Competing Risks

1 Introduction

To this point, we have only looked at 'single-state' processes i.e. we have assumed that each unit is at risk of only one event at any one time. However, it may be the case that a unit is at risk of experiencing multiple events at any one time. In other words, a unit might be at risk of failing in multiple different ways or, equivalently, transitioning to multiple different states. In the case of government duration, for example, a government may collapse in one of two ways: (1) dissolution (new elections) or (2) replacement (no new elections) (Diermeier & Stevenson 1999). In the case of government duration, we have essentially been treating these two types of failure as the same i.e. we have assumed that dissolution is the same kind of event as replacement. However, this decision to 'pool' these two types of events may be inappropriate. In other words, we might think that the risk of experiencing dissolution is different to the risk of experiencing replacement. Put differently, we can think that governments face 'competing' risks – a risk of dissolution and a risk of replacement. Competing risks (or multiple destination) models deal with cases of multiple failure types (destinations).[1]

It is relatively straightforward to reexpress a hazard rate in the continuous time setting to take account of competing risks. For example, if a unit was at risk of experiencing K events ($k = 1, 2, 3, \ldots, r$), then the hazard rate would be:

$$h_k(t, X) = \lim_{\triangle t \to 0} \frac{Pr(t \leq T \leq t + \triangle t | T \geq t, X)}{\triangle t} \tag{1}$$

The only difference from the hazard rate when we had only one risk is that the hazard rate is now subscripted by k. In other words, we have a different hazard rate for each type of event.

We are going to look at three different approaches to estimating competing risks models:

1. Stratified Cox Model
2. Latent Survivor Time Approach
3. Multinomial Logit

The main difference between these approaches is that the latent survivor time approach and the multinomial logit explain heterogeneity across different event types in terms of covariates whereas the stratified Cox model explains heterogeneity in terms of differing baseline hazards. This point should become clearer as we examine each of the three competing risks models.

2 Stratified Cox Model

The basic assumption of this model is that units are at risk of experiencing k possible (unordered) events and that each unit is at risk of experiencing *all* of the events. In other words, if there are five possible events, then each unit could experience, 1, 2, 3, 4, or all 5 events. This has implications for the structure of the data. In effect, each unit will have at least k records of data. So a unit will have 5 records of data if it is at risk of experiencing five events - one record for each possible event. In his lecture notes, Jones provides an example of the data structure for a competing risks model using the stratified Cox approach. This is shown in Table 1. Units are observed for 14 periods and are at risk of five different events: 1, 2, 3, 4, 5. As you

[1] The following notes are based heavily on Zorn and Jones.

can see, unit 1 experiences three events – it experiences event 1 in period 6, event 3 in period 8, and event 5 in period 11. At the end of the last time period (14), unit one had still not experienced events 2 or 4. Unit 2 only experiences one event – it experiences event 5 in period 2. At the end of the last time period (14), unit two had still not experienced events 1, 2, 3, or 4.

Table 1: Example of Data Structure for a Competing Risks Model – Stratified Cox Model

Case ID	Event Time	Event Occurrence	Event Type
1	6	1	1
1	8	1	3
1	11	1	5
1	14	0	2
1	14	0	4
2	2	1	5
2	14	0	1
2	14	0	2
2	14	0	3
2	14	0	4

The central assumption behind the stratified Cox model is that the covariate effects are the same for each event type. In other words, the coefficients on the covariates are the same irrespective of whether we are looking at the risk of event 1, the risk of event 2, the risk of event 3, the risk of event 4, or the risk of event 5; only one set of parameters is estimated. The only way in which the stratified Cox model allows for heterogeneity in terms of the risk of different events is through different baseline hazards. In other words, the stratified Cox model 'stratifies' on the baseline hazards and thereby allows the baseline hazards to differ across events; there is a unique baseline hazard for each type of event. This is why people say that the stratified Cox model 'sweeps' heterogeneity into the baseline hazards. To estimate the stratified Cox model, you would type something like the following:

```
stcox X, exactp nohr strata(event type);
```

You would then interpret the coefficients from the stratified Cox model in exactly the same way as you would the coefficients from a normal Cox model.

In sum, the stratified Cox model approach is probably most useful when you do not think that the covariate effects will differ significantly across the different event types. Of course, you should probably evaluate whether the covariate effects do differ across event types and not simply assume this. As we will see, you can do this by adopting a latent survivor time or multinomial logit approach to competing risks.[2]

3 Latent Survivor Time Approach

Let's assume that observation i is at risk of k different kinds of events. Each event type has a corresponding duration $T_{1i}, T_{2i} \ldots T_{ki}$ associated with it, a corresponding hazard function $h_{ki}(t)$, and a corresponding survivor function $S_{ki}(t)$. We don't observe all of the duration times - we only observe the shortest one: $T_i = \min(T_{1i}, T_{2i} \ldots T_{ki})$. In effect, T_{ki} are potential or latent failure times. Although latent, these unobserved

[2]Unfortunately, I do not have a data set in the correct format for you to do exercises using the stratified Cox model. If anyone has a data set in the correct format and would like to share, please send it to me at `mgolder@fsu.edu`.

failure times are assumed to exist and would be observed if time went on long enough without the unit failing from some other type of event first. In this approach, we also observe an indicator of which event the observation experiences i.e. $R_i = k$ iff $T_i = T_{ki}$. To keep things simple, we tend to assume that the duration times for different events are not *exactly* the same i.e. units can't fail in two different ways at exactly the same time.

If the k various risks are conditionally independent, estimation is easy. The contribution of each uncensored observation to the likelihood is:

$$\mathcal{L}_i = f_k(t_i, X_{ki}, \beta_k) \prod_{k \neq r} S_r(t_i | X_{ki}, \beta_r) \tag{2}$$

where the subscript k denotes the k^{th} event and the r in the product term implies that the product is taken over the survivor times for all states except k. In other words, the contribution of a given observation with failure due to risk k to the likelihood function is identical to its contribution in a model where only failures due to risk k are observed and all other cases are treated as censored. Thus, if risks j and k are independent of one another, then we can analyze durations resulting from risk k by treating those failures due to risk j as censored - in the sense that they have not (yet) reached their theoretical time to failure from risk. Note that β is subscripted by k and so you get a different set of coefficients for each type of failure. This marks a big difference with the stratified Cox model where you only got one set of parameters. By providing a different set of coefficients for each type of failure, the latent survivor time approach captures heterogeneity across different types of events in terms of the covariates.

The likelihood for the full sample can be written as:

$$\mathcal{L} = \prod_{i=1}^{N} f_k(t_i, X_{ki}, \beta_k) \prod_{k=1}^{r} S_k(t_i, X_{ki}, \beta_k) \tag{3}$$

However, since only one failure among the k possible outcomes is observed per unit, the overall likelihood can be partitioned in terms of the number of units failing by each of the k outcomes:

$$\mathcal{L} = \prod_{k=1}^{r} \prod_{i=1}^{N_k} f_k(t_i, X_{ki}, \beta_k) S_k(t_i, X_{ki}, \beta_j) \tag{4}$$

The partitioning is easier to see if we define a censoring indicator δ_{ki} such that $\delta_{ki} = 1$ if i failed due to k and 0 otherwise. Incorporating δ_{ki} into the likelihood function, we have:

$$L = \prod_{k=1}^{r} \prod_{i=1}^{N_k} f_k(t_i, X_{ki}, \beta_k)^{\delta_{ki}} S_k(t_i, X_{ki}, \beta_j)^{1-\delta_{ki}} \tag{5}$$

3.1 Estimation

The latent survivor time approach to competing risks essentially involves estimating K models where all events other than k are assumed to be randomly censored. How does this work exactly? Let's estimate a competing risks model of government survival. Government's can fail by being replaced (no elections) or through dissolution (elections). Say we are interested in modeling the survival of governments that form after elections with the following covariates: PEC MAJORITY2 SINGLEPARTY RANGE CARETAKER INCBARGAIN ENPP2 INVESTITURE. We first need to STSET the data. For our data, this means typing:

```
stset duration, failure(ciep365==0);
```

Let's start by assuming that all risks are the same and estimate a normal Cox model.

```
stcox pec  majority2 singleparty range  caretaker incbargain enpp2
    investiture if postelection==1, efron nohr cluster(country1);
```

These results are shown in column 1 of Table 2.

Table 2: Determinants of Government Survival

Dependent Variable: Duration of Government Survival in Days
for Governments Forming after Elections (Cox Model)

Independent Variables	Pooled PDDA	Dissolution PDDA	Replacement PDDA
Government Attributes			
Pre-Electoral Coalition	0.44	1.06*	0.29
	(0.27)	(0.62)	(0.38)
Majority Government	-0.90***	-1.54***	-0.71***
	(0.30)	(0.52)	(0.24)
Single Party Government	0.11	0.68	0.10
	(0.33)	(0.67)	(0.38)
Ideological Range	0.001	-0.001	0.01
	(0.01)	(0.01)	(0.01)
Caretaker Government	2.53***	———	2.33***
	(0.52)		(0.65)
Formation Attempts	0.09	0.10	0.10
	(0.09)	(0.16)	(0.06)
Legislature Attributes			
Legislative Fragmentation	0.14	0.15	0.19*
	(0.11)	(0.14)	(0.11)
Country Attributes			
Investiture	0.16	0.15	0.04
	(0.20)	(0.43)	(0.30)
Log likelihood	-629.62	-217.75	-599.66
Observations	208	208	208

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)

NOTES: Results are Cox proportional hazards estimates where the Efron method was employed for handling ties; robust standard errors clustered by country in parentheses. **Replacement**: The government ended in a replacement government without new elections. **Pooled**: The government ended in dissolution or replacement.

In order to estimate a competing risks model, we have to re-STSET the data for each of the possible ways of failing. Thus, we could look at dissolution by typing:

```
stset duration, failure(event == 1);
```

We'd then estimate the exact same Cox model as before. These results are shown in column 2. To look at replacement governments, we'd type:

```
stset duration, failure(event == 2);
```

and then estimate the same Cox model as before. These results are shown in column 3.

We interpret these results exactly as we would from a normal Cox model. Thus, a majority government lowers the hazard rate (increases government survival) compared to a minority government. This is true whether the government ends in dissolution or replacement. It appears that pre-electoral coalitions increase the hazard rate (reduce government survival) for those governments that end in dissolution but they have no effect on the hazard rate for governments that end in replacement.[3] If we exponentiated the coefficients, we would see the percentage change in the risk associated with a one unit increase in our covariates. For example, going from a minority to a majority government reduces the hazard of dissolution by $100 \times [e^{-1.54} - 1] = 78\%$ and reduces the hazard of replacement by $100 \times [e^{-0.71} - 1] = 51\%$.

4 Multinomial Logit

A third approach to dealing with competing risks is based on the multinomial logit (MNL) model. This model is the workhorse for those scholars who have unordered multichotomous dependent variables. Earlier we examined discrete time models (logit, probit etc.) for situations in which a unit was at risk for only one type of event. The MNL model is essentially the discrete-time analog to situations in which units are at risk for different types of events i.e. the dependent variable is coded 0, 1, 2, 3, 4 etc. where the numbers indicate different types of events. If you are familiar with MNL models, then what follows will be a review.

What is a multinomial logit model? Basically, it is a series of linked logits. It's really no more complicated than that. Some people have argued that MNL models contain more information than separate logit models. However, this is incorrect (Alvarez & Nagler 1998). Consider the situation where our dependent variable takes on three values – 0, 1, 2 – indicating three different types of events that could occur. In this situation, we could either run a multinomial logit on this dependent variable or run a series of logits where we compare 1 versus 0, 1 versus 2, and 2 versus 0. Obviously, we would have to recode the dependent variables in each of these separate logits so that they were 0s or 1s. But the bottom line is that the separate logits would give us basically the same results as the MNL model where we do everything all at once. The only difference is that the MNL model is more efficient and the standard errors are smaller because the MNL model uses all the observations whereas the separate logits have fewer observations because they drop cases that are not relevant to their comparison. The point here is that MNL is basically the same as a bunch of linked logit models. So, what's the setup of a multinomial logit model?

4.1 Probability Model Setup

There are different ways of motivating a multinomial logit model. In what follows, we use a pure probability model setup. Let y be a dependent variable with J nominal events.[4] Let P_{im} be the probability that unit i experiences event m.

1. Assume that P_{im} is a function of the linear combination, $X_i\beta_m$.
2. To ensure non-negative probabilities, we use the exponential of $X_i\beta_m$.
3. To make the probabilities sum to 1, we divide $e^{X_i\beta_m}$ by $\sum_{j=1}^{J} e^{X_i\beta_j}$ to get

$$P_{im} = \frac{e^{X_i\beta_m}}{\sum_{j=1}^{J} e^{X_i\beta_j}} \tag{6}$$

[3] The '—-' symbol for the risk of dissolution indicates that the coefficient on *Caretaker Government* tends towards infinity because there is only one case (PDDA data) of a caretaker government that forms after an election ending in dissolution (Box-Steffensmeier & Jones 2004, 171).

[4] The multinomial logit model can be derived from a random utility model setup.

This is the setup for the standard multinomial logit model. To put this model in the duration context, we would say that the hazard probability for unit i experiencing event m is given by:

$$\lambda_{im} = \frac{e^{X_i\beta_m}}{\sum_k^K e^{X_i\beta_k}} \tag{7}$$

where K is once again the set of all possible events and m refers to the m^{th} type of event. Note that the coefficients are subscripted by k i.e. the type of event. This means that we will estimate a different set of coefficients for each possible event type.

4.2 Identification

One issue with the MNL model as it stands is that our event-specific coefficients $\hat{\beta}_k$ are underidentified. To restate the problem slightly differently, for any vector of constants q, we can get the exact same hazard probabilities whether we use β_k or β^*, where $\beta^* = \beta_k + q$. In other words, we could add an arbitrary constant to all the coefficients in the model and we would get exactly the same hazard probabilities. To see this, consider the following example where we have three possible events:

$$\lambda_{i1} = \frac{e^{X_i\beta_1}}{\sum_{k=1}^3 e^{X_i\beta_k}} \tag{8}$$

Now add a vector of constants q so that we have:

$$\lambda_{i1} = \frac{e^{X_i(\beta_1+q)}}{\sum_{k=1}^3 e^{X_i(\beta_k+q)}} \tag{9}$$

With a little manipulation, this can be rewritten as:

$$\begin{aligned}
\lambda_{i1} &= \frac{e^{X_i\beta_1}e^{X_iq}}{e^{X_i(\beta_1+q)} + e^{X_i(\beta_2+q)} + e^{X_i(\beta_3+q)}} \\
&= \frac{e^{X_i\beta_1}e^{X_iq}}{e^{X_i\beta_1}e^{X_iq} + e^{X_i\beta_2}e^{X_iq} + e^{X_i\beta_3}e^{X_iq}} \\
&= \frac{e^{X_i\beta_1}e^{X_iq}}{(\sum_{k=1}^3 e^{X_i\beta_k})e^{X_iq}} \\
&= \frac{e^{X_i\beta_1}}{\sum_{k=1}^3 e^{X_i\beta_k}} \qquad (10)
\end{aligned}$$

which is what we started with. Therefore, the model, as written, is underidentified.

A convenient normalization that solves the identification problem is to assume that one of the sets of coefficients – the coefficients for one of the events – are all zero. As we'll see, this then becomes the reference event against which all of the results are compared. Say that we set $\beta_1 = 0$. In other words, say we set all of the coefficients for event 1 to be 0. By doing this, we are assuming that the first event type is the base category event against which all other event types are compared.[5] Once this constraint is added, we

[5] Note that we can arbitrarily choose any event type as our base category. You should be very careful in that different statistical programs automatically choose different events as the base category - some choose the lowest category, some the highest. STATA's default is to select the event that is chosen most often as the baseline category.

now have:

$$\lambda_{im} = \frac{e^{X_i\beta_m}}{\sum_{k=1}^{K} e^{X_i\beta_k}} \text{ where } \beta_1 = 0 \tag{11}$$

Since $e^{X_i\beta_1} = e^{X_i 0} = 1$, we can rewrite Eq. (11) as two separate equations:

$$\lambda_{i1} = \frac{e^{X_i 0}}{e^{X_i 0} + \sum_{k=2}^{K} e^{X_i\beta_k}} = \frac{1}{1 + \sum_{k=2}^{K} e^{X_i\beta_k}} \tag{12}$$

and

$$\lambda_{im} = \frac{e^{X_i\beta_m}}{1 + \sum_{k=2}^{K} e^{X_i\beta_k}} \text{ for } m > 1 \tag{13}$$

Note that this is the formulation used in Alvarez and Nagler (1998). As we'll see, the way to think about the use of a base category is that the logic when it comes to interpretation is exactly the same as when you break a $K+1$ category variable into K dummy variables.

The bottom line is that the multinomial logit (hazard) probability is:

$$\lambda_{im} = \frac{e^{X_i\beta_m}}{\sum_{k}^{K} e^{X_i\beta_k}} \tag{14}$$

but we have to constrain the coefficients for one of the events to be zero. We will then interpret our results with respect to this baseline event.

4.3 Estimation

Estimation of this model is relatively easy since the log likelihood function is globally concave. To specify the likelihood, first define $d_{ik} = 1$ if unit i experiences event k, $d_{ik} = 0$ otherwise. This means that there are K lots of d_{ik}, each indicating an event. We can then use these indicators to select the appropriate terms in the likelihood function. Thus, the likelihood function for unit i is:

$$\mathcal{L}_i = \lambda_{i1}^{d_{i1}} \times \lambda_{i2}^{d_{i2}} \times \lambda_{i3}^{d_{i3}} \times \ldots \times \lambda_{iK}^{d_{iK}} \tag{15}$$

where λ_{ik} is the hazard probability that unit i experiences event k. The likelihood function for the entire sample is:

$$\mathcal{L} = \prod_{i=1}^{N} \left(\lambda_{i1}^{d_{i1}} \times \lambda_{i2}^{d_{i2}} \times \lambda_{i3}^{d_{i3}} \times \ldots \times \lambda_{iK}^{d_{iK}} \right) \tag{16}$$

Thus, the log-likelihood function is just:

$$\ln\mathcal{L} = \sum_{i=1}^{N}\sum_{k=1}^{K} d_{ik}\ln(\lambda_{ik})$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} d_{ik}\ln\left(\frac{e^{X_i\beta_k}}{\sum_{k=1}^{K} e^{X_i\beta_k}}\right) \tag{17}$$

To estimate the MNL model in STATA you simply type:

```
mlogit Y X, base(some event)
```

where you choose what the base 'event' will be against which you compare all other risks.

4.4 Temporal Dependence

One issue that comes up with MNL models in the context of competing risks models but not in the normal use of MNL models is obviously temporal dependence. As in the discrete time duration models that we have already looked at, we need to take account of temporal dependence. We can do this as we did before by correcting the standard errors, by using time dummies, by using functions of time, or by using lowess functions or splines.

4.5 Interpretation

Let's look at some results from an MNL competing risks model of congressional careers using data from Box-Steffensmeier and Jones (2004). A congressman can end his legislative career in four different ways. He might retire (retirement), he might be ambitious and seek an alternative office (ambition), he might lose a primary (primary), or he might lose a general election (general). The dependent variable is coded as 0 if the congressman is reelected, 1 = General, 2 = Primary, 3 = Retirement, and 4 = Ambition. To capture temporal dependence, we will use the log of duration; Box-Steffensmeier and Jones found this to be the most appropriate functional form in their book. To estimate the model, type:

```
mlogit event  rep redist scandal opengov opensen leader age priorm
        logdur if duration~=., robust cluster(memberid) base(0);
```

The results are shown in Table 3.

Table 3: Competing Risks Model of Congressional Careers: MNL Approach

	Reelection (non-event) is the Reference Category (0)			
Regressor	General	Primary	Retire	Ambition
Party	-0.13	-0.01	0.19	0.28*
	(0.12)	(0.25)	(0.14)	(0.15)
Redistrict	1.61***	1.42**	1.36***	1.52***
	(0.31)	(0.59)	(0.28)	(0.32)
Scandal	2.76***	3.29***	1.23***	-45.18
	(0.38)	(0.44)	(0.42)	——
Open Gub.	0.08	0.23	0.04	0.51***
	(0.16)	(0.30)	(0.17)	(0.16)
Open Sen.	-0.28	-0.46	0.09	1.03***
	(0.21)	(0.44)	(0.20)	(0.16)
Leadership	-0.59	-35.78***	-0.38	-1.56
	(0.55)	(0.37)	(0.30)	(1.03)
Age	0.04***	0.04*	0.08***	-0.06***
	(0.01)	(0.02)	(0.01)	(0.01)
Prior Margin	-0.06***	-0.005	-0.01***	-0.0001
	(0.01)	(0.006)	(0.003)	(0.003)
Log(Duration)	-0.31***	-0.17	0.51***	0.49***
	(0.09)	(0.19)	(0.12)	(0.12)
Constant	-3.05***	-6.23***	-8.02***	-1.18***
	(0.39)	(0.73)	(0.49)	(0.40)
Log likelihood	-3064.5			
Observations	5429			

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed).
Robust standard errors clustered by unit are given in parentheses

4.5.1 Interpreting Coefficients

As you can see, there are four sets of coefficients for each type of event that is not the base category. You can interpret the signs of the coefficients in the expected manner. For example, a negative (positive) coefficient indicates that the independent variable reduces (increases) the probability (risk) of that particular event relative to the baseline category. In this particular case, the baseline category is that the congressman is reelected. Thus, a negative coefficient can be interpreted as indicating that the covariate reduces the risk of a particular event relative to being reelected. For example, we can see that being in a party leadership position reduces the risk of losing in a primary when compared to being reelected i.e. in a comparison between losing in a primary and being reelected, being in a leadership position makes it less likely that you will lose in a primary and more likely that you will be reelected.

4.5.2 Odds Ratios

It is possible to give the multinomial logit model an odds interpretation. The odds of event a versus event b, where neither a nor b are the baseline category, is:

$$\text{ODDS}_{ab} = \frac{\lambda_{ia}}{\lambda_{ib}} = \frac{\frac{e^{X_i\beta_a}}{\sum_k e^{X_i\beta_k}}}{\frac{e^{X_i\beta_b}}{\sum_k e^{X_i\beta_k}}} = \frac{e^{X_i\beta_a}}{e^{X_i\beta_b}} = e^{X_i[\beta_a-\beta_b]} \tag{18}$$

If you wanted to look at the odds ratio of experiencing some event a compared to the baseline event (1), then this simplifies to:

$$\text{ODDS}_{a1} = e^{X_i[\beta_a-\beta_1]} = e^{X_i\beta_a} \tag{19}$$

This will tell you the factor change in the odds associated with experiencing event a compared to the baseline event if unit i had characteristics as given by X_i. Thus, if $e^{X_i\beta_a} = 2.92$, then we saw that the odds of unit i experiencing event a is 2.92 times greater than the odds that it will experience the baseline event. Alternatively, we say that the odds of unit i experiencing event a is 192% higher than the odds of experiencing the baseline event.

You can also think about how a change in a particular variable affects the odds of one event compared to the baseline event.

$$\frac{\text{ODDS}_{a1}(X_i, X_{ij}+\delta)}{\text{ODDS}_{a1}(X_i, X_{ij})} = e^{\beta_{ja}\times\delta} \tag{20}$$

where X_{ij} is the j^{th} independent variable for unit i and β_{ja} is the coefficient on the j^{th} independent variable for event a. Thus, a one unit increase in the REDISTRICT variable increases the odds of losing in a general election as opposed to being reelected by a factor of $e^{1.61} = 5.00$ or, alternatively, by $100 \times [e^{1.61} - 1] = 400\%$.

It is easy to get STATA to give you exponentiated coefficients by typing:

```
mlogit event  rep redist scandal opengov opensen leader age priorm
        logdur if duration~=., robust cluster(memberid) base(0) rrr;
```

4.5.3 Predicted Probabilities and First Differences

You can also calculate the predicted probabilities of experiencing particular events using the equation for the hazard probability from earlier:

$$\lambda_{im} = \frac{e^{X_i\beta_m}}{1+\sum_{k=2}^{K} e^{X_i\beta_k}} \quad \text{for } m > 1 \tag{21}$$

5 Latent Survivor Time Approach and MNL Models

As you can see, both the latent survivor time approach and the MNL models capture heterogeneity by allowing the effects of covariates to vary across the different possible events. As a result, you could use either of these models to evaluate the assumption in the stratified Cox model that the effect of covariates is the same across the different events.

Although the latent survivor time approach and MNL models should produce similar results, there may be some differences. The reason for this has to do with the baseline event against which the coefficients should be compared in the two approaches. In the MNL model, the reference or baseline event is 0 i.e. the coefficients indicate the effect of the covariates on a particular risk relative to the specific risk of being reelected. In contrast, in the latent survivor time approach, the reference category, as it were, is comprised of all the other events except the event of interest (recall that we had the assumption of random censoring, so that any outcome other than k is treated as randomly censored). The analog to this in the logit setting is to estimate four separate logits where the reference category was comprised of every event except the one being directly modeled. In the MNL approach, the reference category is just the particular event of being reelected. As a result, you may see some differences in the coefficients across these models.

6 Conditional Independence

Recall that a key assumption to the latent survivor time model is that that the duration times for risk k are independent of the duration times for risk r. In other words, the risk of event k should not be related to the event of risk r after conditioning on the included covariates. If this assumption does not hold, then the assumption that there is random censoring of the remaining r events will not hold. This assumption is not as unrealistic as you might think. All we need is for the risk to be independent *conditional* on the effects of the covariates. This means that if a particular covariate affects the hazard or more than one event and it is in the model, then its effect is 'controlled for'. In other words, it is the the baseline hazards that need to be independent of each other. If your model is 'good', this is probably not too strong an assumption. Note, though, that there are no good tests for whether you meet this assumption in this setting.

It turns out that an equivalent assumption – the assumption of the Independence of Irrelevant Alternatives (IIA) – is also required for the MNL model. It is possible to test to see if IIA is violated in the MNL setting using a standard Hasuman test.

If you really think that your risks are conditionally dependent, then there are a couple of approaches that you can take.

1. Model dependent risks using frailty terms (random effects).

   The idea is to assume that dependence is due to a common unit effect (Gordon 2002).

2. You could use a model like mixed logit or MNP that avoid the IIA assumption.

   As Box-Steffensmeier and Jones (2004, 173-175) note, you can estimate competing risk models using a discrete time setup with MNL. In this approach, it becomes clear that conditional dependence is equivalent to violating the IIA assumption. Thus, just use a model liked mixed logit or multinomial probit that don't make the IIA assumption.

References

Alvarez, R. Michael & Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42:55–96.

Box-Steffensmeier, Janet & Bradford Jones. 2004. *Timing and Political Change: Event History Modeling in Political Science*. Ann Arbor: University of Michigan Press.

Diermeier, Daniel & Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science* 43:1051–1068.

Gordon, Sanford C. 2002. "Stochastic Dependence in Competing Risks." *American Journal of Political Science* 46:200–217.