Microeconometría I

Maestría en Econometría

Lecture 6

Duration Models

Agenda

Introduction

- Econometric models of durations are models of the length of time spent in a given state before transition to another state.
- A **state** is a classification of an individual entity at a point in time, **transition** is movement from one state to another, and a **spell length or duration** is the time spent in a given state.
- A typical regression example is determining the effect of higher unemployment benefit levels on the average length of an unemployment spell or the probability of transition out of unemployment.

Introduction

- Main problems/characteristics:
  1. Several related distributional functions are of interest and either the duration or probability of transition may be modeled
  2. Many different sampling schemes are possible and statistical inference depends on both the duration model and the sampling scheme (**flow sampling vs. stock sampling**).
  3. Data on spell duration are often censored.
  4. Transition data can have several states, such as unemployment, part-time employment, full-time employment, and out-of-the labor force, and data for a given individual may be available on multiple transitions among these states.

Moving into Employment

- Suppose that we are interested in modelling the process of movement into employment for someone who is looking for a job.
- One way of building a model for this transition is to suppose that as a result of his search he receives offers for jobs from time to time.
- Of the possible offers he should get, some he would find worth accepting and some not.
- Then, whether he moves into employment on any day depends upon he receives an offer that day and whether, if he does, he deems it worth accepting.

Moving into Employment

- An economist could then develop this approach by asking, and solving, the questions:
  - What set of wage offers it is optimal for a given person to accept? and,
  - How much resources should optimally be devoted to search?
- Answers to these questions depend not only on the criterion of optimality chosen by the economist but also on the circumstances of the unemployed person, in particular, the resources he possesses and the constraints he faces.

Moving into Employment

- Suppose that the relevant circumstances of a person who has been looking for a job for t days are assembled in a vector $x(t)$.
- Elements of this vector might include, for example, the level of unemployment benefit available after t, or the average wage payable in jobs that might be offered to him.
- For a person described at t by $x(t)$ suppose his optimal amount of search produces a probability, say, $\lambda(x(t))\,dt$, that a job offer will be made to him in the interval of time from t to $t + dt$, and there is a probability $P(x(t))$ that if such offer is received it will be worth accepting.

Moving into Employment

- Then, the outcome of this model-building effort is a quantity

$$\lambda(x(t))P(x(t))\,dt = \theta(x(t))\,dt$$

describing the probability of a transition out of unemployment in the time interval $(t,\ t + dt)$.

- The function $\theta(x(t))$ is called a hazard function, when, as in our example, there is only a single destination state (employment).
- There is an analogous set of functions when there are multiple destinations and in that context they are called transition intensities.
- Notice that, the choice variables in the transition model are the rate of search and the set of acceptable wage offers and these are not necessarily observable.

Moving into Employment

- Even if they are not observable we will be able to use the theory because of its implications for the hazard function.
- Even in the case where all relevant x were known to the investigator he still would not be able to say with certainty whether a transition will occur.
- Neither the econometrician nor the unemployed person can say for sure whether a transition will occur today. Thus the transition model is intrinsically stochastic.

Agenda

Definitions

- Let $T \geq 0$ denote the duration, which is the time elapsed until a certain event occurs, with some distribution in the population.
- t denotes a particular value for T.
- T is the time at which a person leaves the initial state. For example, if the initial state is unemployment, as in our motivation example, T would be the time, measured in, say, days, weeks or months, until a person becomes employed.
- The cumulative distribution function (cdf) of T is defined as

$$F(t) = P(T < t), \quad t \geq 0$$

Definitions

- The survivor function is defined as

$$S(t) \equiv 1 - F(t) = P(T \geq t), \qquad t \geq 0$$

  and this is the probability of "surviving" past time t.
- Denote the density of T by

$$f(t) = \frac{dF}{dt}(t)$$

Definitions

- For $dt > 0$, $P(t \leq T < t + dt | T \geq t)$ is the probability of leaving the initial state in the interval $[t, t + dt)$ given survival up until time t.
- If we divide this probability by dt we get the average probability of leaving per unit time period over a short interval after t.
- Considering shorter and shorter intervals we formally define

$$\theta(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt} \tag{1}$$

  as the hazard function.
- The hazard function is the instantaneous rate of leaving per unit time period at t.

Definitions

- From equation (1) it follows that for "small" dt,

$$P(t \leq T < t + dt | T \geq t) \approx \theta(t)\, dt$$

- The rough interpretation of the function θ is that $\theta(t)\, dt$ is the probability of exit from a state in a short interval of length dt after t conditional on the state still being occupied at t.
- We can express the hazard function in terms of the density and cdf of T using the law of conditional probability,

$$P(t \leq T < t + dt | T \geq t) = \frac{P(t \leq T < t + dt)}{P(T \geq t)} = \frac{F(t + dt) - F(t)}{1 - F(t)}$$

Definitions

- Dividing by dt and letting dt goes to zero, we get

$$\begin{aligned}\theta(t) &= \lim_{dt \to 0} \frac{F(t+dt)-F(t)}{dt}\frac{1}{1-F(t)} = F'(t)\frac{1}{1-F(t)} \\ &= \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}\end{aligned}$$

- Notice that the derivative of $S(t)$ is $-f(t)$, therefore we have

$$\theta(t) = -\frac{d\, log\, S(t)}{dt} \tag{2}$$

Definitions

- Equation (2) is a differential equation in t whose solution, subject to the initial condition $S(0) = 1$ (or $F(0) = 0$) is

$$S(t) = e^{-\int_0^t \theta(s)\, ds}, \tag{3}$$

- Equation (3) shows how one can calculate the probability distribution of duration of state occupancy given the hazard function.
- From the definition of the survivor function and equation (3) we have,

$$F(t) = 1 - e^{-\int_0^t \theta(s)\, ds}, \tag{4}$$

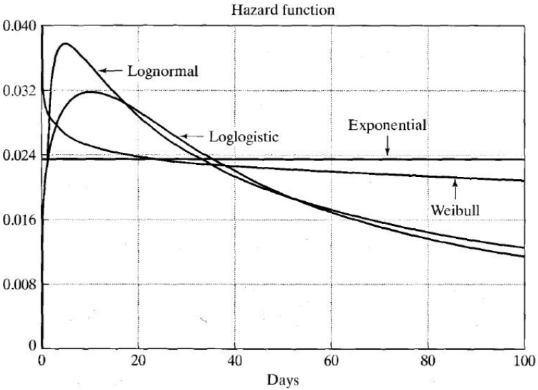
Definitions

- Straightforward differentiation of equation (4) gives the density of T as,

$$f(t) = \theta(t)\, e^{-\int_0^t \theta(s)\, ds}, \tag{5}$$

- Therefore, all probabilities can be computed using hazard functions.
- θ, f and S are alternative ways of describing the distribution of the probability of exit.
- Next figure shows some examples of hazard functions.

Definitions



Definitions

- The hazard function provides a convenient definition of duration dependence.
- Positive duration dependence exists at some point, say, t^* if $d\theta(t)/dt > 0$ at $t = t^*$.
- Positive duration dependence means that the probability of exiting the initial state increases the longer one is in the initial state.
- Negative duration dependence exists at some point, say, t^* if $d\theta(t)/dt < 0$ at $t = t^*$.
- Negative duration dependence means that the probability of exiting the initial state decreases the longer one is in the initial state.

Definitions

- The integrated hazard

$$\Theta(t) = \int_0^t \theta(s)\, ds \tag{6}$$

is also a useful function in practice. however, it does not have a convenient interpretation.

Agenda

Discrete data

- It is very common for a duration to be measured as an interval. For example, data may indicate that a transition occurred in a particular week, but the exact time in the week is not given.
- In such cases the transition times are said to be grouped and it is assumed that the hazard within the interval is constant.
- **Discrete-time hazard models** deal with such data.

Discrete data

- The starting point is to define the **discrete-time hazard function** as the probability of transition at discrete time $t_j, j = 1, 2, \ldots$, given survival to time t_j :

$$\begin{aligned} \lambda_j &= \Pr\left[T = t_j \mid T \geq t_j\right] \\ &= f^{\mathrm{d}}\left(t_j\right) / S^{\mathrm{d}}\left(t_{j_-}\right) \end{aligned} \tag{7}$$

where the superscript d denotes discrete, and where $S^{\mathrm{d}}\left(a_-\right) = \lim_{t \to a_-} S^{\mathrm{d}}\left(t_j\right)$, an adjustment made because formally $S^{\mathrm{d}}(t)$ equals $\Pr[T > t]$ rather than $\Pr[T \geq t]$.

Discrete data

- The **discrete-time survivor function** is obtained recursively from the hazard function as

$$
\begin{aligned}
S^{\mathrm{d}}(t) &= \Pr[T \geq t] \\
&= \prod_{j|t_j \leq t} (1-\lambda_j)\,.
\end{aligned}
\tag{8}
$$

- For example, $\Pr[T > t_2]$ equals the probability of no transition at time t_1 times the probability of no transition at time t_2 conditional on surviving to just before t_2, so that $\Pr[T > t_2] = (1-\lambda_1) \times (1-\lambda_2)$. The function $S^{\mathrm{d}}(t)$ is a decreasing step function with steps at $t_j, j = 1, 2, \ldots$

Discrete data

- The **discrete-time cumulative hazard function** is

$$\Lambda^{\mathrm{d}}(t) = \sum_{j | t_j \leq t} \lambda_j \tag{9}$$

Using (7), we have that the discrete probability that the spell ends at t_j is $\lambda_j S^{\mathrm{d}}(t_j)$

Agenda

Censoring mechanisms

- Survival data are usually censored, as some spells are incompletely observed.
- As an example, instead of observing the length of completed spell of unemployment, data may come from a survey of the currently unemployed, so that only the length of an incomplete spell of unemployment is observed.
- In practice data may be **right-censored, left-censored, or interval-censored**.
- For **right-censoring or censoring from above**, we observe spells from time 0 until a censoring time c. Some spells will have ended by this time anyway (completed spells), but others will be incomplete and all we know is that they will end some time in the interval (c, ∞).

Censoring mechanisms

- **Left-censoring or censoring from below** occurs when spells are known to end at some time in the interval $(0, c)$ but the exact time is unknown.
- The classical Tobit model is an example, where data on some spells are lost and the censoring time is unknown.
- **Interval-censoring** occurs when the completed spell length is observed but only in interval form such as in $[t_1^*, t_2^*]$.
- **Random censoring or exogenous censoring** means that each individual in the sample has a completed duration T_i^* and censoring time C_i^* that are independent of each other.
- We observe the completed duration T_i^* if the spell ends before the censoring time and the censoring time C_i^* if the spell ends after the censoring time.

Censoring mechanisms

- For standard survival analysis methods to be valid in the presence of censoring the censoring mechanism needs to be one with **independent (noninformative) censoring.**
- This means that parameters of the distribution of C_i^* are not informative about the parameters of the distribution of the duration T_i^*.
- Then one may treat the censoring indicator as exogenous, and it is then not necessary to model the censoring mechanism if interest lies in the duration model parameters.

Agenda

Nonparametric estimation

- Nonparametric estimation of survival functions are very useful for descriptive purposes.
- These methods are used to know the shape of the raw (unconditional) hazard or survival function before considering introducing regressors.
- Let $t_1 < t_2 < \cdots < t_j < \cdots < t_k$ denote the observed discrete failure times of the spells in a sample of size N, $N \geq k$.
- Define d_j to be the number of spells that end at time t_j (since the data are discrete d_j may exceed one).
- Define m_j to be the number of spells right-censored in the interval $[t_j, t_{j+1})$.
- The censoring mechanism is assumed to be independent censoring, so the only thing known about a spell censored in $[t_j, t_{j+1})$ is that the failure time is greater than t_j.

Nonparametric estimation

- Spells are **at risk** of failure if they have not yet failed or been censored.
- Define r_j to equal the number of spells at risk at time t_{j-}, that is, just before time t_j. Then $r_j = (d_j + m_j) + \cdots + (d_k + m_k) = \sum_{l|l \geq j} (d_l + m_l)$. Note that $r_1 = N$.
- In summary
  $d_j = \#$ spells ending at time t_j,
  $m_j = \#$ spells censored in $[t_j, t_{j+1})$,
  $r_j = \#$ spells at risk at time $t_{j-} = \sum_{l|l \geq j} (d_l + m_l)$.

Nonparametric estimation

- Using equation (7) an obvious estimator of the hazard function is the number of spells ending at time t_j divided by the number at risk of failure at time t_{j-}, or

$$\widehat{\lambda}_j = \frac{d_j}{r_j}$$

- The Kaplan-Meier estimator or product limit estimator of the survivor function is the sample analogue of equation (8)

$$\widehat{S}(t) = \prod_{j|t_j \le t} \left(1 - \widehat{\lambda}_j\right) = \prod_{j|t_j \le t} \frac{r_j - d_j}{r_j}. \tag{10}$$

- his is a decreasing step function with jump at each discrete failure time.

Nonparametric estimation

- In the case of no censoring $\widehat{S}(t)$ in (10) simplifies to $\widehat{S}(t) = r/N$, the number still at risk at time t divided by the sample size, which is one minus the empirical cdf.
- The discrete-time cumulative hazard function is defined in (9). The **Nelson-Aalen estimator** of the cumulative hazard function is the obvious sample analogue

$$\widehat{\Lambda}(t) = \sum_{j|t_j \leq t} \widehat{\lambda}_j = \sum_{j|t_j \leq t} \frac{d_j}{r_j}$$

Agenda

Some Distributions

Exponential Distribution

- For the exponential distribution with parameter $\theta > 0$ we have,

$$
\begin{aligned}
F(t) &= 1 - e^{-\theta\, t} \\
S(t) &= e^{-\theta\, t} \\
f(t) &= \theta e^{-\theta\, t} \\
\theta(t) &= \theta \\
\Theta(t) &= \theta\, t
\end{aligned}
$$

  - It is termed memoryless, because the hazard function is constant and so reflects no duration dependence.
  - The probability of exit from a state does not depends on how long it has been occupied.

Some Distributions

Weibull Distribution

- For the Weibull distribution with parameters $\alpha > 0$ and $\theta > 0$,

$$
\begin{aligned}
F(t) &= 1 - e^{-\theta\, t^{\alpha}} \\
S(t) &= e^{-\theta\, t^{\alpha}} \\
f(t) &= \theta \alpha t^{\alpha-1} e^{-\theta\, t^{\alpha}} \\
\theta(t) &= \theta \alpha t^{\alpha-1} \\
\Theta(t) &= \theta\, t^{\alpha}
\end{aligned}
$$

- It can be thought of as an exponential distribution on a re-scaled time axis (i.e. t^{α} has an exponential distribution with parameter θ).

Some Distributions

- Any hazard function can be transformed into the constant hazard by a transformation of the time scale.

Log-Logistic Distribution

- For the Log-Logistic distribution with parameters $\alpha > 0$ and $\theta > 0$,

$$
\begin{aligned}
F(t) &= 1 - [1/(1+\theta t^{\alpha})] \\
S(t) &= 1/(1+\theta t^{\alpha}) \\
f(t) &= \theta \alpha t^{\alpha-1}/(1+\theta t^{\alpha})^{2} \\
\theta(t) &= \theta \alpha t^{\alpha-1}/(1+\theta t^{\alpha}) \\
\Theta(t) &= \ln(1+\theta t^{\alpha})
\end{aligned}
$$

Agenda

Time-Invariant Covariates

- Usually in economics one is interested in hazard functions conditional on a set of covariates or regressors.
- When covariates do not change over time the hazard and all other features of T can be specified conditional on the covariates.
- The Conditional Hazard is

$$\theta(t;\mathbf{x}) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t, \mathbf{x})}{dt} \tag{11}$$

  where $\mathbf{x}$ is a vector of explanatory variables.
- All of the formulas defined previously continue to hold provided the cdf and density are defined conditional on $\mathbf{x}$. For example,

$$\theta(t;\mathbf{x}) = \frac{f(t|\mathbf{x})}{1 - F(t|\mathbf{x})}$$

  where $f(\cdot|\mathbf{x})$ is the density of T given $\mathbf{x}$.

Time-Invariant Covariates

- An important class of models with time-invariant regressors are the proportional hazard models.
- A proportional hazard can be written as

$$\theta(t; \mathbf{x}) = \kappa(\mathbf{x})\theta_0(t)$$

  where $\kappa(\cdot)$ is a nonnegative function of $\mathbf{x}$ and $\theta_0(t) > 0$ is called the baseline hazard.
- The baseline hazard is common to all units in the population.
- Individual hazard function differ proportionately based on a function $\kappa(\mathbf{x})$ of observed covariates.

Time-Invariant Covariates

- Typically, $\kappa(\cdot)$ is parameterized as $\kappa(\mathbf{x}) = e^{\mathbf{x}\beta}$, where β is a vector of parameters.
- Then,

$$\ln \theta(t; \mathbf{x}) = \mathbf{x}\beta + \ln \theta_0(t)$$

  and $\beta_{\mathbf{j}}$ measures the semi-elasticity of the hazard with respect to $\mathbf{x_j}$.
- One of the most used proportional hazard (PH) model is the Cox PH model
  - is a semiparametric model
  - makes no assumptions about the form of $\theta_0(t)$ (nonparametric part of model)
  - assumes parametric form for the effect of the predictors on the hazard
  - In most situations, we are more interested in the parameter estimates than the shape of the hazard

Agenda

Time-Varying Covariates

- Studying hazard functions is more complicated when we wish to model the effects of time-varying covariates on the hazard.
- It makes no sense to specify the distribution of the duration T conditional on the covariates at only one time period.
- Let $\mathbf{x(t)}$ denote the vector of regressors at time t.
- For $t \geq 0$, let $\mathbf{X(t)}$ denote the covariate path up through time t: $\mathbf{X(t)} \equiv \{\mathbf{x(s)} : 0 \leq s \leq t\}$.
- Following Lancaster (1990, chapter 2), we define the conditional hazard function at time t by

$$\theta(t; \mathbf{X(t)}) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t, \mathbf{X(t+dt)})}{dt} \tag{12}$$

assuming that this limit exists.

Time-Varying Covariates

- By definition, the covariates are sequentially exogenous because, by specifying $\theta(t; \mathbf{X(t)})$ we are conditioning on current and past covariates.
- One case where this limit exists very generally occurs when T is continuous and for each t, $\mathbf{x(t+dt)}$ is constant for all $dt \in [0, \eta(t)]$ for some function $\eta(t) > 0$.
- In this case we can replace $\mathbf{X(t+dt)}$ with $\mathbf{X(t)}$ in equation (12) because $\mathbf{X(t+dt)} = \mathbf{X(t)}$ for dt sufficiently small.

Time-Varying Covariates

- In practice, since the interval of observation is discrete we will have to assume that time varying covariates are constant over this interval in which case there is no problem in defining equation (12).
- It is important to know if time-varying covariates are strictly exogenous. Lancaster (1990) provides a definition that rules out feedback from the duration to future values of the covariates.
- If $\mathbf{X(t, t + dt)}$ denotes the covariate path from time t to $t + dt$, then Lancaster's exogeneity condition is

$$P[\mathbf{X(t, t + dt)} | T \geq t + dt, \mathbf{X(t)}] = P[\mathbf{X(t, t + dt)} | \mathbf{X(t)}] \tag{13}$$

for all $t \geq 0$, $dt \geq 0$.

Time-Varying Covariates

- The definition of strict exogeneity applies to covariates whose entire path is well-defined whether or not the person is in the initial state.
- One example are the so called external covariates (Kalbfleisch and Prentice, 1980) having the feature that the covariance path is independent of whether any person has or has not left the initial state. The city labor force participation in the moving into employment case.
- Other covariates are not external to each individual but have paths that are still defined after the person leaves the initial state. For example, marital status is well defined after someone becomes employed, but it is possibly related to whether someone has been unemployed. Whether marital status satisfy condition (13) is an empirical issue.

Time-Varying Covariates

- The definition of strict exogeneity cannot be applied to time-varying covariates whose path is not defined once the person leaves the initial state.
- These are internal covariates. For example consider an example of job tenure duration where a time-varying covariate is wage paid in that job. It makes no sense to define the future wage path in that job.
- It is clear that internal covariates cannot satisfy any reasonable strict exogeneity assumption.
- This fact is important when estimating duration models with unobserved heterogeneity.

Agenda

Single-Spell data

- Assume the population of interest is individuals entering the initial state during a given interval of time, say $[0, b]$, where $b > 0$ is a known constant.
- By convention, let zero denote the earliest calendar date that an individual can enter the initial state, and b is the last possible date.
- For example, if we are interested in the population of workers who became unemployed at any time during 2005, and unemployment duration is measured in years (with 0.5 meaning half a year) then $b = 1$. If duration is measured in weeks, then $b = 52$; if duration is measured in days, then $b = 365$.

Single-Spell data

- We restrict attention to single-spell data. That is, we use, at most, one completed spell per individual. If after leaving the initial state, an individual subsequently reenters the initial state in the interval $[0, b]$, we ignore this information.
- Covariates in this analysis are time invariant, meaning that we collect covariates on individuals at a given point in time -usually, at the beginning of the spell- and we do not re-collect data on the covariates during the course of the spell.
- With flow sampling we sample individuals who enter the state at some point during the interval $[0, b]$, and we record the length of time each individual is in the initial state.

Single-Spell data and Flow sampling

- We collect data on covariates known at the time the individual entered the initial state.
- Suppose we are interested in the population of workers who became unemployed at any time during 2005. Then, we randomly sample from the, say, population of male workers who became unemployed during 2005.
- At the beginning of the unemployment spell we might obtain information on tenure in last job, wage on last job, gender, marital status, unemployment benefits etc.

Single-Spell data and Flow sampling

- There are two common ways to collect data on unemployment spells.
- First: we may randomly sample individuals from a large population, say, all working-age individuals in the country for a given year, say, 2005.
- Some fraction of these people will be in the labor force and will become unemployed during 2005 -and this group of people who become unemployed is our random sample of all workers who become unemployed during 2005.
- Second: retrospective sampling.
- Suppose that, for a given province in the country we have access to unemployment records for 2005. We can obtain a random sample of all workers who become unemployed during 2005.

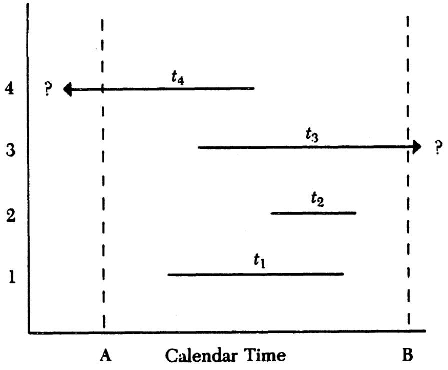
Single-Spell data and Flow sampling

- Flow data are usually subject to right censoring.
- That is, after a certain amount of time, we stop following individuals in the sample.
- For individuals who have completed their spells in the initial state, we observe the exact duration. But for those still in the initial state, we only know that the duration lasted as long as the tracking period.
- In the unemployment duration example, we might follow each individual for a fixed length of time, say, two years. If unemployment spells are measured in weeks, we would have censoring at 104 weeks.

Single-Spell data and Flow sampling

- Alternatively, we might stop tracking individuals at a fixed calendar date, say the last week of 2006. Because individuals can become unemployed at any time during 2005, calendar-date censoring results in censoring times that differ across individuals.

Flow sampling



A:	Beginning of study period (e.g., March survey)
B:	End of study period (e.g., April survey)
t_1, t_2:	Completed spells
t_3:	Right-censored spell
t_4:	Left-censored spell

Single-Spell data and Flow sampling

- For a random draw i from the population, let $a_i \in [0, b]$ denote the time at which individual i enters the initial state (the "starting time")
- Let t_i^* denote the length of time in the initial state (the duration)
- Let $\mathbf{x_i}$ denote the vector of observed covariates.
- We assume that t_i^* has a continuous conditional density $f(t|\mathbf{x_i}; \beta)$, $t \geq 0$, where β is a vector of unknown parameters.
- Without right censoring we would observe a random sample on $(a_i, t_i^*, \mathbf{x_i})$, and estimation would be a standard exercise on conditional maximum likelihood.

Single-Spell data and Flow sampling

- To account for right censoring, we assume that the observed duration, t_i, is obtained as

$$t_i = \min(t_i^*, \ c_i) \tag{14}$$

where c_i is the censoring time for individual i. Notice that in some cases c_i is constant across i.

- We assume that, conditional on the covariates, the true duration is independent of the starting point a_i, and the censoring time c_i:

$$D(t_i^*|\mathbf{x_i}, a_i, c_i) = D(t_i^*|\mathbf{x_i}) \tag{15}$$

where $D(\cdot|\cdot)$ denotes the conditional distribution.

Single-Spell data and Flow sampling

- Assumption (15) clearly holds when a_i and c_i are constant for all i.
- Under assumption (15) the distribution of t_i^* given $(\mathbf{x_i}, a_i, c_i)$ does not depend on (a_i, c_i).
- Therefore, if the duration is not censored, the density of $t_i = t_i^*$ given $(\mathbf{x_i}, a_i, c_i)$ is simply $f(t|\mathbf{x_i}, \beta)$.
- The probability that t_i is censored is

$$P(t_i^* \geq c_i|\mathbf{x_i}) = 1 - F(c_i|\mathbf{x_i}, \beta) \tag{16}$$

where $F(t|\mathbf{x_i}, \beta)$ is the conditional cdf of t_i^* given $\mathbf{x_i}$.

Single-Spell data and Flow sampling

- Let d_i be a censoring indicator ($d_i = 1$ if uncensored, $d_i = 0$ if censored).
- The conditional likelihood for observation i can be written as

$$f(t_i|\mathbf{x_i}, \beta)^{d_i}[1 - F(t_i|\mathbf{x_i}, \beta)]^{1-d_i} \tag{17}$$

- Given data on $(t_i, d_i, \mathbf{x_i})$ for a random sample of size N, the maximum likelihood estimator of β is obtained by maximizing

$$\sum_{i=1}^{N}\{d_i \ln f(t_i|\mathbf{x_i}, \beta) + (1 - d_i) \ln [1 - F(t_i|\mathbf{x_i}, \beta)]\} \tag{18}$$

- With the maximum likelihood estimations we can compute the hazard functions with the formulas above.

Single-Spell data and Stock sampling

- With stock sampling we randomly sample from individuals that are in the initial state at a given point in time.
- The population is again individuals who enter the initial state during a specified interval $[0, b]$.
- Rather than observe a random sample of people flowing into the initial state, we can only obtain a random sample of individuals that are in the initial state at time b.
- In addition to the possibility of right censoring, we may face the problem of left censoring, which occurs when some or all of the starting times a_i are not observed.

Single-Spell data and Stock sampling

- For now assume: (1) we observe the starting times a_i for all individuals we sample at time b; and (2) we can follow sampled individuals for certain lenght of time after we observed them at time b.
- In the unemployment duration example, where the population comprises workers who became unemployed at some point during 2005, stock sampling would occur if we randomly sampled from workers who were unemployed during the last week of 2005.
- This kind of sampling causes a clear sample selection problem: we necessarily exclude from our sample any individual whose unemployment spell ended before the last week of 2005.

Single-Spell data and Stock sampling

- Because these spells were shorter than a year, we cannot just assume that the missing observations are randomly missing.
- The sample selection problem caused by stock sampling is just the same as the left truncation problem.
- Under the assumptions that we observe the a_i and can observe some spells past sampling date b, left truncation is fairly easy to deal with.

Single-Spell data and Stock sampling

- To account for the truncated sampling, we must modify the density in equation (17) to reflect the fact that part of the population is systematically omitted from the sample.
- Let $(a_i, c_i, \mathbf{x_i}, t_i)$ denote a random draw from the population of all spells starting in $[0,\ b]$.
- We observe this vector if and only if the person is still in the initial state at time b, that is, if and only if $a_i + t_i^* \geq b$ or $t_i^* \geq b - a_i$, where t_i^* is the true duration.
- Under the conditional independence assumption (15),

$$P(t_i^* \geq b - a_i | a_i, c_i, \mathbf{x_i}) = 1 - F(b - a_i | \mathbf{x_i}; \beta) \tag{19}$$

where $F(\cdot|\mathbf{x_i}; \beta)$ is the cdf of t_i^* given $\mathbf{x_i}$ as before.

Single-Spell data and Stock sampling

- The log-likelihood function can be written as

$$\sum_{i=1}^{N}\{d_i \ln f(t_i|\mathbf{x_i}, \beta) + (1-d_i)\ln[1-F(t_i|\mathbf{x_i}, \beta)] - \ln[1-F(b-a_i|\mathbf{x_i}; \beta)]\} \quad (20)$$

  where again, $t_i = c_i$ when $d_i = 0$.
- Unlike the case of flow sampling, with stock sampling both the starting dates, a_i, and the length of the sampling interval, b, appear in the conditional likelihood function.
- Their presence makes it clear that specifying the interval $[0, b]$ is important for analyzing stock data.

Agenda

Single-Spell data and Unobserved Heterogeneity

- It may be the case that some elements of $\mathbf{x_i}$ are unknown to the investigator and must be supposed to vary over the population. Example: reservation wage in a job search model.
- This gives a second source of stochastic variation in the model in the form of unobserved heterogeneity.
- The key assumptions used in most models that incorporate unobserved heterogeneity are:
  (1) the heterogeneity is independent of the observed covariates, as well as starting times and censoring times.
  (2) the heterogeneity has a distribution known up to a finite number of parameters.
  (3) the heterogeneity enters the hazard function multiplicatively.

Single-Spell data and Unobserved Heterogeneity

- For a random draw i from the population, a Weibull hazard function conditional on observed covariates $\mathbf{x_i}$ and unobserved heterogeneity ν_i is

$$\theta(t; \mathbf{x_i}; \nu_i) = \nu_i \alpha t^{\alpha-1} e^{\mathbf{x_i}\beta} \tag{21}$$

  where $x_{i1} \equiv 1$ and $\nu_i > 0$.
- To identify the parameters α and β we need a normalization on the distribution of ν_i. The most common is $E(\nu_i) = 1$.
- This implies that, for a given vector $\mathbf{x}$, the average hazard is $\alpha t^{\alpha-1} e^{\mathbf{x}\beta}$.
- In the general case, where the cdf of t_i^* given $(\mathbf{x_i}, \nu_i)$ is $F(t|\mathbf{x_i}, \nu_i; \beta)$, we can obtain the distribution of t_i^* given $\mathbf{x_i}$ by integrating out the unobserved effect.

Single-Spell data and Unobserved Heterogeneity

- Because ν_i and $\mathbf{x_i}$ are independent, the cdf of t_i^* given $\mathbf{x_i}$ is

$$G(t|\mathbf{x_i};\beta,\rho)=\int_0^\infty F(t|\mathbf{x_i},\,\nu_i;\beta)h(\nu;\rho)\,d\nu \tag{22}$$

  where, for concreteness, the density of ν_i, $h(\cdot;\rho)$, is assumed continuous and depends on the unknown parameters ρ.
- For flow data, the log-likelihood function is

$$\sum_{i=1}^{N}\{d_i \ln g(t_i|\mathbf{x_i},\beta,\rho)+(1-d_i)\ln[1-G(t_i|\mathbf{x_i},\beta,\rho)]\} \tag{23}$$

  where $g(t|\mathbf{x_i},\beta,\rho)$ is the density of t_i^* given $\mathbf{x_i}$.
- We should assume that $D(t_i^*|\mathbf{x_i},\nu_i,a_i,c_i)=D(t_i^*|\mathbf{x_i},\nu_i)$ and $D(\nu_i|\mathbf{x_i},a_i,c_i)=D(\nu_i)$. These assumptions ensure that the conditional independence condition (15) holds.