Econometrics 6027
Lecture 6
Duration Models Continued

1 Incorporating Covariates

In reality, duration may depend on observed characteristics. It seems perfectly reasonable, for example, that the duration of unemployment will depend on human capital and skill characteristics. A model which does not admit this as a possibility is likely to be misspecified to some degree. We thus introduce covariates into our model to control for other factors that can affect duration.

There are two kinds of covariates. The first class of covariates are termed *time-invariant covariates,* and are so named because the values do not depend on the length of duration in a state. So, for time-invariant characteristics x_i, the value of t does not influence the value of the covariate - one could describe these covariates as exogenous to the duration process. In a study of unemployment, one might include the gender of the individual, or the level of school qualifications. Each are clearly fixed in relation to the period of unemployment.

On the other hand, *time-varying covariates* are much harder to deal with. A time-varying covariate x_{it} is one for which the level of the covariate depends on the duration in the state on question. Examples might include socio-demographic characteristics, or the human capital/skills base of the subject, all of which may vary over the period of unemployment.

1.1 The proportional hazard specification

A common mechanism by which to introduce covariates into duration and survival analysis is through the so-called proportional hazards specification, which adjusts the *conventional hazard specification* according to the following rule:

$$h(t) = g(x)h_0(t) > 0 \text{ (***)}$$

for some parametric baseline hazard $h_0(t)$ and covariate function $g(x)$, where x represents a vector of covariates thought to influence the duration in a state and the instantaneous exit rate. Typically, $g(x) = \exp(x'\beta)$ is used to make sure the predicted duration is positive. Then, $h(t) = e^{x\beta}h_0(t)$.

Based on (***), we can define the survival function in the normal fashion as

$$\begin{aligned} S(t) &= \exp[-\int_{s=0}^{t} \exp(x'\beta)h_0(s)ds] \\ &= \exp[-\exp(x'\beta)\int_{s=0}^{t} h_0(s)ds] \\ &= \exp[-\exp(x'\beta)\Lambda_0(t)] \end{aligned}$$

with the standard formulation for the likelihood.

1.2 The Accelerated hazard function

If the proportional hazard does not depend on duration, we have the accelerated hazard function. There is no duration dependence in this model. The general form for the accelerated hazard function introduces covariates through the h_0 multiplier in the parametric hazard function defined earlier. Typically,

$$h_0 = \exp(x'\beta).$$

Given how the baseline hazard is defined, the survival function must be:

$$\begin{aligned} S(t) &= e^{\int_0^t e^{x'\beta} ds} \\ &= e^{e^{x'\beta}} t \\ &= exp(exp(x'\beta))t \end{aligned}$$

In which case our PDF must be:

$$\begin{aligned} f(t) &= h(t) * S(t) \\ &= e^{x'\beta} e^{e^{x'\beta}} t \end{aligned}$$

It can be shown that the expected duration for such a model is:

$$\begin{aligned} E(t) &= \frac{1}{\exp(x'\beta)} \\ &= \int_0^\infty sf(s)ds \end{aligned}$$

We can transform this expectation, and estimate this transformed model using OLS. Specifically, if we take the log of the accelerated failure we obtain $E(ln(t)) = -X'\beta$. This linearizes the accelerated model and now we can just run a simple OLS regression, ie. on $\log(t) = -(x'\beta) + v$ where v is a stochastic term.

1.2.1 Accelerated Weibull hazard

The general form for the accelerated Weibull would be as above,

$$\begin{aligned} h(t) &= h_0\lambda(h_0 t)^{\lambda-1} \\ &= \exp(x'\beta)\lambda(\exp(x'\beta)t)^{\lambda-1} \end{aligned}$$

Estimation of the accelerated Weibull follows as before, substituting $h_0 = \exp(x'\beta)$ into the standard likelihood expression. For monotone hazard functions such as the Weibull, the signs of the coefficients on β will therefore show the effects of covariates on the exit rate. For non-monotonic functions, however, interpretation is less straightforward.

2 Unobserved Heterogeneity

If there are individual-specific effects that affect duration, we say that there is unobserved heterogeneity in the duration data. If this is the case, we should include a γ_i term in our model which is different for each individual, but doesn't change over time.

This might be relevant, for example, in research on programs that try to reduce the duration of unemployment. Since participation in the program (the treatment) is a matter of choice, there are selection issues, even if randomization was used in allocating individuals to the opportunity to choose treatment. Selection bias could arise if those individuals who believe they would get the most benefit from the program are disproportionately the ones that choose to take advantage of the program. If we include an individual-specific term, we address these selection issues by controlling for unobserved heterogeneity between treatment and control groups.

2.1 The Mixed Proportional Hazard Model

Mixed Proportional Hazard models incorporate an individual-specific effect γ_i:

$$h(t) = \gamma_i g(x_i) h_0(t_i),$$

where γ_i is unobserved, t_i is the spell of eg. employment for individual i, $g(x_i) = e^{x_i'\beta}$ and $h_0(t_i)$ is the baseline hazard for individual i.

Here, an individual has a given value of γ_i, and his or her spell durations are independent drawings from the univariate duration distribution $F(t_i|x_i;\gamma_i)$, where γ_i is unobserved, so that the durations given only x_i are not independent. Conditional on x_i and γ_i, the durations t_i are independent. Given independence, the integrated hazard for the spell of employment of a given individual is:

$$\begin{aligned} \Lambda_i &= \gamma_i g(x_i) \Lambda_0(t_i), \text{ with} \\ \Lambda_0(t_i) &= \int_0^t h(u)\, du. \end{aligned}$$

There are three options for estimation. The M-transformation integrates out the heterogeneity, assuming that γ_i and x_i are independent; the Gamma method uses a parametric assumption; and the mass-points method uses a nonparametric method.

2.1.1 Estimating the Mixed Proportional Hazard: M-transformation

We can find the PDF of the model, and thereby estimate, using the fact that $h(t) = \frac{f(t)}{S(t)}$:

$$\begin{aligned} h\left(t_i|\gamma_i, x_i\right) &= \frac{f\left(t_i|\gamma_i, x_i\right)}{1 - F\left(t_i|\gamma_i, x_i\right)} \\ f\left(t_i|\gamma_i, x_i\right) &= h\left(t_i|\gamma_i, x_i\right)\left(1 - F\left(t_i|\gamma_i, x_i\right)\right) \\ &\text{where} \\ 1 - F\left(t_i|\gamma_i, x_i\right) &= e^{-\gamma_i g(x_i)\Lambda_0(t_i)} \\ &\text{so} \\ f\left(t_i|x_i\right) &= \gamma_i g(x_i) h_0(t_i) * e^{-\gamma_i g(x_i)\Lambda_0(t_i)} \\ &= \int_0^\infty f\left(t_i|v_i, x_i\right) dG\left(\gamma_i\right) \end{aligned}$$

Here, we are effectively integrating out the γ_i to obtain the the PDF of the distribution independent of the γ, implicitly assuming that γ_i are independent of x_i. For convenience, we consider $h_0\left(t_i\right)$, $g\left(x_i\right)$, γ_i, and the distribution G of γ_i in the population to satisfy some regularity assumptions.

We can construct a likelihood estimator using the density $f\left(t_i|x_i\right)$. We estimate by maximizing the logarithm of the joint likelihood with respect to the parameters of interest.

$$\left\{\widehat{h_0}, \widehat{\beta}\right\} = \arg\max_{h_0,\ \beta} L\left(h_0, \beta; t_i\right)$$

2.1.2 Estimating the Mixed Proportional Hazard: Gamma distribution

We can assume that γ_i follows a Gamma distribution. Indeed, we can simply say that the γ_i follow the normal.

2.1.3 Estimating the Mixed Proportional Hazard: Mass points

Alternatively, we can say γ follows a nonparametric specification, ie using mass points:

$$\gamma_i = \begin{cases} \gamma_{l(low)} & \text{with Prob L} \\ \gamma_{h(high)} & \text{with prob H} = 1 - PrL \end{cases}$$

This is semi-parametric. The procedure is like adding a mass (weight) that nudges the function from the parametric specification towards the actual data.

3 The Cox Mixed Proportional Hazard

This is a semi-parametric model with a parameter γ that is not a function of time, a baseline hazard h_0 that is non-parametric, covariates x_i, and change over time.

$$h(t) = \gamma e^{x'\beta} h_0(t)$$

The non-parametric baseline hazard $h_0(t)$ enables us to estimate zig-zags and other odd shapes in the duration data. By including the parametric γ, we have a regular baseline shape. If the shape of the duration data is smooth, we don't need the nonparametric term.

There are several options for estimating the Cox model, each of which use different frameworks to estimate the γ. The gamma method assumes the γ follow the Gamma distribution. The Mass points Heckman-Singer type method makes the parameter take on one of two values (a.k.a. points). The Correlated Random effects model allows the parameter to take on many different values, and allows for correlation with x.

3.1 Gamma distribution

$$G(\gamma_i) = \gamma_i^{\rho-1} \frac{e^{-\frac{\gamma_i}{\alpha}}}{\Gamma(\alpha)\rho^\alpha}$$

Where Γ is the Gamma function.[1] The support for gamma is $0 \rightarrow \infty$, as required.

$$\begin{aligned} h(t_i|\gamma_i x \beta) &= \gamma e^{x\beta} h(t) \\ &= \int_0^\infty \gamma e^{x\beta} h_0 * t * \gamma G(\gamma_i) \\ &= \int_0^\infty \gamma_0 e^{x'\beta} (h_0 t)^{\gamma-1} \frac{\gamma_i^{\rho-1} e^{-\frac{\gamma_i}{\alpha}}}{\Gamma(\alpha)\rho^\alpha} d\gamma_i \end{aligned}$$

where

- γ disappears, because you integrate on it
- assume $E(\gamma) < \infty$
- assume $\gamma > 0$
- assume no correlation between the distribution of γ and x
- $g(x) = e^{X'\beta}$
- $h_0(t) = (h_0 t)^{\gamma-1}$

[1] The Gamma function is the continuous function that connects the integer factorials. For integers the function is $\Gamma(m) = (m-1)!$, i.e, $\Gamma(3) = 1 * 2$.

3.2 Heckman-Singer Mass Points

The parameter can take on one of two values:

$$\gamma_i = \begin{cases} \gamma_l^* & \text{with probability } P_L, \\ \gamma_h^* & \text{with probability } 1 - P_L, \end{cases}$$

$$h(t_i|x_i, \beta_i, \gamma_i, \gamma_H, \gamma_L, \rho_H, h_0) = \rho_H \gamma_H e^{x\beta}(h_0 t)^{\gamma-1} + (1-\rho_H)\gamma_L e^{x\beta}(h_0 t)^{\gamma-1}$$

Here, the distribution collapses to two mass points (in addition to the distribution taken by the h_0).

3.3 Correlated Random Effects

This approach is like the mass points approach, but it allows for many points, each of which could follow their own distribution. For example,

$$\gamma_i = \begin{cases} \gamma_1^* & \text{with probability } P_1 \sim \textit{normal}, \\ \gamma_2^* & \text{with probability } P_2 \sim \textit{logistic}, \\ \gamma_3^* & \text{with probability } P_3 \sim \textit{gamma}, \\ \gamma_4^* & \text{with probability } P_4 \sim 1 - P_1 - P_2 - P_3. \end{cases}$$

This approach allows for some correlation between the γ and the x covariates. If this approach is adopted, and we allow correlation with x, estimation has to be done jointly, with simultaneous estimation of all estimators at once. With Mass points and gamma approaches, estimation can be done separately.

4 Estimation & Censoring in Duration Data

To estimate the parameters of the various models of survival and exit detailed above requires a sample of observed durations for the particular state of interest. So, for example, to model unemployment duration requires that we collect a sample of observations of time spent in unemployment from the beginning of a period of unemployment to its end.

A number of complications can arise in the collection of data which can complicate the treatment of the duration analysis.

- One may not observe a completed duration for a subset of the sample, leading to a problem of "right censoring".
- One may not observe the start of events. So, individuals may enter the period of the sample in a state of unemployment, before exiting to unemployment at some point within the period of observation (so called "left-censoring").

- The reported duration of unemployment may not be a continous event. One may, for example, observe multiple spells of unemployment for a given individual over the period of the sample.

Some of these problems are particularly difficult to accomodate in a statistical model of duration, as we now explore. In the following, assume that $\theta =$ all parameters $= \beta, \alpha, \gamma, h_0, etc.$

4.1 Estimation with no censoring

Consider first a sample of n observed durations $t_1, t_2, ..., t_n$ within a given sample period. Furthermore, suppose that all observations t_i represent a completed duration. Given a parametric hazard function $h(_i; \theta)$ based on a set of parameters θ, the general density function for the completed duration t_i is

$$f(t_i; \theta) = h(t_i; \theta) S(t_i; \theta)$$

and the likelihood function for completed durations is

$$L(\theta) = \prod_{i=1}^{n} f(t_i; \theta) = \prod_{i=1}^{n} h(t_i; \theta) S(t_i; \theta)$$

and a corresponding log-likelihood

$$\ln L^T(\theta) = \sum_{i=1}^{n} \ln f(t_i, \theta) = \sum_{i=1}^{n} \ln h(t_i; \theta) + \sum_{i=1}^{n} \ln S(t_i; \theta)$$

Maximum likelihood estimates $\widehat{\theta}$ for a particular parametric hazard rate specification then follow straightforwardly, via the optimisation of the log-likelihood function.

Parametric models require joint estimation. Semi-parametric models, such as the mass points model, require sequential estimation. When there is correlation between the parameter and x, we have to estimate jointly. It is possible to estimate semi-parametric models jointly, but this is quite computer-intensive. Most work on duration models today is done with semi or non-parametric models because parametric approaches sometime ignore the shape of the data and tell us very little.

For example, estimation of the Cox model using the Gamma distribution approach would be done jointly using a likelihood function as follows:

$$
\begin{aligned}
L_i &= l(\theta|t_i) \\
L^T &= \prod^{N} l(\theta|t_i) \\
\hat{\theta} &= \arg\max_{\theta} \ln(L^T)
\end{aligned}
$$

with complete durations

$$s(t_i|x,\theta) = e^{-\iint_{\gamma} \gamma_t X'\beta(h_0\beta)^{\gamma-1}\left(\frac{\gamma\beta e^{\frac{\gamma}{\alpha}}}{\Gamma(\alpha)\rho^{\alpha}}\right) d\gamma ds}$$

4.2 The problem of right-censoring

Say we have a sample of data about unemployment insurance, but we don't observe when unemployment ends for some individuals. That is, suppose that the sample $t_1, t_2, ..., t_n$ are right-censored:

$$t_i = \begin{cases} t_h & \text{if } t_i^* \geq h \\ t_i^* & \text{otherwise.} \end{cases}$$

To continue our example, if we have duration data on unemployment (in weeks), we might have individuals whose spell of unemployment runs from 0 to 36; from 0 to 12; 0 to 23; 0 to 36; 0 to 36; and 0 to 2. The people who are recorded at 36 might get a job in period 36, or perhaps afterwards: we only observe 36 weeks, and can't distinguish between people who exit unemployment at 36 weeks (=), on the one hand, and those that exit afterwards (>) but who are censored, on the other.

Then, we know that, for the censored observations, $T > t_i$ with probability $S(t_i)$. In the same way as for the treatment of censored observations in a Tobit analysis, we can include this piece of information in the likelihood for all observations for which the durations is incomplete. So, define the following indicators:

$$
\begin{aligned}
\delta_i &= 1 \text{ if the observed duration is completed} \\
&= 0 \text{ if the observed duration is right-censored.}
\end{aligned}
$$

Then, the likelihood for a parameteric hazard function in the presence of right-censored data becomes

$$
\begin{aligned}
L_c(\theta) &= \prod_{\delta_i=1} f(t_i;\theta) \prod_{\delta_i=0} S(t_i;\theta) \\
&= \prod_{\delta_i=1} h(t_i;\theta) S(t_i;\theta) \prod_{\delta_i=0} S(t_i;\theta)
\end{aligned}
$$

with corresponding log-likelihood

$$
\begin{aligned}
l_c(\theta) &= \sum_{\delta_i=1} \ln f(t_i;\theta) + \sum_{\delta_i=0} \ln S(t_i;\theta) \\
&= \sum_{\delta_i=1} \ln h(t_i;\theta) + \sum_{i=1}^{n} \ln S(t_i;\theta).
\end{aligned}
$$

4.3 The problem of left-censoring

Left censoring is inherently more difficult to deal with. This would apply if, for example, we don't see the beginning of unemployment, just the time when people apply for unemployment insurance.

$$
t_i = \begin{cases} t_l & \text{if } t_i^* \leq l_i \\ t_i^* & \text{otherwise.} \end{cases}
$$

With left censoring, durations t are only observed conditional on $t_i > l_i$ for some $l_i > 0$. If we knew the value of l_i for each left-censored observation, then the appropriate treatment in estimation would require that the density be conditioned on the event $t_i > l_i$. The conditional density for left-censored data would be:

$$
f(t_i|t_i > l_i;\theta) = \frac{h(t_i;\theta)S(t_i;\theta)}{S(l_i;\theta)}
$$

leading to a log-likelihood (in the absence of right-censoring) of the form

$$
\ln L_{lc}^{T}(\theta) = \sum_{i=1}^{n} \ln f(t_i;\theta) + \sum_{i=1}^{n} [\ln S(t_i;\theta) - \ln S(l_i;\theta)]
$$

Where the last term takes out the contribution of the censored to the likelihood. Of course, this likelihood can only be evaluated if l_i is known for each observed duration, a circumstance which is unlikely in most cases.