Duration Models
Introduction to Single Spell Models

James J. Heckman
University of Chicago

Econ 312, Spring 2023

- The hazard function gives the probability that a spell, denoted by the nonnegative random variable T with distribution $g(t)$, will end at t, given that it has lasted until t:

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

- Integrated hazard function (using $G(0) = 0$ to eliminate c):

$$H(t) = \int_0^t h(u)du = -\ln(1 - G(t))|_0^t + c = -\ln(1 - G(t)).$$

- Working backwards, we can derive g from h:

$$\begin{aligned} G(t) &= 1 - e^{-\int_0^t h(u)du} = 1 - e^{-H(t)}, \\ g(t) &= h(t)[1 - G(t)] = h(t)e^{-H(t)}, \\ H(t) &= \int_0^t h(u)\, du. \end{aligned}$$

- So the survival function, the probability that the spell lasts until t, i.e., $T \geq t$, is

$$S(t) = 1 - G(t) = e^{-H(t)}.$$

- The density and hazard function for T may have a number of qualities. If T has a nondefective duration density, then

$$\lim_{t \to \infty} \int_0^t h(u)du \to \infty \iff S(\infty) = 0$$

- Duration dependence arises when $\dfrac{\partial h(t)}{\partial t} \neq 0.$
- If $\dfrac{\partial h(t)}{\partial t} > 0$ (< 0), then we have positive (negative) duration dependence

- In constructing estimable models, we will often work with the conditional hazard function

$$h(t|x(t), \theta(t)),$$

where the regressor vector $x(t)$ may include
    - Entire past: $x_1(t) = \int_{\infty}^{t} k_1(z_1(u))du$
    - or future: $x_2(t) = \int_{t}^{\infty} k_2(z_2(u))du$
    - or both: $x_3(t) = \int_{-\infty}^{\infty} k_3(z_3(u), t)du$

of some variables.

- Associated with the conditional hazard function is the conditional survival function

$$S(t|x(t),\theta(t)) = 1 - G(t|x(t),\theta(t)) = e^{-\int_0^t h(u|x(u),\theta(u))du}$$

and the conditional density of T

$$\begin{aligned} g(t|x(t),\theta(t)) &= h(t|x(t),\theta(t)) \cdot [1 - G(t|x(t),\theta(t))] \\ &= h(t|x(t),\theta(t)) \cdot e^{-\int_0^t h(u|x(u),\theta(u))du}. \end{aligned}$$

- In these models we will assume
  1. $\theta(t)$ independent of $x(t)$and $\theta \sim \mu(\theta)$, $x \sim D(x)$
  2. No functional restrictions connecting the conditional distribution of $T|\theta, x$ and the marginal distribution of θ, x.

- A common specification of the conditional hazard is the proportional hazard specification:

$$h(t|x(t), \theta(t)) = \psi(t)\phi(x(t))\eta(\theta(t))$$

$$\ln h(t|x(t), \theta(t)) = \ln \psi(t) + \ln \phi(x(t)) + \ln \eta(\theta(t))$$
$$\psi(t) \geq 0, \quad \phi(x(t)) > 0, \quad \eta(\theta(t)) \geq 0 \quad \forall t.$$

Sampling Plans and Initial Condition Problems

- For interrupted spells, one of the following duration times may be observed:
    - time in state up to sampling date (T_b)
    - time in state after sampling date (T_a)
    - total time in completed spell observed at origin of sample ($T_c = T_a + T_b$)
- Duration of spells beginning after the origin date of the sample, denoted T_d, are not subject to initial condition problems.
- The intake rate, $k(-t_b)$, is the proportion of the population entering a spell at $-t_b$.

- Assume
  - a time homogenous environment, i.e. constant intake rate, $k(-t_b) = k, \forall b$
  - a model without observed or unobserved explanatory variables.
  - no right censoring, so $T_c = T_a + T_b$
  - underlying distribution is nondefective
  - $m = \int_0^{\infty} x g(x) dx < \infty$

- The proportion of the population experiencing a spell at $t = 0$, the origin date of the sample, is

$$\begin{aligned} P_0 &= \int_0^{\infty} k(-t_b)(1 - G(t_b))dt_b = k\int_0^{\infty}(1 - G(t_b))dt_b \\ &= k\left[t_b(1 - G(t_b))\big|_0^{\infty} - \int t_b d(1 - G(t_b))\right] \\ &= k\int t_b g(t_b)dt_b = km, \end{aligned}$$

where $1 - G(t_b)$ is the probability the spell lasts from $-t_b$ to 0 (or equivalently, from 0 to $-t_b$).

- So the density of a spell of length t_b interrupted at the beginning of the sample $(t = 0)$ is

$$
\begin{aligned}
f(t_b) &= \frac{\text{proportion surviving til } t = 0 \text{ from batch } t_b}{\text{total surviving til } t = 0} \\
&= \frac{k(-t_b)(1 - G(t_b))}{P_0} = \frac{1 - G(t_b)}{m} \neq g(t_b)
\end{aligned}
$$

- The probability that a spell lasts until t_c given that it has lasted from $-t_b$ to 0, is

$$g(t_c|t_b) = \frac{g(t_c)}{1 - G(t_b)}$$

- So the density of a spell that lasts for t_c is

$$\begin{aligned} f(t_c) &= \int_0^{t_c} g(t_c|t_b) f(t_b) dt_b \\ &= \int_0^{t_c} \frac{g(t_c)}{m} dt_b = \frac{g(t_c)t_c}{m} \end{aligned}$$

- Likewise, the density of a spell that lasts until t_a is

$$\begin{aligned} f(t_a) &= \int_0^\infty g(t_a + t_b | t_b) f(t_b) dt_b \\ &= \int_0^\infty \frac{g(t_a + t_b)}{m} dt_b \\ &= \frac{1}{m} \int_{t_a}^\infty g(t_b) dt_b \\ &= \frac{1 - G(t_a)}{m} \end{aligned}$$

- So the functional form of $f(t_b) \approx f(t_a)$.

- Some useful results that follow from this model:
  1. If $g(t) = \theta e^{-t\theta}$, then $f(t_b) = \theta e^{-t_b\theta}$ and $f(t_a) = \theta e^{-t_a\theta}$.
     Proof:
     $$g(t) = \theta e^{-t\theta} \to m = \frac{1}{\theta},$$
     $$G(t) = 1 - e^{-t\theta} \to f(t_a) = \frac{1 - G(t)}{m} = \theta e^{-t\theta}$$
  2. $E(T_a) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$.

- **Proof**:

$$
\begin{aligned}
E(T_a) &= \int t_a f(t_a) dt_a = \int t_a \frac{1 - G(t_a)}{m} dt_a \\
&= \frac{1}{m} \left[\frac{1}{2} t_a^2 (1 - G(t_a))|_0^\infty - \int \frac{1}{2} t_a^2 d(1 - G(t_a)) \right] \\
&= \frac{1}{m} \int \frac{1}{2} t_a^2 g(t_a) dt_a = \frac{1}{2m} [var(t_a) + E^2(t_a)] \\
&= \frac{1}{2m} [\sigma^2 + m^2]
\end{aligned}
$$

- $E(T_b) = \frac{m}{2}(1 + \frac{\sigma^2}{m^2})$.
- **Proof**: See proof of Proposition 2.
- $E(T_c) = m(1 + \frac{\sigma^2}{m^2})$.
- **Proof**:

$$E(T_c) = \int \frac{t_c^2 g(t_c)}{m} dt_c = \frac{1}{m}(var(t_c) + E^2(t_c))$$

$$\rightarrow E(T_c) = 2E(T_a) = 2E(T_b), E(T_c) > m \text{ unless } \sigma^2 = 0$$

- $h'(t) > 0 \to E(T_a) = E(T_b) > m$.
- **Proof:** See Barlow and Proschan.
- $h'(t) < 0 \to E(T_a) = E(T_b) < m$.
- **Proof**: See Barlow and Proschan.

Pitfalls in Using Regression Methods to Analyze Duration Data

1. Density of duration in a spell (T) for an individual with fixed characteristics Z is $f(t|Z)$.
2. Assume
   1. No time elapses between end of one spell and beginning of another,
   2. No unobserved heterogeneity components,
   3. $f(t|Z) = \theta(Z)e^{-\theta(Z)t}, \theta(Z) = \frac{1}{\beta Z} > 0,$
   4. At origin, $t = 0$, of sample of length K, everyone begins a spell.
3. The expected length of spell in the population given Z is

$$E(T|Z) = \int_0^{\infty} tf(t|Z)dt = \frac{1}{\theta(Z)} = \beta Z.$$

1. The expected length of a spell in a sample frame of length K, however, is

$$\begin{aligned}
E(T|Z,K) &= \int_0^K tf(t|Z)dt + K\int_K^\infty f(t|Z)dt \\
&= \int_0^K t\theta e^{-\theta t}dt + K\int_K^\infty \theta e^{-\theta t}dt \\
&= \left[-te^{-\theta t}|_0^K + \int_0^K e^{-\theta t}dt\right] + K\left[\int_K^\infty \theta e^{-t\theta}dt\right] \\
&= \left[-Ke^{-\theta K} + \left(-\frac{1}{\theta}e^{-\theta t}\right)\Bigg|_0^K\right] + K\left[-e^{-\theta t}|_K^\infty\right] \\
&= -Ke^{-\theta K} - \frac{1}{\theta}e^{-\theta K} + \frac{1}{\theta} + Ke^{-\theta K} \\
&= \beta Z(1-e^{-\frac{K}{\beta Z}}) \neq \beta Z.
\end{aligned}$$

- So OLS of T on Z will not estimate β. But as $K \to \infty$, the selection bias term $(\beta Z e^{-\frac{K}{\beta Z}})$ disappears.
- A widely used method to avoid this bias is to use only completed first spells.
- This results in another sort of selection bias,

$$\begin{aligned} E\left(T \mid Z, K, T < K\right) &= \frac{\int_0^K tf(t|Z)dt}{\int_0^K f(t|Z)dt} \\ &= \frac{-Ke^{-\theta K} - \frac{1}{\theta}e^{-\theta K} + \frac{1}{\theta}}{1 - e^{-\theta K}}, \end{aligned}$$

where recall that

$$\beta Z = \frac{1}{\theta}.$$

- As $K \to \infty$,

$$E\left(T \mid Z, K, T < K\right) = \beta Z.$$

THE UNIVERSITY OF CHICAGO