Dynamic Models
Part 1

Christopher Taber

University of Wisconsin

December 5, 2016

Survival analysis

This is especially useful for variables of interest measured in lengths of time:

- Length of life after a medical procedure
- Unemployment spell
- Life of a company
- How long someone has health insurance

Let T_i be the dependent variable we are trying to explain (kind of annoying notation in this class-it is not a treatment it is typically an outcome)

There are a whole bunch of different-and equivalent ways to characterize the distribution of T_i.

- Distribution function

$$F(t) \equiv Pr(T_i \leq t)$$

- Density function

$$f(t) \equiv \frac{\partial F(t)}{\partial t}$$

- Survivor function

$$S(t) \equiv Pr(T_i > t) = 1 - F(t)$$

- Hazard Function

$$\begin{aligned} \lambda(t) &\equiv \lim_{\delta \to 0} \frac{\Pr(T_i \leq t + \delta \mid T_i \geq t)}{\delta} \\ &= \frac{f(t)}{S(t)} = \frac{-dlogS(t)}{dt} \end{aligned}$$

- Integrated Hazard

$$\begin{aligned} \Lambda(t) &\equiv \int_0^t \lambda(s) ds \\ &= - log(S(t)) \end{aligned}$$

Hazard Function

In some ways it would be natural to just look at the data like we typically do and analyze $log(T_i)$

It is nicer to think about the model in terms of the hazard function

Reasons:

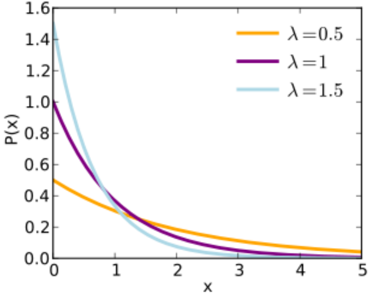
- Right truncation
- Time varying covariates
- Just easier to think about

Specifying the model

The easiest place to start is just a constant hazard rate λ which gives an exponential distribution

$$S(t) = e^{-\lambda t}$$

with pdf $f(t) = \lambda e^{-\lambda t}$



We can easily add $X's$ into this model and the most common specification is

$$\lambda_i = \exp(X_i'\beta)$$

We can allow the $X's$ to change over time but lets not worry about that yet

Cox Proportional Hazard

An issue here is that we have still restricted the model so that the hazard rate has to be constant

We relax this by letting the hazard rate take the proportional hazard form:

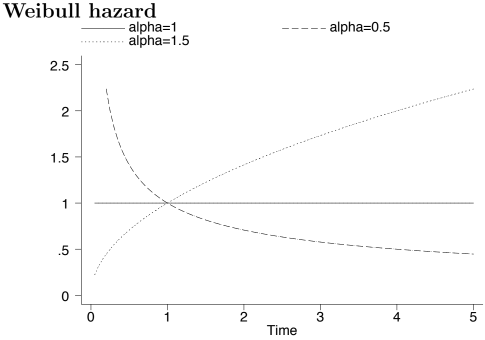
$$\lambda_i(t) = \lambda_0(t) \exp(X_i' \beta)$$

We would then typically specificy a parametric functional form for $\lambda_0(t)$

Weibull

The most common is the Weibull

$$\lambda_0(t) = pt^{p-1}$$



This also has the nice property that we can write

$$\log(T_i) = X_i' \left(\frac{-\beta}{p} \right) + \frac{\varepsilon_i}{p}$$

where ε_i is extreme value

This makes parameters easy to interpret

To see why note that the integrated hazard is

$$\Lambda(t) = e^{X_i'\beta} \int_0^t p s^{p-1} ds = e^{X_i'\beta} t^p$$

so

$$\begin{aligned} S(t) &= e^{-\Lambda(t)} \\ &= e^{-e^{X_i'\beta} t^p} \\ &= e^{-e^{X_i'\beta + p\log(t)}} \end{aligned}$$

The extreme value distribution is

$$Pr\left(\varepsilon_i < x\right) = 1 - e^{-e^{x}}$$

so if

$$log(T_i) = -\frac{X_i'\beta}{p} + \frac{\varepsilon_i}{p}$$

Then

$$\begin{aligned} Pr\left(T_i < t\right) &= Pr\left(-\frac{X_i'\beta}{p} + \frac{\varepsilon_i}{p} < \log(t)\right) \\ &= Pr\left(\varepsilon_i < p\log(t) + X_i'\beta\right) \\ &= 1 - e^{-e^{p\log(t) + X_i'\beta}} \end{aligned}$$

Estimation

Estimate the model by maximum likelihood

- When we see the end of the spell likelihood is

$$f(T_i \mid X_i; \theta) = \lambda(T_i \mid X_i; \theta) S(T_i \mid X_i; \theta)$$

- If data is right sensored at τ, likelihood is

$$Pr(T_i > \tau \mid X_i; \theta) = S(\tau \mid X_i; \theta)$$

Note that we have allowed observed heterogeneity across individuals, but no unobserved heterogeneity

Unobserved Heterogeneity

We can write the mixed proportional hazard model as

$$\begin{aligned}\lambda_i(t) =& \lambda_0(t)\exp(X_i'\beta + v_i)\\ =& \lambda_0(t)\exp(X_i'\beta)V_i\end{aligned}$$

For example hazard out of unemployment falls with t this can be due to:

- Duration dependence $\lambda_0(t)$ falls with t
- Unobserved heterogeneity

Identification

Seems subtle, but as long as V_i is independent of X_i the model

$$\lambda_i(t) = \lambda_0(t)\phi(x)V_i$$

is identified

To see why define the integrated baseline hazard as

$$\Lambda_0(t) \equiv \int_0^t \lambda_0(t)dt.$$

$$Pr(T_i > t \mid X_i, V_i) = e^{-\Lambda_0(t)\phi(X_i)V_i}.$$

Let F_V to be the distribution of V_i and define $g(\cdot) = -\log(\phi(\cdot))$

Then generalizing what we did before with the Weibull

$$
\begin{aligned}
Pr(T_i \leq t \mid X_i = x) &= \int 1 - e^{-\Lambda_0(t)\phi(x)V} dF_v \\
&= \int 1 - exp(-exp(log(\Lambda_0(t)) - g(x) + log(V))dF_v \\
&\equiv F_{v^*}(log(\Lambda_0(t)) - g(x))
\end{aligned}
$$

where F_{v^*} is defined implicitly by this relationship.

Given the separability between $log(\Lambda_0(t))$ and $g(x)$ one can show that with sufficient scale/location normalizations the model is identified

Likelihood

The likelihood function is analogous we just need to integrate across the distribution of V:

$$\int \lambda_0(T_i)\phi(X_i)Ve^{-\Lambda_0(T_i)\phi(X_i)V}dF_v$$

For data right truncated at τ

$$\int e^{-\Lambda_0(\tau)\phi(X_i)V}dF_v$$

Left truncation

We have showed that right truncation (ongoing spells at the end of the sample) is not a problem

However left truncation or ongoing spells at the beginning of the sample are a big deal

The problem is that even if I know the length of the ongoing spells, I don't observe spells that began at the same time

The distribution of V here is selected: we are going to oversample small values of V for any length of ongoing spell

You either need to throw out ongoing spells or somehow model the initial state.

(this is a case where Indirect Inference might be easier to compute then maximum likelihood)

Time Varying X's

In principle its easy, in practice it could be a pain

Let $X_i(t)$ denote X over time

Let the integrated hazard (apart from V) be

$$\Lambda_i(t) \equiv \int_0^t \lambda_0(s)\phi(X_i(s))ds$$

Then the likelihood function is:

$$\int \lambda_0(T_i)\phi(X_i(T_i))Ve^{-\Lambda_i(T_i)V}dF_v$$

For data right truncated at τ

$$\int e^{-\Lambda_i(\tau)V}dF_v$$

Competing Risk Model

Sometimes can leave a spell for more than two reasons (other than data ending):

- Can die of cancer or something else
- Can leave unemployment to OLF rather than employment
- Can die before you become married

This is easy to deal with, just define two different hazards:

- $\lambda_i^1(t)$
- $\lambda_i^2(t)$

We get to see the minimum of T_i^1, T_i^2

A lot like a straight Roy model

Lets think about this with non time varying X's but allow for unobserved heterogeneity

$$\lambda_i^1(t) = \lambda_{01}(t)\phi_1(X_{i1})V_{i1}$$
$$\lambda_i^2(t) = \lambda_{02}(t)\phi_2(X_{i1})V_{i2}$$

with $\Lambda_{01}(t)$ and $\Lambda_{02}(t)$ the corresponding integrated hazard

So we can write the likelihood as

- if leave because of risk 1

$$\int \lambda_{01}(T_{i1})\phi_1(X_{i1})V_1 e^{-\Lambda_{01}(T_{i1})\phi_1(X_{i1})V_1} e^{-\Lambda_{02}(T_{i1})\phi(X_{i2})V_2} dF_v$$

- if leave because of risk 2

$$\int \lambda_{02}(T_{i2})\phi_2(X_{i2})V_2 e^{-\Lambda_{01}(T_{i2})\phi_1(X_{i1})V_1} e^{-\Lambda_{02}(T_{i2})\phi(X_{i2})V_2} dF_v$$

- if right truncated at τ

$$\int e^{-\Lambda_{01}(\tau)\phi_1(X_{i1})V_1} e^{-\Lambda_{02}(\tau)\phi(X_{i2})V_2} dF_v$$

Example 1: Gibbons and Katz

Gibbons and Katz, “Layoffs and Lemons", Journal of Labor Economics, Oct. 1991

They write down a model of Asymmetric Information in the labor market

They compare displaced workers who lost their job either due to layoffs or plant closings

Idea is that layoff is a bad signal relative to plant closing

Table 6
Effects of Selected Variables on the Duration of the First Spell of Joblessness following Displacement from January 1986 CPS Displaced Workers Survey, Males with Only One Spell of Joblessness since Displacement
Dependent Variable = Log (Weeks of Joblessness)
Weibull Duration Model Specification

Variable	(1)	(2)	(3)	(4)
Layoff = 1	.248	.244	.352	.323
	(.086)	(.108)	(.106)	(.126)
Layoff × white collar	. . .	−.049	. . .	. . .
		(.168)		
Layoff × high union	. . .	. . .	−.299	. . .
			(.147)	
Layoff × fraction union	. . .	. . .	. . .	−.358
				(.345)
Fraction union	. . .	1.173	1.363	1.326
		(.266)	(.294)	(.033)
Previous tenure in years	.037	.034	.034	.033
	(.007)	(.007)	(.007)	(.007)
Log of previous real weekly earnings	−.301	−.339	−.331	−.333
	(.100)	(.099)	(.099)	(.099)
Weibull scale parameter (σ)	1.146	1.139	1.137	1.139
	(.033)	(.032)	(.032)	(.032)
Log likelihood	−1,831.3	−1,822.2	−1,820.2	−1,821.7

Example 2: Meyer

Meyer, “Unemployment Insurance and Unemployment Spells," Econometrica, July, 1990

Meyer looks at unemployment spells right before benefits run out

After a period of time benefits run out

He estimates baseline hazard flexibly

TABLE V

HAZARD MODEL ESTIMATES[a]

Variable	Specification (1)	(2)	(3)	(4)	(5)
Number of dependents	−.0418	−.0422	−.0416	−.0386	−.0386
	(0.0169)	(0.0171)	(0.0168)	(0.0239)	(0.0242)
1 = married, spouse present	.1302	.1221	.1315	.1006	.1001
	(0.0508)	(0.0515)	(0.0507)	(0.0722)	(0.0730)
1 = white	.2097	.2230	.2171	.2337	.2364
	(0.0572)	(0.0579)	(0.0568)	(0.0834)	(0.0841)
Years of schooling	−.0276	−.0275	−.0272	−.0177	−.0176
	(0.0083)	(0.0084)	(0.0083)	(0.0123)	(0.0124)
Log UI benefit level	−.8782	−.8157	−.8478	−.8685	−.8757
	(0.1091)	(0.1096)	(0.1088)	(0.2042)	(0.2065)
Log pre-UI after tax wage	.5630	.5651	.5530	.7289	.7411
	(0.0855)	(0.0860)	(0.0848)	(0.1415)	(0.1433)
Age 17–24	.2596	.2613	.2636	.2664	.2670
	(0.0855)	(0.0865)	(0.0855)	(0.1242)	(0.1256)
Age 25–34	.1545	.1542	.1529	.1080	.1068
	(0.0750)	(0.0759)	(0.0749)	(0.1066)	(0.1078)
Age 35–44	.1642	.1594	.1621	.1466	.1492
	(0.0776)	(0.0787)	(0.0774)	(0.1110)	(0.1122)
Age 45–54	.0473	.0417	.0460	.0234	.0239
	(0.0828)	(0.0837)	(0.0827)	(0.1156)	(0.1169)
State unemployment rate	−.0237	.0019	−.0234	.0967	.0993
	(0.0133)	(0.0126)	(0.0134)	(0.0216)	(0.0218)
Exhaustion spline:[b]					
UI 1	.6772	.6473	.5977	.7379	.6670
	(0.2470)	(0.1996)	(0.2479)	(0.2499)	(0.2513)
UI 2–5	.1288	.1468	.1665	.1448	.1847
	(0.0612)	(0.0519)	(0.0618)	(0.0625)	(0.0634)
UI 6–10	.0054	.0183	.0012	.0054	.0052
	(0.0317)	(0.0280)	(0.0317)	(0.0334)	(0.0336)
UI 11—25	−.0052	.0074	−.0067	−.0093	−.0102
	(0.0068)	(0.0063)	(0.0068)	(0.0078)	(0.0078)
UI 26–40	−.0018	.0016	−.0008	−.0001	.0015
	(0.0064)	(0.0063)	(0.0064)	(0.0074)	(0.0075)
UI 41–54	.0211	.0264	.0209	.0291	.0289
	(0.0133)	(0.0133)	(0.0134)	(0.0152)	(0.0152)
Benefits previously expected to lapse[c]			1.4643		1.6280
			(0.1876)		(0.2006)
State fixed effects	no	no	no	yes	yes
Nonparametric baseline	yes[d]	no	yes[d]	yes	yes[d]
Heterogeneity variance	[e]	[e]	[e]	.7560	.7901
				(0.1943)	(0.1953)
Sample size	3365	3365	3365	3365	3365
Log-likelihood value	−9038.07	−9085.06	−9015.68	−8927.80	−8901.94

Discrete Time Duration Models

An alternative way to model duration data is to treat time as discrete

For some of us, this is a more intuitive way to understand these models

Now T_i is an integer

We can do something analogous to the constant hazard with $X's$ by modeling the discrete hazard with a logit (or a probit or something else)

$$\begin{aligned}
\lambda_{it} &= Pr(T_i = t \mid T_i > t - 1, X_i) \\
&= \frac{1}{1 + e^{X_i'\beta}}
\end{aligned}$$

To make it like the proportional hazard model we could allow the intercept to vary with t.

More generally we can easily see how to allow both X_i and β to vary with t

$$\begin{aligned}\lambda_{it} &= Pr(T_i = t \mid T_i > t-1, X_{it}) \\ &= \frac{1}{1+e^{X_{it}'\beta_t}}\end{aligned}$$

And then adding unobserved heterogeneity is also straight forward

$$Pr(T_i = t \mid T_i > t-1, X_{it}) = \frac{1}{1+e^{X_{it}'\beta_t + v_i}}$$

We can write the likelihood without censoring as

$$\int \left[\prod_{t=1}^{T_i-1} \frac{e^{X_{it}'\beta_t+v}}{1+e^{X_{it}'\beta_t+v}} \right] \frac{1}{1+e^{X_{iT_i}'\beta_{T_i}+v}} dF_v(v)$$

and with censoring as

$$\int \prod_{t=1}^{\tau} \frac{e^{X_{it}'\beta_t+v}}{1+e^{X_{it}'\beta_t+v}} dF_v(v)$$

Identification

Cameron and Heckman, "Lifecycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," JPE April 1998

This is easiest to think about with two periods. Suppose T_i takes on three values $0, 1, 2$

$$T_i > 0 \Longleftrightarrow 1(g_1(X_{i1}) + v_{i1} > 0)$$

Conditional on $T_i > 0$,

$$T_i = 2 \Longleftrightarrow 1(g_2(X_{i2}) + v_{i2} > 0)$$

This is like the standard selection model

Can identify the model in the following way (assuming sufficient normalizations):

1. From first period identify g_1
2. Use identification at infinite to get g_2
3. Given those construct joint distribution of (v_{i1}, v_{i2})

Example: Cameron and Heckman

Same Cameron and Heckman Paper

They estimate a dynamic duration model for schooling accounting for selection in a flexible way

TABLE 4

Educational Transition Probabilities for OCG White Males Born 1937–46 (Aged 26–35 in 1973): Estimated Coefficients of Logistic Probability with a Nonparametric Heterogeneity Correction

	Complete Elementary (1)	Attend High School (2)	Graduate High School (3)	Attend College (4)	Graduate College (5)	Attend 17+ (6)
1. Number of siblings	−.101 (3.8)	−.159 (5.1)	−.175 (10.0)	−.163 (8.5)	−.175 (5.6)	−.036 (.8)
2. Family income at age 16	.201 (7.7)	.156 (5.5)	.075 (6.6)	.082 (9.6)	.054 (5.2)	.024 (1.6)
3. HGC father	.145 (4.6)	.130 (3.8)	.110 (6.3)	.120 (7.7)	.064 (3.1)	.069 (2.4)
4. HGC mother	.201 (5.9)	.117 (3.2)	.144 (7.3)	.175 (9.2)	.190 (7.0)	.207 (5.5)
5. Broken home at age 16	−.124 (.5)	−.037 (.1)	−.304 (2.2)	−.134 (1.0)	−.309 (1.5)	−.786 (2.7)
6. Farm residence at age 16	−.039 (.2)	−.089 (.4)	.432 (3.0)	−.209 (1.6)	−.109 (.6)	−.147 (.5)
7. Southern birth	−.055 (.3)	.368 (1.8)	.038 (.4)	.012 (.1)	−.235 (1.9)	−.762 (4.1)

Note.—Family income is denominated in thousands of 1995 dollars. A two-point model was deemed sufficient to characterize the heterogeneity distribution. Variable definitions: Family income at age 16 is the income of the individual's parents in the individual's sixteenth year; HGC father and HGC mother are the highest grades attained by the individual's father and mother; broken home is a binary variable indicating whether one or more of the individual's parents were absent from his household most of the time up to age 16; farm residence is an indicator recording whether the individual lived on a farm at age 16; southern birth records whether or not the individual was born in the southern census region. *t*-values are in parentheses.