rV(\theta)$, the admissible support set for $\Theta$ is

$$\{\Theta : 0 \leq rV(\theta) \leq w\}.$$

# Pitfalls In Using Regression Methods To Analyze Duration Data

To focus on essential ideals, 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

$$f(t \mid Z).$$

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

$$E(t \mid Z) = \int_0^\infty t f(t \mid Z) dt = g(Z). \tag{76}$$

From a regression analysis, we seek to estimate the parameters of $g(Z)$. For example, if

$$f(t \mid Z) = \theta(Z) \exp[-\theta(Z)t], \theta(Z) > 0$$

$$E(t \mid Z) = \frac{1}{\theta(Z)}. \tag{77}$$

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

$$E(t \mid Z) = \beta Z. \tag{78}$$

Under ideal conditions, a regression of $t$ on $Z$ will estimate $\beta$. 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 $E(t \mid Z)$.

But in our panel sample, the expected value of the length of the first spell is *not* $E(t \mid Z)$ but is rather

$$E(t \mid Z, T) = \int_0^T tf(t \mid Z)dt + T\int_T^\infty f(t \mid Z)dt \leq E(t \mid Z). \quad (79)$$

Thus, in the exponential example

$$E(t \mid Z, T) = \beta Z\{1 - \exp(-T/\beta Z)\}. \quad (80)$$

Clearly, at least squares regression of $t$ on $Z$ *will not* estimate $\beta$. 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

$$E(t \mid Z, T, t < T) = \frac{\displaystyle\int_0^T tf(t \mid Z)dt}{\displaystyle\int_0^T f(t \mid Z)dt}. \tag{81}$$

In our exponential example

$$E(t \mid Z, T, t < T) = \beta Z \left\{ 1 - \frac{(T/\beta Z)\exp(-T/\beta Z)}{1 - \exp(-T/\beta Z)} \right\}. \qquad (82)$$

Again, a simple least squares regression of $t$ on $Z$ does not estimate $\beta$ for this sample. As $T \to \infty$, the bias disappears. (Note that the model could be consistently estimated by nonlinear least squares.)

Clearly there is no 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

$$E(t_2 \mid Z, T, t_1 + t_2 < T) = \frac{\int_0^T \int_0^{T-t_2} t_2 f(t_2 \mid Z) f(t_1 \mid Z) dt_1 dt_2}{\int_0^T \int_0^{T-t_2} f(t_2 \mid Z) f(t_1 \mid Z) dt_1 dt_2}. \tag{83}$$

Because $t_1$ and $t_2$ are conditionally independent, and hence the subscripts 1 and 2 can be interchanged without affecting the validity of the expression, this is also the conditional expectation of the length of the first spell. For a sample of individuals with at least two completed spells of the event

$$\begin{aligned} & \mathrm{E}(\mathrm{t}_2 \,|\mathrm{Z},\mathrm{T},\mathrm{t}_1{+}\mathrm{t}_2 < \mathrm{T}) \\ = \;& \beta \mathrm{Z}\left[\frac{1\text{-} \exp(\text{-T}/\beta Z)(1{+}T/\beta \mathrm{Z})\text{-}(\beta \mathrm{Z}/2)\exp(\text{-T}/\beta \mathrm{Z})}{1-\exp(-T/\beta Z)-(T/\beta Z)\exp(-T/\beta Z)}\right]. \end{aligned} \tag{84}$$

Note further that

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

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 (72). Least squares estimators are critically dependent on both the sample selection rule and the length of the panel.

A common functional form for the hazard function, $h(\cdot)$, 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

$$\text{First spell: } h[t_1, Z(t_1), V] \exp\left\{-\int_0^{t_1} h[u, Z(u), V] du\right\}.$$

The density for the duration time in the second spell $t_2$ *given that the first spell ends at calendar time* $t_1$ is

**Conditional second spell density:**

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

The marginal second spell density is obtained by integrating out $t_1$. Thus $f^*(t_2, Z, V) =$

$$\int\limits_0^\infty \left( h[t_2, Z(t_2 + t_1), V] \exp\left\{ - \int_0^{t_2} h[u, Z(u + t_1), V] du \right\} \right) \left( h[t_1, Z(t_1), V] \exp\left\{ - \int_0^{t_1} h[u, Z(u), V\right.\right.$$

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

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

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 the 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

$$f(t_1 \mid Z, V) = \theta(Z, V) \exp[-\theta(Z, V) t_1] \qquad 0 < t_1 < \infty$$

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

$$E(t_1 \mid Z, V) = \frac{1}{\theta(Z, V)} = \beta Z + V. \tag{85}$$

Suppose we consider another individual who is subject to a different value of $Z$ before and after calendar time $\tau_1$. The density of $t_1$ for this person is derived most simply from the conditional density before and after $\tau_1$, *etc*.

$$f(t_1 \mid Z, V, t_1 < \tau_1) = \frac{\theta(Z_1, V)\exp[-\theta(Z_1, V)t_1]}{1 - \exp[-\theta(Z_1, V)\tau_1]} \quad 0 < t_1 < \tau_1$$

and

$$f(t_1 \mid Z, V, t_1 > \tau_1) = \frac{\theta(Z_2, V)\exp[-\theta(Z_2, V)t_1]}{\exp[-\theta(Z_2, V)\tau_1]} \quad t_1 \geq \tau_1.$$

The conditional expectation of duration in the first spell is

$$E(t_1 \mid Z_1, Z_2, \tau_1, V) \tag{86}$$
$$= \frac{1}{\theta(Z_1, V)} + \exp[-\theta(Z_1, V)\tau_1]\left(\frac{1}{\theta(Z_2, V)} - \frac{1}{\theta(Z_1, V)}\right).$$

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 *after the end of the first spell*, is, for a case of no time varying variables

$$E(t_2 \mid Z, V) = 1/\theta(Z, V). \tag{87}$$

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

$$\begin{aligned} E(t_2 \quad &| \quad Z_1, Z_2, t_1 < \tau_1) = \frac{1}{\theta(Z_1, V)} \\ & + \exp[-\theta(Z_1, V)(\tau_1 - t_1)] \left[ \frac{1}{\theta(Z_2, V)} - \frac{1}{\theta(Z_1, V)} \right] \quad (88) \\ t_1 \quad &\leq \quad \tau_1. \end{aligned}$$

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

$$E(t_2 \mid Z_1, t_1, V, t_1 > \tau_1) = \frac{1}{\theta(Z_2, V)} \quad t_1 > \tau_1. \tag{89}$$

Although equations (83) and (85) are of the same functional form, there is one important difference: in equation (85) $t_1$ 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 $t_t < \tau_1$ raises further problems.

By Bayes theorem, the conditional mean of $V$ in equations (85) and (86) depends on $t_1$ and the explanatory variables so that the error term (inclusive of $V$) associated with regression specifications for equations (85) or (86) 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.

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 (68) is the density. In the presence of nonstationary explanatory variables, the two distributions differ.

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

$$E(t_2 \mid Z_1, Z_2, \tau_1, V) \\ = \frac{1}{\theta(Z_1, V)} + \exp\left[-\theta(Z_1, V)\tau_1\right] \left[\frac{\theta(Z_1, V) - \theta(Z_2, V)}{\theta(Z_2, V)}\right] \left[\frac{1}{\theta(Z_1, V)} + \right. \tag{90}$$

Table 1: Ln Employment Durations (Based on Two Complete

| | Within spell averages of exogenous variables | | | | | Start exogenou |
|---|---|---|---|---|---|---|
| | Spell one | t | Spell two | t | Spell one | t |
| Intercept | 1.232 | (2.16) | -1.052 | (1.25) | 7.094 | (2.1) |
| Marital status (1 if married) | .624 | (1.43) | -.449 | (.72) | -.409 | (.81) |
| National unemployment | -2.020 | (5.40) | -.975 | (.21) | -6.523 | (2.5) |
| | | | **Difference specifications** | | | |
| Intercept | | 2.502 | (1.95) | | | 12.423 |
| Δ Marital status | | -.138 | (.17) | | | 0.112 |
| Δ Unemployment | | 185 | (.56) | | | .858 |
| Marital status (first spell) | | -.570 | (.49) | | | -.496 |
| National | | | | | | |

Table 2: Ln Employment Durations (Based on Two Completed

| | Within spell averages of exogenous variables | | | | Start o exogenous | |
|---|---|---|---|---|---|---|
| | Spell one | t | Spell two | t | Spell one | t |
| Intercept | -.398 | (.97) | .44 | (.65) | -1.051 | (2.28) |
| Marital status (1 if married, spouse present) | -.117 | (.24) | .16 | (.44) | -.375 | (.764) |
| National unemployment | -.335 | (1.46) | -.53 | (1.74) | .057 | (.21) |
| | | | **Difference specifications** | | | |
| Intercept | | -.074 | (0.84) | | | 12.423 |
| Δ Marital status | | -.342 | (.645) | | | 0.112 |
| Δ Unemployment | | -.402 | (.85) | | | .858 |
| Marital status (first spell) | | [illegible] | [illegible] | | | [illegible] |

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

| | Employment to Nonemployment | Nonemployment to Employm |
|---|---|---|
| Panel A: Regressors Fixed at Average Value Over Spell | | |
| Intercept | .971 | -.093 |
| | (1.535) | (.221) |
| ln Duration ($\gamma$) | -.137 | -.287 |
| | (1.571) | (2.976) |
| MSP | -1.093 | .347 |
| | (2.679) | (1.134) |
| Unemployment | -1.800 | -.577 |
| | (6.286) | (3.119) |
| $\$^2 = -711.457$ | | |

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

| | Employment to Nonemployment | Nonemployment to Employm |
|---|---|---|
| | Panel B: Regressors Fixed at Value for First Month of Spell | |
| Intercept | -3.743 | -1.054 |
| | (12.074) | (3.464) |
| ln Duration ($\gamma$) | -.230 | -.363 |
| | (2.888) | (4.049) |
| MSP | -.921 | .297 |
| | (2.310) | (.902) |
| Unemployment | .569 | -.130 |
| | (3.951) | (.900) |
| $\$ = -740.998$ | | |

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

| | Employment to Nonemployment | Nonemployment to Employm |
|---|---|---|
| | Panel C: Regressors Free to Vary Over the Spell | |
| Intercept | -3.078 | -.899 |
| | (8.670) | (2.742) |
| ln Duration ($\gamma$) | -.341 | -.316 |
| | (3.941) | (3.279) |
| MSP | -.610 | .362 |
| | (1.971) | (1.131) |
| Unemployment | .209 | -.204 |
| | (1.194) | (1.321) |
| $\mathcal{L} = -746.515$ | | |

## Table 4. Maximum Likelihood Estimates with Time Varying Variables and Heterogeneity[1]

| | Employment to Nonemployment | Nonemployment to Em |
|---|---|---|
| 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 | |

common in many economic data sets.

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

| | Without heterogeneity | | With heterogeneity | |
|---|---|---|---|---|
| | $E \rightarrow N$ | $N \rightarrow E$ | $E \rightarrow N$ | $N \rightarrow E$ |
| Const. | -3.271 | -.762 | -3.565 | -.748 |
| | (8.901) | (2.425) | (8.537) | (2.247) |
| Tenure/10 | -.806 | -1.714 | .045 | -1.704 |
| | (1.858) | (2.731) | (.085) | (2.685) |
| $Tenure^2/100$ | .028 | .602 | -.120 | .603 |
| | (.145) | (1.673) | (.607) | (1.666) |
| MSP | -.568 | .349 | -.490 | .313 |
| | (1.731) | (1.089) | (1.353) | (.956) |