LECTURE 11
DURATION MODELS[1]

Duration models allow us to understand how long someone may stay in a particular state (e.g. unemployment), what factors about them affect how long they will stay, and whether the probability of staying increases, decreases or remains constant with their duration. There are two key dimensions of the analysis: number of states and number of spells. All applications must have at least two states, but the duration in one of those states may not be of interest. For example, some researchers have studied women's durations between births. The state of interest here is how long women wait until the next birth, but the complementary state will usually last a fixed time period and would therefore not be properly estimated by a duration analysis. The number of spells is something that, while dependent on the application itself, often depends on the availability of data.

A single spell model is what we would think of as a cross-section study, where we only have a single observation for each individual. This is typically the result of a question to unemployed people, such as "How long have you been unemployed?" This differs from a study in which individuals are observed over some fixed time period. In this time period we may find some individuals remaining in a single state the entire period, while other individuals may move in and out of various states. The resulting data will be a single spell within a single state for the former, while the latter individuals will provide observations of multiple durations.

1. Single Spell Duration Models

The duration of an episode (e.g. of unemployment) is represented by the variable Y, assumed in our analysis to be a continuous positive variable. The density and distribution functions of the duration Y are denoted by $f(y)$ and $F(y)$ respectively. The **survivor function**, defined as

$$S(y) = 1 - F(y)$$

expresses the probability that an individual remains in the state ("survives") until time y, that is, that an event has not yet occurred and the episode is still continuing.

[1] See chapter 11 of Amemiya (1985).

[1] This subsection and the following are based on Blossfeld et. al. (1989) and on Lancaster (1990).

The **hazard rate** (**hazard function** or **failure rate**), defined as

$$h(y) = \lim_{\Delta y \downarrow 0} \frac{1}{\Delta y} \Pr(y \leq Y < y + \Delta y | Y \geq y) = \frac{f(y)}{S(y)} = -\frac{d \ln S(y)}{dy}$$

may be interpreted as the instantaneous probability that the episode in the interval $[y, y + \Delta y]$ is terminating, conditional on having survived for length y. One can derive from this expression the relationship between the survivor function and the hazard function:

$$\begin{aligned} S(y) &= \exp\left(-\int_0^y h(u)\, du\right) \\ &= \exp(-H(y)) \end{aligned}$$

where

$$H(y) = \int_0^y h(u)\, du$$

is called the **integrated hazard.** We can see that the density function can also be obtained as a function of the hazard function

$$f(y) = h(y)\, S(y) = h(y) \exp\left(-\int_0^y h(u)\, du\right)$$

Then, each of the three quantities $f(y)$, $S(y)$, and $h(y)$ may be used to describe the duration of an episode.

An important concept to the study of duration models is whether or not the probability of leaving a particular state increases, decreases, or remains constant throughout the duration. This concept, **duration dependence**, is one of the most important determinants of the specification of the hazard function.

Below we review common parametric duration models.

- **Exponential Distribution:** It is characterized by a constant hazard rate

  $$h(y) = \theta \qquad \theta > 0$$

  Mean duration implied by this hazard rate is just

  $$E(Y) = \frac{1}{\theta}$$

  This is one of the most commonly applied distributions for the waiting time and lifetime. The respective survivor and density functions are

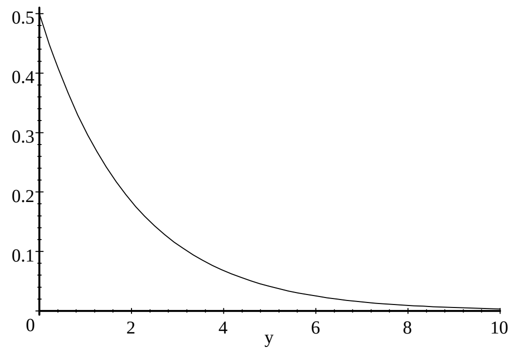
  $$\begin{aligned} S(y) &= e^{-\theta y} \\ f(y) &= \theta e^{-\theta y} \end{aligned}$$

Although in this case we are assuming that the hazard rate does not increase or decrease with the length of duration (i.e. there is no duration dependence), we could allow it to be different for each person. A common specification is to set

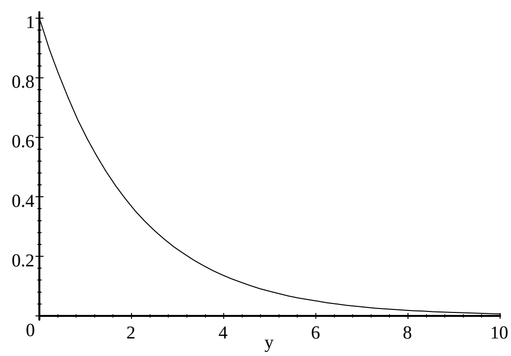
$$\theta_i = e^{x_i \beta}$$

where x_i can represent personal and/or aggregate characteristics associated with each individual.[2]

Using $\theta = 0.5$, the density, survivor and hazard exponential functions are represented by the following graphs:

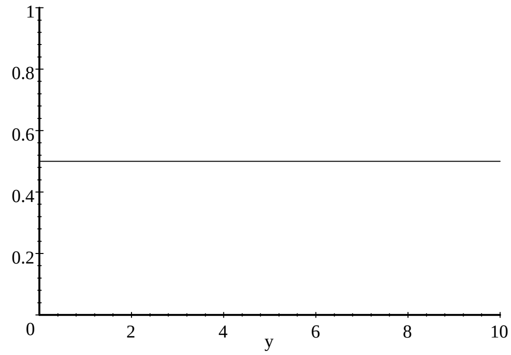


Exponential Density Function



Exponential Survivor Function

[2]Making h a function of a set of regressors is equivalent to changing the units of measurement (accelerating or decelerating) on the time axis. For this reason such models are called **accelerated failure time models**.



Exponential Hazard Function

- **Weibull distribution:** It is a generalization of the exponential distribution that is characterized by the hazard rate

$$h(y) = \gamma\theta^{\gamma}y^{\gamma-1}$$

The latter is constant for $\gamma = 1$, increasing in y for $\gamma > 1$ and decreasing in y for $\gamma < 1$. The mean duration implied by this hazard rate is:

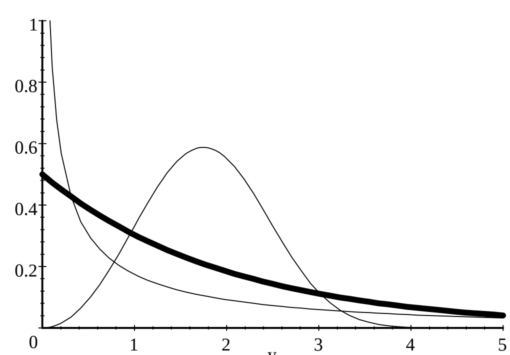
$$E\,(Y) = \theta^{-1}\Gamma(1 + 1/\gamma)$$

where $\Gamma(\cdot)$ is the Gamma function. The survivor and density functions are

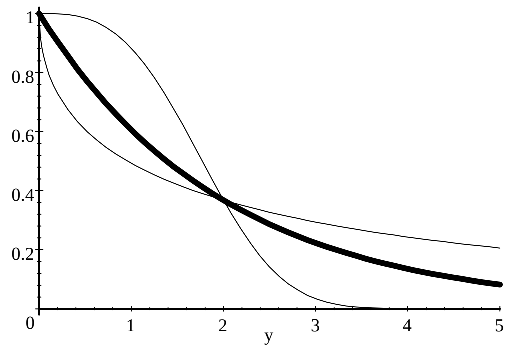
$$S(y) = \exp(-(\theta y)^{\gamma})$$
$$f(y) = \gamma\theta(\theta y)^{\gamma-1}\exp(-(\theta y)^{\gamma})$$

If we want to allow the hazard function to depend on individual characteristics, we may specify $\theta_i = e^{x_i\beta}$.

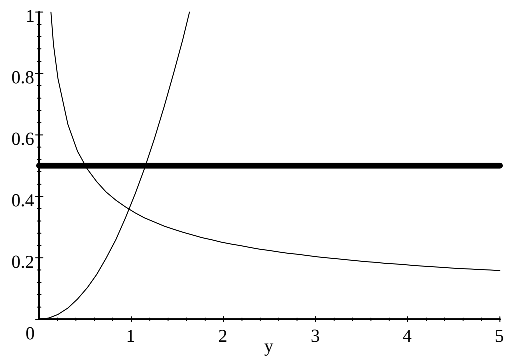
Using θ=0.5, the density, survivor and hazard Weibull functions are represented by the following graphs (assuming γ=3, thin solid lines; γ=1, thick lines; and γ=0.5, dotted lines):



Weibull Density Function



Weibull Survivor Function



Weibull Hazard Function

- **Extreme Value Distribution:** It is characterized by the hazard rate

$$h(y) = \frac{1}{\sigma} \exp\left(\frac{y-\mu}{\sigma}\right)$$

$-\infty < y < \infty$, and is closely related to the Weibull distribution. If Y^* possesses a Weibull distribution, then $Y = \ln Y^*$ possesses an extreme value distribution with $\sigma = \gamma^{-1}$ and $\mu = -\ln \theta$.

- **Log-Logistic Distribution**: It is characterized by the hazard rate

$$h(y) = \frac{\theta\gamma(\theta y)^{\gamma-1}}{1+(\theta y)^{\gamma}}$$

and has the feature that, for $\gamma > 1$, it has a first increasing and then decreasing hazard rate over time.

- **Log-Normal Distribution:** It is characterized by

$$\begin{aligned} f(y) &= \frac{1}{\sqrt{2\pi}\sigma y}\exp\left(-\frac{1}{2}\left(\frac{\ln y-\mu}{\sigma}\right)^2\right) = \left(\frac{1}{\sigma y}\right)\phi\left(\frac{\ln y-\mu}{\sigma}\right) \\ S(y) &= \Phi\left(-\frac{\ln y-\mu}{\sigma}\right) \\ h(y) &= \frac{\left(\frac{1}{\sigma y}\right)\phi\left(\frac{\ln(\lambda y)}{\sigma}\right)}{\Phi\left(-\frac{\ln(\lambda y)}{\sigma}\right)} \end{aligned}$$

  If Y possesses a log-normal distribution, then $Y^* = \ln Y$ possesses a normal distribution with mean $\mu = -\ln\lambda$ and standard deviation σ.

1.1. Estimation of Single Spell Duration Models

In parametric duration models we can estimate the parameters of interest by Maximum Likelihood. The construction of the likelihood function depends on the information available to us. For example, when the exact durations y_i are measured for each person, the likelihood takes the usual form

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta)$$

However, in other cases, when the full durations of all individuals are not available, we may be required to use different expressions, to correct for *censoring problems*. For example in the case where some observations are *right censored* (i.e. survey stops before spell is completed), the likelihood has the form

$$L(\theta) = \prod_{i\text{: uncensored}} f(y_i|\theta) \prod_{i\text{: censored}} S(y_i|\theta)$$

Another variation occurs when we do not know the durations, but we know the state in which the person is. In this case, we need to know what is the probability that an individual is in a specific state. An example is when we observe individuals either employed or unemployed and durations are unknown. In this simple world, to get the likelihood function, we need to know the probability that each individual is employed and the probability that is unemployed. Calling Z the duration of employment and Y the duration of unemployment we have

$$\Pr(\text{Employed}) = \frac{E(Z)}{E(Z)+E(Y)}$$

$$\Pr(\text{Unemployed}) = \frac{E(Y)}{E(Z)+E(Y)}$$

where $E(Y)$ and $E(Z)$ are defined by the hazard rates (e.g. if Z and Y are exponentially distributed with the hazard rates θ_E and θ_U, respectively, then $E(Z) = 1/\theta_E$ and $E(Y) = 1/\theta_U$). Then the likelihood function will be:

$$L(\theta) = \prod_{i=1}^{n} \Pr\left(\text{i=E}\right)^{1\{\text{i=E}\}} \Pr\left(\text{i=U}\right)^{1\{\text{i=U}\}}$$

For example we may assume $\theta_E = e^{x\beta_E}$ and $\theta_U = e^{x\beta_U}$ as in the exponential case. Then we can estimate β_E and β_U by maximizing this likelihood function, even though we do not observe any durations. This is because the hazard function and the associated variables and parameters imply expected durations for each state, and hence, which portion of the sample should be in each state at any given point in time.

More complicated likelihoods arise when the observations are also subject to *left censoring*, that is when the spells have possibly started before the survey.

1.1.1. The Exponential Model

We will next consider estimation in the context of the exponential model when all observed spells are complete. The hazard rate, conditional on individual characteristics x_i is assumed to be given by $h_i = \exp\left(x_i\beta_0\right)$. Since $f(y) = h(y)\,S(y) = h(y)\exp\left(-\int_0^y h(u)\,du\right)$, the likelihood takes the form

$$L(\beta) = \prod_{i=1}^{n} \exp\left(x_i\beta_0\right)\exp\left(-\exp\left(x_i\beta_0\right)y_i\right)$$

The MLE of β_0 is asymptotically efficient. We will calculate its asymptotic covariance matrix. We can easily get

$$\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\sum_i y_i \exp\left(x_i\beta\right) x_i' x_i$$

Since

$$E\left(y_i|x_i\right) = \int_0^\infty z h_i \exp\left(-h_i z\right) dz = \frac{1}{h_i}$$

we have that

$$AV\left(\hat{\beta}_{ML}\right) = \left[E\left(x_i' x_i\right)\right]^{-1}$$

As we will next see however it is possible to estimate β by two other methods. First, observe that since

$$E\left(y_i|x_i\right) = \frac{1}{h_i} = \exp\left(-x_i\beta_0\right)$$

we can write

$$\begin{aligned} y_i &= E(y_i|x_i) + [y_i - E(y_i|x_i)] \\ &= \exp(-x_i\beta_0) + u_i \\ &= g(x_i, \beta_0) + u_i \end{aligned}$$

where by construction $E(u_i|x_i) = 0$. This constitutes a NLR model and β_0 can be therefore consistently estimated by NLS. Note however that since

$$E\left(y_i^2|x_i\right) = \int_0^\infty z^2 h_i \exp(-h_i z)\, dz = \frac{2}{h_i^2}$$

we have that

$$V(y_i|x_i) = \frac{1}{h_i^2} = \exp(-2x_i\beta_0)$$

In other words, the model is heteroskedastic and therefore WNLS should be more efficient asymptotically. In fact here WNLS is asymptotically as efficient as ML since it can be easily shown that

$$AV\left(\hat{\beta}_{WNLS}\right) = \left[E\left(\exp(2x_i\beta_0)\left.\frac{\partial g}{\partial \beta}\right|_{\beta_0}\left.\frac{\partial g}{\partial \beta'}\right|_{\beta_0}\right)\right]^{-1} = \left[Ex_i'x_i\right]^{-1}$$

A second estimator results from observing that

$$\begin{aligned} E(\ln y_i|x_i) &= h_i\int_0^\infty \ln(z)\exp(-h_i z)\, dz \\ &= -c - \ln(h_i) \\ &= -c - x_i\beta_0 \end{aligned}$$

where $c \approx 0.577$ is **Euler's constan**t. Hence the log duration constitutes a linear regression model

$$\ln y_i + c = -x_i\beta_0 + u_i$$

where $E(u_i|x_i) = 0$ and

$$E\left(u_i^2|x_i\right) = \frac{\pi^2}{6} + (c + \ln h_i)^2$$

which implies that

$$V(u_i|x_i) = \frac{\pi^2}{6}$$

Hence OLS should be consistent. However OLS is less efficient than either ML or WNLS since

$$AV\left(\hat{\beta}_{WNLS}\right) = \frac{\pi^2}{6}\left[E\left(x_i'x_i\right)\right]^{-1}$$

REFERENCES

Amemiya, T. (1985): *Advanced Econometrics.* Harvard University Press.