Unobserved Heterogeneity and Frailty Models

1 Types of Duration Dependence

We have talked quite a bit about duration or time dependence so far.[1] However, we have tended to talk about it in terms of *state dependence* - the value of the baseline hazard depends in some way on previous values and/or the amount of time that has passed in a particular state. As Zorn notes, there are many examples of this type of state dependence.

- Institutionalization - institutions become 'sticky' and so the hazard of their termination drops. In this situation, we observe *positive state dependence* - the longer you are in a state, the more likely you are to stay there. This is equivalent to *negative duration dependence* - the longer that you are in the state, the less likely you are to fail

- Coalition Governments - this is the idea that as coalitions continue in office, they slowly anger the subgroups that got them elected. Over time, these groups defect, leading to higher hazards of, say, votes of no confidence. This is *negative state dependence* - the longer you've been in a state, the more likely you are to leave it. This is equivalent to *positive duration dependence*.

However, state dependence is just one of the reasons for observing duration dependence. The other is unobserved heterogeneity. In effect, there are two possible sources for duration dependence (Zorn 2000).[2]

1. State dependence or 'true' duration dependence

2. Unobserved heterogeneity or 'spurious' duration dependence

The notion of unobserved heterogeneity amounts to observations being conditionally different (heterogeneous) in terms of their hazards in ways that are unaccounted for in the systematic part of our models. All of our models so far have assumed 'exchangeability' - that observations with the same values for all covariates are otherwise identical. If this is not the case, then we have unobserved heterogeneity in the sense that some units are more prone to experience events or failures than others. This leads to a misspecification of our model, which can be particularly bad in duration models.

Zorn offers the following example for getting a better handle on exactly what goes wrong when we have unobserved heterogeneity that we ignore. Think of two groups (X_0 and X_1) with different (exponential) hazards. If we include X in our model, then this would tell us the magnitude of the difference between these two groups. However, if we don't include X, then, over time, the observations with the higher hazards (X_1) will experience the event (and exit) at a higher rate than those with the lower hazards (X_0). Thus, the estimated hazard will appear to decline over time. We might say that those observations with the higher hazard rate are more 'frail' than those with the lower hazard rates.[3] In this particular example, the unobserved heterogeneity leads to more negative (declining) hazards - this is referred to as *spurious* duration dependence. Note that this apparent duration dependence will occur even if the omitted covariate is independent of all the other covariates and/or time. Note also that this spurious duration dependence will *always* make duration dependence appear more negative than it really is. The result is that flat hazards will look as if they

[1]Notes are based on notes from Zorn, Beck, and Jones.

[2]To some extent, I take issue with calling state dependence 'true' dependence. As we have noted before, it seems reasonable to think that if we had all of the right covariates in our model, then there would be no state dependence – the hazard rate would be flat with respect to time. That we typically have state dependence in our models, then, is more the result of us never having the perfect model and not the result of there being *true* duration dependence in the real world. This is largely a semantic issue.

[3]This is where the name 'frailty' models come from.

are declining and rising hazards will appear flat or even non-monotonic. The bottom line is that if there is unobserved heterogeneity that is ignored, then model estimates will be inconsistent.

A point to take away from here is that observing, say, negative duration dependence (a Weibull model with $\hat{\rho} < 1$) may or may not indicate positive state dependence. This indicates, once again, that there is very little we can tell by simply looking at the shape of the baseline hazard. This supports our earlier statement that we might not really want to draw substantive inferences about the shape of the baseline hazard.

2 Possible Solutions

Zorn (2000) offers a number of different solutions for dealing with issues of duration dependence and unobserved heterogeneity. Below, I summarize some of them.

2.1 Model Specification

The better the model, the less unobserved heterogeneity there will be. Zorn argues that the less heterogeneity in the model, the more appropriate it is to interpret any observed duration dependence in substantive terms. However, remember the point we made earlier that if we include the right variables, it seems reasonable to think that the baseline hazard rate will be flat (no duration dependence). So, still be careful interpreting baseline hazards in substantive terms.

2.2 Models Addressing Unit-Level Heterogeneity

Since we have some observations that are more frail than others, you might think of including fixed effects i.e. dummy variables for the units. By doing so, we could get a distinct hazard rate for each unit. However, it has been shown that fixed effects are not a particularly good option in the duration model context. In particular, there is an incidental parameter problem that leads to inconsistency and a deflation of our standard errors. As a result, people don't generally use fixed effects. However, an alternative is to use random effects. Doing this leads to what are called frailty models. The basic idea is that frailty models introduce an additional random parameter into the hazard rate that accounts for unobserved heterogeneity. These models are, therefore, random-effects models just like those that you may have come across in TSCS or panel settings. These frailties may be individual-specific or group-specific. Models constructed in terms of group-level frailties are sometimes referred to as 'shared' frailty models because observations within a subgroup share unmeasured 'risk factors' that prompt them to exit earlier than other subgroups. Models based on individual-level frailties are simply called frailty models or individual-level frailty models. One assumption for these frailty models is that the random frailty terms must be independent of your covariates (it is not clear that this is actually plausible in many cases but . . .). A slightly different approach to these frailty models are so-called 'split-population' or 'cure' models that take into account that not all observations will fail; observations come from a population that is essentially split, with one group at risk of experiencing an event and another group that is not.[4] We'll evaluate frailty and split-population models in more detail in a moment.

2.3 Model the Duration Dependence

Zorn (2000) argues that if you are really interested in duration dependence, you might have hypotheses about it. For example, the democratic peace literature might lead you to think that democracies should have greater duration dependence than dictatorships because their (s)electorates are more responsive to losing

[4] Split-population models are sometimes called 'cure' models because part of the population will not fail - they are 'cured'.

military efforts than dictatorial ones. In other words, you think that duration dependence is not random due to unobserved heterogeneity but is systematic and related to state dependence. If this is the case, you could model the duration dependence directly.

One way to do this would be to allow, say, the ρ shape parameter in the Weibull model to vary as a function of some covariates. For example, you might specify $\rho = e^{X_i\beta}$ since we know that ρ is strictly greater than 0. We can then replace ρ with this new term in the usual Weibull likelihood function. This approach is equivalent to heteroskedastic models developed in the context of OLS, probit, ordered probit, and event count models that you may have come across. Variables that increase ρ cause the hazard to rise more quickly or drop more slowly. Note that it is also possible to model dependence in this way while also using a frailty model. This allows you to separate out the duration dependence that is due to unobserved heterogeneity from the duration dependence that is due to state dependence. For an empirical example of this, see Zorn (2000). You can use a model like this in STATA by using the ancillary parameter option. For example, you could type:

```
streg X, dist(weibull) nohr ancillary(Z)
```

where X and Z may be the same or different. Essentially what you are doing is allowing each observation to have a slightly different shaped (monotonic in this case) hazard rate that varies depending on the value of Z_i.[5]

3 Frailty Models

3.1 Individual Frailty

Suppose we have a sample of j observations where some observations are more failure prone due to unobserved heterogeneity (reasons unknown or unmeasurable).[6] Say, we estimated a PH duration model:

$$h_j(t) = h_0(t)e^{X_j\beta} \tag{1}$$

In this proportional hazards model, the hazard rate increases or decreases with the covariates as we have seen before. The problem is that if there are unmeasured or unobserved 'frailties', then the hazard rate will be a function of the covariates *and* the frailties.

$$h_j(t) = h_0(t)e^{X_j\beta + W_j\psi} \tag{2}$$

where W_j is a frailty term from a probability distribution with a mean of 0 and a variance of 1. Obviously, if $\psi = 0$, then the standard PH model is obtained. Also, if we could measure and include W_j in our model, then ψ would again go to 0.

It is perhaps useful to think of the frailty model as being a misspecified model. If we had the correct specification, then ψ would be zero in expectation. If some assumptions are made about the distribution of the frailty term, we can attempt to take account of the misspecification. Let's start by rewriting the hazard in the following form:

$$h(t)_j = h_0\nu_j(t)e^{X_j} \tag{3}$$

where $\nu_j = e^{W_j\psi}$. You can see from this that the frailty term acts multiplicatively on the hazard rate. The hazard rate is now conditional on both the covariates and the frailty term. For identification purposes, we

[5] Simply putting Z in the main bit of the model rather than the ancillary bit would control for Z but would mean that the shape of the hazard is the same for all observations.

[6] The following set of notes follows Brad Jones' notes and Keele (2007).

assume that the mean of $\nu = 1$ and that the variance is unknown and equal to some parameter θ. As Jones notes, we always make assumptions about ν whether we use frailty models or not. When we don't take account of frailty, we are essentially assuming that $\nu = 1$ with probability 1.

If the hazard is a function of the frailties, then the survivor function is also conditional on both the covariates and the frailty term. Thus, we have the conditional survivor function as:[7]

$$\begin{aligned} S(t, X, \nu) &= \exp\left(-\int_0^t h(u|\nu)du\right) \\ &= \exp\left(-\nu\int_0^t h(u)du\right) \end{aligned} \tag{4}$$

To derive the expected value of the survivor function, we need to specify a probability distribution for ν_i. Let's call this $g(\nu)$. We could use any continuous distribution with positive support, a unit mean, and a finite variance θ – the gamma, inverse Gaussian, log-normal etc. The gamma is the most commonly used distribution in the literature. As is true with any random effects model, integrating over the assumed distribution provides an expectation that is no longer conditional on the unobserved portion of the model. The same is true for frailty models. With the adoption of a distribution $g(\nu)$, the expected survivor function can be derived from the hazard rate as follows:

$$\begin{aligned} S(t) &= E[S(t, X, \nu)] \\ &= E\left[\exp\left(-\nu\int_0^t h(u)du\right)\right] \\ &= \mathrm{L}\left[\exp\left(\int_0^t h(u)du\right)\right] \end{aligned} \tag{5}$$

where L is the Laplace transformation. The use of the Laplace transformation is required to integrate out the distribution of the unobserved frailty. The function shown in Eq. (5) is commonly referred to as the 'marginal survivor function'. As Keele (2007) notes, once the frailty is integrated out, accounting for unobserved heterogeneity is reduced to estimating θ, the variance of the frailty term. This is the marginal survivor function because it is the observed survivor function after the ν has been integrated out. The bottom line is that so long as we assume that ν_i has some distribution, then we can estimate the frailty model. In effect, all we have to do is estimate the frailty variance term θ.

To sum up, with unobserved heterogeneity, you have a mixture of hazards across different units. It is this that might lead you to estimate a frailty model. Frailty models are essentially random effects models for the survival setting. Frailty 'terms' explicitly account for the extra variance associated with unmeasured risk factors. To obtain a model, we need to make the usual assumptions about which model to pursue and then make the added assumption about $g(\nu)$. The problem with ignoring frailty is seen in the hazard. In the PH models, the hazard is a multiplicative function of the covariates. If there is frailty, the hazard is also a function of ν. To make our problem tractable, we integrate out ν and so we're left with the problem of estimating the variance, θ.[8]

3.1.1 Weibull

Consider a Weibull frailty model. The conditional survivor function is:

$$S(t|\nu) = e^{-(\nu\lambda t)^{\rho}} \tag{6}$$

[7]This equation is one that we have seen earlier and flows from the fact that $S(t) = \exp[-H(t)]$.

[8]The exact process by which frailty models are estimated is quite complicated. For more information see, Therneau and Grambsch (2000).

With the exception of the frailty term, ν, this is identical to the normal Weibull survivor function that we have seen before.

Now suppose that the gamma distribution is specified for $g(\nu)$. With gamma frailty, the marginal Weibull survivor function is:

$$S(t) = [1 + \theta(\lambda t)^{\rho}]^{\frac{-1}{\theta}} \tag{7}$$

and the hazard rate is now:

$$h(t) = \lambda\rho(\lambda t)^{\rho-1}[S(t)]^{\theta} \tag{8}$$

When the variance of the frailty is 0, the model reduces to the normal Weibull model.

We can evaluate the hypothesis that $\theta = 0$ and determine whether we need to worry about unobserved heterogeneity using a likelihood ratio test. The likelihood ratio test statistic for the frailty term is twice the difference between the log partial-likelihood with the frailty integrated out (often referred to as the integrated likelihood) and the loglikelihood for the model without a frailty term. The usual practice is to drop the frailty term if the likelihood ratio test statistic is not greater than the typical test critical values for p-values of 0.05.

3.1.2 Results

Let's now look at some results from an individual frailty model using our government duration data from before. Suppose that the correctly specified model is a Weibull model where we type:

```
streg investiture polarization formation_attempts postelection
        caretaker opposition_party , dist(weib) nohr;
```

These results are shown in column 1 of Table 1.

Table 1: Various Weibull Frailty Models

	Full	Full (gamma)	Omit POSTELEC (gamma)	Omit POSTELEC (inverse gaussian)
Investiture	0.27*	0.28*	0.44**	0.45**
	(0.14)	(0.15)	(0.18)	(0.19)
Polarization	0.03***	0.03***	0.04***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Formation Attempts	0.13**	0.13**	0.11*	0.10*
	(0.04)	(0.05)	(0.06)	(0.06)
Post-Election	-0.85***	-0.87***		
	(0.14)	(0.15)		
Caretaker Gov.	1.91***	1.94***	2.17***	2.21***
	(0.23)	(0.31)	(0.38)	(0.40)
Opposition Strength	-0.22	-0.20	-0.32	-0.39
	(0.38)	(0.43)	(0.52)	(0.53)
Constant	-3.95***	-4.04***	-4.71***	-4.68***
	(0.39)	(0.53)	(0.62)	(0.63)
Shape parameter ρ	1.24	1.27	1.36	1.40
	(0.06)	(0.11)	(0.12)	(0.17)
Frailty parameter θ		0.05	0.38	0.67
		0.18	(0.21)	(0.65)
Observations	314	314	314	314
Log-Likelihood	424.33	424.28	-440.77	-440.87

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed); Standard errors are given in parentheses

We could check for individual level frailty in this model. Supposed we used the gamma distribution to model our frailty. We would type:

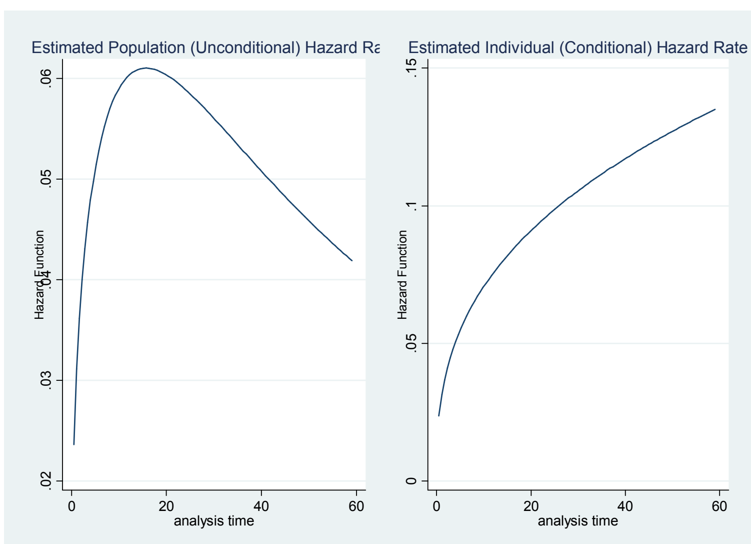
```
streg investiture polarization formation_attempts postelection
        caretaker opposition_party, dist(weib) nohr frailty(gamma);
```

These results are shown in column 2 of Table 1. To check to see if we have any unobserved heterogeneity at the individual level, we would look at the results from a likelihood ratio test that STATA reports (results not shown here) - we would see that there is no evidence of frailty. Now, let's induce a model misspecification by omitting the POSTEL variable. In columns 3 and 4, I show the results from this model where I use a gamma and then an inverse Gaussian distribution to model the frailty. A likelihood ratio test for both models indicate that there is now unobserved frailty. The results do not differ markedly if I use the gamma instead of the inverse Gaussian or vice versa.

3.1.3 Individual vs. Population Hazards

In frailty models, there is a distinction between the hazard rates that individuals face and the population hazard rate that arises by averaging over all the survivors. In a normal PH model, these hazards are the same thing since all individuals are assumed to be identical. In a heterogeneous population, though, it turns out that the population hazard can fall while the individual hazards all rise. This is because, over time, the population becomes populated by more and more robust individuals as the frail members fail. This is the frailty effect we talked about earlier; it virtually assures that population hazards decline over time regardless of the shape of hazards that individuals face. One thing it means, though, is that you need to be careful if you decide to graph the hazard - you need to know which one (individual or population) that you have graphed.

Figure 1: Individual and Population Hazards from a Weibull Model with Gamma Frailty



To obtain the population (or unconditional) hazard, you would type:

```
stcurve, hazard unconditional;
```

To obtain the mean individual (or conditional) hazard, you would type:

```
stcurve, hazard alpha1;
```

These figures for the gamma frailty model are shown in Figure 1. As you can see, the population hazard does decline over time, whereas the individual hazard continues to climb.

3.2 Shared Frailty

The main difference between 'shared' and 'unshared' frailty models is the assumption of how frailty is distributed in the data. Shared frailty models assume that similar observations share frailty, even though frailty may vary from group to group. For example, some countries might be more prone to war than others for unobserved reasons. In effect, shared frailty causes observations within the same group to be correlated.

Suppose we have j observations and i subgroups. The hazard rate for the j^{th} individual in the i^{th} subgroup is:

$$h_{ij}(t) = h_0(t)e^{(X_{ij}\beta + W_i\psi)} \tag{9}$$

where W_i are the subgroup frailties which, as before, are assumed to be independently distributed with a mean of 0 and a variance of 1. Again, if $\psi = 0$, then we have the normal PH model. We can reexpress the hazard as:

$$h_{ij}(t) = h_0(t)\nu_i e^{X_{ij}\beta} \tag{10}$$

where $\nu_i = e^{W_i\psi}$. The only difference with the individual frailty models is that frailty is now shared among the j observations in the i^{th} group. To get a model, we do exactly as before i.e. we make assumptions about $g(\nu)$.

3.2.1 Results

The idea with shared frailty is that groups have different frailties. For example, different regime types such as dictatorships and democracies may have different frailties. Let's look at data on states' adoption of anti-obscenity legislation. How many legislative sessions does a state go before it adopts a certain kind of policy? Let's start with a basic Weibull model where the covariates are FUNDAM SOUTH COLLC CITIDEO MURDERRATE LEGISLATIVEIDEOLOGY WOMLEG. The results from this model are shown in column 1 of Table 2. One might think that each state has different frailties that are not being captured by our covariates. Thus, we could estimate a shared frailty Weibull model where the frailties are shared on the states. Thus, we would type:

```
streg fundam south collc citideo murderrate legislativeideology
        womleg, nohr dist(weib) shared(statnam) frailty(gamma);
```

These results are shown in column 2 of Table 2. Columns 3 and 4 show the results from a Cox model and from a shared frailty Cox model. Again, you would determine the existence of shared frailty using a likelihood ratio test.

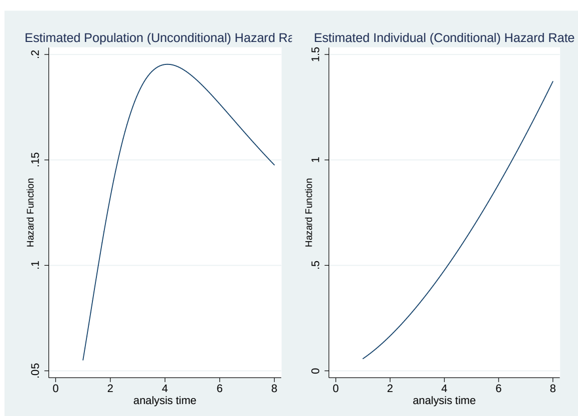
Table 2: Various Shared Frailty Models

	Weibull	Weibull (gamma)	Cox	Cox (gamma)
Fundamentalists(%)	0.20**	0.13	0.18**	0.16
	(0.09)	(0.19)	(0.09)	(0.20)
South	0.20	0.95	0.19	0.83
	(0.27)	(0.70)	(0.27)	(0.69)
College Graduates	0.13***	0.20**	0.12***	0.19**
	(0.04)	(0.08)	(0.04)	(0.08)
Ideology	0.06***	0.05**	0.05***	0.04**
	(0.01)	(0.02)	(0.01)	(0.02)
Murder Rate	0.19***	0.15**	0.16***	0.14**
	(0.03)	(0.07)	(0.03)	(0.07)
Legislative Ideology	-0.04***	-0.04***	-0.03***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
Female Legislators (%)	-0.01	0.01	-0.01	0.02
	(0.02)	(0.03)	(0.02)	(0.03)
Constant	-7.84***	-9.41***		
	(0.90)	(1.92)		
Shape parameter ρ	1.78	2.53		
	(0.14)	(0.21)		
Frailty parameter θ		1.91		1.92
		0.52		(0.58)
Observations	400	400	400	400
Log-Likelihood	-270.48	-189.88	-477.21	-454.30

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed); Standard errors are given in parentheses

We could also plot the individual and population hazards from these models. Figure 2 shows the figures from the Weibull model with shared gamma frailty.

Figure 2: Individual and Population Hazards from a Weibull Model with Shared Gamma Frailty



You can also ask STATA to create a variable that captures the frailty estimates (log frailties) by adding , EFFECTS(NAME) to the STCOX command:

```
stcox X, efron shared(var) frailty(dist) nohr effects(name);
```

```
graph twoway scatter frailty stcd21, mlabel(statnam) yline(0);
```

For example, the frailty terms for the 50 states in our anti-obscenity legislation example are shown in Figure 3. Cases above the 0 line are the most failure-prone.

Figure 3: Frailty Terms for 50 States

3.3 Some Conclusions on Frailty Models

Frailty models can help us when we have unobserved heterogeneity. However, there are some issues to be aware of. First, neither theory nor data typically provides much guidance for choosing a specific distribution from which to draw the frailty term and the parameter estimates can be very sensitive to the assumed parametric form. Second, you should also note that if the frailty effect is real, then the usual interpretation given to your models may not be appropriate. In effect, we lose the normal proportional hazards property because the hazard ratios are now conditional on the unobserved frailty. As θ approaches 0, the proportional hazards property returns. Very few scholars actually provide much substantive interpretation or calculations of quantities of interest from frailty models above and beyond simply interpreting the sign and significance of the coefficients. I, and others who I have spoken to, have little advice on how to provide much substantive interpretation beyond this.

4 Split-Population or Cure Models

Split-population or cure models do not assume that all observations will eventually fail or experience an event. This is an assumption that we have implicitly been making in all of the models to this point. The split-population 'splits' the population into two groups - one that will experience the event and the other that will not (the cured). Let's start with some notation.[9] As we've had before, let our duration variable be t_i, our censoring variable be C_i[10], the density be $f(t)$, the CDF be $F(t)$, $S(t)$ be $1 - F(t)$, and $h(t)$ be $\frac{f(t)}{S(t)}$. Now we are going to add some more terms. Think of an unobserved variable Y_i such that $Y_i = 1$ for those who will eventually experience an event and $Y_i = 0$ for those who will not. Let $\Pr(Y_i = 1) = \delta_i$. Now let

- $f(t_i|Y_i = 1) = g(t)$

[9]Here I follow Zorn's notes.

[10]Box-Steffensmeier and Jones have this as R

- $F(t_i|Y_i = 1) = G(t)$
- $S(t_i|Y_i = 1) = S'(t)$

These are just the density, distribution, and survival functions conditional on being in the subpopulation that will experience an event.[11].

We don't actually observe Y_i, but we do observe whether or not some observation has an event - this is our censoring variable C_i.

1. For those observations which experience the event, we observe $C_i = 1$ and t_i.
   - For these observations, we know that $Y_i = 1$
   - We can write the unconditional density for the uncensored observations as:

   $$\text{Pr}(Y_i = 1)\text{Pr}(t_i \leq t_C|Y_i = 1) = \delta_i g(t_i) \tag{11}$$

   where t_C is the censoring time

2. For the other observations, we don't observe a failure time (C_i=0).
   - This may occur either (i) because $Y_i = 0$ (the observation will never have the event) or (ii) because $t_i > t_C$ (the observation is censored).
   - The unconditional density for these observations ($C_i = 0$) is:

   $$\text{Pr}(Y_i = 0) + \text{Pr}(Y_i = 1)\text{Pr}(t_i > t_C|Y_i = 1) = (1 - \delta_i) + \delta_i S'(t_i) \tag{12}$$

Combining these unconditional probabilities, we have

$$\mathcal{L} = \prod_{i=1}^{N}[\delta_i g(t_i)]^{C_i}[(1 - \delta_i) + \delta_i S'(t_i)]^{1-C_i} \tag{13}$$

All we've done is mix the probabilities across the two sets of populations according to a mixing probability defined by δ. The probability of δ_i is typically modeled as a logit or probit i.e.

$$\delta_i = \frac{e^{Z_i\gamma}}{1 + e^{Z_i\gamma}} \tag{14}$$

We estimate our model as normal. The $\hat{\beta}$'s are the usual estimates, but conditional on the observation being in the subset that will eventually experience the event. The $\hat{\gamma}$'s are the effect of the Z's on the probability that the observation will either be 'cured' or 'not cured'. When $\delta_i = 1$, the equation reduces to the standard duration model with censoring. You can test to see if you should be running a split-population model or not simply by seeing whether δ is significantly different from 1.

This model has been used by Milan Svolik in a recent APSR article to examine democratic consolidation and democratic collapse. Over the years, much has been written about when we know that a democracy has been consolidated. One way to answer this question is, in some sense, to let the data speak for themselves. A consolidated democracy is one that has been 'cured' and is no longer at risk of democratic collapse. In effect, democracies can be split into two populations - those that might collapse and those that are consolidated. By ignoring the fact that there are two types of democracies in the world, previous studies of democratic

[11] $f(t_i|Y_i = 0)$, $F(t_i|Y_i = 0)$, and $S(t_i|Y_i = 0)$ are undefined

collapse may have drawn incorrect inferences. Svolik uses a split-population model to examine those factors that influence democratic consolidation and democratic collapse in the same model.

The split-population model is not canned in STATA. You could program it yourself. For help doing this, see the replication files provided by Svolik on his webpage. Alternatively, there are two ado files that you can download. First, there is the SPSURV ado file from Jenkins (2001). This is a discrete time split-population model that uses a cloglog link function. To download this, go to `http://ideas.repec.org/c/boc/bocode/s418601.html`. Second, there is the LNCURE ado file from Cleves (2000). This is a continuous time log-normal model. To download this, go to `http://www.stata.com/users/mcleves/lncure/`. Take a look at the actual ado files to see how they are programmed and read the help files.

References

Keele, Luke. 2007. "Cross Validation Tests for Frailty Models." Unpublished manuscript, Ohio State University.

Therneau, Terry M. & Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.

Zorn, Christopher J. W. 2000. "Modeling Duration Dependence." *Political Analysis* 8:367–380.