Models for Discrete Time and the Inclusion of Time-Varying Covariates

Sarah Hunter

University of Georgia

April 10, 2018
Objectives

By the end of this meeting, participants should be able to:

- Explain the theory behind and estimate a discrete-time process model
- Apply the Cox proportional hazards model to discretized data using the exact discrete method.
- Describe the usefulness of a flexible parametric model.
- Incorporate time-varying covariates into a duration model.
- Explain the concerns associated with including time-varying covariates.
Structuring the Data
Two Equivalent Options

Option A  Treat time as continuous and simply report the time until failure.

Option B  Choose a unit for time to create a panel data set.
- The outcome variable is a dummy for whether an event has occurred yet.
- Usually coded 0 until its last observation, when it is 1.
- The coding can be reversed, but take software algorithms into account.
- Here, time is being broken into discrete units.
## Structuring the Data

**Option A**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Intervenor</th>
<th>Target</th>
<th>Duration</th>
<th>Contiguity</th>
<th>Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>U.K.</td>
<td>Albania</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>El Salvador</td>
<td>Honduras</td>
<td>657</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>81</td>
<td>U.S.</td>
<td>Panama</td>
<td>274</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>184</td>
<td>Bulgaria</td>
<td>Greece</td>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>236</td>
<td>Taiwan</td>
<td>China</td>
<td>7456</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>278</td>
<td>Botswana</td>
<td>South Africa</td>
<td>1097</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>332</td>
<td>Uganda</td>
<td>Kenya</td>
<td>409</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>467</td>
<td>Israel</td>
<td>Egypt</td>
<td>357</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>621</td>
<td>Malawi</td>
<td>Mozambique</td>
<td>631</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>672</td>
<td>India</td>
<td>Pakistan</td>
<td>173</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
## Structuring the Data

**Option B**

### TABLE 5.1: Example of Discrete-Time Event History Data

<table>
<thead>
<tr>
<th>Case I.D.</th>
<th>Event Occurrence</th>
<th>Event Year</th>
<th>Time Elapsed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1974</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1975</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1986</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1987</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1974</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>1974</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>1975</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>1992</td>
<td>19</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>1993</td>
<td>20</td>
</tr>
</tbody>
</table>
Components of the Discrete-Time Model

- Probability mass function: \( f(t) = P(T = t_i) \)
- Survivor function: \( S(t) = P(T \geq t_i) = \sum_{j \geq i} f(t_j) \)
- The discrete-time hazard rate: \( h(t) = \frac{f(t)}{S(t)} = P(T = t_i | T \geq t_i) \)
- Conditional probability of survival: \( 1 - h(t) = P(T > t_i | T \geq t_i) \)
- We can work-out the p.m.f. as:

\[
\begin{align*}
    f(t) &= P(T = t_i | T \geq t_i) \times P(T > t_{i-1} | T \geq t_{i-1}) \times \cdots \\
    &\quad \times P(T > t_2 | T \geq t_2) \times P(T > t_1 | T \geq t_1) \\
    &= h(t_i) \times (1 - h(t_{i-1})) \times \cdots \times (1 - h(t_2)) \times (1 - h(t_1)) \\
    &= h(t_i) \prod_{i=1}^{t-1} (1 - h(t_i))
\end{align*}
\]
Estimation

- Therefore the survivor function is:

\[ S(t) = (1 - h(t_1)) \times (1 - h(t_{i-1})) \times \cdots \times (1 - h(t_2)) \times (1 - h(t_1)) \]

\[ = \prod_{i=1}^{t} (1 - h(t_i)) \]

- The likelihood function:

\[ L = \prod_{i}^{n} \left[ h(t_i) \prod_{i=1}^{t-1} (1 - h(t_i)) \right]^{y_{it}} \left[ \prod_{i=1}^{t} (1 - h(t_i)) \right]^{1-y_{it}} \]

\[ = \prod_{i=1}^{n} \left\{ f(t) \right\}^{y_{it}} \left\{ S(t) \right\}^{1-y_{it}} \]

- Final decision: How will you model the probability of an event? Logit, probit, or cloglog are reasonable.
Modeling the Baseline Hazard

- Ignore it (exponential implication)
- Linear
- Quadratic
- :
- Splines
- Logarithms
- Lowess

Model Interpretation. These are models of hazards. Therefore:

**Positive coefficients** imply the hazard rate is increasing; hence, the survival time is shortened.

**Negative coefficients** imply the hazard rate is decreasing; hence, the survival time is lengthened.
Review: What does a conditional logit model do?

At a specific time, failures are grouped with non-failures still in the risk set.

Probability of response pattern $y_k$:

$$P(y_k | \sum_{i=1}^{J} y_{ki} = n_{1k}) = \frac{\exp(\beta' \sum_{i=1}^{J} x_{ki} y_{ki})}{\sum_{d_k \in R_k} \exp(\beta' \sum_{i=1}^{J} x_{ki} d_{ki})}$$
Advantages and Disadvantages of Modeling Strategies

- **Parametric Models**
  - Easy-to-interpret baseline hazard.
  - Clearer handling of unequal spacing.
  - Ease of implementation (MLE).
    - Usually arbitrary decisions on the baseline hazard rate. This can bias estimates.

- **Discrete-Time Models**
  - Common and well-understood in social science.
  - Inaccurate parameterization of baseline hazard can be problematic.

- **Cox Models**
  - Does not estimate the baseline hazard (assuming it is a nuisance).
    - Extracted baseline hazard is overfitted and noisy.
    - Tougher to interpret the baseline hazard function.
Flexible Parametric Models

- Spline focused.
- Log of the integrated hazard:
  \[ \log H(t; x) = \log H_0(t) + \beta' x = s(x) + \beta' x \]
- Log of the cumulative proportional odds:
  \[ \log O(t; x) = \log O_0(t) + \beta' x = s(x) + \beta' x \]
Time-Varying Covariates

Terminology and Assumptions

- Fixed covariates: time-independent.
- Defined covariates: non-stochastic.
- Ancillary: stochastic, but independent of the event history process.
- Recall: internal v. external.
- Exogenous v. endogenous. Exogeneity:

\[ P(X(t, t + \Delta t) | T \geq t + \Delta t, X(t)) = P(X(t, t + \Delta t) | X(t)) \]

- Jump process.
## Time Varying Covariates

### Data Structure

**TABLE 7.3: Example of Event History Data Set with TVCs**

<table>
<thead>
<tr>
<th>Case I.D.</th>
<th>Weeks to Event</th>
<th>Southern District</th>
<th>Incumbent’s Party</th>
<th>1988 Vote</th>
<th>“War Chest” (in Millions)</th>
<th>Censoring Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>.62</td>
<td>.003442</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>.62</td>
<td>.010986</td>
<td>1</td>
</tr>
<tr>
<td>201</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>.59</td>
<td>.142588</td>
<td>0</td>
</tr>
<tr>
<td>201</td>
<td>53</td>
<td>1</td>
<td>0</td>
<td>.59</td>
<td>.158570</td>
<td>0</td>
</tr>
<tr>
<td>201</td>
<td>65</td>
<td>1</td>
<td>0</td>
<td>.59</td>
<td>.202871</td>
<td>0</td>
</tr>
<tr>
<td>201</td>
<td>75</td>
<td>1</td>
<td>0</td>
<td>.59</td>
<td>.217207</td>
<td>0</td>
</tr>
<tr>
<td>516</td>
<td>26</td>
<td>0</td>
<td>1</td>
<td>.79</td>
<td>.167969</td>
<td>0</td>
</tr>
<tr>
<td>516</td>
<td>53</td>
<td>0</td>
<td>1</td>
<td>.79</td>
<td>.147037</td>
<td>0</td>
</tr>
<tr>
<td>516</td>
<td>65</td>
<td>0</td>
<td>1</td>
<td>.79</td>
<td>.164970</td>
<td>0</td>
</tr>
<tr>
<td>516</td>
<td>72</td>
<td>0</td>
<td>1</td>
<td>.79</td>
<td>.198608</td>
<td>1</td>
</tr>
<tr>
<td>706</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>.66</td>
<td>.139028</td>
<td>0</td>
</tr>
<tr>
<td>706</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>.66</td>
<td>.225633</td>
<td>0</td>
</tr>
<tr>
<td>706</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>.66</td>
<td>.225817</td>
<td>0</td>
</tr>
<tr>
<td>706</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>.66</td>
<td>.342801</td>
<td>0</td>
</tr>
<tr>
<td>706</td>
<td>83</td>
<td>0</td>
<td>0</td>
<td>.66</td>
<td>.262563</td>
<td>1</td>
</tr>
<tr>
<td>905</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.211122</td>
<td>0</td>
</tr>
<tr>
<td>905</td>
<td>53</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.270816</td>
<td>0</td>
</tr>
<tr>
<td>905</td>
<td>65</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.262153</td>
<td>0</td>
</tr>
</tbody>
</table>
Survivor function:

\[ S(t_k) = \prod_{j=1}^{k} P(T > t_j | T \geq t_{j-1}) \]

Where:

\[ P(T > t_j | T \geq t_{j-1}) = \exp \left( - \int_{t_{j-1}}^{t_j} h(u | x_j) \, du \right) = \exp(-H(t_k)) \]

Discrete-Time Models: Easiest way to include TVCs.
Time-Varying Covariates

Incorporating into the Cox model

- Cox Model, partial likelihood:

\[
L_p = \prod_{i=1}^{K} \left[ \frac{e^{\beta' x_i(t_i)}}{\sum_{j \in R(t_i)} e^{\beta' x_j(t_i)}} \right]^{\delta_i}
\]

- Cox model, log-partial likelihood:

\[
\log L_p = \sum_{i=1}^{K} \delta_i \left[ \beta' x_i(t_i) - \log \sum_{j \in R(t_i)} e^{\beta' x_j(t_i)} \right]
\]
Time-Varying Covariates

Incorporating into parametric models

- Compile data in a way that allows episodes to be split into successive intervals that correspond to changes in the TVC.
- Weibull model, hazard rate:
  \[
  h[t|x(t^-)] = \exp\{-\beta'x(t^-)\}^p \times (\exp\{-\beta'x(t^-)t\})^{p-1}
  \]
- Weibull model, survivor function: \( S[t|x(t^-)] = \exp\{-\lambda t^p\} \)
- Weibull density: \( f[t|x(t^-)] = \lambda p(\lambda t)^{p-1} \exp\{-\lambda t^p\} \)
- Weibull model, likelihood function:
  \[
  \mathcal{L} = \prod_{j=1}^{K} \left\{ \lambda p(\lambda t)^{p-1} e^{-(\lambda t)^p} \right\}^{\delta_i} \left\{ e^{-(\lambda t)^p} \right\}^{1-\delta_i}
  \]
Concerns when Including TVCs

- Temporal ordering and measurement error.
- Endogenous TVCs: loss of interpretation.
- Serial dependence. Solutions:
  - Robust covariance matrices.
  - Frailty terms.
For Next Time

- Read *Event History Modeling*, Chapters 8-11.
- Analyze the depression data (depression_pp.csv):
  - Model time until an initial depressive event (EVENT) as a function of whether a person has divorced parents (PD).
  - Your time variable is age (age_18). Estimate three models to account for the baseline hazard by age: A linear function of age, a smoothed spline, and a conditional logit using age as strata.
  - Include frailties by individual (ID) in the first two models.
  - Present all three models in a professional table (or separate tables).
  - Interpret the effect of divorced parents.
  - Which model would you be most inclined to stake your reputation on?
- Remember that papers are due on April 24.