Duration Data

Nathaniel Beck
Department of Politics
NYU
New York, NY 10012
nathaniel.beck@nyu.edu
http://www.nyu.edu/gsas/dept/politics/faculty/beck/beck_home.html

Jan. 2004

Introduction and analysis of exponential and Weibull models

Duration models (event history analysis) have a dependent variable which measures how long it takes for something to happen. Note this dependent variable must be positive.

Duration models are related to binary dependent variable models (which ask whether something happens) and event count models (which ask how many things happen). Thus a duration model treats how long it takes before the first event occurs, an event count model treats the number of events that occur and a binary model is appropriate for whether a single event occurs in some time interval (or for a binary variable which is one if one or more events occur). See the Alt, King and Signorino paper.

Applications

- Recidivism in criminal justice
- Length of time to complete Ph.D.
- How long an individual is unemployed
- How long an individual is married
- Lengths of coalitions
- Length of tenure on a Congressional committee
- Time until announce support of a bill
- How long a leader is in power
- How long a war lasts

- How long a peace lasts (the democratic peace)

REFERENCES

a. General

- Alison, P. Event History Analysis (Sage, 1984), easy
- Blossfeld, et al, Event History Analysis (may have new edition)
- Kalbfleish and Prentice, The Statistical Analysis of Failure Time Data (Wiley, 1980)
- Collett, D., The Analysis of Survival Data in Medical Research (Chapman Hall, 1994)
- Box-Steffensmeir and Jones, Time is of the Essence (American Journal of Political Science, Oct. 1997)
- Beck, Modelling Space and Time (in Scarbrough and Tanenbaum, Research Strategies in the Social Science, Oxford, forthcoming)

b. Econometric

- Lancaster, T., The Econometric Analysis of Transition Data (Cambridge, 1990)
- Kiefer, N., Economic Duration Data and Hazard Functions, Journal of Economic Literature, 1988
- Sueyoshi, G., A Class of Binary Response Models for Grouped Duration Data, Journal of Applied Econometrics, 1996
- Schmidt, P. and Witte, A., Predicting Criminal Recidivism Using Split Population Survival Models, Journal of Econometrics, 1989

c. Applications to Pol. Sci.

- King, Alt, Laver and Burns, A Unified Model of Cabinet Dissolution in Parliamentary Democracies, American Journal of Political Science, 1990 (34:846–871)
- Warwick, Paul, Government Survival in Parliamentary Democracies, Cambridge, 1995.
- Berry and Berry, State Lottery Adoptions as Policy Innovations: An Event History Analysis, American Political Science Review, 1990 (84:395–415)
- Bennett, Testing Alternative Models of Alliance Duration, 1816–1984 (American Journal of Political Science, July, 1997)
- Box-Steffensmeir, Arnold and Zorn, The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement (American Political Science Review, 1997) Trade

What goes wrong with linear regression

$$y_i = \vec{x_i}\beta + \epsilon_i \tag{1}$$

1. Predicted Value Could be Negative

2. Prediction Duration increases linearly with X, doubling any x doubles predicted duration

3. Like probit, effect of a change in x depends on the current expected value of y. If this is large, x has little effect. If x is small, may also have little impact, or may have huge impact.

4. (Right) Censoring. Some spells not complete at observation time

5. (Left) Censoring. Do not observe beginning of spell. Have to think how to model

The simple (exponential) parametric model

Let y_i be the duration of, say, cabinet i with characteristics $\vec{x_i}$. In regression the x's are called independent variables, in statistics (and most duration models) they are called covariates. Assume the covariates are stable over the cabinet (no time-varying covariates.) We deal with the issue of time varying covariates later.

You sometimes see t_i used for the dependent variable since the dependent variable is time. On the other hand, sometimes t is a subscript. It seems less confusing to use y.

Assume that y_i has a conditional (on its covariates) exponential distribution

$$y_i \sim fe(y_i|\lambda_i) = \lambda_i e^{-\lambda_i y_i}; \quad \lambda_i > 0. \tag{2}$$

$\frac{1}{\lambda}$ is the mean of the distribution, and hence must be positive. (Variance of distribution is square of mean.) In English, this means that coalition i will, on average last $\frac{1}{\lambda}$ months.

One way of parameterizing λ is

$$\lambda = e^{-\vec{x_i}\beta} \tag{3}$$

where β is a vector of parameters to be estimated. This is same as parameterizing expected duration

$$E(y_i) = e^{\vec{x_i}\beta} \tag{4}$$

Why this way?

Any function returning $\lambda > 0$ would do. Why this parameterization?

Think about how changes in the covariates affect expected duration. What is the effect of the same change in x when expected duration is high and low? It should be greater in the former case. (It is “harder” to move durations when they are short. For cabinets, a change from 6 months to 1 year is in some sense a bigger move than from 30 months to 3 years. Note that for some given individual

$$E(y_i) = \frac{1}{\lambda_i} = e^{\vec{x_i}\beta} \tag{5}$$

so

$$\frac{\partial E(y_i)}{\partial x_k} = \beta_k e^{\vec{x_i}\beta} \tag{6}$$

$$= \beta_k E(y_i) \tag{7}$$

which is what is wanted.
A nice way to display results is to calculate expected durations and derivatives at interesting combinations of the independent variables (or perhaps for the countries in the coalition example).

CENSORING

Right Censoring
Sometimes the data is (right) censored. Coalitions might last forever, but law demands a new coalition form every three or four or five years. Thus cannot observe a coalition longer than this length. The data are CENSORED.
(If we are studying unemployment duration, someone may still be unemployed when we do the study. That observation is censored.)
Suppose for coalition i, the censoring point (inter-election period) is C_i which is fixed and known to the investigator. Suppose an observation is censored, that is, it reaches its maximum. What is the probability of this occurring (for an exponential)?

$$\text{Prob}(y_i \geq C_i) = \int_{t=C_i}^{\infty} \lambda_i e^{-\lambda_i t} dt \tag{8}$$

$$= e^{-\lambda_i C_i} \tag{9}$$

As Warwick notes, right censoring usually occurs because of measurement issues (that is, when the study was done some units still had not failed). The King method of censoring, while I think sensible, is non-standard.

Left censoring

Suppose we have a panel and for each period (month) we ask whether someone is on welfare. Some people will start and stop a welfare spell during the analysis, so we have complete duration data. Other people will begin off welfare, but end on welfare - they are right censored and can be handled as above.

Others begin on welfare but end off welfare. They are left censored. They can be handled exactly like the right censored, since all that we know is their duration is at least how long we observed them on welfare. If we observe them being on welfare for T months, their contribution to the likelihood is just $\int_T^\infty f(t)d(t)$. (For those who start and end on welfare, all we know is that their T is the length of the panel.)

This assumes censoring is independent of everything else.

Likelihood

Thus the likelihood of the sample $y_1, y_2, \ldots\ y_n$ with known censoring points C_i is

$$\left[\prod_{y_i < C_i} f(y_i|\lambda_i)\right]\left[\prod_{y_i \geq C_i} P(y_i \geq C_i)\right] \tag{10}$$

which, for the exponential case is

$$\left[\prod_{y_i < C_i} \lambda_i e^{-\lambda_i y_i}\right]\left[\prod_{y_i \geq C_i} e^{-\lambda_i C_i}\right] \tag{11}$$

which can be rewritten

$$\prod \lambda_i^{d_i} e^{-\lambda_i y_i} \tag{12}$$

where d_i is one if the observation is not censored and zero otherwise. We can then take the log likelihood and maximize it by our usual methods, putting in

$$\lambda_i = e^{-x_i \beta}. \tag{13}$$

The Weibull model

Exponential has only one parameter. (It is a lot like the Poisson.) Its variance is just the square of its mean - cannot separately estimate a variance once we have estimated a mean.

Exponential is actually derived from a Poisson. If the number of events is described by a a Poisson, the time until the FIRST event (failure) is exponential. Thus all that you like or dislike about the Poisson carries over to the exponential. (What you like is that it is simple; what you dislike is that it is probably too simple!)

The probability of a coalition breaking up in the interval $(t_1, t_1 + t_2)$ *given that it has survived until* t_1 is the same as the probability of a breakup in the interval $(0, t_2)$. This is the "memoryless" property of exponentials, that is, probability of surviving t_2 more months is the same everywhere. This is a STRONG assumption. (Is the probability of a human surviving 20 more years the same at age 30 and age 80, conditional on making it to that age?)

The memoryless feature derives from $S(t) = e^{-\lambda t}$. Hence

$$P(S > t_2 | S | t_1) = e^{-\lambda(t_2 - t_1)} \tag{14}$$

which depends only on $t_2 - t_1$, not the specific t_1.

A more flexible model is that y has a WEIBULL distribution.

$$y \sim f_W(y|\lambda, p) = \lambda^p p y^{p-1} e^{-\lambda y^p} \quad (y, \lambda, p > 0) \tag{15}$$

where

$$\lambda = e^{-\vec{x}\beta} \tag{16}$$

$$E(y) = \frac{\Gamma(1 + \frac{1}{p})}{\lambda} \tag{17}$$

The Weibull is not symmetric, it easier to understand and compute percentiles

$$t(p) = \left[\frac{1}{\lambda}\log\left(\frac{100}{100 - p}\right)\right]^{\frac{1}{p}} \tag{18}$$

for the p'th percentile, so median is

$$= \left[\frac{1}{\lambda}\log 2\right]^{\frac{1}{p}} \tag{19}$$

At this point can do maximum likelihood just as with exponential, substituting Weibull density for exponential density for non-censored observations and the integral of the Weibull for the integral of the exponential for censored observation.

Note that the exponential is special case of ('nested in') the Weibull. When $p = 1$ the Weibull is same as exponential. Can test $H_0 : p = 1$ by usual methods ($\frac{p-1}{se.(p)} \sim N(0,1)$).

Can gauge substantive impact of Weibull by noting how expected durations (as a function of the covariates, perhaps at their mean or median) change in moving from the exponential to the Weibull. The formulae above are most useful.

The effect of a small change in any covariate is similar to what we saw with the exponential:

$$\frac{\partial E(y_i)}{\partial x_k} = \beta_k e^{\vec{x}_i \beta} \Gamma(1 + \frac{1}{p}) \tag{20}$$

$$= \beta_k E(y_i) \tag{21}$$

Note: Be careful, some programs report p, some report $\frac{1}{p} = \sigma$, some report β, some $-\beta$, some $\frac{\beta}{\sigma}$ and so forth. SO BE CAREFUL!

Hazard rate modelling

Maximum likelihood is the simplest way to look at duration models, but not the most common. Often we work with the hazard function, $h(t)$. $h(t)\Delta t$ is the probability of a coalition breaking up in the interval $(t, t + \Delta t)$, given that it survived up to t (actually in the limit as $\Delta t \to 0$). Sometimes the hazard rate is easier to model. Think about human mortality. Hazard is initially high, then declines, then rises. This is easier than thinking about the distribution of lifetimes. Same thing for coalitions and probably most other durations.

Exponential model implies constant hazard rate, failure rate doesn't change over time. Weibull has monotonically increasing or decreasing hazard, depending on p.

The hazard rate for the exponential is just $\lambda = \frac{1}{E(y)}$. For the Weibull it is

$$h(t) = \lambda p (\lambda t)^{p-1}. \tag{22}$$

Math of hazard rates

Let $F(t)$ be the distribution of coalition 'failure' times, and f(t) the associated density. (Note that here we use t for time, rather than y. This is standard, and I hope causes no confusion.) Yesterday we worked with f being the exponential or Weibull. While the distribution can be discrete or continuous, for now think of continuous distributions; we return to discrete time models later. Also assume that F is differentiable so that f exists.

$$S(t) = 1 - F(t) \tag{23}$$

is called the SURVIVOR function, that is the proportion of coalitions (or whatever) surviving past t. The hazard rate is then

$$h(t) = \frac{f(t)}{S(t)}. \tag{24}$$

Since $f(t) = F'(t)$, the numerator is simply $\lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t}$ which is the rate of failure per unit time in the infinitesimal interval starting at t. The denominator, $S(t)$ makes for a conditional rates of failure, conditional on surviving to t.

(Beware: Some authors use $\lambda(t)$ to be the hazard rate, but that is confusing because they often use $\lambda = e^{-\vec{x}\beta}$ which is only the hazard rate for an exponential model.)

Note: Since $f = F'$ and $S = 1 - F$ and $h = f/S$, there is only one free component here. Once you have defined F, the survivor and hazard functions are also defined. Conversely, once you define h, the failure and survivor functions are also defined.

To see this, we need to look at the integrated hazard. (Some authors notate by $\Lambda(t)$. A better notation is $H(t)$.)

$$H(t) = \int_0^t h(\tau) d\tau. \tag{25}$$

It is then easy to show that

$$H(t) = -\log S(t). \tag{26}$$

To see this, just remember that $\int \frac{1}{t} dt = \log(t)$, $f = F' = -S'$

$$\int h(\tau) d\tau = \int \frac{f(\tau)}{S(\tau)} d\tau \tag{27}$$

$$= -\log S(\tau) \tag{28}$$

and since $S(0) = 1$, Equation 26 follows.

Since we can also produce H if we know h, Equation 26 allows us to produce S if we know h. (With finite data or discrete distributions we will replace the integral with a sum. Some authors call H the cumulative hazard.)

Kaplan-Meier plots

If we are just interested in observing $S(t)$ or $h(t)$ for some duration variable (say coalition lengths) *without any covariates* we can use the Kaplan-Meier estimates.These are non-parametric estimates of the survivor and hazard functions.

Let exits occur at $T_1, \ldots, T_K$ and let n_i be the number "at risk" (that is, the number still "alive" and not censored at T_{i-1}) in the interval from T_{i-1} to T_i ($T_0 = 0$). Let d_i be the number that "die" in that interval (remember we only observe deaths at discrete time points).

The the probability of surviving beyond T_i is the the product of the conditional probabilities of surviving through each interval, conditional on having made it to that interval (with the probability of surviving to 0 being 1). So the KM ("product-limit") estimate of the survivor and hazard functions are

$$\hat{S}_{t_i} = \prod_{i=1}^{k} \frac{n_i - d_i}{n_i} \tag{29}$$

$$\hat{h}(t_i) = \frac{\frac{d_i}{n_i}}{t_i - t_{i-1}}. \tag{30}$$

Note: Standard errors can be calculated for these estimates (they are, after all, maximum likelihood). There are a number of different standard errors floating around, all of them more or less equally good. (There are also a number of estimators of S floating around, again all more or less equally good, differing in how they handle simultaneous (tied) exits.) STATA does a good job with all this.

Alternative Parametric Models

The monotonicity of the Weibull may be troublesome. A simple non-monotonic hazards function is the log-logistic

$$S(t) = \frac{1}{1 + (\lambda t)^p} \tag{31}$$

$$h(t) = \frac{\lambda p (\lambda t)^{p-1}}{1 + (\lambda t)^p}. \tag{32}$$

This distribution first rises and then falls if $p > 1$ and monotonically falls if $p \leq 1$ (it can not monotonically rise!). While it looks more flexible than th monotonic Weibull, it is not, since both are two parameter distributions. Once you have estimated the mean and variance of the log-logistic, its shape is fixed. (It is a bit like estimated $y = a + bx^2$ without the linear term, so where the bend occurs is fixed, not estimated.)

There are many other distributions that are used: Pareto, Gompertz, log normal, gamma. In general you can cover most possibilities with the Weibull and the log logistic. Also remember that each book seems to have its own parameterization.

Generalized Gamma

Finally, note that there are some very general distributions, like the generalized gamma, of which gamma, Weibull and exponential are special cases. The generalized gamma has two free parameters (note that the gamma, even though it has a bend, only has one, and density must be concave) with

$$f(t) = \frac{(\lambda p)(\lambda t)^{p\theta - 1} e^{(-(\lambda t)^p)}}{\Gamma(\theta)} \tag{33}$$

where λ is parameterized as usual.

If $\theta = 1$ this reduces to a a Weibull, and if $p = 1$ it reduces to a simple gamma (which in practice looks like a Weibull), and if both parameters are one, it is an exponential. Thus the generalized gamma nests the other survival distributions of interest, and can, *in principal*, be used to discriminate among them.

If $\theta = 1$ this reduces to a a Weibull, and if $p = 1$ it reduces to a simple gamma (which in practice looks like a Weibull), and if both parameters are one, it is an exponential. Thus the generalized gamma nests the other survival distributions of interest, and can, *in principal*, be used to discriminate among them.

Some Stories

Pareto results if duration is the mixture of exponentials (that is, heterogeneity in λ where the heterogeneity follows a gamma distribution. It is the length of time to failure if number of failures (event count) is modelled as negative binomial.

The gamma results as the sum of n independent exponentials with common parameter λ.

Weibull results if have n exponential components, failure when first one fails.

Which, of any of these, makes most sense for cabinet durations?

Cox Partial Likelihood

Another possibility is the Cox Partial Likelihood for proportional hazard models

$$h_i(t) = h_0(t)e^{-\vec{x}_i\beta} \tag{34}$$

where $h_0(t)$ is an unknown 'baseline' hazard which varies with time but not across individuals. This is a member of the 'proportional hazards' family since,

$$\frac{h_i(t)}{h_j(t)} = e^{(\vec{x}_j - \vec{x}_i)\beta} \tag{35}$$

and so is independent of t (always remains in same proportion for all t.

How do we estimate such a model? Cox came up with a non-parametric method called partial likelihood which turns out to be easy to estimate and is almost fully efficient. The idea is that suppose we have n individuals at risk at time t, and that i exits at time t. Given the parameters and covariates, then, conditional on either someone exiting at time t, the probability that i is the exiter is

$$PL(i) = \frac{e^{-\vec{x}_i\beta}}{\sum_{k=1}^{n} e^{-\vec{x}_k\beta}} \tag{36}$$

which then can be maximized as a function of β.

It turns out that this estimate yields asymptotically normal estimates with a covariance matrix which yields, as in maximum likelihood, asymptotic standard errors. Maximizing the partial likelihood is a lot lot maximizing the full likelihood.

Why is this so? Because the only information that Cox ignores is the timing of the exits; at the exits the method uses all the information about which unit is most likely to exit. It turns out that this is most of the information in a duration model.

Note that the Weibull model is a proportional hazards model, since

$$\frac{h_i(t)}{h_j(t)} = \left[\frac{\lambda_i}{\lambda_j}\right]^p \tag{37}$$

which depends only on the covariates but not time. Thus we note that the Cox estimates of the β are usually similar to the Weibull estimates.

Thus while proportional hazards is a strong assumption, it must be weaker than the assumptions which underlie the Weibull.

More on Ties

In continuous time there can be no tied exits. But we frequently record data discretely, so ties are very common in practice. There is some conventional wisdom that a non-trivial proportion of ties causes problems for Cox P-H, but that is not obviously the case.

What goes wrong?

Say units 1 and 2 both exit at time t. If we knew that 2 exited after 1 (say with 3 left), the contribution to partial likelihood would be

$$PL(1) = \frac{e^{-\vec{x}_1\beta}}{e^{-\vec{x}_1\beta} + e^{-\vec{x}_2\beta} + e^{-\vec{x}_3\beta}} \tag{38}$$

$$PL(2) = \frac{e^{-\vec{x}_2\beta}}{e^{-\vec{x}_2\beta} + e^{-\vec{x}_3\beta}} \tag{39}$$

$$\tag{40}$$

whereas if 2 had really exited first we would have

$$PL(1) = \frac{e^{-\vec{x}_1\beta}}{e^{-\vec{x}_1\beta} + e^{-\vec{x}_3\beta}} \tag{41}$$

$$PL(2) = \frac{e^{-\vec{x}_2\beta}}{e^{-\vec{x}_1\beta} + e^{-\vec{x}_2\beta} + e^{-\vec{x}_3\beta}} \tag{42}$$

$$\tag{43}$$

As you see, the only problem is caused by the "second" denominator. The quick and dirty method, due to Breslow, is to use the same denominator (the sum of all the $e^{-\vec{x}_i\beta}$) for all the tied exits. This is quick, and with few ties it isn't bad.

A better method is due to Efron. He argued that either 1 or 2 could have equally left first. Thus he takes the denominator for the second exit (say 2) as $.5e^{-\vec{x}_1\beta} + .5e^{-\vec{x}_2\beta} + e^{-\vec{x}_3\beta}$ with similar weightings for more than two ties.

It used to be too hard to compute the exact partial likelihood, but with modern computers and good software, this can now be done. Here we would just compute the probability of 1 and 2 both exiting at time t, given the risk set.

Testing the proportionality assumption

We can test the assumption of proportionality by seeing whether the impact of the covariates changes over time. The simplest way is to estimate a model which has

$$h_i(t) = h_0(t)e^{(\vec{x_i}\beta + \vec{x}_i t\gamma)} \tag{44}$$

which can be estimated using time varying methods to be discussed later.

If γ is not zero by a standard t-test, then we know the effect of some covariates changes with time. Here we impose a strong linear form, but we could test to see if a period dummy changes things, or any other function of time.

Estimating the baseline hazard

Once we have estimate the β we can estimate the components of h_0 (or we could do jointly, but never is done that way). Given the hazard function, and the general result on survivor functions. we have

$$S(t) = S_0(t)^{e^{-\vec{x}\beta}} \tag{45}$$

since

$$S(t) = e^{-H(t)} \tag{46}$$

but

$$H(t) = \int_0^t h(\tau)d\tau \tag{47}$$

$$= e^{\vec{x}\beta}\int_0^t h_0(\tau)d\tau \tag{48}$$

$$= e^{\vec{x}\beta}H_0(t) \tag{49}$$

where $H_0(t)$ is the "baseline integrated hazard" and $S_0(t)$ is the "baseline survivor function."

To estimate the baseline function, since exits occur only at $t_1, \ldots, t_K$, we note that maximum likelihood estimates of the baseline hazard must be zero at other than these exit points. (We thus really have a discrete failure time model, which will be discussed later.)

Letting α_k be the baseline probability of surviving at t_k, given survival through t_{k-1}, $1 - \alpha_k$ is the baseline hazard at t_k. The baselines survivor is then

$$F_0(t_i) = \prod_{j=0}^{i-1} \alpha^j \tag{50}$$

where α_0=1.

Thus, given the β, the likelihood of the data as a function of the α_k is

$$\prod_{i=1}^{K} \left[\prod_{j \in \mathcal{D}_j} (1 - \alpha_i^{e^{-\vec{x}\beta}}) \prod_{j \in \mathcal{R}_j - \mathcal{D}_i} \alpha_i^{e^{-\vec{x}\beta}} \right] \tag{51}$$

where $\mathcal{D}_i$ denotes the units dying at t_i and $\mathcal{R}_i$ denoting the units still at risk just before t_i.

This can then be maximized as a function of the α to allow for estimation of the baseline hazard function. The baseline survival function is then obtained by computing the baseline integrated hazard and exponentiating. Since the hazards are only non-zero for discrete times, the integrated hazard is obtained by simple addition (and is a step function).

In the absence of covariates, the estimated baseline survival is identical to the Kaplan-Meier estimate.

Time Varying Covariates

In the King model, duration is a function of characteristics of coalition which don't change. What if you think that a coalition fails partly as a function of the state of the economy. Then we have time varying covariates (TVC's). Usually easiest to handle in terms of Cox model, though can do with full likelihood. (May be quite tedious to setup in dataset and handle in software, but this is getting easier.)

STATA is quite good for these types of data sets. Both use a "counting process" setup, based on Flemming and Harrington, *Counting Processes and Survival Analysis*, Wiley, 1991.

In this setup each record (line of data) gives the value of covariates that are constant between a beginning and ending time point, and whether an event (failure) has occurred by the ending time point. Thus we can keep "count" of the number of failures for multiple failure data, and otherwise handle complicated data sets.

Note for relatively continuously varying measures, such as the economy, we might wish to take them as constant over a year or so in order to simplify data handling (or not, the counting process notation would have no trouble with monthly varying covariates). But covariates do have to be constant over some discrete time interval. This is not a serious limitation since data comes to us in discrete from in any event. (You can have variables which are a continuous function of time, such as time itself, but such variables are handled differently and are not TVC's.)

Beware of endogeneity or spurious causation

Once we allow time varying covariates, we must be very careful of using endogenous or jointly causal covariates. For example, in a model of marriage duration, we might use the time varying covariate, whether there are children of the marriage. But we might expect (hope!) that people decide to have children if they think their marriage is stable, so finding that children increases marriage duration might well be spurious.

Similarly, in the cabinet data, we might have a variable as to whether a key minister has resigned. But such resignations might come because the minister recognizes an imminent breakup, and so in spite of a large coefficient, the effect of this variable is spurious, not causal.

One simply must be careful here. Is the variable under the control of the unit? The economy is a pretty good exogenous measure (usually), but any strategic variable must be suspect. There is little difference between thinking about this in a duration and a simple multiple regression context. There are no statistical means to assess the suitability of time varying covariates.

We also must be careful about exactly what we mean by time varying covariates. Consider a study of length of unemployment spells, and assume that in some states people have 26 weeks of benefits, in others 39 weeks. We might thing that weeks left on benefits is a good time varying covariate. But it is not. People know the rules throughout their spell. Thus there is nothing that is changing from week to week in terms of the rules. Better would be to have a dummy variable which marks which rule a person is under. This would be a non-time varying covariate.

Estimation

Cox proportional hazards

In the Cox setup there is no reason that P(exit by i at time t_i) should not be a function of $\vec{x}_{t_i}$ rather than just a function of $\vec{x}_i$. One needs to be a bit careful here, since maybe the economy at time t has an effect with a lag, or maybe it is not the value of the covariate at time t but rather a discounted sum of past values that should be entered, but these are standard modelling issues whenever we have temporal data. So once one gets the data setup AND THE MODEL right, TVC's present no special problem. Note that with TVC's hazards are no longer proportional (that is $\frac{h_i(t)}{h_j(t)}$ will vary over time, as the covariates change over time. This is not a problem. STATA handles this setup quite well.

Parametric Methods

The TVC setup is much more complicated in the fully parametric models. This is because we don't get to just make probability calculations at exit times.

Remembering that once we know the survivor function everything we need for likelihood can be obtained, all we need do is handle the survivor function. Assume that the time varying covariates are measured discretely at intervals $t_0, t_1, \ldots, t_k$ (which might vary by individual, but let us ignore that - it is only a complication of the data setup, and the counting process data setup handles this easily enough).

By the laws of conditional probability

$$S(t_k) = \prod_{j=1}^{k} P(T > t_j | T > t_{j-1}). \tag{52}$$

Letting the hazard in the interval (t_{j-1}, t_j) depend on the value of the covariates in that interval, x_j (they are constant during that interval), each of the terms in the product in Equation 52 has the form (from the basic integrated hazard equation)

$$P(T > t_j | T > t_{j-1}) = e^{-\int_{t_{j-1}}^{t_j} h(s|x_j)ds} \tag{53}$$

where $h(s|x_j)$ is the hazard function (which depends on the covariates for the relevant subinterval).

For those units that exit at t_k, the contribution to the likelihood is $f(t_k) = S(t_k)h(t_k)$ while for those that are censored at t_k, the contribution is just $S(t_k)$.

Letting δ_i be the usual (non)censoring indicator, we thus have an individual who exits or is censored at t_k contributing

$$\log L_i = \delta_i \log h(t_k) - \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} h(s)ds \tag{54}$$

to the log likelihood. (For censored observations, with $\delta_i = 0$, this reduces to the log of the survivor function; for non-censored observations it is $\log h(t_k) + \log S(t_k)$.)

Depending on the the distribution, these integrals are more or less easy to evaluate. They are particularly easy for the Weibull and log-logistic (and Weibull with gamma heterogeneity).

Generalized Residuals

We can see how our model fits by looking at the difference between observed and predicted durations (using the formulae for $E(y_i|\vec{x}_i)$. These can be very helpful, but these differences do not have the nice mathematical properties that ordinary OLS residuals have. Cox (with Snell) has defined a "generalized residual" as some function of the data and estimated coefficients such that if the model is correct these generalized residuals should have a known distribution.

For duration models the generalized residuals are the estimated integrated hazards (that is, the log of the estimated survival function). If the model is correct, these generalized residuals should look like draws from a unit exponential distribution (that is, the observed survivals should look like draws from a uniform distribution on the unit interval). WHY?

First look at the mathematics. Say we have ANY duration random variable, t, and its true survival function, $S(t)$. Then by laws of transforming random variables we have

$$f_S(s) = \frac{f_T(S^{-1}(S(t)))}{\left|\frac{ds}{dt}\right|}. \tag{55}$$

But, by definition, $\left|\frac{ds}{dt}\right| = -f_T(t)$ so $f_S(s) = 1$ on the unit interval, that is, it has a uniform distribution.

It is easier to see this without a derivation. What value of t corresponds to $F(t) = .75$, say. It is the time at which all but 25% of subjects are dead. What is the probability of drawing a t, and hence an $F(t)$, which has $F(t) < .75$. By definition it is .75. Thus the distribution function of $F(t)$ is simply t, so F has a uniform density function. But $S = 1 - F$, so it too must have a uniform distribution on the unit interval.

Now suppose we estimate S by $\hat{S}$. If we have a good estimator, the properties of $\hat{S}$ should resemble those of S, that is, $\hat{S}$ should have a uniform density. If $\hat{S}$ has a uniform density, $\hat{H} = -\log \hat{S}$ has an unit exponential density (that is, with mean of one).

Note that this means the generalized residuals, which are simply the $\hat{H}$'s, should be draws from a unit exponential if the model is correctly specified. We can check this by looking at the moments of the generalized residuals. By construction the mean of the generalized residuals is one. But the next three central moments of the unit exponential are 1, 2 and 9. Thus we compare these to the moments of our $\hat{H}$ to test whether they are drawn from a unit exponential. The only difficulty is that some observations are censored, and so the corresponding integrated hazard is an underestimate. Jaggia, *Economics Letters*, 1991, 37:35-8 provides a test which takes into account censoring.

But the most common method for checking specification is graphical. Note that if $S(t)$

follows a Weibull,

$$S(t) = e^{-\lambda t^p} \tag{56}$$

$$\log(-\log(S(t)) = \log \lambda + p \log(t) \tag{57}$$

. For a unit exponential, both λ and p are one. Remembering that the integrated hazard is the negative log of the estimated survivor function, we thus can test for whether S follows a unit exponential by plotting the log of the integrated hazard (remembering the integrated hazard is the negative log of the survivor) against the log of time, and seeing whether the intercept is zero and the slope is one.

Here we wish to see if the generalized residuals follow a unit exponential. So we use the Kaplan-Meier estimate of the survivor function of the integrated hazards, and plot the log negative log of those estimates against the log integrated hazards, looking for a straight line with a slope of one and an intercept of zero. Departures from this indicate misspecification, and can tell us for what types of durations we have misspecification. This works for checking any duration model.

We could also plot these residuals against possible independent variables to see if they relate to survival - could also pick up non-linear effects this way.

There are lots of other residuals - modified Cox residuals which account for censoring, martingale residuals, deviance residuals and score residuals. They all are transformations of the Cox residuals. One can also use the residuals to pick up influential points (Delta beta's or dfbetas). See Collett, ch. 5.

Weibull with Gamma heterogeneity

Note that if we have heterogeneous exponential durations, then it will appear that probability of exit is declining over time. Why? Because those with highest λ will exit earliest. As time goes on, our risk set contains only those with smaller and smaller λ's.

Suppose we think that on average the Weibull model is, on average, correct, but that the survival rate randomly varies from individual to individual in an unobserved manner (that is, we won't model as a function of covariates). We can write

$$S(t|\nu) = \nu e^{-(\lambda t)p} \tag{58}$$

where ν measures random variation. ν must be non-negative and should have mean one. A distribution which leads to nice math is that ν is Gamma with mean one and variance θ. Since we only know the distribution of ν, all we can do is take expected values (that is, "integrate it out"). (THIS IS A NICE EXAMPLE OF IDEAS ABOUT MIXING DISTRIBUTIONS.) Letting someone else do the work, we get

$$S(t) = [1 + \theta(\lambda t)^p]^{-\frac{1}{\theta})} \tag{59}$$

which yields a hazard rate

$$\lambda p(\lambda t)^{p-1}[S(t)]^{\theta} \tag{60}$$

which nicely reduces to the Weibull when there is no variance in ν, that is, $\theta = 0$.

Note that using the arguments about pseudo-maximum likelihood (GMT), if you ignore the heterogeneity your parameter estimates will be right but your standard errors can be wrong. These errors can be fixed using the usual robust maximum likelihood standard errors.

Truncation

Suppose we only observe durations of a certain length, e.g. suppose unemployment spells are only recorded if they exceed four weeks. Then we know that $f(t) = 0, t <= 4$. We can use standard truncation methods to fix up the likelihood, that is, if say we think that unemployment durations are Weibull (f_W), the truncated density,

$$f_{wtrunc(t)} = \frac{f_W(t)}{1 - \int_0^4 f_W(s)ds}, t > 4. \tag{61}$$

Split Population

An interesting model has been studied by Schmidt and Witte. Modelling recidivism, they assume that of those people who have not returned to jail, some will (that is are censored) but some never will. Note that those who are already back in jail are clearly of the former type! It is easy to set this model up (and easy to estimate in LIMDEP).

Let R_i be the probability that individual i is the type who might eventually return to prison. Thus an individual will not have returned either if they are censored or of the non-returning type. The probability of this (which goes in the likelihood) is

$$(1 - R_i) + R_i\ P(\text{return after censoring point}) \tag{62}$$

and the density for those whose return is observed is just R_i times the original duration density.

It then remains to estimate the R equation as a function of covariates; this is just a logit (or probit).

Multiple Destination (Competing Risks)

So far have assumed that only possibility is that something ends (coalition breaks up, divorce, death). Only other possibility so far is right censoring.

But there may be a variety of "destinations." A member of the House and retire, be defeated for re-election, die, run for the Senate, etc. A coalition could break-up, add new members, change at election time, etc. A marriage could end in divorce, annulment or death. These are all multiple destination problems (sometimes called "competing risks" though Lancaster prefers to reserve that term for a simple subset of multiple destination problems.

How to model? Say the destinations are labelled $1, 2, \ldots, K$.

Let $h_k(t)$ be the "transition intensity" for exiting to state K at time t. (It could also be called a state specific hazard, which is why we use λ.)

These can be defined just like hazard functions, except in terms of exit to destination k in a small interval after time t. All of the other definitions in terms of survivor functions goes through analogously, and all the destination specific functions can be composed to generate unconditional hazard and survivor functions. (See Lancaster.)

The Medical Model

The simplest multiple destination model is the medical "competing risk" model. A patient could die of lots of things. Let T_k represent a patient's time until death from cause k. Then we observe the minimum of all the T_k, and we can treat the various T_k of independent (how long would it be until the patient died of prostate cancer if he hadn't died of lung cancer first seems meaningful; can we model how long a Member of the House would have served before being defeated if said Member hadn't been elected to the Senate?).

For competing risks

$$P(\text{exit to k in interval t, t+dt}) \tag{63}$$

$$= h_k(t) e^{\left[-\sum_{j=1}^{K} \int_0^t h_k(u) du\right] dt} \tag{64}$$

which is a pretty simple model.