

I. Data management

A recommended STATA book:

Microeconometrics Using Stata, by Colin Cameron and Pravin Trivedi (revised edition)

Download data sets in the book:

Open Stata when internet is connected.

```
.net install mus
```

```
.net get mus
```

Search where the files installed in your computer and put into a folder of your choice.

I stored the data sets under:

```
C:\DurationWorkshop\MUS
```

Data Format Needed to Run a Discrete Time Duration Model

Logistic discrete time hazard function

$$\lambda_i(t) = \frac{1}{1 + \exp(-Z_{it}\eta - f(t))}$$

Where Z_{it} is a vector of (possibly) time-changing explanatory variables, η is a vector of parameters, and $f(t)$ represents duration dependence.

In a single spell setting, likelihood contribution for individuals with a completed spell of K periods:

$$L_i(K) = \lambda_i(K) \prod_{t=1}^{K-1} (1 - \lambda_i(t))$$

For individuals with a right-censored spell of K periods:

$$L_i(K) = \prod_{t=1}^K (1 - \lambda_i(t))$$

- This structure of likelihood function implies we need a panel data (or long form data, the panel certainly is not balanced) to estimate a discrete time duration model.
- Usually survey data come in wide-form, each observation occupying one line.
- For discrete time model, data preparation is the same with or without time varying covariate.

Stata commands used to organize data for estimating a continuous time model (**stset** when no time-varying covariates and **stspl** to handle presence of time varying covariates for episode-splitting) not applicable to discrete time model.

How to create a panel (long form) data in STATA

Example #1 (on page 280 of the Colin Cameron and Pravin Trivedi book)

C:\DurationWorkshop\STATA programs\Example1.do

Example #2: a single spell example

- A dataset I created. For each individual, we observe a single unemployment spell.
- A randomly assigned treatment, D_i , before sample starts.
- A time constant variable, X_i
- A time changing variable, W_{it} – local unemployment rates.
- Duration dependence $f(t)$

$$\lambda_i(t) = \frac{1}{1 + \exp(-\theta - D_i\gamma - X_i\beta - W_{it}\delta - f(t))}$$

- γ is the coefficient for the treatment dummy variable. γ is an “intent-to-treat” (ITT) parameter and it captures the effect of being assigned to a treatment not the actual treatment effect on employment status.
- W_{it} is the Canadian provincial prime age male monthly unemployment rate from January 2000 to December 2007.
- Each individual enters the spell at different calendar time (from January 2000 to June 2001) that is determined by a random draw.
- Right censored at 40 periods.

C:\DurationWorkshop\STATA programs\Example2.do

II. Estimate a single-spell duration model

2.1 A single-spell duration model without unobserved heterogeneity

This can be done easily in STATA using “logit” command.

Duration dependence $f(t)$: the step function we just created can be incorporated as additional regressors.

C:\DurationWorkshop\STATA programs\EstimationNoUH.do

Hazard coefficient interpretation

Effects of covariate on hazard rate

Directional (not magnitude) effect of covariates on expected duration.

Usually researchers and policy makers care about:

Expected duration

Magnitude effect of changing covariates (most interestingly policy change) on expected duration

Additional computation is needed, and we will cover them tomorrow.

Break

2.2 A single-spell duration model with unobserved heterogeneity

For the model with unobserved heterogeneity, we will rely on maximum likelihood estimators and need to handle numerical optimization. Some do this with Fortran or C, but R and Matlab are becoming much more popular and also they are much easier to use. I will first give you a brief introduction to R and then we consider a simple example of using one of R's optimization routines.

- **What is R?**

- **R** is an open source programming language and software environment for statistical computing and graphics.
- It is originally used in academics as the predominant language for statisticians, but also widely used by engineers and scientists.
- R is rapidly gaining ground in the business world. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it.

- **Main features – a flexible and versatile language**
 - R uses a command line interface similar to Stata. You may type your command at the command prompt and see side effects immediately. Thus with a bit of preparation it is as easy to use as Stata, but
 - Programs are much better vetted for mistakes
 - Always backward compatible unlike Stata
 - Much better than Stata for non-parametric estimation.
 - R has quickly found a following because statisticians, engineers and scientists who find it much easier to use than C, C++ and Fortran 90.
 - For computationally intensive tasks, C, C++, and Fortran code can be linked and called at run time.
 - Another strength of R is static graphics, which can produce publication-quality graphs, including mathematical symbols.
 - R's matrix features and performance are comparable to Matlab and Gauss. (there is an online manual on how to translate between R and Matlab).

R – Why it could be very useful for economists

I assume you already have a language such as Stata to clean data and run some basic models.

R will be extremely helpful in the following aspects if:

- You want to implement a more complicated model requiring numerical optimization (such as duration models we are covering now).
 - R has extremely reliable and relatively user-friendly optimization algorithms and a very active online forum where you can seek help if a problem occurs.
 - Here you get fast responses from technically strong statisticians and mathematicians who routinely handle numerical maximization. Along this direction the resources are more abundant on the R forum than on the Stata forum.

- You want to adopt a cutting-edge nonparametric approach, such as matching, LATE, regression discontinuity. R's reliability in nonparametric implementation has probably the highest standard among all statistical softwares (I would avoid Stata for such tasks as simple as kernel density estimation).

R Official website: <http://www.r-project.org/>

One of R's optimization routine: optim

- R is well-suited for programming your own maximum likelihood estimators.
- There are several routines for numerical optimization. I will focus on the “optim” command, which implements the BFGS and simulated annealing (which we will discuss later), among others.
- Optimization through optim is relatively straightforward, since it only requires user-provided likelihood functions.

BFGS

The BFGS method approximates Newton method. It belongs to a class of hill-climbing optimization techniques that seek a stationary point of a (preferably twice continuously differentiable) function. It relies on gradients and hessian to determine the next move in the process of searching for an optimum.

Define your log-likelihood function

Syntax:

```
Name = function(pars,object){  
  declarations  
  logl = loglikelihood function  
  return(-logl)  
}
```

A simple example

Suppose we have a sample that was drawn from a normal distribution with unknown mean μ and variance σ^2 . The objective is to estimate these parameters relying on numerical maximization. The normal log-likelihood function (of a sample with n observations) is given by:

$$l = -0.5n \ln(2\pi) - 0.5n \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \mu)^2$$

Let's consider this numerical maximization problem in R.

C:\DurationWorkshop\R programs\SimpleMLE.r

Now consider the estimation of our single-spell duration model with unobserved heterogeneity

As John discussed, we often assume a nonparametric distribution with finite support point for unobserved heterogeneity:

θ follows a discrete distribution with points of support $\theta_1, \theta_2, \dots, \theta_J$ and associated probabilities p_1, p_2, \dots, p_J respectively, where $p_J = 1 - \sum_{k=1}^{J-1} p_k$.

It is distributed independently across individuals and is fixed over time for a given individual.

Individual likelihood contribution with a completed spell:

$$l_i(K) = \log \left\{ \sum_{j=1}^J p_j \cdot [\lambda_i(K | \theta_j)] \prod_{t=1}^{K-1} [1 - \lambda_i(t | \theta_j)] \right\}$$

This functional form creates a very difficult numerical maximization problem (multiple local optima).

Our simple single spell data actually has unobserved heterogeneity and individuals belong to one of the two (unobserved) types with the following parameter values:

$$\theta_1 = -0.8, \theta_2 = -2.8, p_1 = p_2 = 0.5, D \sim \text{Bernoulli}(p = 0.5), \gamma = 1,$$

$$\beta = 1, \delta = -0.05, X \sim N(0, \sigma^2 = 0.25), f(t) = \exp\left(\frac{1-t}{5}\right) - 1$$

Given this setting, the hazard becomes:

$$\lambda_i(t|\theta_j) = \frac{1}{1 + \exp(-\theta_j - \gamma D_i - \beta x_i - \delta W_{it} - f(t))} \quad j = 1, 2$$

Now let's consider estimating this model in R.

C:\DurationWorkshop\R programs\EstimationSingleSpell.r

For continuous time duration models in Stata, the Gamma and Inverse Gaussian distribution are most commonly used to capture unobserved heterogeneity because they lead to close form solution to the integration problem.

An alternative numerical optimization algorithm: simulated annealing

Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60, 65–99.

Gaure, S., Røed k., Zhang, T., 2007. Time and causality: A Monte Carlo Assessment of the Timing-of-events Approach. *Journal of Econometrics*, 141, 1159-1195

Li, X., Smith, B., 2011. Diagnostic Analysis and Computational Strategies for Estimating Discrete Time Duration Models – A Monte Carlo Study.

IV. Estimate expected duration and effects of changing covariates on expected duration

Conditional on the unobserved heterogeneity θ_j , the probability that a non-employment spell lasts longer than $t - 1$ months is given by the survivor function

$$S(t-1|\theta_j) = \prod_{\tau=1}^{t-1} [1 - \lambda(\tau|\theta_j)].$$

The density function for a non-employment spell that lasts t months is given by $f(t|\theta_j) = \lambda(t|\theta_j)S(t-1|\theta_j)$.

The expected duration for a non-employment spell (averaged over the estimated unobserved heterogeneity distribution)

$$ED = \sum_{j=1}^J \left[\sum_{t=1}^{\infty} t \cdot f(t|\theta_j) \right] \cdot p_j.$$

Since there is no guarantee this expected duration will be finite, we instead calculate a truncated mean as follows:

$$ED^{trunc} = \sum_j \left[\sum_{t=1}^{T^*} t \cdot f(t|\theta_j) + S(T^*|\theta_j) \cdot T^* \right] \cdot p_j.$$

We may choose $T^* = 60$.

Estimating expected duration:

First calculate the truncated mean for each individual, and then take the sample average.

Estimating effects of changing a covariate on expected duration:

- Set the target explanatory variable at two different levels and calculate the corresponding expected durations;
- Take the difference between the two expected durations, and the difference represents the effect of changing that particular explanatory variable on expected duration.

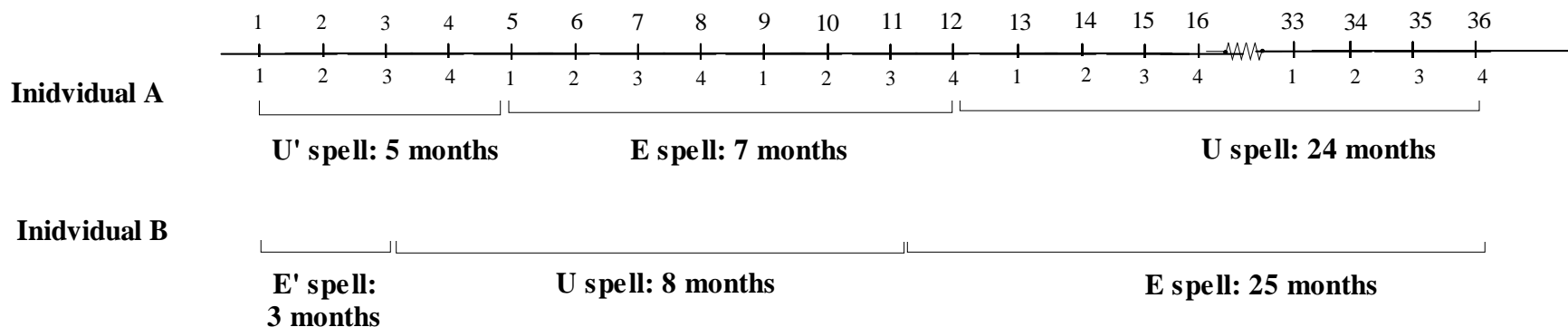
Standard errors:

- Evaluate numerical partial derivatives of the expected duration function with respect to each parameter;
- Calculate standard errors using delta method

C:\DurationWorkshop\R programs\ExpectedDurationSingleSpell.r

V. Multiple-spell, multiple-state data

- Two states: employment and nonemployment
- Four types of spell: left-censored employment spell, left-censored nonemployment spell, fresh employment spell, and fresh nonemployment spell.



$$L_A = \sum_{j=1}^J p_j \left(\begin{array}{l} \left[\lambda_{U'}(5|\theta_{U'j}) \prod_{r=1}^4 (1 - \lambda_{U'}(r|\theta_{U'j})) \right] \cdot \\ \left[\lambda_E(7|\theta_{Ej}) \prod_{r=1}^6 (1 - \lambda_E(r|\theta_{Ej})) \right] \cdot \\ \left[\prod_{r=1}^{24} (1 - \lambda_U(r|\theta_{Uj})) \right] \end{array} \right)$$

$$L_B = \sum_{j=1}^J p_j \left(\begin{array}{l} \left[\lambda_{E'}(3|\theta_{E'j}) \prod_{r=1}^2 (1 - \lambda_{E'}(r|\theta_{E'j})) \right] \cdot \\ \left[\lambda_U(8|\theta_{Uj}) \prod_{r=1}^7 (1 - \lambda_U(r|\theta_{Uj})) \right] \cdot \\ \left[\prod_{r=1}^{25} (1 - \lambda_E(r|\theta_{Ej})) \right] \end{array} \right)$$

A multi-state, multi-spell data: Example 3

You may set indicators such as the following to help you code the likelihood function.

Fresh/Left-censored employment/nonemployment spell

Indicator

EMPFRINT = 1 Fresh Employment Spell

EMPFRINT = 2 Fresh Unemployment Spell

EMPFRINT = 3 Left-censored Employment Spell

EMPFRINT = 4 Left-censored Unemployment Spell

Spell Transition Indicator

EMPCENS = 0 Spell not censored

EMPCENS = 1 Spell is censored

C:\DurationWorkshop\STATA programs\Example3.do

Estimating the fraction of time spent in employment

- The effect of changing an explanatory variable on the fraction of time that an individual is employed is another important parameter for policy makers.
- We may use simulations to predict the fraction of time spent in employment over certain horizons, say 3 or 5 years, as well as to examine how these fractions change with changes in macro, demographic and public policy variables.
- To get the point estimates of interest, we simulate an employment/non-employment history for each individual over a particular time horizon and then calculate the sample fraction of time employed based on the individual simulated histories.
- Since the simulated fractions are discontinuous functions of the parameter estimates, the delta method cannot be used to calculate consistent estimates of the confidence intervals.

- We can form $(1 - \alpha)$ confidence intervals by taking a large number of draws from the asymptotic distribution of the parameter estimates, simulating the model using each draw of the parameter estimates.

The following step by step simulation procedure covers both fraction of time and standard error estimation.

Simulation procedure step by step:

1. For each individual, we simulate her, say 72-period, employment history by unobserved heterogeneity type. For example, if the model indicates that there are two points of support for unobserved heterogeneity, we simulate two employment histories for each individual conditional on belonging to each of the two types and take the expected value with respect to the estimated unobserved heterogeneity distribution. Specifically:
 - The starting point of an employment history is determined by the data. If a person was in a left-censored non-employment spell at the beginning of the sample, her simulated employment history will start with a left-censored non-employment spell.

- From the starting month, an individual monthly hazard rate is calculated based on unobserved heterogeneity type, observed heterogeneity (individual means over the sample period are used), spell type (e.g. left-censored employment spell) and spell duration.
- A uniform random number is drawn to compare with the calculated hazard and determine whether the individual exits into the next spell. The previous two steps are repeated for 72 periods for each individual and unobserved heterogeneity type. If individual i of unobserved type j is employed in month t , then $E_{ij}^t = 1$; otherwise $E_{ij}^t = 0$.
- For individual i , we average simulated histories across unobserved heterogeneity type according to the estimated probabilities p_j to yield weighted employment history

$$E_i^t = \sum_{j=1}^J p_j E_{ij}^t .$$

2. We average the simulated histories over the sample and over the 72 periods to estimate the fraction of time employed

$$ER = \frac{\sum_{i=1}^N \sum_{t=1}^{72} E_i^t}{72 \cdot N}$$

3. Use the CI-bootstrap to calculate the 95% confidence interval for the fraction of time employed estimates as follows:
 - Generate 20,000 parameter vectors from the estimated asymptotic distribution.
 - Keep the parameter vectors that are within the 95% confidence interval of the parameter estimates.
 - Calculate ER for each parameter vector from the previous step.
 - Take the minimum and maximum of ER estimates from the previous step and report them as the lower bound and upper bound of the estimated 95% confidence interval.
4. To predict the effect of changing one variable on the fraction, follow the previous steps by setting the variable of interest at different levels for each person and then taking the difference in estimated fractions. For example, to estimate the effect of implementing positive incentives to leave welfare

(“carrot” waiver policy), first set the carrot waiver dummy to 1 for each person (i.e. carrot waivers are implemented) to calculate ER^1 , and then set the carrot waiver dummy to 0 for each person (i.e. carrot waivers are not implemented) to calculate ER^0 . Then $ER^1 - ER^0$ is the estimated effect of implementing a positive incentive to leave welfare.

5. Estimate 95% confidence intervals for the effects of changing variables following step 3 at each level of those variables.