Lectures Notes for John Ham's Section of the Course on Duration Models for Social Scientists Given in June 2012 at DIW, IFAU, and IFS by John Ham and Xianghong Li

Applications of Hazard Models that are of interest to economists and sociologists:

1. Time until first birth
2. Time until marriage for single individuals
3. Time until divorce for married couples
4. Time until relapse for graduates of an alcohol/drug treatment program
5. Time until return to jail for those released from jail

6. Time until failure for a new firm
7. Time until retirement
8. Time until entering employment for unemployed
9. Time until drilling for oil on a site that a firm has an option on

A few commonly used terms in the duration literature: (defined in class):

Spell
Left-censoring and right-censoring
Hazard rate
Duration dependence
Unobserved heterogeneity
Expected duration

What questions do we ask in duration (hazard) models?

What is the probability in a given week that one leaves a state (e.g. unemployment) given you have been in a state for t weeks? Or (equivalently) what is the probability an individual will stay in a state for T weeks.

What is the expected length of a spell and how does it change with a change in the explanatory variable – an important policy effect.

Why the probability of leaving a state changing with the length of the spell – duration dependence – is important for interventions?

In a two state model – e.g. employment and nonemployment – We will also consider what fraction of time does an individual spend in employment and how does it change with a change in the explanatory variable – a second important policy effect.

We use time aggregated (e.g. weekly or monthly) outcome data and the most complicated models covered by this workshop will contain the following components:

- Explanatory variables that may change with Calendar time
- Duration dependence
- Unobserved differences across individuals
- Correlated unobserved differences across different types of spells for the same individual, e.g. employment spells and unemployment spells.

Duration models have a reputation for being difficult, and it is certainly true that one's intuition from a linear model does not carry over well. However, if one builds the models up one step at a time, they become much more accessible. On my website I list papers by former students and their papers often involve duration models, and are published in good journals.

We especially want to worry about time changing explanatory variables – for example, unemployment rate in a study of employment or unemployment duration. This is not done in much European empirical work right now, but there is really no reason for this omission. Time changing explanatory variables provide a nice source of independent variation in more complicated models, e.g. models with endogenous participation in training.

- Why the emphasis on time changing explanatory variables - if we don't have them, we could simply run regression, for example Tobit Models where the Tobit Index function is

$$Dur_i^* = X_i \beta + \varepsilon_i$$

and the relation to the true data is given by

$Dur_i^* = Dur_i$ for complete spells,
$Dur_i^* > Dur_i$ for right-censored spells.

(Right-censoring occurs if someone runs out of benefits and drops out of the data, or in survey data if they are still in the spell at the end of the period.) I am sure you all know how to estimate a Tobit model in Stata.

However, one cannot use Tobit Model to do this and let the explanatory variables change over the spell.

A second, but in my view less strong, motivation for hazard models is that they may be more closely linked to theory.

Example of a simple search model in unemployment. Here the probability someone leaves unemployment after a spell of t days is

$$\lambda(t) = \mu(t)(1 - F(w^r(t))$$

where $F(w(t))$ is the distribution function for wages

and $w^r(t)$ is the reservation wage.

How $\lambda(t)$ changes with t may let us learn about $\mu(t)$ and $w^r(t))$

As a simple example, set $\mu(t) = 1$ and learn about how $\mathrm{w}^r(t)$ changes with time.

We will work in discrete time. Why?

It is much more intuitive – our goal is to take away impression that duration models are inherently very hard or complicated.

As we discuss below transitions (outside financial markets) are often observed in discrete time (e.g. German socio-economic panel where duration is monthly).

The discrete time model will approach the continuous time model as the time unit decreases.

Further, time changing explanatory variables change discretely (e.g. by month or quarter). The advantages of continuous time disappear once one has to account for this data structure.

In much of what follows we will consider a flow sample – look at those entering unemployment within a certain calendar window, i.e. Jan 2011-May 2012. An alternative to this is a stock sample, all of those currently unemployed at a given calendar time.

Single State Duration Models without Unobserved Heterogeneity

Case 1 (Simplest): λ is constant over time and across individuals

- The probability of a transition is given by

  $\lambda = 1/(1+e^{-\delta})$.

We use logit structure to insure that hazard stays between 0 and 1. It's easy to change the functional form to something that keeps the probability between 0 and 1; however a nice feature of the logit specification is that it lets you compare results from your programs against those from logit routines in standard statistical packages, such as Stata.

We will want to consider the

1. Probability that a spell lasts T weeks (density function)

$$\Pr(Dur = T) = \lambda(1-\lambda)^{T-1}.$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur>K)}=(1-\lambda)^{K}.$$

(Note: in discrete time $\text{Prob (Dur>K)} \neq \text{Prob (Dur} \geq K).$)

3. How long will a spell last on average (expected duration)

$$E(Dur) = \sum_{t=1}^{\infty} t\lambda(1-\lambda)^{t-1} = \lambda^{-1} = (1+e^{-\delta}).$$

Comparison to a continuous time Hazard $\lambda = e^{\theta}$

(Logit converges to this as time period approaches 0 in length.)

1. Density for a spell lasting exactly t periods

$$h(t) = \lambda e\text{xp}\{-\int_0^t \lambda d\tau\}.$$

2. Survivor function

$$S(t) = e\text{xp}\{-\int_0^t \lambda d\tau\}.$$

3. Expected duration

$$E(dur) = \int_0^\infty \tau h(\tau) d\tau.$$

But in common used survey data such as the German Socio-Economic Panel or the Survey of Income and Program Participation (SIPP) we observe that spell ends between month T and T+1. The probability of this event with a continuous time hazard function is

$$\Pr(T \le dur \le T+1) = \Pr(dur \ge T) - \Pr(dur \ge T+1) = S(T) - S(T+1).$$

This will be the contribution of such a spell to the likelihood function, and with this type of (common) time aggregation the continuous time model starts to look like the discrete time model; sometimes people use the approximation that the transition takes place at time T+.5. This grouping can be incorporated in Stata programs for estimating continuous time hazard models (which have a number of limitations noted below).

Case 2: λ is constant over time but differs across individuals

$$\lambda_i = \frac{1}{1+\exp\left(-\alpha X_i\right)};$$

note *if* $\alpha_k > 0$, then $\dfrac{\partial \lambda_i}{\partial X_{ki}} > 0.$

Why would we expect the hazard to differ across individuals?

- Job arrival rate may be higher for highly educated, workers in their 20's and 30's, men vs. women etc.

Continuous time analogue is $\lambda_i = e^{\alpha X_i}$.

Again we will want to consider the

1. Probability a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h(T) = \lambda_i (1-\lambda_i)^{T-1}.$$

2. Probability a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob } (\text{Dur}_i > \text{K}) = (1-\lambda_i)^K.$$

3. How long will a spell last on average (expected duration)

$$E(Dur_i) = \sum_{t=1}^{\infty} t\lambda_i (1-\lambda_i)^{t-1} = \lambda_i^{-1} = 1 + \exp\left(-\alpha X_i\right)..$$

Note: Expected Duration for the sample

$$ED^{*} = \frac{1}{N}\sum_{i=1}^{N} E(Dur_i) = \frac{1}{N}\sum_{i=1}^{N} 1 + \exp\left(-\alpha X_i\right) \neq 1 + \exp\left(-\alpha \bullet \left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right).$$

In other words it is better to calculate the ED for each person and then average these.

ED's are very useful for calculating policy effects:

The effect of increasing one of the X's, say X_k, by one unit, is

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \mathrm{E}(\mathrm{dur}_i)}{\Delta X_{ki}} \ .$$

We often will take the derivative numerically.

For dummy variables we look at

$$\frac{1}{N}\left[\sum_{i=1}^{N}\mathrm{E}(\mathrm{dur}_i|\mathrm{X}_{ik}=1)-\sum_{i=1}^{N}\mathrm{E}(\mathrm{dur}_i|\mathrm{X}_{ik}=0)\right].$$

Of course, again $\frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \mathrm{E}(\mathrm{dur}_i)}{\Delta X_{ki}} \neq \frac{\Delta \mathrm{E}(\mathrm{dur}|\bar{\mathrm{X}})}{\Delta X_{ki}}$, where $\bar{\mathrm{X}}$ is the mean of the explanatory variables in the sample. Calculating expected durations is much more informative that simply looking at the hazard coefficient on X_k.

Note that since we assume that there are no unobservables (for now), there is no possibility of an endogenous variable unless it is a function of duration. Also it doesn't make sense to cluster by individual when calculating the standard errors. The problem with clustering is that ignoring unobserved heterogeneity in estimation will lead to inconsistent parameter estimates, so it doesn't make sense to account for it in calculating the standard errors.

Case 3: The hazard depends on calendar time

$$\lambda_i(t) = \frac{1}{1+\exp(-Z_{i\tau}\eta)}, \;\; \tau = \tau_0 + t \quad \text{where}$$

τ_0 is the calendar time start date of the spell.

Continuous time analogue $\quad \lambda_i(t) = e^{Z_{i\tau}\eta}, \; \tau = \tau_0 + t.$

Why might the hazard function depend on calendar time?

Variables such as the overall unemployment rate in employment or unemployment duration. Monte Carlo work indicates that ignoring the time changing nature of variables such as the unemployment rate biases the effect of the business cycle substantially.

Note we are *not* considering time changing X's that are controlled (to some extent) by the individual – these will not help in identification, as we discuss below.

We have

1. Prob a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h_i(T) = \lambda_i(T+\tau_0)\prod_{r=1}^{T-1}(1-\lambda_i(r+\tau_0)).$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob }(\text{Dur}_i > \text{K}) = \prod_{r=1}^{K}(1-\lambda_i(r+\tau_0)).$$

How long will a spell last on average (expected duration)

$$\text{E}(\text{dur}_i) = \sum_{r=1}^{\infty} r h_i(r).$$

An empirical issue: how to set the time changing X's after an individual gets out of the sample period in the expected duration calculation?

What happens in continuous time model with time changing explanatory variables?

Suppose $X_{it+\tau_0}$ changes only once at duration time t_1, calendar time $t_1+\tau_0$. Then even if outcome is continuously measured, the contribution to the likelihood for a completed spell

$$f(t)=\lambda_i(t+\tau_0)\exp\{-\int_0^{t_1}\lambda(r+\tau_0\,|\,X(\tau_0))dr+\int_0^{t-t_1}\lambda(r+\tau_0+t_1)\,|\,X(\tau_0+t_1)dr\}.$$

In other words, the continuous time approach loses its nice form. Stata's program will let you break up the interval a limited number of times, but it's very easy to make a mistake here. With general duration dependence it gets even trickier. *With discrete time you take care of this in the data step.*

Case 4: In discrete time let λ_i also depend on duration

$$\lambda_i(t) = \frac{1}{1 + \exp\left(-Z_{i\tau}\eta - f(t)\right)}$$

Define

negative duration dependence: $\partial\lambda_i(t)/\partial t < 0$

positive duration dependence: $\partial\lambda_i(t)/\partial t > 0$.

Why might we expect negative duration dependence in unemployment duration?

Discouraged workers, negative signaling to employers

Why might we expect positive duration dependence?

Assets dwindling, learning about the wage distribution and shifting down reservation wages, running out of benefits, being subject to 'sticks' (penalties) if they stay unemployed too long.

Here the important question is how to specify duration dependence function $f(t)$. Early studies took simple functional forms for $f(t)$ – linear in t, quadratic in t, or t^{α}.

However, you will want to be more flexible. Can either use a polynomial in t or a step function in t. Latter seems to be better if you are only considering one type of spell, the former may be better if you are estimating the parameters of several types of spells simultaneously since it tends to lead to more parsimonious specifications.

Important point: suppose you have weekly data, and the highest duration you can see is 52 weeks. Temptation put in 52 dummies for duration. That will lead to very noisy estimates and make it impossible to allow for unobserved heterogeneity. Instead make sure each dummy corresponds to 3-5% of the transitions. Suppose we choose 3% i.e., make a dummy for 30-35 weeks if 3% of the transitions occur in these weeks. A smaller percentage (e.g. 1%) may be chosen if you have a larger sample.

Over-specified duration dependence (one step for each period) is a problem in Meyer (1990), which otherwise covers much of the same material as Ham and Rea (1987).

Relevant formulae

1. Prob a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h_i(T) = \lambda_i(T + \tau_0)\prod_{r=1}^{T-1}(1-\lambda_i(r+\tau_0)).$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur}_i\text{ >K)=}\prod_{r=1}^{K}(1-\lambda_i(r+\tau_0)).$$

3. How long will a spell last on average (expected duration)

$$\text{E(dur}_i\text{)=}\sum_{r=1}^{\infty} r h_i(r).$$

One possible problem – a defective distribution occurs if the duration dependence drives the hazard to 0 for t>t*; in this case the expected duration won't exist. A couple of alternatives – use median duration, or use a truncated duration:

$$\text{TruncE}(\text{dur}_i) = \sum_{r=1}^{T^*} r f_i(r) + S_i(T^*)T^*.$$

For more on this issue, see Eberwein, Ham and LaLonde (2002).

In discrete time with duration dependence and time changing X's one can still uses a logit program, and still the only difference across the models is in the data step. It is straight-forward to estimate this type of duration model in discrete time as long as you set up the data as Xianghong will suggest.

It is tricky to have general duration dependence and time changing X's in continuous time, but not to have general duration dependence and only time constant X's in continuous time.

Session II starts about here

Case 5: Unobserved differences across individuals

This will lead to a considerable increase in difficulty conceptually and in estimation. Xianghong's R programs should be extremely helpful for you here. The hazard is

$$\lambda_i\left(t \middle| \theta_i\right) = \frac{1}{1+\exp\left(-\theta_i - X_{i\tau}\eta - f\left(t\right)\right)}$$

What does θ_i capture

-optimism

-work ethic

-ability

Note: need $\text{cov}(\theta_i, X_{i\tau})=0$ at the start of the spell, i.e. $\tau=\tau_0$. We will talk later about what one can do if $\text{cov}(\theta_i, X_{i\tau}) \neq 0$ at the start of the spell (i.e. $\tau=\tau_0$), e.g. endogenous training participation.

Note that we discuss below why we will have $\text{cov}(\theta_i, X_{i\tau}) \neq 0$ for t > 1, We will assume that the heterogeneity is drawn from a discrete distribution with points of support $\theta_1, \theta_2, ..., \theta_{J-1}, \theta_J$ and associated probabilities $P_1, P_2, ..., P_{J-1}, P_J$, where $P_J = 1 - \sum_{j=1}^{J-1} P_j$ (Heckman and Singer Econometrica 1984). Assuming this type of flexible specification of the distribution is preferred to a simpler parametric distribution.

Need to control for unobserved heterogeneity if we want to measure duration dependence $f(t)$ and the coefficients. Why?

Even if $\operatorname{cov}(\theta_i, X_{i\tau})=0$ at the start of the spell, i.e. $\tau=\tau_0$ there will be dynamic selection it won't be true that $\operatorname{cov}(\theta_i, X_{i\tau})=0$ for $\tau > \tau_0$. Here is the intuition: if a high education person is still in unemployment after several periods, they probably have a bad draw on the unobserved heterogeneity, assuming education leads to a faster exit rate from an unemployment spell holding θ_i constant. This correlation will lead to biased parameter estimates.

With unobserved heterogeneity our relevant formulae become more complicated. *Basic idea* – derive everything for a given value of the heterogeneity, and then average the heterogeneity out as your last step.

Formulae

Probability a spell lasts T weeks (density function) *conditional* on the person being of type θ_j

$$h\left(T \mid \theta_j\right) = \lambda_i(T \mid \theta_j)\prod_{r=1}^{T-1}(1-\lambda_i(r \mid \theta_j)).$$

Unconditional probability of the a spell lasting T weeks

$$h(T) = \sum_{j=1}^{J} P_j \lambda_i(T \mid \theta_j)\prod_{r=1}^{T-1}(1-\lambda_i(r \mid \theta_j)),$$

i.e. average or integrate out θ_j.

Probability a spell is still ongoing after K weeks (Survivor function) *conditional* on the person being of type θ_j

$$S\left(K \mid \theta_j\right) = \prod_{r=1}^{K} (1 - \lambda_i (r \mid \theta_j)).$$

The *unconditional* Survivor function is given by $S(K) = \sum_{j=1}^{J} P_j \prod_{r=1}^{K-1} (1 - \lambda_i (r \mid \theta_j))$,

i.e. average or integrate out θ_j.

How long will a spell last on average (expected duration)

The *conditional* expression is

$$E\left(dur \mid \theta_j\right) = \sum_{r=1}^{\infty} rh\left(r \mid \theta_j\right),$$

and the *unconditional* expression is

$$E\left(dur\right) = \sum_{j=1}^{J} P_j E\left(dur \mid \theta_j\right).$$

Identification

Most of the results are for the mixed proportional hazard (MPH) model

$$\lambda_i\left(t|\theta_i\right)=\exp(\theta_i)\exp(X_i\eta)\exp(f\left(t\right)=\exp\left(\theta_i+X_i\eta+f\left(t\right)\right).$$

Basic result – the distribution of θ_i and $f\left(t\right)$ are nonparametrically identified under certainly conditions given the proportional hazard function. *However, note the MPH assumption, as well as the assumption of no time aggregation in terms of observing the exact time that a spell ends*, are identifying the model. There is no exclusion restriction here.

Time changing X's will help in the sense of getting smaller standard errors, as will having two or more spells for the same individual (as we will describe later).

Essentially the past and future time changing X's are excluded from the current hazard, giving us many exclusion restrictions. For whatever reason European studies have avoided using time changing X's. But since the model is identified formally without either of these factors, we are getting basic identification off functional form assumptions of MPH.

Hausman and Woutersen (2012) show that in *discrete time* sufficient conditions for nonparametric identification are i) weak restrictions on the form of the dependence (e.g. for a step function there are two months share the same value of the step function or ii) time changing X's. Fortunately Monte Carlo evidence suggests that discrete approach works relatively well with time constant X's. We would expect that a continuous time model with any time aggregation would have the same identification conditions.

Estimation with Time Changing X's (within a spell) and unobserved heterogeneity

External Time Changing X's – those outside the individual's control – local unemployment rate or local vacancy rate

Internal Time Changing X's – those within the individual's control, e.g.1

remaining weeks of entitlement for unemployment insurance or e.g.2 getting a penalty – have to report to a training program, when the probability of getting assigned grows with the duration of unemployment. These are likely to be correlated with the unobservables and thus can be thought of as endogenous.

We are emphasizing the value of variation in the external time changing X's. Variation over a spell in internal variables often makes things more complicated – see the discussion surrounding identifying the effect of remaining entitlement (which is a function of duration dependence) from duration dependence in Ham and Rea (1987).

Empirical Hazard and Survivor Functions

Implicitly assume only duration matters –estimate empirical hazard as

$\hat{\mu}(t)$ =number leaving at week t/number still in unemployment entering t.

Get standard errors for $\{V(\hat{\mu}(t)) = \hat{\mu}(t)[1-\hat{\mu}(t)]/\mathrm{N(t)}\}$. N(t) is the number still in unemployment entering t.

Empirical Survivor function

$$S(T) = \prod_{t=1}^{T}\left(1-\hat{\mu}(t)\right).$$

To get a confidence interval for this function we will need $Var(S(r))$ for all *r*. Can use the *delta method* to get this variance, as we will show below. In this we will want to exploit the fact that $Cov(\hat{\mu}(t), \hat{\mu}(t')) \approx 0$ - see Lancaster's book.

Estimation

Without unobserved heterogeneity estimating the model in discrete time is easy with something like Stata, no matter how many time the X's change over time as long as there is no unobserved heterogeneity. Standard errors for parameter estimates – use the standard approach in MLE.

Estimating the model with unobserved heterogeneity complicates the analysis and raises the following problems.

1. As Xianghong will show you this is a nontrivial nonlinear optimization problem. It's easy to over fit the unobserved heterogeneity. Some argue that this implies one should use objective criterion like Schwartz or AIC that penalizes parameterization to choose points of support for unobserved heterogeneity, others argue that the Log Likelihood Ratio Statistic with 2 degrees of freedom is best in serving this purpose.

2. There is an incidental parameter problem since the number of support points J goes up with the sample size.

3. Hypothesis testing – Suppose we have two points of support

And want to test $H_0 : \theta_1 = \theta_2$ then P= $\Pr(\theta = \theta_1)$ is not identified, so you end up with a non-standard testing problem see Davies (1987).

To get standard errors for the parameter estimates – again use the standard approach in MLE.

A nice simplification for policy analysis – some Monte Carlo evidence suggests that if you estimate the model with and without unobserved heterogeneity, you will get different parameters but the estimates of $\frac{\Delta ED^*}{\Delta X_k}$ will be very similar.

We will want standard errors for our estimated policy effects

$$E(dur_i \mid \theta_j) = \sum_{t=1}^{\infty} t \cdot h_i(t \mid \theta_j)$$

$$E(dur_i) = \sum_{j=1}^{J} P_j E(dur_i \mid \theta_j).$$

$$ED^* = \frac{1}{N} \sum_{i=1}^{N} E(dur_i)$$

$$\frac{\Delta ED^*}{\Delta X_k} = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta E(dur_i)}{\Delta X_k}.$$

Again it is important to note that

$$E(dur_i) \neq (dur_i \mid E(\theta)),$$

although some empirical studies do use this.

In all models we use the second derivative matrix to get standard errors – without unobserved heterogeneity, Stata will do this for you with the logit program. For standard errors of the policy effect, the expected durations are differentiable functions of the estimated parameters, with bounded and non-zero derivatives, so we can use the *delta method*

If $y = g(Z)$ and Var(Z)=Ω, then Var(y)=$\left[\frac{\partial g(Z)}{\partial Z}\right]' \Omega \left[\frac{\partial g(Z)}{\partial Z}\right]$.

Lecture 3 starts about here

Multiple Spells and Multiple Spells

Assume for *individual A*, first unemployment spell starts at the beginning of the sample, and we see two unemployment spells and one employment spell.

Assume for *individual B*, first unemployment spell starts at the beginning of the sample, and we see two unemployment spells and one employment spell.

Assume unemployment hazard

$$\lambda_i\left(t|\theta_i\right)=\frac{1}{1+\exp\left(-\theta_i-Z_{i\tau}\eta-f\left(t\right)\right)}$$

Assume employment hazard

$$\mu_i\left(t|u_i\right)=\frac{1}{1+\exp\left(-u_i-Z_{i\tau}\beta-g\left(t\right)\right)}$$

If we allow for unobserved heterogeneity, for now, assume that the unobserved heterogeneity in the employment hazard, u_i, is independent of the unobserved heterogeneity in the employment hazard, θ_i, we can analyze employment spells separately from unemployment spells.

If we assume independent heterogeneity across spells for same person there is no complication for computing since a new spell is like a new person, but this is not realistic.

A much better strategy: assume that unobserved heterogeneity term is the same across spells, then we do the history conditional on the heterogeneity and average out across the history. This really helps with identification.

Conditional on the unobserved heterogeneity we have

$$L\left(t_1,t_2\,|\,\theta_j\right)=\lambda_i(t_1\,|\,\theta_j)\prod_{r=1}^{t_1-1}(1-\lambda_i(r\,|\,\theta_j))\lambda_i(t_2\,|\,\theta_j)\prod_{r=1}^{t_2-1}(1-\lambda_i(r\,|\,\theta_j));$$

After eliminating the unobserved heterogeneity, we have the following *unconditional* contribution to the likelihood function

$$L\left(t_1,t_2\right)=\sum_{j=1}^{J}P_j\lambda_i(t_1\,|\,\theta_j)\prod_{r=1}^{t_1-1}(1-\lambda_i(r\,|\,\theta_j))\lambda_i(t_2\,|\,\theta_j)\prod_{r=1}^{t_2-1}(1-\lambda_i(r\,|\,\theta_j)).$$

An interesting special case: The Partial Likelihood and Fixed Effects

Partial likelihood is used in duration analysis. It lets you do things like estimate the coefficients on the explanatory variables without specifying the form of the duration dependence. Described in some detail in Ch. 8 of Lancaster,

I have not used it in my own work since I have not found a case previously when it would help me, given I want to allow for explanatory variables that change within a spell, flexible unobserved heterogeneity, and flexible duration dependence, not to mention multiple states, and calculate policy effects using expected durations.

However relying on this identification strategy, there is a case where it will let you relax the assumption that the explanatory variables and the unobservables are independent at the beginning of the spell, as described on p. 286 of Lancaster's book.

Specifically, if we have

i) At least two complete spells for each person;

ii) Explanatory variables that *change between spells but not within spells*.

(Note that these assumptions violate the rules that we set down for the approaches we would consider.)

Then we can estimate the coefficients on the explanatory variables that change between spells using a fixed effect model for the heterogeneity by essentially differencing out the fixed effect. This is very unusual for a nonlinear model and avoids the incidental parameter problem. You can't recover the distribution of the heterogeneity so you can't estimate the effect of a variable on expected durations, but can test the null hypothesis that the coefficient on a time changing (across spells) variable.

Left censored spells

Up until now we have been considering *flow samples*, so everyone starts in a fresh spell, i.e. the spell starts at the beginning of sampling period.

As noted above, we may look at the *stock of unemployed* – look at all those unemployed at a given time.

Also, we may simply have a *random sample,* in the US SIPP, PSID–everyone will be in the middle of either an employment or nonemployment spell (or in a marriage or out of a marriage).

In general people in the middle of spell will have a hazard that is different from fresh spells (e.g. in the flow sample) unless there is no unobserved heterogeneity and no duration dependence.

Consider first the case of no unobserved heterogeneity but duration dependence. What if we don't have unobserved heterogeneity, and know the start date of the spell, we can adjust the hazard. Suppose in the U' spell we know it started 8 months before the start of the sample. Then U' spill will contribute

$$f'(t_1) = \lambda_i(5+8)\prod_{r=1}^{4}(1-\lambda_i(r+8))$$

i.e. solve the problem by conditioning on the length of time in the spell prior to the start of the sample.

Note we don’t want to use left censored spells including time before the start of the sample to calculate average duration – will oversample long spells and this leads to ‘length biased sampling’. Long spells dominate those in unemployment at a moment of time.

Why does this oversample long spells - Consider people who start an unemployment spell 10 periods before the start of the sample. The only ones who are in a left censored unemployment spell at the start of the sample are those with duration greater than 10. Next consider people who start an unemployment spell 9 periods before the start of the sample. The only ones who are in a left censored unemployment spell at the start of the sample are those with duration greater than 9.

If we don't know how long they have been in U' before the start of the sample there is nothing much you can do to use these spells to estimate fresh unemployment spell hazard if there is duration dependence, essentially have to use a different hazard for the left censored spells. (Can't solve the problem by conditioning on the start date of the spell because you don't have this information.)

Thus if you are involved in data collection, make sure you get the start date of the spell.

With unobserved heterogeneity, can't condition on time in spell prior to the start of the sample even if you know it. The problem is that the heterogeneity distribution for a person will be conditional on time in spell prior to the start of the sample, and hence different for everyone.

One solution – throw out the left censored spells. Problem – if you are interested in long term unemployed, most of the action is in the left censored spells. Also if you want to simulate a counterfactual history for this person, you need to be able to account for the fact they start in a left censored spell.

Some studies go even further – only take fresh spells for those who have not been unemployed in the last 3 years – now you are oversampling those who do not get unemployed often, doesn't make sense.

To account for left censored spells follow Heckman-Singer's JOE84 suggestion – use a different hazard with a different heterogeneity term for the left censored spells.

Thus we will analyze 4 types of spells for these histories –

left censored employment spells,

left censored nonemployment spells,

fresh employment spells,

fresh nonemployment spells.

If the heterogeneity is independent across spells, we will be able analyze the different types of spells separately – much easier estimation problem. However, seems unreasonable that this assumption holds, especially for say left censored and fresh unemployment spells. Eberwein, Ham, and LaLonde (1997) find that treating the different types of spells as independent doesn't change their results, but more research is needed on this topic.

Competing risk models

When you can leave the current state for two or more destinations, e.g. an individual can leave nonemployment for employment and retirement.

Define the probability for leaving nonemployment for employment as

$$\lambda_{nei}\left(t\middle|\theta_{nei}\right)=\frac{1}{1+\exp\left(-\theta_{nei}-Z_{i\tau}\eta_{ne}-f_{ne}\left(t\right)\right)},$$

the probability for leaving nonemployment for retirement as

$$\lambda_{nri}\left(t\middle|\theta_{nri}\right)=\frac{1}{1+\exp\left(-\theta_{nri}-Z_{i\tau}\eta_{nr}-f_{nr}\left(t\right)\right)}.$$

Assume you can't leave nonemployment for both employment and retirement in the same period, i.e. events mutually exclusive.

Remember, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Then the leaving nonemployment is probability you leave for employment or leave for retirement – remember they are mutually exclusive so here $P(A \cap B) = 0$. Thus

$$\begin{aligned}\lambda_{ni}\left(t \middle| \theta_{nei}, \theta_{nri}\right) &= \lambda_{nei}\left(t \middle| \theta_{nei}\right) + \lambda_{nri}\left(t \middle| \theta_{nri}\right) \\ &= \frac{1}{1+\exp\left(-\theta_{nei} - Z_{i\tau}\eta_{ne} - f_{ne}(t)\right)} + \frac{1}{1+\exp\left(-\theta_{nri} - Z_{i\tau}\eta_{nr} - f_{nr}(t)\right)}.\end{aligned}$$

Note that the overall hazard does not have the logit form.

Assume that $\theta_{nei}, \theta_{nri}$ are drawn from a discrete distribution (McCall 1996) with support points $(\theta_{nej}, \theta_{nrj})$ and associated probabilities P_j. Then the conditional probability of making a transition out of nonemployment in week T is

$$h(T \mid \theta_{nej}, \theta_{nrj}) = \lambda_n(T \mid \theta_{nej}, \theta_{nrj}) \prod_{r=1}^{T-1} \left[1 - \lambda_n(T \mid \theta_{nej}, \theta_{nrj})\right].$$

The unconditional probability of the a spell lasting T weeks

$$h(T) = \sum_{j=1}^{J} P_j \lambda_n(T \mid \theta_{nej}, \theta_{nrj}) \prod_{r=1}^{T-1} \left[1 - \lambda_n(r \mid \theta_{nej}, \theta_{nrj})\right].$$

The contribution of a spell that is censored after T’ weeks is

$$S(T)=\sum_{j=1}^{J}P_j\prod_{r=1}^{T'}\left[1-\lambda_n\left(r\mid\theta_{nej},\theta_{nrj}\right)\right].$$

The contribution of a spell where the *individual leaves for employment* after T_e weeks is

$$h(T_e)=\sum_{j=1}^{J}P_j\lambda_{nei}\left(T_e\middle|\theta_{nei}\right)\prod_{r=1}^{T_e-1}\left[1-\lambda_n\left(r\mid\theta_{nej},\theta_{nrj}\right)\right].$$

Finally the contribution of a spell where the *individual leaves for retirement* after T_r weeks is

$$h(T_r)=\sum_{j=1}^{J}P_j\lambda_{nri}\left(T_r\middle|\theta_{nri}\right)\prod_{r=1}^{T_r-1}\left[1-\lambda_n\left(r\mid\theta_{nej},\theta_{nrj}\right)\right].$$

Problem: even if with no unobserved heterogeneity, or independent unobserved heterogeneity, must use nonlinear estimation not available in programs like Stata –

Competing Risks is the one example I know of where computation is easier in a continuous time model.

However you can use the approximation to simplify estimation for model selection:

$$\lambda_{ni}(t) \approx \lambda_{nei}(t) + \lambda_{nri}(t) - \left[\lambda_{nei}(t)\lambda_{nri}(t)\right],$$

i.e. ignore the fact that the transitions are mutually exclusive. Then in the absence of unobserved heterogeneity, you can estimate the hazards separately using Stata. Even adding unobserved heterogeneity that is independent across hazards, you can analyze spells separately.

Why? The likelihood will segment into a term for the transition rate to employment and the transition rate to retirement. Contribution of transition to employment is

$$h\left(T_e\right)=\lambda_{nei}\left(T_e\right)\prod_{r=1}^{T_e-1}\left[1-\lambda_n\left(r\right)\right]=\lambda_{nei}\left(T_e\right)\prod_{r=1}^{T_e-1}\left[1-\lambda_{nei}\left(r\right)\right]\left[1-\lambda_{nri}\left(r\right)\right]$$

For example, the spells where you transit to a new job are completed spells, and the spells where you go to retirement are treated as censored as are the truly censored spells. Of course, this is only approximation, but I have found it useful for exploratory data analysis and to get starting values.

Analyzing Different Types of Spells Simultaneously

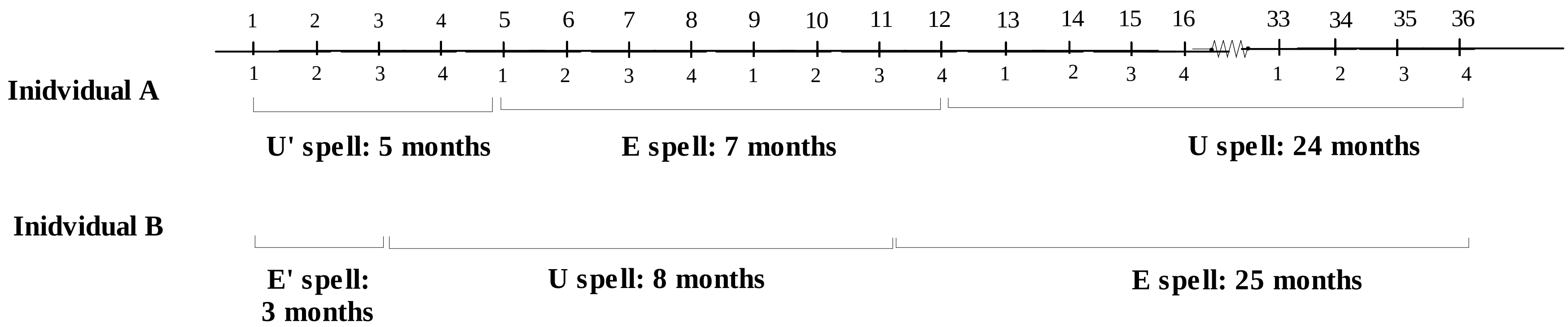
Let U' and U represent left-censored and fresh non-employment spells respectively, and let $E'$and E represent left-censored and fresh employment spells respectively.

Our hazards will take the form

$$\lambda_{ki}(t \mid X_i(\tau+t), \theta_{ki}) = \frac{1}{1+\exp\left\{-h_k(t) - X_i(\tau+t)\beta_k - \theta_{ki}\right\}},\ k = E', U', E,\ \mathrm{U}$$

As we noted above, if the $\theta_U, \theta_{U'}, \theta_E, \theta_{E'}$ are mutually independent then you can estimate the parameters of the different hazard rates separately. In the absence of this independence, we must analyze spells jointly. Following McCall's (1996) multivariate generalization of the Heckman-Singer (1984) approach, we let θ follow a discrete distribution with points of support $\theta_1, \theta_2, ..., \theta_J$, (where, e.g., $\theta_1 = (\theta_{U1}, \theta_{U'1}, \theta_{E1}, \theta_{E'1})$) and associated probabilities $p_1, p_2, ..., p_J$ respectively, where $p_J = 1 - \sum_{k=1}^{J-1} p_k$.

Example of Employment History - consider individual A below



We need to form the likelihood for A. Conditional on the unobserved heterogeneity we have

$$L_A(\cdot \mid \theta_{U'j}, \theta_{Ej}, \theta_{Uj}) = \left[\lambda_{U'}\left(5 \middle| \theta_{U'j}\right) \prod_{r=1}^{4} \left(1 - \lambda_{U'}\left(r \mid \theta_{U'j}\right)\right) \right] \cdot \left[\lambda_E\left(7 \mid \theta_{Ej}\right) \prod_{r=1}^{6} \left(1 - \lambda_E\left(r \mid \theta_{Ej}\right)\right) \right]$$

$$\left[\prod_{r=1}^{24} \left(1 - \lambda_U\left(r \mid \theta_{Uj}\right)\right) \right].$$

After averaging out the unobserved heterogeneity we have

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

Next, we need to form the likelihood for the first history B. Conditional on the unobserved heterogeneity

$$L_B(\cdot \mid \theta_{E'j}, \theta_{Uj}, \theta_{Ej}) = \lambda_{E'}\left(3 \middle| \theta_{E'j}\right) \prod_{r=1}^{2} \left(1 - \lambda_{E'}\left(r \mid \theta_{E'j}\right)\right) \left[\lambda_U\left(8 \mid \theta_{Uj}\right) \prod_{r=1}^{7} \left(1 - \lambda_U\left(r \mid \theta_{Uj}\right)\right) \right]$$
$$\prod_{r=1}^{25} \left(1 - \lambda_E\left(r \mid \theta_{Ej}\right)\right).$$

After averaging out the unobserved heterogeneity

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

Selection Issue

In principle if we treat the θ_{ki} as independent for each i when they are not, in principle there will be selection bias. Ham and LaLonde (1996) consider a case where making this independence assumption results in dramatically biased results. However, they have a very special case. As noted above Eberwein, Ham and LaLonde (1997) find that in the usual case there doesn't seem to be much bias caused by treating the θ_{ki} as independent for each i.

Policy Effects

1. Expected durations

Remember

$$\lambda_{ki}(t \mid X_i(\tau+t),\theta_{ki}) = \frac{1}{1+\exp\left\{-h_k(t) - X_i(\tau+t)\beta_k - \theta_{ki}\right\}},\ k = E',U',E,\ \text{U}.$$

So we know how to calculate expected durations:

Probability a spell lasts T weeks (density function) conditional on the person being of type θ_j

$$h_k\left(T \mid \theta_j\right) = \lambda_{ki}(T \mid \theta_{kj})\prod_{r=1}^{T-1}(1-\lambda_{ki}(r \mid \theta_{kj})).$$

How long will a spell last on average (expected duration)?

The conditional expression is

$$E\left(dur_k \mid \theta_{kj}\right) = \sum_{r=1}^{\infty} r \cdot h_k\left(r \mid \theta_{kj}\right),$$

and the unconditional expression

$$E\left(dur_k\right) = \sum_{j=1}^{J} P_j E\left(dur_k \mid \theta_{kj}\right).$$

We can differentiate each expected spell duration with respect to the explanatory variables.

The steady state fraction of time in employment *without time changing X's, duration dependence or unobserved heterogeneity is*

$$\text{FRACTE} = \text{ED}_e / (\text{ED}_e + \text{ED}_u).$$

(Initial conditions do not matter in the steady state.)

Can use this *with time changing X's, duration dependence and unobserved heterogeneity* as an approximation, *i.e. allow for these factors when calculating the expected durations.*

Again we can differentiate this with respect to an explanatory variable, and then use the *delta method* to get the standard error.

Lecture 4 starts about here

Simulating the fraction of time an individual is in employment

For a correct answer we need to simulate our model, but this will also let us look at the expected fraction of time in employment 3, 6 and 10 years out.

To do this we simulate the model to predict employment status for each person in each period. Suppose individual is in employment at the start of the spell. Start with $\theta = \theta_1$.

1. Set $\text{Emp}_i(1|\theta_1) = 1$ and calculate for individual i

$$\lambda_{E'i}(1 \mid X_i(\tau + t), \theta_{E'1}) = \frac{1}{1 + \exp\left\{-h_{E'}(1) + X_i(\tau + t)\beta_{E'} + \theta_{E'1}\right\}}.$$

2. Draw a uniform random number r from [0,1]. If

$r \le \lambda_{E'i}(1 | X_i(\tau + t), \theta_{E'1})$, move to unemployment and $\text{Emp}_i(2|\theta_1) = 0$,
$r > \lambda_{E'i}(1 | X_i(\tau + t), \theta_{E'1})$, stay in employment and $\text{Emp}_i(2|\theta_1) = 1$.

3. Do the analogous procedure for the case that arises, e.g. if you move to unemployment in 2. calculate the unemployment hazard, draw a random number from [0,1], see if it's bigger than the calculated hazard. If so stay in unemployment, if not move back to employment. Continue until all 10 years are calculated. (Same issues arise as with expected duration in terms of out of sample duration dependence and time changing explanatory variables out-of-sample.)

4. Save the employment history for individual i *and denote it by* $\text{VEmp}_{i1}(t \mid \theta_1)$

5. Repeat this process M times to obtain the histories

   $\text{VEmp}_{i1}(t \mid \theta_1), ..., \text{VEmp}_{iM}(t \mid \theta_1)$.

6. Repeat 1.-5. for individual i for $\theta_2, ..., \theta_J$.

   Next take the average

$$\text{VEmp}_i^*(t) = \frac{1}{M} \sum_{m=1}^{M} \sum_{j=1}^{J} \text{P}_j \text{VEmp}_{im}(t \mid \theta_j).$$

7. Repeat 1.-6 . for each individual i and take the average

$$\mathrm{VEmp}^{*}(t) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{VEmp}_i^{*}(t).$$

How to get standard errors? Can't use delta method since simulation is not a differentiable function. Denote the simulation as $G(\delta)$, where δ represents the full vector of parameters. Let $\hat{\delta}$ be the MLE estimates with variance-covariance matrix $\hat{\Omega}$.

Two possible Approaches:

First Approach

1. Take a draw δ_k from $N(\hat{\delta},\hat{\Omega})$.
2. Calculate the function value G_k, i.e. evaluate the function at δ_k
3. Drop the bottom 2.5% and top 2.5% of the resulting $G_k\text{'}s$.

This is used in many studies and advocated by many text books, but does not give correct confidence intervals (Ham and Woutersen 2012).

Instead Ham and Woutersen 2012 propose a **Second Approach** which is correct.

1. Take a draw δ_k from $N(\hat{\delta},\hat{\Omega})$.

2. Keep the draw if δ_k is in the 95% confidence interval of $\hat{\delta}$,

otherwise discard the draw.

3. Calculate the function value G_k.

4. The resulting distribution of G_k's is your 95% confidence interval

for $G(\hat{\delta})$, don't drop any of the G_k's.

If G is monotonic, and δ is a scaler, the two approaches will give the same answer.

In other cases, the problem with the first approach can be understood as one will probably discard extreme values of G_k that result from reasonable values of δ_k, i.e. those in the 95% confidence interval of $\hat{\delta}$. These values of G_k should be in the confidence interval for $G(\hat{\delta})$ since they occur for a reasonable value of δ_k.

The second approach can also be used when the function G is differentiable but has zero or unbounded derivatives (with respect to the parameters), i.e. the case when the delta method fails for differentiable functions. It is also appropriate when the delta method is appropriate.

Treatment Effects – Experimental and Non-Experimental Data

Ham and LaLonde (1996) and Eberwein, Ham and LaLonde (1997) consider case where there is a treatment group randomly assigned to training with $DR_{it} = 1$ for all t, and a control group $DR_{it} = 0$ for all t. Note that this insures that $DR_{it} = 1$ and θ_i are independent. We can estimate the hazards

$$\lambda_{ki}(t \mid X_i(\tau + t), \theta_{ki}) = \frac{1}{1 + \exp\left\{-h_k(t) - \gamma DR_{it} + X_i(\tau + t)\beta_k - \theta_{ki}\right\}},$$

$k = E', U', E,$ U

and there is no new issue. This is an Intent to Treat Parameter, since it won't reflect effect of actual participation. If one has non–compliance, i.e.

some of the treatment group don't take training and/or some of the control group manage to get training, it does not measure the actual effect of training. If individuals have (some) control over when they enter training (or if they enter training), we may focus on estimating

$$\lambda_{ki}(t \mid X_i(\tau+t),\theta_{ki}) = \frac{1}{1+\exp\left\{-h_k(t)-\gamma ACT_{it}+X_i(\tau+t)\beta_k-\theta_{ki}\right\}},$$

$$k = E', U', E, \text{ U}$$

where $ACT_{it}=1$ if individual i has entered training by t. Here we would consider the decision to enter training to be endogenous in the sense that there is a hazard for entering training

$$\lambda_{eti}(t \mid X_i(\tau+t), \theta_{eti}) = \frac{1}{1+\exp\left\{-h_{et}(t) - \delta DR_{it} + X_i(\tau+t)\beta_{et} - \theta_{eti}\right\}},$$

the unobservable θ_{eti} is not independent of the unobservables in the labor market hazards θ_{ki}.

One can allow for the endogeneity of ACT_{it} by jointly analyzing entering training with the labor market histories and assuming a joint distribution for $\theta_{eti}, \theta_{E'i}, \theta_{Ei}, \theta_{U'i}, \theta_{Ui}$.

What will identify the model?

1. It's a nonlinear model, so functional form assumptions are sufficient to identify the model.

2. Also, we have random assignment in our empirical setting, and we assume whether you are in the treatment or control group, only affects the labor market transition rates through actually participating in training. This is a testable assumption.

3. Finally we assume that only the current value of the time changing X's affect the labor market hazards, while the past values of the time changing X's will affect ACT_{it}.

Abbring and van den Berg model (2003) consider a similar setup where an individual will not enter training if he leaves unemployment before he is assigned to training (which is something of a black box). They work in continuous time, but it is easy to put their model in discrete time for comparability. They have a transition rate from unemployment

$$\lambda_{ki}(t \mid X_i(\tau+t), \theta_{ki}) = \frac{1}{1+\exp\left\{-h_k(t) - \gamma ACT_{it} + X_i\beta_k - \theta_{ki}\right\}}, \text{ for } k = \mathrm{U}$$

and a transition rate into training

$$\lambda_{eti}(t \mid X_i(\tau+t), \theta_{ei}) = \frac{1}{1+\exp\left\{-h_{et}(t) - \delta DR_{it} + X_i\beta_{et} - \theta_{ei}\right\}}.$$

However, only those still unemployed are eligible for training in period t. If you have a short spell of unemployment, you are less likely to be assigned to training. This creates a selection problem in addition to the endogeneity problem considered by Eberwein, Ham and LaLonde (1997). (They only consider 1 fresh spell of unemployment.)

Identification

They don't have random assignment or time changing X's. They also must assume that individuals don't anticipate when they will be assigned to training. They prove identification, but from the proof it is clear that identification is coming from functional form assumptions.

Their paper has had an extremely unfortunate effect on European empirical work in that it has lead researchers to ignore the variation provided by time changing X's, and adding time changing X's here is no more difficult than in Eberwein, Ham and LaLonde. Thus while researchers often aggressively assert that their model is identified, they don't seem to realize it is based strictly on function form assumptions. Since basing identification on functional form assumptions has proved to be a bad idea in other contexts, e.g. the Heckman-Lee selection rule without exclusion restrictions, there is no reason to believe it will work better in this more complicated setting. The one Monte Carlo Study here is not very informative, since it looks at their approach when the treatment effect is zero. The upshot is that one must take much of recent European empirical work as being of limited usefulness.

References

Abbring, J. and G. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." Econometrica, **71** 1491-1517

Davies, R.B. (1987). " Hypothesis Testing When A Nuisance Parameter Is Present Only Under The Alternative." *Biometrika* **74** 33-43.

Eberwein, C., J. Ham and R. LaLonde (2002). "Alternative Methods of Estimating Program Effects in Event History Models." *Labour Economics* **9** 249-278.

______________________________ (1997). "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence From Experimental Data." *The Review of Economic Studies* **64** 655-682.

Ham, J. and R. LaLonde (1996). "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training." *Econometrica* **64** 175-205.

Ham, J. and S. Rea (1987). "Unemployment Insurance and Male Unemployment Duration in Canada." *Journal of Labor Economics* **5** 325-353.

Hausman, J.A., and T. Woutersen (2010). "Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity." Mimeo, Johns Hopkins.

Heckman, J. J., and B. Singer (1984a). "Econometric Duration Analysis." *Journal of Econometrics* **24** 63-132.

__________ (1984b). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* **52** 271-320.

Lancaster T. (1992). *The Econometric Analysis of Transition Data* Cambridge University Press.

McCall, Brian P. (1996). “ Unemployment Insurance Rules, Joblessness and Part-Time Work.” *Econometrica* **52** 647-82.

Meyer, B. (1990). “Unemployment Insurance and Unemployment Spells.” *Econometrica* **58** 757-782