Discrete-Time Data and Models

"Discretized" duration data *are still* duration data!

Consider Table 1 (Note connection to "start-stop" process).

Table 1: Example of Discrete-Time Event History Data

Case I.D.	Event Occurrence	Year	Time Elapsed
1	0	1974	1
1	0	1975	2
⋮	⋮	⋮	⋮
1	0	1986	13
1	1	1987	14
5	1	1974	1
45	0	1974	1
45	0	1975	2
⋮	⋮	⋮	⋮
45	0	1992	19
45	0	1993	20

These data are a portion of a data set originally analyzed in Brace, Hall, and Langer (1999). I thank Laura Langer for letting us use them.

Discrete Models and Math

$$f(t) = \Pr(T = t_i) \tag{1}$$

$$S(t) = \Pr(T \geq t_i) = \sum_{j \geq i} f(t_j) \tag{2}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{3}$$

Likelihood

$$\mathcal{L} = \prod_{i}^{n} \Big[h(t_i) \prod_{i=1}^{t-1} (1 - h(t_i)) \Big]^{y_{it}} \Big[\prod_{i=1}^{t} (1 - h(t_i)) \Big]^{1 - y_{it}} \tag{4}$$

$$\mathcal{L} = \prod_{i=1}^{n} \{f(t)\}^{y_{it}} \{S(t)\}^{1 - y_{it}}, \tag{5}$$

A Generic Model

$$h(t) = \Pr(T = t_i \mid T \geq t_i, \mathbf{x}) \tag{6}$$

$$\lambda_{it} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}. \tag{7}$$

Some Binary Link Functions

Logit

$$\log\Big(\frac{\lambda_i}{1-\pi}\Big) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}. \tag{8}$$

$$\hat{\lambda}_i = \frac{e^{\beta' \mathbf{x}}}{1 + e^{\beta' \mathbf{x}}}, \tag{9}$$

Probit

$$\Phi^{-1}[\lambda_i] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}, \tag{10}$$

$$\hat{\lambda}_i = \Phi(\beta' \mathbf{x}). \tag{11}$$

Complementary Log-Log

$$\log[-\log(1-\lambda_i)] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}. \tag{12}$$

$$\hat{\lambda}_i = 1 - \exp[-\exp(\beta' \mathbf{x})]. \tag{13}$$

The Exponential Equivalent of a Logit EH Model

Recall the exponential model. There is a close connection between that model's parameterization of the baseline hazard and the "parameterization" obtained by garden variety binary link models.

$$\log\left(\frac{\lambda_i}{1-\lambda_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},$$

where x_{ki} are two covariates of interest that have a mean of 0, and β_0 is the constant term. The "baseline" hazard under this model would be equivalent to

$$\hat{\lambda}_i = h_0(t) = \exp(\beta_0),$$

which is a constant. The baseline hazard is thus an exponential equivalent.

Transformations on T

The previous result suggest we may want to try different functional forms for time.

There are a variety of choices:

- Ignore it (not usually recommended!)

- Piecewise Functions (perhaps with restrictions over sets of parameters)

- Standard transformations (i.e. logs, polynomials, etc.)

- Smoothing Functions (splines, lowess, etc.)

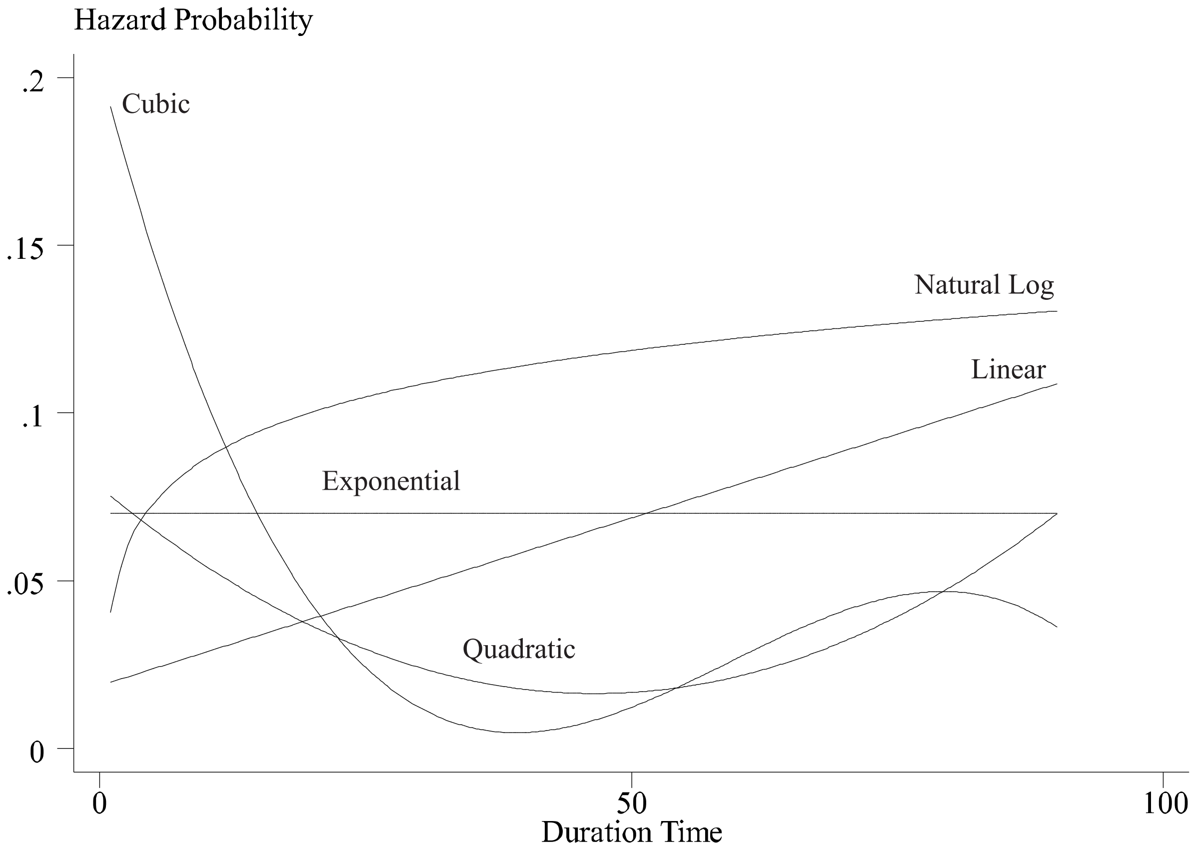

Figure 1: *This figure illustrates different ways that the duration time may be transformed to account for duration dependency in the discrete-time model. The lines in the graph represent a cubic transformation, natural log transformation, time entered linearly, no time dependency (exponential), and a quadratic transformation.*

Smoothing Functions

Let's consider smoothing options. With smoothers, we want to find a function that provides a smooth characterization of the survival times.

Splines

At issue with spline functions is the determination of the location of so-called *knot* points—that is, the point at which two segments of T are joined together.

The choice of knot placement as well as the number of knots used in the spline function can be an issue. It is common that knots are placed at given percentiles of T, or that cubic spline functions are implemented.

What makes the use of splines (as well as other smoothing methods) appealing is that they provide a very general way to characterize duration dependency.

Few assumptions regarding duration dependency need to be made, as the shape of the function is empirically determined (conditional on the location and number of knot point placements).

This flexibility is desirable insofar as it avoids having to make (perhaps wrong) assumptions about the shape of the baseline hazard rate or having to make (perhaps arbitrary) decisions regarding transformations of T.

Locally Weighted Scatterplot Smoothing (Lowesss)

Lowess proceeds by estimating the relationship between y_i and x_i at several target points over the range of x. For each value of y_i, a smoothed value of y_i is estimated.

The resulting smoothed function, in the context of the discrete-time duration model, can be used to characterize the baseline hazard function.

As with the use of spline functions, there are important issues.

Among them is the "bandwidth" decision.

Bandwidth denotes the fraction or proportion of the data that is used in smoothing each point. The larger the bandwidth, the greater the data reduction (and the more smooth the function becomes); the smaller the bandwidth, the closer the lowess function follows the actual data.

Note: splines were estimated using the `btscs.do` module. You can get it at:

`http://www.vanderbilt.edu/~rtucker/papers/btscs/ajps98/`

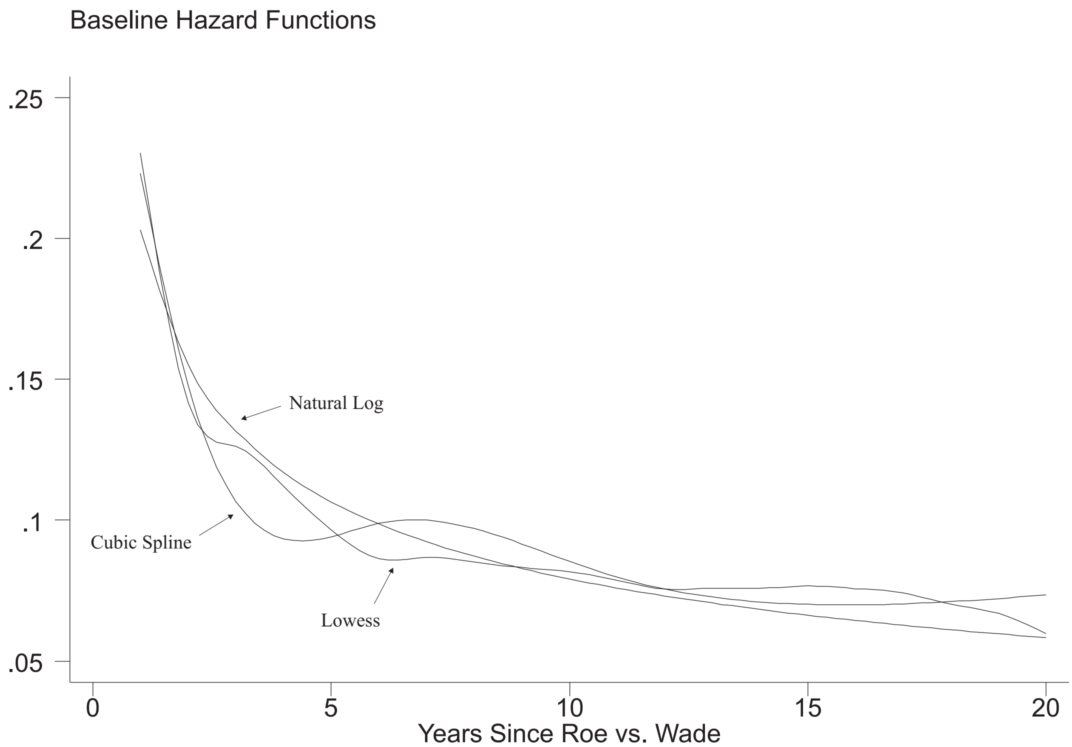

Figure 2: *This figure illustrates a cubic spline function, lowess, and natural log transformation of the baseline hazard function from a model of restrictive abortion policy adoption.*

Some Similarity Between the Cox Model and the Conditional Logit Model

Let's (quickly) get to the conditional logit.

$\exists$ i individuals who choose among J alternatives ($J \geq 3$)

e.g. Voters= i, Parties=J, $j = 1, 2, \ldots J$

RUM Formulation: $U_{ij} = \beta x_i + \epsilon_{ij}$

Example:

$$U_{ij} = \beta_{j0} + \beta_{j1} Ideology_i + \beta_{j2} Gender_1 + \beta_{j3} Education_i.$$

- Multinomial Logit

x_i can have different "effects" for each J

Gives rise to multinomial logit estimator:

$$\Pr(y_i = J) = \frac{\exp(\beta_j x_i)}{1 + \exp(\beta_K x_i)}$$

After normalization, the log-linear model is obtained:

$$\log \left[\frac{P_{ij}}{P_{i1}} \right] = \Sigma(\beta_j x_i)$$

For $J - 1$ alternatives, $\exists$ $J - 1$ non-redundant logits (sometimes called "baseline category" logit).

Conditional Logit

Under MNL, we are concerned with x_i. These are attributes of the ith individual—the chooser.

Choice may be conditioned on "choosers" and "choice" attributes:

x_{ij} are covariates that vary across choosers and choices: i.e. party proximity scores will vary across the J parties.

w_i are covariates that vary across choices but not across individuals: i.e. ideological self-placement, gender, education.

This may lead to consideration of a utility model like

$U_{ij} = \beta_1 Proximity_{ij} + \gamma_{j0} + \gamma_{j1} Ideology_i + \gamma_{j2} Gender_i + \gamma_{j3} Education_i.$

This shows dependency on x_{ij} and w_i and leads to consideration of the conditional logit estimator:

$$P_{ij} = \frac{\exp(\beta_j x_{ij} + \gamma_j w_i)}{\Sigma_{k=1}^{J} \exp(\beta_j x_{ij} + \gamma_j w_i)}$$

which is equivalent to

$$P_{ij} = \frac{\exp(\beta_j x_{ij}) \exp(\gamma_j w_i)}{\Sigma_{k=1}^{J} \exp(\beta_j x_{ij}) \exp(\gamma_j w_i)}$$

which yields the important result:

$$P_{ij} = \frac{\exp(\beta_j x_{ij})}{\Sigma_{k=1}^{J} \exp(\beta_j x_{ij})}$$

Important implication: any factor that is constant is dropped out of the model. Here, all w_i are "fixed" within individuals; they cannot be estimated.

Some “data”

Choice set consists of 4 parties (i.e. $J = 4$).

“Proximity” measures the proximity of chooser i to party j: it is an x_{ij} variable.

Gender, education, and ideology are all constants within the chooser (i.e. they vary across choices, but not choosers): it is a w_i variable.

	ID	choice set	choice	proximity	gender	educ.	ideology
1.	1	1	0	7	1	16	7
2.	1	2	1	9	1	16	7
3.	1	3	0	3	1	16	7
4.	1	4	0	1	1	16	7
5.	2	1	0	1	0	13	2
6.	2	2	0	1	0	13	2
7.	2	3	0	7	0	13	2
8.	2	4	1	10	0	13	2
9.	3	1	1	9	1	10	10
10.	3	2	0	6	1	10	10
11.	3	3	0	2	1	10	10
12.	3	4	0	2	1	10	10
13.	4	1	0	3	0	14	8
14.	4	2	1	7	0	14	8
15.	4	3	0	3	0	14	8
16.	4	4	0	1	0	14	8
17.	5	1	0	2	1	12	3
18.	5	2	0	3	1	12	3
19.	5	3	1	7	1	12	3
20.	5	4	0	5	1	12	3
21.	6	1	0	8	0	18	9
22.	6	2	1	10	0	18	9
23.	6	3	0	2	0	18	9
24.	6	4	0	1	0	18	9

Let's apply conditional logit:

```
. clogit  choice proximity gender education ideology, group(ID)
nolog
note: ideology omitted due to no within-group variance.
note: education omitted due to no within-group variance.
note: gender omitted due to no within-group variance.

Conditional (fixed-effects) logistic regression  Number of obs   =         80
                                                 LR chi2(1)      =      50.84
                                                 Prob > chi2     =     0.0000
Log likelihood =  -2.308022                      Pseudo R2       =     0.9168

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   proximity |   1.645505    .741898     2.22   0.027     .1914115    3.099598
------------------------------------------------------------------------------
```

Note what happens if we specify the model with the w_i covariates: They drop out.

Implications:

Repeating: fixed factors within groups are dropped out.

For this reason, the conditional logit is sometimes called a "fixed effects" model (group-level factors are fixed).

There is an equivalency to the MNL model (though I won't talk about it here).

…there is also an equivalency to a Cox model …and I will talk about that!

Table 2: Example of Matched Case-Control Duration Data

Observation Number	Risk Period	Event Occurrence	Duration Time
Risk Period 1			
1	1	0	5
2	1	0	5
3	1	0	5
4	1	1	5
5	1	1	5
Risk Period 2			
1	2	0	11
2	2	1	11
3	2	1	11

Some Similarites:

This kind of fixed-effects model is also known as a "matched case-control" model.

Observations are "matched" on some attribute; in the party choice example, the matching is on the individual.

"Cases" record whether or not some outcome has occurred within the grouping and "controls" record the non-occurrence of an outcome (i.e. these are the 1s and 0s)

Above, we have "1 : 3" matching for each grouping: an individual can only chose 1 party (the case); the other 3 are not selected (the controls).

We can think of duration data as matched case-control data.

What are cases matched on?

THE ORDERED FAILURE TIMES ... or equivalently, the periods of risk.

Sound like a Cox model?

Take a look at Table 1. In risk period 1, there are five matched observations. Of these five observations, there are two “cases” and three “controls” which elicits a ratio of 2 : 3 matching. For risk period two, two cases are observed and one control, for 2 : 1 matching.

The “case” is the event's occurrence. In standard choice models, the “case” is the choice made from the available choice-set.

There is a fundamental equivalency in the structure of the data.

Suppose there are $k = 1, 2, \ldots K$ distinct risk periods in the data set (equivalently, there are k ordered failure times observed), and $i = 1, 2, \ldots J_k$ observations at risk in the kth period.

The response pattern of the dependent variable for the kth risk period is $\mathbf{y_k} = (y_{k1} + y_{k2} + \ldots + y_{kJ_k})$.

The total number of events, n_{1k} (or “cases”) observed in the kth risk period is:

$$n_{1k} = \sum_{i=1}^{J} y_{ki}. \tag{14}$$

The total number of non-events or "controls" is:

$$n_{0k} = J_k - n_{1k}. \tag{15}$$

The probability of the response pattern $\mathbf{y_k}$ is then estimated, conditional on n_{1k}. It is:

$$\Pr(\mathbf{y_k} \mid \sum_{i=1}^{J} y_{ki} = n_{1k}) = \frac{\exp(\beta' \sum_{i=1}^{J} \mathbf{x_{ki}} \mathbf{y_{ki}})}{\sum_{\mathbf{d_k} \in R_k} \exp(\beta' \sum_{i=1}^{J} \mathbf{x_{ki}} d_{ki})}, \tag{16}$$

where R_k denotes all possible combinations of ones and zeroes, i.e., cases and controls, in the kth risk period, $\mathbf{d_k} = (d_{k1}, d_{k2}, \ldots, d_{kJ})$, $d_{ki} = 0$ or 1.

Look familiar? It's a very slightly modified version of the conditional logit estimator. Why the modifications? We have to account for the response pattern at each failure time and we have to account for the fact that the risk pool decreases in size after each failure time.

What we are doing is estimating the probability of the response pattern $\mathbf{y_k}$ for each risk period R_k, for all possible combinations of events and non-events, i.e., 1s and 0s, observed among the J observations at risk.