Survival Data Analysis

Kosuke Imai

Princeton University

POL573 Quantitative Analysis III
Fall 2016

Readings

- (Required) Freedman. (2008). “Survival Analysis: A Primer” *The American Statistician*, Vol. 62, pp. 110–119.
- (Required) Various articles on cabinet dissolution (references given in later slides)

- (Suggested) Wooldridge, Chapter 20.
- (Suggested) Box-Steffensmeier and Jones (1997). “Time is of the Essence: Event History Models in Political Science,” *American Journal of Political Science*, Vol. 41, pp. 1414–1461.

- (Reference) Box-Steffensmeier and Jones. (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge UP.
- (Reference) Kalbfleisch and Prentice. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. Wiley & Sons.

What is Survival Data?

- Duration data consisting of start time and end time
- A running example: Cabinet duration
- Other examples: Congressional career, Peace agreement etc.
- Predicting how long does something last?
- Estimating the causal effect of one factor on duration
- Key methodological issues: censoring, shape of hazard function

Survival Function

- $Y_i \in [0, \infty)$: time to an event or failure time
- Survival function: Probability of surviving at least up to time y

$$S(y) \equiv \Pr(Y_i > y) = 1 - \underbrace{\Pr(Y_i \leq y)}_{CDF}$$

1. Monotonically decreasing
2. Right-continuous
3. $\lim_{y \to \infty} S(y) = 0$ and $\lim_{y \to 0} S(y) = 1$

- Density function of Y_i (if it exists):

$$f(y) = -\frac{d}{dy} S(y) \quad \text{and} \quad S(y) = \int_y^\infty f(t) dt$$

- For sufficiently small h and if $f(y)$ is continuous at y,

$$h \times f(y) \approx \Pr(y \leq Y_i < y + h) = S(y) - S(y + h)$$

Hazard Function

- What is it?: The instantaneous rate at which events occur at time y given that the observation "survived" up to time y

$$\begin{aligned}\lambda(y) &\equiv \lim_{h \downarrow 0} \frac{\Pr(y \leq Y_i < y + h \mid Y_i \geq y)}{h} \\ &= \frac{f(y)}{S(y)} = -\frac{d}{dy}\log S(y)\end{aligned}$$

- It's not probability!
- Hazard function completely determines the distribution of Y_i:

$$S(y) = \exp\left(-\int_0^y \lambda(t)dt\right)$$

- Cumulative hazard function:

$$\Lambda(y) \equiv \int_0^y \lambda(t)dt = -\log S(y)$$

Quantities of Interest

- Survival curve
- Hazard function is difficult to interpret
- Expected time to an event:

$$\begin{aligned} \mu(y) &\equiv \mathbb{E}(Y_i - y \mid Y_i > y) \\ &= \int_y^\infty (t-y)\frac{f(t)}{S(y)}dt \;=\; \frac{1}{S(y)}\int_y^\infty S(t)dt \end{aligned}$$

- Note that $\mu(0) = \mathbb{E}(Y_i) = \int_0^\infty S(t)dt$
- $\mu(y)$ also completely determines distribution of Y_i:

$$S(y) = \frac{\mu(0)}{\mu(y)}\exp\left(-\int_0^y \frac{1}{\mu(t)}dt\right)$$

Censoring

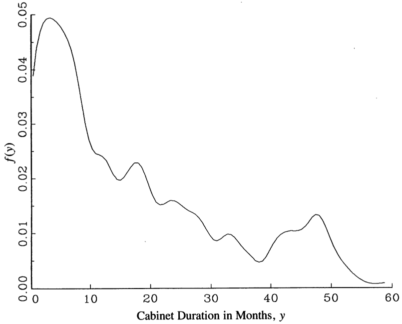
- Right-censoring $Y_i \in (t, \infty)$: an observation does not experience an event during the study period
- Independence assumption: (Given X_i) Hazard rates of those who are censored do not systematically differ from those who are not

$$\lim_{h \downarrow 0} \frac{\Pr(y \leq Y_i < y + h \mid Y_i \geq y, X_i)}{h} = \lim_{h \downarrow 0} \frac{\Pr(y \leq Y_i < y + h \mid Y_i \geq y, C_i \geq y, X_i)}{h}$$

  for all y where C_i be a time to censoring
- Implied by $Y_i \perp\!\!\!\perp C_i \mid X_i$: Missing at random; censoring is ignorable
- Other types of censoring:
  1. Left-censoring $Y_i \in [0, t]$
  2. Interval-censoring $Y_i \in (t_1, t_2)$
- Left-truncation: units failing before the study period

Cabinet Duration and Censoring

- King, Alt, Burns, and Laver (1990). *AJPS*
- Censoring: 12 months prior to the end of the constitutional inter-election period



- Is censoring independent?

Discrete Time Approximation

- Time is continuous but we observe discrete time: $t_1 < t_2 < \cdots$
- Density function: $f(t_j) = \Pr(Y_i = t_j)$
- Survival function: $S(y) = \sum_{\{j:\ t_j > y\}} f(t_j)$
- Hazard function:

$$\lambda(t_j) \ = \ \Pr(Y_i = t_j \mid Y_i \geq t_j) \ = \ \frac{f(t_j)}{\lim_{t \downarrow t_j} S(t)} \ = \ \frac{f(t_j)}{S(t_{j-1})}$$

- Cumulative hazard function: $\Lambda(y) = \sum_{\{j:\ t_j \leq y\}} \lambda(t_j)$
- Key relationships:

$$S(t_j) = \prod_{k=1}^{j} (1 - \lambda(t_k)) \quad \text{and} \quad f(t_j) = \lambda(t_j) \prod_{k=1}^{j-1} (1 - \lambda(t_k))$$

- Expected time to an event:

$$\mu(t_j) \ = \ \frac{1}{S(t_j)} \sum_{k=j}^{\infty} (t_{k+1} - t_k) S(t_k)$$

Nonparametric ML Estimation of Survival Function

- Idea: Use one minus the empirical CDF to estimate $S(y)$
- Observed failure times: $t_1 < t_2 < \cdots < t_J$
- Ties are allowed, i.e., $J \leq n$, and d_j units failed at t_j
- m_j units are right-censored in the interval $[t_j, t_{j+1})$ and fail in the interval $[t_{j+1}, \infty)$
- Under the independence assumption,

$$\begin{aligned} \Pr(Y_i = t_j) &= S(t_{j-1}) - S(t_j) = f(t_j) \quad \text{uncensored} \\ \Pr(Y_i > t_j) &= S(t_j) \quad \text{censored} \end{aligned}$$

- The likelihood function:

$$L(\lambda \mid Y) = \prod_{j=1}^{J} f(t_j)^{d_j} S(t_j)^{m_j} = \prod_{j=1}^{J} \lambda(t_j)^{d_j} (1 - \lambda(t_j))^{n_j - d_j}$$

where $n_j = \sum_{k=j}^{J} (d_k + m_k)$ is the number of units "at risk"

Kaplan-Meier Estimator

- The log-likelihood function:

$$l_n(\lambda \mid Y) = \sum_{j=1}^{J} \{d_j \log \lambda(t_j) + (n_j - d_j) \log(1 - \lambda(t_j))\}$$

- The Kaplan-Meier estimator:

$$\widehat{S}(y) = \prod_{\{j:\ t_j \leq y\}} (1 - \hat{\lambda}(t_j)) = \prod_{\{j:\ t_j \leq y\}} \frac{n_j - d_j}{n_j}$$

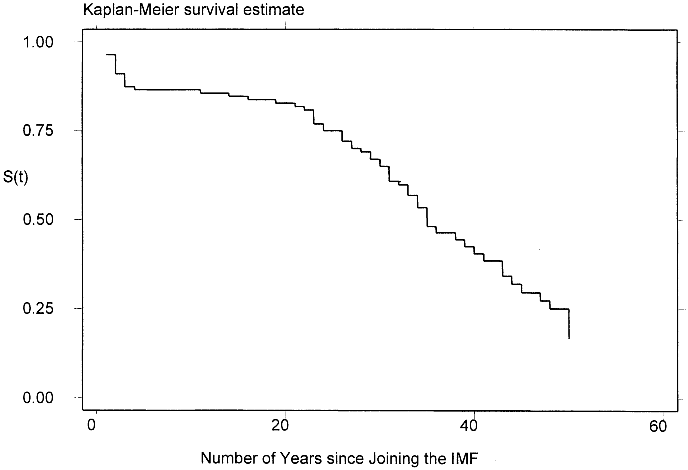
- The Asymptotic variance estimator:

$$\widehat{\mathbb{V}(\widehat{S}(y))} = \widehat{S}^2(y) \sum_{\{j:\ t_j \leq y\}} \frac{d_j}{n_j(n_j - d_j)}$$

- The Nelson-Aalen estimator: $\widehat{\Lambda}(y) = \sum_{\{j:\ t_j \leq y\}} d_j / n_j$

Compliance in International Law

FIGURE 2. The Kaplan-Meier Survival Function Duration of Article XIV Status over Time



B. Simmons (2000) *APSR*

Exponential Regression Model

- Model: $Y_i \mid X_i \overset{\text{indep.}}{\sim} \text{Exponential}(\mu_i)$ where $\mu_i = \exp(X_i^\top \beta)$
- Density: $f(y \mid \mu_i) = \frac{1}{\mu_i} \exp(-y/\mu_i)$
- Mean $\mathbb{E}(Y_i \mid \mu_i) = \mu_i$ and Variance $\mathbb{V}(Y_i \mid \mu_i) = \mu_i^2$
- Survival function: $S(y) = \exp(-y/\mu_i)$
- Constant Hazard function: $\lambda(y) = 1/\mu_i$
- Memory-less Property:

$$\Pr(Y_i > y + h \mid Y_i > y) \;=\; \exp(-h/\mu_i)$$

  which does not depend on y

- Poisson process with rate $1/\lambda$:

$$S(y) = \Pr(Y_i(y) = 0) = \exp(-y/\lambda)$$

  "waiting time" for the next event

Exponential Likelihood Function with Right-censoring

- Censoring indicator $C_i = 1$ for being censored
- Y_i is the censoring time (rather than failure time) when $C_i = 1$
- Likelihood function:

$$\begin{aligned} L_n(\beta \mid Y, X, C) &= \prod_{i=1}^{n} \underbrace{\left\{ \frac{1}{\mu_i} \exp(-Y_i/\mu_i) \right\}^{1-C_i}}_{uncensored\ obs.} \underbrace{\{\exp(-Y_i/\mu_i)\}^{C_i}}_{censored\ obs.} \\ &= \prod_{i=1}^{n} \exp\left\{-(1-C_i)X_i^\top \beta\right\} \exp\left\{-\exp(-X_i^\top \beta) Y_i\right\} \end{aligned}$$

- Log-likelihood function:

$$l_n(\beta \mid Y, X, C) = \sum_{i=1}^{n} -(1-C_i)X_i^\top \beta - \exp(-X_i^\top \beta) Y_i$$

- Score and Hessian can be calculated in the usual manner

EM Algorithm for the Exponential Regression Model

- *E*-step:

$$Q(\beta \mid \beta^{(t)}) = -\sum_{i=1}^{n} \left\{ X_i^\top \beta + \exp(X_i^\top \beta) Y_i^{*(t)} \right\}$$

  where

$$Y_i^{*(t)} = \begin{cases} Y_i & \text{if} \quad C_i = 0 \\ Y_i + \exp(X_i^\top \beta^{(t)}) & \text{if} \quad C_i = 1 \end{cases}$$

- *M*-step: Use the iterated weighted least squares algorithm!

- The *EM* algorithm makes the complex optimization problem with missing data into an easy one without missing data

Weibull Regression Model

- Model: $Y_i \mid X_i \overset{\text{indep.}}{\sim} \text{Weibull}(\mu_i, \gamma)$ where $\mu_i = \exp(X_i^\top \beta)$ and $\gamma > 0$
- Density: $f(y \mid \mu_i, \gamma) = \frac{\gamma}{\mu_i^\gamma} y^{\gamma-1} \exp\{-(y/\mu_i)^\gamma\}$
- A generalization of the exponential regression model (when $\gamma = 1$)
- Mean: $\mathbb{E}(Y_i \mid X_i) = \mu_i \Gamma(1 + 1/\gamma)$
- Variance: $\mathbb{V}(Y_i \mid X_i) = \mu_i^2[\Gamma(1 + 2/\gamma) - \Gamma(1 + 1/\gamma)^2]$
- Survival function: $S(y) = \exp\{-(y/\mu_i)^\gamma\}$
- Monotone hazard function:

$$\lambda(y) = \frac{\gamma}{\mu_i^\gamma} y^{\gamma-1}$$

  increasing (decreasing) if $\gamma > 1$ (if $\gamma < 1$)

Proportional Hazard and Log-Linear Models

- Exponential model $\in$ Weibull model $\in \begin{cases} \text{Proportional hazards models} \\ \text{Log-linear models} \end{cases}$
- Proportional hazard:

$$\lambda(y; x) = \underbrace{\lambda_0(y)}_{\textit{baseline hazard}} r(x)$$

- Exponential model: $\lambda_0(y) = 1$ and $r(x) = \exp(-x^\top \beta)$
- Weibull model: $\lambda_0(y) = \gamma y^{\gamma - 1}$ and $r(x) = \exp(-x^\top \beta)$

- Log-linear models:

$$\log Y_i \;=\; X_i^\top \beta + \epsilon_i$$

  - Exponential model: extreme value distribution
  - Weibull model: scaled extreme value distribution
  - Log-normal model: normal distribution

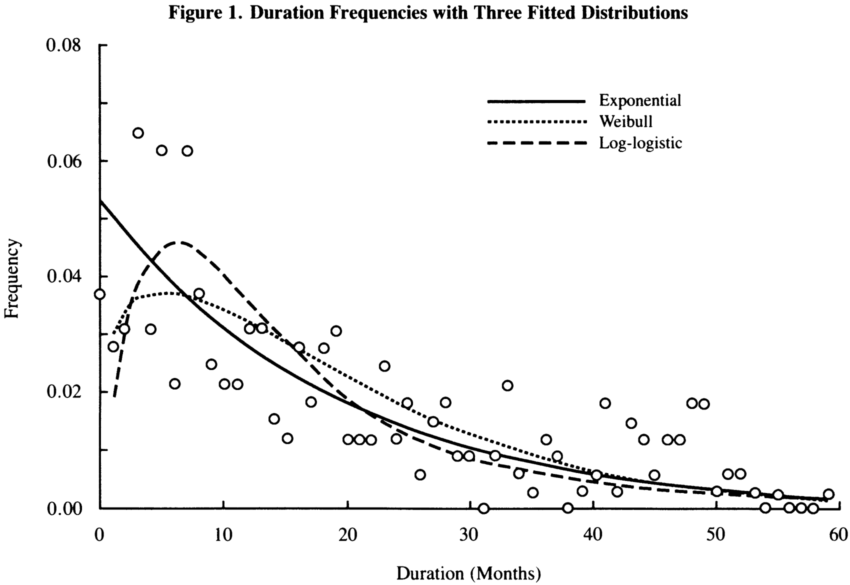
- Other distributions: Gamma, Log-logistic etc.

Log-Normal Regression Model

- Model: $\log Y_i \mid X_i \stackrel{\text{indep.}}{\sim} \mathcal{N}(X_i^\top \beta, \sigma^2)$
- Density: $f(Y_i = y \mid X_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2} y} \exp\left\{-\frac{1}{2\sigma^2}(\log y - X_i^\top \beta)^2\right\}$
- Mean: $\mathbb{E}(Y_i \mid X_i) = \exp(X_i^\top \beta + \sigma^2/2)$
- Variance: $\mathbb{V}(Y_i \mid X_i) = \exp(2X_i^\top \beta + \sigma^2)\{\exp(\sigma^2) - 1\}$
- Survival function: $S(y) = \Phi\left(\frac{X_i^\top \beta - \log y}{\sigma}\right)$
- Hazard function: $\lambda(y) = \frac{\phi\left(\frac{X_i^\top \beta - \log y}{\sigma}\right)}{\Phi\left(\frac{X_i^\top \beta - \log y}{\sigma}\right) y\sigma}$
- Hazard function is not monotone

Debate on the Hazard Function for Cabinet Dissolution

- King, Alt, Burns, and Laver (1990): Exponential model
  - Systematic factors: country attributes, party structure attributes, coalition attributes, etc. measured at the beginning of the cabinet
  - Stochastic events: time-varying economic shocks, scandals, etc.
  - Constant hazard rate: stochastic risk is constant over time
- Warwick and Easton (1992, *AJPS*): Weibull model
  - Increasing hazard rate
  - Warwick (1992, *AJPS*)
- Alt and King (1994, *CPS*): only slightly increasing
- Similar debate in leadership duration:
  - Londregan and Poole (1990, *WP*): constant hazard
  - Bienen and van de Walle (1991, *Stanford UP*): decreasing hazard



Warwick and Easton (1992, *AJPS*)

Going Beyond the Parametric Regression Models

- Less restriction assumption on the hazard function
- Time-varying covariates to further model stochastic risks

- Cox Proportional Hazard Model (1972, *JRSS B*):

$$\underbrace{\lambda(y, X_i(y))}_{\textit{semiparametric}} = \underbrace{\lambda_0(y)}_{\textit{nonparametric}} \underbrace{\exp(X_i(y)^\top \beta)}_{\textit{parametric}}$$

- Baseline hazard $\lambda_0(y)$:
  - Underlying hazard common to all observations varying across time
  - Unspecified allowing for flexible functional form
- Time-varying covariates $X_i(y)$: Modeling the hazard rate as a function of both time-invariant and time-varying covariates

Partial Likelihood Function

- J observed *distinct* failure times: $0 = t_0 < t_1 < t_2 < \cdots < t_J < \infty$
- Assume independent right-censoring as before
- Alternative representation: $\{V_j, W_j\}_{j=1}^J$
  1. V_j: the event that particular observations get censored in $[t_{j-1}, t_j)$
  2. W_j: the event that particular observations fail at t_j
- Likelihood function:

$$\prod_{j=1}^{J} f(V_j \mid V^{(j-1)}, W^{(j-1)}, \theta) \times \underbrace{\prod_{j=1}^{J} f(W_j \mid V^{(j)}, W^{(j-1)}, \theta)}_{partial\ likelihood\ function}$$

where $V^{(j)} = \{V_1, \ldots, V_j\}$, and $W^{(j)} = \{W_1, \ldots, W_j\}$

- Under the no ties assumption,

$$\prod_{j=1}^{J} f(W_j \mid V^{(j)}, W^{(j-1)}, \theta) = \prod_{j=1}^{J} \frac{\lambda(t_j, X_i(t_j))}{\sum_{i' \in R(t_j)} \lambda(t_j, X_{i'}(t_j))}$$
$$= \prod_{j=1}^{J} \frac{\exp(X_i(t_j)^\top \beta)}{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta)}$$

  where i represents the observation which failed at time t_j and $R(t)$ is the risk set at time t

- Risk set: all observations who have not failed and have not been censored just prior to time t

Properties of Maximum Partial Likelihood Estimator

- Consistency: $\hat{\theta}_J \xrightarrow{P} \theta$ where

$$\hat{\theta}_J \ = \ \underset{\theta \in \Theta}{\text{argmax}} \sum_{j=1}^{J} \log f(W_j \mid V^{(j)}, W^{(j-1)}, \theta)$$

- Asymptotic normality:

$$\sqrt{J}(\hat{\theta}_J - \theta) \ \xrightarrow{D} \ \mathcal{N}(0, \ I(\theta)^{-1})$$

  where $I(\theta)$ is the information matrix based on the partial likelihood
- Usual hypothesis testing (Wald, Partial LRT)
- Unlike MLE, not efficient: the information about θ which is contained in $f(V_j \mid V^{(j-1)}, W^{(j-1)}, \theta)$ is being ignored
- But, in the Cox model, this term is silent about who fails $\rightarrow$ little information about β

Treatment of Ties

- Continuous time means no ties, but, rounding creates ties
- How do we incorporate ties into the partial likelihood framework?
- Idea: Break the ties in all possible ways and average the likelihood
- A special case of 2 ties:

$$\underbrace{\frac{\exp(X_1(t_j)^\top \beta)}{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta)}}_{obs.\ 1\ fails\ first} \times \underbrace{\frac{\exp(X_2(t_j)^\top \beta)}{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta) - \exp(X_1(t_j)^\top \beta)}}_{obs.\ 2\ fails\ next}$$

$$+ \underbrace{\frac{\exp(X_2(t_j)^\top \beta)}{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta)}}_{obs.\ 2\ fails\ first} \times \underbrace{\frac{\exp(X_1(t_j)^\top \beta)}{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta) - \exp(X_2(t_j)^\top \beta)}}_{obs.\ 1\ fails\ next}$$

- In general, if d_j is the number of observations failing at time t_j,

$$\prod_{j=1}^{J}\left[\exp\left(\sum_{i=1}^{d_j} X_i(t_j)^\top \beta\right)\left\{\sum_{P\in Q_j}\prod_{r=1}^{d_j}\frac{1}{\sum_{i'\in R(t_j,P,r)}\exp(X_{i'}(t_j)^\top\beta)}\right\}\right]$$

  where Q_j is the set of $d_j!$ permutations of $\{1,2,\ldots,d_j\}$, $P=\{1,\ldots,p_{d_j}\}$ is an element of Q_j, and $R(t_j,P,r)=R(t_j)\setminus\{p_1,\ldots,p_{r-1}\}$
- Discrete-time approximation:

$$\prod_{j=1}^{J}\frac{\exp\left(\sum_{i=1}^{d_j} X_i(t_j)^\top\beta\right)}{\sum_{S'\in S_j}\exp\left(\sum_{i'\in S'} X_{i'}(t_j)^\top\beta\right)}$$

  where S_j is the set of all possible combinations of d_j units selected from the risk set $R(t_j)$
- This corresponds to `method = "exact"` of the `coxph()` function in R

Some Popular Approximations

- But these are hard to compute if the number of ties is large
- Breslow (Peto):

$$\prod_{j=1}^{J} \frac{\exp\left(\sum_{i=1}^{d_j} X_i(t_j)^\top \beta\right)}{\left\{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta)\right\}^{d_j}}$$

- Efron:

$$\prod_{j=1}^{J} \frac{\exp\left(\sum_{i=1}^{d_j} X_i(t_j)^\top \beta\right)}{\prod_{r=0}^{d_j-1}\left\{\sum_{i' \in R(t_j)} \exp(X_{i'}(t_j)^\top \beta) - \frac{r}{d_j}\sum_{i' \in D(t_j)} \exp(X_{i'}(t_j)^\top \beta)\right\}}$$

  where $D(t_j)$ is the set of observations failing at t_j

- No tie reduces to the original partial likelihood

Estimation of the Baseline Hazard Function

- Partial likelihood allows us to estimate β but what about $\lambda_0(y)$?
- Idea: Use the Kaplan-Meier estimator given the PMLE of β
- The key relationship with the time-invariant covariates:

$$S(t_j) \;=\; \exp\left(-\int_0^{t_j} \lambda(y, X_i) dy\right) \;=\; S_0(t_j)^{\exp(X_i^\top \beta)}$$

- Apply the Kaplan-Meier nonparametric likelihood framework:

$$\begin{aligned} & \prod_{j=1}^{J} \left\{ \prod_{i \in D(t_j)} \lambda(t_j, X_i) \prod_{i \in R(t_j) \setminus D(t_j)} \{1 - \lambda(t_j, X_i)\} \right\} \\ = & \prod_{j=1}^{J} \left\{ \prod_{i \in D(t_j)} \left(1 - \alpha_j^{\exp(X_i^\top \beta)}\right) \prod_{i \in R(t_j) \setminus D(t_j)} \alpha_j^{\exp(X_i^\top \beta)} \right\} \end{aligned}$$

where $\alpha_j = S_0(t_j)/S_0(t_{j-1})$

- Given $\hat{\beta}$, nonparametric MLE of α_j is a solution to:

$$\sum_{i \in D(t_j)} \frac{\exp(X_i^\top \hat{\beta}) \alpha_j^{\exp(X_i^\top \hat{\beta})}}{1 - \alpha_j^{\exp(X_i^\top \hat{\beta})}} = \sum_{i \in R(t_j) \setminus D(t_j)} \exp(X_i^\top \hat{\beta})$$

- When $d_j = 1$, we have a closed-form solution,

$$\hat{\alpha}_j = \left\{ 1 - \frac{\exp(X_i^\top \hat{\beta})}{\sum_{i \in R(t_j)} \exp(X_i^\top \hat{\beta})} \right\}^{\exp(-X_i^\top \hat{\beta})}$$

- Finally, we have,

$$\widehat{S}_0(t_j) = \prod_{j' : t_{j'} \leq t_j} \hat{\alpha}_{j'}, \quad \widehat{S}(t_j) = \widehat{S}_0(t_j)^{\exp(X_i^\top \hat{\beta})}$$

$$\hat{\lambda}_0(t_j) = \left\{ 1 - \hat{\alpha}_j^{\exp(X_i^\top \hat{\beta})} \right\} \exp(-X_i^\top \hat{\beta})$$

- With time-varying covariates, it will be more complex but possible

Relaxing the Proportionality Assumption

- The key assumption of the Cox model: proportionality
- Stratification: for observation i in strata $s[i]$

$$\lambda_{0s[i]}(t_j) \exp(X_i(t_j)^\top \beta)$$

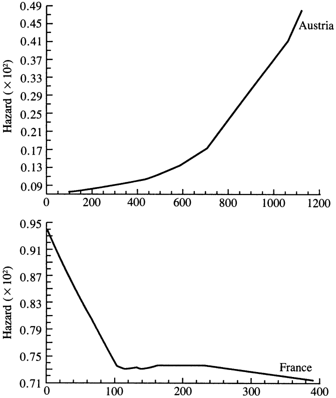
  Note that the partial likelihood function will be different from the no-stratification case because the risk set will be different across strata
- Time-varying coefficients:

$$\lambda_{0s[i]}(t_j) \exp(X_i(t_j)^\top \beta(t_j))$$

- Tests and diagnostics based on residuals (see Therneau and Grambsch , 2000; Box-Stefensmeier and Zorn, 2001 AJPS)

Back to the Cabinet Dissolution Example

- Warwick (1992 *AJPS*) applies the Cox model
- Some evidence of heterogeneity in the baseline hazard?



- Perhaps, stratify by countries?

Beyond a Single Failure Type

- So far, we assumed that each observation has a single failure type
- Multiple distinct failure types, i.e., *competing risks*
- Cabinet dissolution example (Diermeier & Stevenson, 1999 *AJPS* and 2000 *APSR*):
    1. Dissolution followed by an early election
    2. Replacement followed by a new cabinet
- Hazard rates may be different across competing risks
- Models based on cause-specific hazard functions
    1. Relationship between covariates and each cause-specific hazard
    2. Interrelation between failure types
    3. Estimation of hazard rates for certain failure types given the removal of some or all other failure types

Difference in Raw Hazard Rates

FIGURE 7. Estimates—Dissolution Hazard

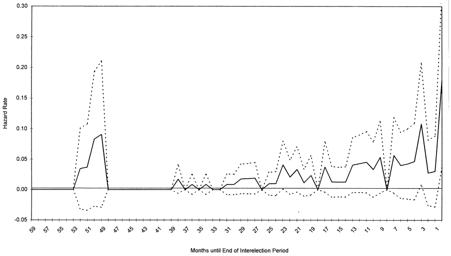
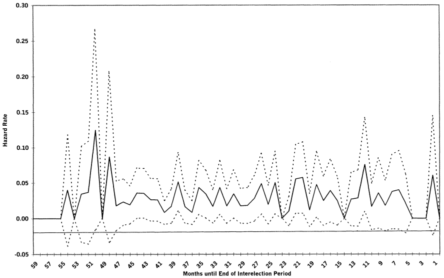


FIGURE 8. Estimates—Replacement Hazard



Competing-Risks Model

- Cause-specific hazard function:

$$\lambda_k(y) \;=\; \lim_{h \downarrow 0} \frac{\Pr(y \leq Y_i < y + h, F_i = k \mid Y_i \geq y)}{h}$$

  where F_i represents the cause of failure
- This is identifiable from the data; for example,

$$\hat{\lambda}_k(t_j) = \frac{d_{jk}}{n_j}$$

  where d_{jk} is the number of units who failed at t_j from cause k and n_j is the number of those who are at risk at time t_j
- Assuming that only one failure type can occur at once:

$$\lambda(y) \;=\; \sum_{k=1}^{K} \lambda_k(y)$$

Cabinet Dissolution, One Last Time

- Diermeier and Stevenson (1999, *AJPS*)
- Failure types: dissolution vs. replacement
- The Cox model with the following likelihood function,

$$\prod_{j=1}^{J} \left\{ \lambda_{F_i}(t_j, X_i(t_j)) \prod_{k=1}^{K} \exp\left(- \int_0^{t_j} \lambda_k(s; X_i(s)) ds \right) \right\}$$

  where $\lambda_k(y, X_i(y)) = \lambda_{k0}(y) \exp(X_i(y)^\top \beta_k)$
- Obtain the maximum partial likelihood estimate of β_k for each k
- The authors do not report quantities of interest, e.g., cumulative incidence function

Survival Function with Multiple Causes

- Survival function:

$$S(y) = \exp\left(-\sum_{k=1}^{K}\int_0^y \lambda_k(t)dt\right) = \prod_{k=1}^{K}\exp\left(-\int_0^y \lambda_k(t)dt\right)$$

  which does not generally equal the product of "marginal" (cause-specific) survival function (more on this later)
- Cumulative incidence function:

$$I_k(y) \equiv \Pr(Y_i \leq y, F_i = k) = \int_0^y \lambda_k(t)S(t)dt$$

  which is *the probability of failure from cause k before time y*
- Maximum Likelihood Estimator:

$$\hat{I}_k(y) = \sum_{j:t_j\leq y}\hat{\lambda}_k(t_j)\widehat{S}(t_{j-1}) = \sum_{j:t_j\leq y}\hat{\lambda}_k(t_j)\left\{\prod_{j'\leq j-1}\left(1-\sum_{k=1}^{K}\hat{\lambda}_k(t_{j'})\right)\right\}$$

- Models (e.g., Cox PH) can be used for $\lambda_k(y, X_i(t))$

No Example in Political Science, Yet...

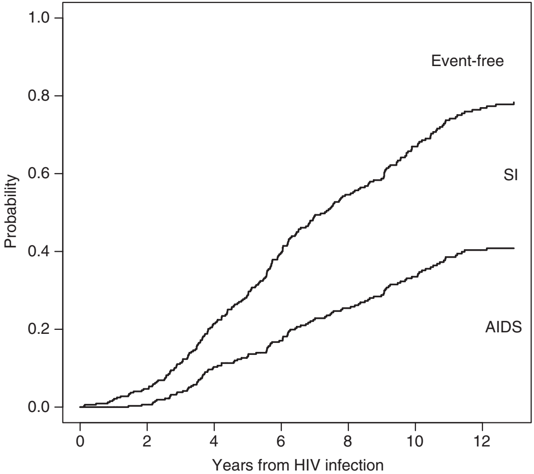


Figure 4. Cumulative incidence curves of AIDS and SI appearance. The cumulative incidence function are stacked; the distance between two curves represent the probabilities of the different events.

Identifiability

- Latent failure time approach: $Y_i = \min(Y_{i1}^*, \ldots, Y_{iK}^*)$
- Can the joint distribution of Y_i^* be identified?
- Joint survival function: $S^*(y_1, \ldots, y_K) = \Pr(Y_{i1}^* > y_1, \ldots, Y_{iK}^* > y_K)$
- Overall survival function: $S(y) = S^*(y, \ldots, y)$
- The data only identifies the cause-specific hazard:

$$\lambda_k(y) \; = \; -\frac{\partial}{\partial y_k} \log S^*(y_1, \ldots, y_K)\Big|_{y_1=\cdots=y_K=y}$$

- "Marginal" survival function, $Q_k(y) = \Pr(Y_i^* > y)$, is not identified unless, for example, the following independence holds

$$S^*(y_1, \ldots, y_K) \; = \; \prod_{k=1}^{K} Q_k(y_k)$$

- Fundamental problem: only a failure of one cause occurs for each unit

Concluding Remarks and Further Topics

- Many data in political science come in the form of survival data
- Do not force it into arbitrary data format (e.g., annual observations with binary event indicator)
- Report quantities of interest rather than a table of coefficients

- The literature on survival analysis has a long history and is quite sophisticated
- Political scientists have never used it until 1990!
- Fairly recent appearance of Cox model and competing risks model

- More interesting topics not covered in POL 573:
  1. Recurrent events
  2. Correlated failure time
  3. Multi-state models