Nonparametric and Semiparametric Estimators

Robert A. Miller

Structural Econometrics

September 2020

- Define the right continuous cumulative probability distribution function $F_x : \mathbb{R}^K \rightarrow [0, 1]$ as:

$$F_x(x) \equiv F_x(x_1, \ldots, x_K) \equiv \Pr\{X_1 \leq x_1, \ldots, X_K \leq x_K\}$$

- Denote by $x_n \equiv (x_{1n}, \ldots, x_{Kn}) \in \mathbb{R}^K$ for $n \in \{1, \ldots, N\}$ a draw of observations from $F(x)$.
- Without making any parametric assumptions, one could estimate $F(x)$ with the *empirical distribution function* defined as:

$$\begin{aligned} F_x^{(N)}(x) &\equiv F_x^{(N)}(x_1, \ldots, x_K) \\ &\equiv \frac{1}{N} \sum_{n=1}^{N} 1\{x_{1n} \leq x_1, \ldots, x_{Kn} \leq x_K\} \end{aligned}$$

- In words the estimator $F^{(N)}(x^*)$ is the sample proportion of observations with a value less than or equal to x^*.
- Note the derivative of $F^{(N)}(x^*)$ is zero wherever it exists, and $F^{(N)}(x^*)$ itself is punctuated by jump points (at each observation).

- The joint distribution of (y, x) can be estimated in the same way.
- We also define the empirical conditional distribution function as:

$$F_{y|x}^{(N)}(y|x) \equiv \frac{F_{y,x}^{(N)}(y,x)}{F_x^{(N)}(x)} = \frac{\sum_{n=1}^{N} 1\{y_n \leq y, x_{1n} \leq x_1, \ldots, x_{Kn} \leq x_K\}}{\sum_{n=1}^{N} 1\{x_{1n} \leq x_1, \ldots, x_{Kn} \leq x_K\}}$$

- Hence the estimated expected value of y conditional on $x \leq x^*$ is:

$$E_{y|x}^{(N)}(y|x \leq x^*) = \frac{\sum_{n=1}^{N} y_n 1\{x_{1n} \leq x_1, \ldots, x_{Kn} \leq x_K\}}{\sum_{n=1}^{N} 1\{x_{1n} \leq x_1, \ldots, x_{Kn} \leq x_K\}}$$

- This result can be extended to obtain expressions for $E_{y|x}^{(N)}(y|x^* < x \leq x^{**})$ and so forth.
- But if $F(x)$ does not have mass at x^* this approach does not yield a sensible estimate of $E_{y|x}(y|x^*)$.

Nonparametric Estimation

A uniform kernel for a univariate probability density function

- Suppose $F_x(x)$ is differentiable with *pdf* $f_x(x)$
- In the univariate case $(K = 1)$ approximate $f_x(x)$ with:

$$
\begin{aligned}
f_x^{(N)}(x) &\equiv \left[F_x^{(N)}\left(x + h^{(N)}\right) - F_x^{(N)}\left(x - h^{(N)}\right)\right] \Big/ 2h^{(N)} \\
&= \frac{1}{Nh^{(N)}} \sum_{n=1}^{N} 1\left\{x - h^{(N)} < x_n \leq x + h^{(N)}\right\} \\
&= \frac{1}{Nh^{(N)}} \sum_{n=1}^{N} k\left(\frac{x_n - x}{h^{(N)}}\right)
\end{aligned}
$$

where $k(z) \equiv \frac{1}{2}$ if $|z| \leq 1$ and $k(z) \equiv 0$ if $|z| > 1$.

- Defined this way $f_x^{(N)}(x)$ is called a uniform kernel estimator because its *kernel*, $k(z)$, is a uniform pdf.
- The rationale for the dependence of $h^{(N)}$, its *bandwidth*, on N is that $h^{(N)} \to 0$ but $Nh^{(N)} \to \infty$ as $N \to \infty$.

- More generally we can define a kernel estimator $f_x^{(N)}(x)$ for the multivariate density $f_x(x)$ of $x \in \mathbb{R}^K$ as:

$$f_x^{(N)}(x) \equiv \frac{1}{N\left(h^{(N)}\right)^K} \sum_{n=1}^{N} k_x\left(\frac{x_n - x}{h^{(N)}}\right)$$

where:

$$k_x(z_1, \ldots, z_K) \equiv \prod_{k=1}^{K} k(z_k)$$

and $k(z) : \mathbb{R} \rightarrow \mathbb{R}^+$ is a symmetric pdf with mean zero and finite variance:

$$k(z) \equiv k(-z) \qquad \int k(z)\, dz = 1 \qquad \int z^2 k(z)\, dz > 0$$

- We can form a kernel estimator for the nonparametric regression function of y on x:

$$E\left[y\left|x\right.\right] \equiv \int y f_{y|x}\left(y\left|x\right.\right) dy \equiv \int y \frac{f_{y,x}\left(y,x\right)}{f_x\left(x\right)} dy \tag{1}$$

by substituting in the kernel estimators for $f\left(x\right)$:

$$f_x^{(N)}\left(x\right) = N^{-1}\left(h^{(N)}\right)^{-K} \sum_{n=1}^{N} k_x\left(\frac{x_n - x}{h^{(N)}}\right)$$

and $f\left(y,x\right)$:

$$f_{y,x}^{(N)}\left(y,x\right) = N^{-1}\left(h^{(N)}\right)^{-(K+1)} \sum_{n=1}^{N} k_y\left(\frac{y_n - y}{h^{(N)}}\right) k_x\left(\frac{x_n - x}{h^{(N)}}\right)$$

into the right side of (1).

- As the next slide shows, this implies the estimator is the weighted average of $\{y_n\}_{n=1}^N$, given by:

$$E^{(N)}[y\,|x] \equiv \int y \frac{f_{y,x}^{(N)}(y,x)}{f_x^{(N)}(x)} dy = \sum\nolimits_{n=1}^N w^{(N)}(x,x_n)\, y_n$$

where the weights $w^{(N)}(x,x_n)$ are defined as:

$$w^{(N)}(x,x_n) \equiv k_x\left(\frac{x_n - x}{h^{(N)}}\right) \Big/ \sum\nolimits_{n=1}^N k_x\left(\frac{x_n - x}{h^{(N)}}\right)$$

- Note that $h^{(N)} \to 0$ as $N \to \infty$ and the weight on observations distant from x disappear (since the variance of the kernel pdf is finite).
- If, as in most practical applications, $k(z)$ is single peaked, then the weight on every observation declines with N.

Nonparametric Estimation

Derivation of kernel estimator

- To obtain the formula use the facts that:

$$\int zk(z)\,dz = 0 \text{ and } \int k(z)\,dz = 1$$

and the change in variables $z = (y_n - y) \Big/ h_y^{(N)}$ to obtain

$$\begin{aligned}
\int y \frac{f_{y,x}^{(N)}(y,x)}{f_x^{(N)}(x)} dy &= \int y \frac{\sum_{n=1}^{N} k_y\left(\frac{y_n - y}{h^{(N)}}\right) k_x\left(\frac{x_n - x}{h^{(N)}}\right)}{h^{(N)} \sum_{n=1}^{N} k_x\left(\frac{x_n - x}{h^{(N)}}\right)} dy \\
&= \sum_{n=1}^{N} w^{(N)}(x, x_n) \int \frac{y}{h_y^{(N)}} k_y\left(\frac{y_n - y}{h_y^{(N)}}\right) dy \\
&= \sum_{n=1}^{N} w^{(N)}(x, x_n) \int \left(y_n + h_y^{(N)} z\right) k_y(z)\,dz \\
&= \sum_{n=1}^{N} w^{(N)}(x, x_n)\, y_n
\end{aligned}$$

Nonparametric Estimation

Hazard function defined and estimated

- Let $t \in (0, T)$ denote continuous time, where $T \leq \infty$.
- Suppose $F(t)$ denotes the probability distribution over the event of stopping, as opposed to continuing.
- Thus $1 - F(t)$, called the *survivor function*, is the probability of continuing until at least time t.
- Suppose stopping occurs at time s. We define the *hazard rate* at t, the hazard of $s = t$ as:

$$h(t) \equiv \lim_{\Delta \to 0} \left[\frac{P(t \leq s < t + \Delta \,|\, s \geq t)}{\Delta} \right]$$

- The next slide shows:

$$h(t) = \frac{F'(t)}{1 - F(t)}$$

- We obtain nonparametric estimates of $h(t)$ by replacing $F'(t)$ and $F(t)$ with their nonparametric estimators.

Nonparametric Estimation

Proof of the functional form describing the hazard function

- From its definition:

$$
\begin{aligned}
h(t) &\equiv \lim_{\Delta \to 0} \left[\frac{P(t \le s < t + \Delta \,| s \ge t)}{\Delta} \right] \\
&= \lim_{\Delta \to 0} \int_t^{t+\Delta} \frac{F'(s \,| s \ge t)}{\Delta} ds \\
&= \lim_{\Delta \to 0} \frac{1}{\Delta} \int_t^{t+\Delta} \frac{F'(s)}{1 - F(t)} ds = \frac{F'(t)}{1 - F(t)}
\end{aligned}
$$

- In words the hazard rate is the probability density for the stopping time divided by the survivor function.
- In discrete time $t \in \{1, 2, \ldots, T\}$ where we denote the probability of stopping at t by $p(t)$ the discrete hazard takes the form:

$$
h(t) \equiv \frac{p(t)}{1 - \sum_{s=1}^{t-1} p(s)}
$$

- To estimate the support of a distribution function, a very different set of procedures is used.
- Suppose $F(x)$ is a cumulative distribution function with support $[\underline{x}, \overline{x}]$ for unknown $\underline{x}$ and $\overline{x}$ satisfying $-\infty \leq \underline{x} < \overline{x} \leq \infty$.
- Assume x_n is independently drawn from $F(x)$ for $n \in \{1, \ldots, N\}$.
- Estimate $\overline{x}_0$ with $\overline{x}^{(N)} \equiv \max\{x_1, \ldots, x_N\}$.
- Similarly estimate $\underline{x}^{(N)}$ with $\underline{x}_0 \equiv \min\{x_1, \ldots, x_N\}$.
- Note $\overline{x}^{(N)}$ is a monotone increasing sequence bounded above by $\overline{x}_0$.
- Also $P(x_n < x^*) = F(x^*) < 1$ for all $x^* < \overline{x}_0$
  $\Rightarrow P\left(\overline{x}^{(N)} < x^*\right) = F(x^*)^{(N)} \rightarrow 0$ for all $x^* < \overline{x}_0$.
- These observations motivate the choice of $\overline{x}^{(N)}$ as an estimator for $\overline{x}_0$.
- Analogous arguments apply to the properties of $\underline{x}^{(N)}$.

- To further develop this line of enquiry, suppose $E\left[y\,|x\right] \equiv g\left(x\right)$ and $g\left(x\right)$ is (weakly) increasing in all its arguments.
- Suppose, as before, $\left(y_n, x_n\right)$ are independent for $n \in \{1, \ldots, N\}$.
- To estimate $g\left(x^*\right)$ at x^* for the partial sums:

$$\max_{x}\left\{\frac{\sum_{n=1}^{N} 1\left\{x < x_n \leq x^*\right\} y_n}{\sum_{n=1}^{N} 1\left\{x < x_n \leq x^*\right\}}\right\} \tag{2}$$

or:

$$\min_{x}\left\{\frac{\sum_{n=1}^{N} 1\left\{x^* < x_n \leq x\right\} y_n}{\sum_{n=1}^{N} 1\left\{x^* < x_n \leq x\right\}}\right\} \tag{3}$$

- Thus the criterion function determining x, justified by the monotonicity assumption, determines the bandwidth automatically, contrasting with the discretion of bandwidth choice for an unrestricted nonparametric regression function estimator.

- Consider the following equation of augmenting a linear framework with an additive nonlinear real valued function:

$$y_n = x_n' \beta_0 + g(z_n) + \epsilon_n \tag{4}$$

where:

  - y is the dependent variable:
  - x and z are explanatory variables with no common components
  - g has an unknown functional form
  - β_0 and $g(z)$ are the parameters to be estimated

- We make the standard assumption that:

$$E[\epsilon_n | x_n] = E[\epsilon_n | z_n] = 0 \tag{5}$$

- The key to identification and estimation in this model is that given z_n there is covariation between y_n and x_n. Let:

$$\begin{aligned} \widetilde{y}_n &\equiv y_n - E\left[y_n \left| z_n \right.\right] \\ \widetilde{x}_n' &\equiv x_n' - E\left[x_n' \left| z_n \right.\right] \end{aligned}$$

1. Nonparametrically estimate $E\left[y_n \left| z_n \right.\right]$ and $E\left[x_n' \left| z_n \right.\right]$.
2. Estimate β_0 by regressing $\widetilde{y}_n$ on $\widetilde{x}_n$, and exploiting the identity:

$$\widetilde{y}_n = \widetilde{x}_n' \beta_0 + \epsilon_n \tag{6}$$

3. Substitute the estimates of β_0, $E\left[y_n \left| z_n \right.\right]$ and $E\left[x_n' \left| z_n \right.\right]$ into:

$$g\left(z_n\right) = E\left[y_n \left| z_n \right.\right] - E\left[x_n' \left| z_n \right.\right] \beta_0$$

Semiparametric Estimation

Identities behind inference in partially linear models

- From (4) and (5):

$$E\left[y_n \left| z_n\right.\right] = E\left[x_n' \left| z_n\right.\right]\beta_0 + g\left(z_n\right)$$

- Then (6) follows from subtracting this equation from (4):

$$\widetilde{y}_n \equiv y_n - E\left[y_n \left| z_n\right.\right] = \left\{x_n' - E\left[x_n' \left| z_n\right.\right]\right\}\beta_0 + \epsilon_n \equiv \widetilde{x}_n'\beta_0 + \epsilon_n$$

- Multiplying (6) by $\widetilde{x}_n$ and taking expectations yields:

$$E\left[\widetilde{x}_n\widetilde{y}_n\right] = E\left[\widetilde{x}_n\widetilde{x}_n'\right]\beta_0$$

- If $E\left[\widetilde{x}_n\widetilde{x}_n'\right]$ is invertible then:

$$\beta_0 = E\left[\widetilde{x}_n\widetilde{x}_n'\right]^{-1} E\left[\widetilde{x}_n\widetilde{y}_n\right]$$

and:

$$g\left(z_n\right) = E\left[y_n \left| z_n\right.\right] - E\left[x_n' \left| z_n\right.\right] E\left[\widetilde{x}_n\widetilde{x}_n'\right]^{-1} E\left[\widetilde{x}_n\widetilde{y}_n\right]$$

- Single index models are defined by the equation:

$$y_n = g\left(x_n'\beta_0\right) + \epsilon_n \tag{7}$$

where $x_n'\beta_0$ is an index and:

  - y is the dependent variable.
  - x is a $k \times 1$ vector of explanatory variables.
  - $g : \mathbb{R} \rightarrow \mathbb{R}$ has an unknown functional form to be estimated.
  - β_0 is a $k \times 1$ parameter vector to be estimated too.

- Yet again we assume $E\left[\epsilon_n \left| x_n \right.\right] = 0$.
- This assumption implies $E\left[y_n \left| x_n \right.\right] = g\left(x_n'\beta_0\right)$.

- If $g(\cdot)$ was known, β_0 could be estimated with NLS.
- Alternatively if β_0 was known, $g(x_n'\beta_0)$ could be estimated with a nonparametric regression.
- The following estimator combines these ideas, choosing β to solve:

$$\arg\min_{\beta} \sum_{n=1}^{N} \left[y_n - \widehat{g}(x_n'\beta)\right]^2 = \arg\min_{\beta} \sum_{n=1}^{N} \left[\widehat{g}(x_n'\beta)^2 - 2y_n\widehat{g}(x_n'\beta)\right]$$

  - with $k \times 1$ FOC:

$$\frac{1}{N}\sum_{n=1}^{N} \widehat{g}(x_n'\beta)\,\widehat{g}'(x_n'\beta)\,x_n = \frac{1}{N}\sum_{n=1}^{N} y_n'\widehat{g}(x_n'\beta)\,x_n$$

  - where $g(\cdot)$ is approximated by the kernel estimator $\widehat{g}(\cdot)$:

$$\widehat{g}(x'\beta) = \sum_{n=1}^{N} w^{(N)}(x'\beta, x_n'\beta)\,y_n$$

  - and weights $w^{(N)}(x, x_n)$ are defined:

$$w^{(N)}(x'\beta, x_n'\beta) \equiv k_x\left(\frac{x_n' - x'}{h^{(N)}}\beta\right) \Big/ \sum_{n=1}^{N} k_x\left(\frac{x_n' - x'}{h^{(N)}}\beta\right)$$

- Suppose:

$$y_n = x_n \beta + \epsilon_n$$

where ϵ_n is distributed normal with mean μ and variance σ^2 and:

$$d_n = \begin{cases} 2 \text{ if } \quad y_n > \overline{y} \\ 1 \text{ if } \underline{y} < \quad y_n \leq \overline{y} \\ 0 \text{ if } \quad y_n < \quad \underline{y} \end{cases}$$

- This model is parametrized by $\theta \equiv \left(\beta, \mu, \sigma, \overline{y}, \ \underline{y}\right)$.
- Conditional on x_n the probability density function of $1\left\{d_n = 1\right\} y_n$ is:

$$\exp\left[-\left(\frac{y_n - x_n\beta}{\sigma}\right)^2\right] \Big/ \int_{\underline{y}}^{\overline{y}} \exp\left[-\left(\frac{y - x_n\beta}{\sigma}\right)^2\right] dy$$

- Suppose we observe (y_n, x_n) for a truncated sample of N observations in which $d_n = 1$ and $\theta_0 \equiv \left(\beta_0, \mu_0, \sigma_0, \overline{y}_0, \underline{y}_0\right)$.
- Rather than estimating θ_0 simultaneously, we estimate:
  1. $\overline{y}_0$ with $\overline{y}^{(N)}$ and $\underline{y}_0$ with $\underline{y}^{(N)}$ where $\overline{y}^{(N)} \equiv \max\{y_1, \ldots, y_N\}$ and $\underline{y}^{(N)} \equiv \min\{y_1, \ldots, y_N\}$.
  2. $(\beta_0, \mu_0, \sigma_0)$ with ML (or NLS) as if $\overline{y}_0 = \overline{y}^{(N)}$ and $\underline{y}_0 = \underline{y}^{(N)}$:

$$
\begin{aligned}
& \underset{\beta,\mu,\sigma}{\arg\min} \prod_{n=1}^{N} \left\{ \frac{\exp\left[\frac{-1}{2\sigma^2}(y_n - x_n\beta)^2\right]}{\int_{\underline{y}^{(N)}}^{\overline{y}^{(N)}} \exp\left[\frac{-1}{2\sigma^2}(y_n - x_n\beta)^2\right] dy} \right\} \\
= \quad & \underset{\beta,\mu,\sigma}{\arg\min} \sum_{n=1}^{N} \left[(y_n - x_n\beta)^2 - \ln\left\{ \int_{\underline{y}^{(N)}}^{\overline{y}^{(N)}} \exp\left[\frac{-1}{2\sigma^2}(y_n - x_n\beta)^2\right] dy \right\} \right]
\end{aligned}
$$

Semiparametric Estimation

Partially linear model with a monotone nonparametric correction term

- In a variation on (4) suppose:

$$y_n = x_n\beta + g(x_n) + \epsilon_n \tag{8}$$

  where:
  - $g(x_n)$ is (weakly) increasing in all its arguments.
  - the nonzero coefficients β may, or might not, correspond to components in $x_n = (x_{1n}, \ldots, x_{kn})$ that affect $g(x_n)$.
  - $E[\epsilon_n | x_n] = 0$ implying x_n is an instrument.
- Thus (8) generalizes (4) in some directions, through a more flexible specification of x_n, and specializes it in others, with the monotonicity assumption.

Semiparametric Estimation

Estimating a partially linear model with a monotone nonparametric correction term

- Exploiting the nonparametric estimator for monotone functions:
  1. Estimate $g(x_n)$ with $\widehat{g}(x_n)$ computed from (2) or (3).
  2. Replace $g(x_n)$ with $\widehat{g}(x_n)$, and form:

$$\widehat{y}_n = y_n - \widehat{g}(x_n)$$

  3. Run OLS (or GLS) on:

$$\widehat{y}_n = x_n\beta + \epsilon_n \tag{9}$$

Choosing between the Alternatives

Factors influencing the choice of an estimation approach

- Should we estimate nolinear models:
  - nonparametrically
  - semiparametrically
  - or parametrically?
- Some factors influencing the choices:
  - number of obervations in the data set . . the number of parameters . . . not that there needs to be more observations than par
  - computational ease (and hence robustness in implementation)
  - the role of imposing restrictions in achieving greater efficiency and precision versus misspecifying the model
  - the nature of the predictions the model is designed to make . . elastiticites . . expected welfare losses

Choosing between the Alternatives

A competing risk model

- Should we estimate nolinear models:
  - nonparametrically
  - semiparametrically
  - or parametrically?
- Some factors influencing the choices:
  - number of obervations in the data set . . the number of parameters . . . not that there needs to be more observations than par
  - computational ease (and hence robustness in implementation)
  - the role of imposing restrictions in achieving greater efficiency and precision versus misspecifying the model
  - the nature of the predictions the model is designed to make . . elastiticites . . expected welfare losses

Discrete choices in continuous time

A competing risk model

- Orders are either cancelled or executed:

$$
\begin{aligned}
& \Pr\left\{ t+\tau_{execute} \leq t+\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right\} \\
= \;& F_{execute}\left(\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right) \\
= \;& 1 - \exp\left[-\exp\left(z_t' \kappa_k^{(b)} \right) \tau^{\beta_k^{(b)}} \right] \\
& \Pr\left\{ t+\tau_{cancel} \leq t+\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right\} \\
= \;& F_{cancel}\left(\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right) \\
= \;& 1 - \exp\left[-\exp\left(z_t' \gamma_k^{(b)} \right) \tau^{\alpha_k^{(b)}} \right]
\end{aligned}
$$

- The hazard rate is:

$$
\begin{aligned}
& \Pr\left\{ t+\tau_{execute} \leq t+\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right\} \\
= \;& F_{execute}\left(\tau \,|z_t \,, d_{kt}^{(b)} = 1 \right)
\end{aligned}
$$

Modeling in continuous time

Failure times and the hazard rate

- Let $T \in \mathbb{R}$ denote the *failure time* of a random event.
- Denote by $f(\tau)$ its pdf, $F(\tau)$ its cdf, and call $1 - F(\tau)$ the *survivor function*.
- Define the *hazard rate* for T at τ as:

$$h(\tau) \equiv \lim_{\Delta\tau \downarrow 0} \frac{\Pr\{T \in [\tau, \tau + \Delta\tau] | T \geq \tau\}}{\Delta\tau} = \frac{f(\tau)}{1 - F(\tau)}$$

- Hence:

$$-\int_0^\tau h(s)\,ds = \log[1 - F(\tau)] \Rightarrow F(\tau) = 1 - \exp\left(-\int_0^\tau h(s)\,ds\right)$$

- Since there is a one-to-one mapping between $h(\tau)$ and $f(\tau)$, we could specify the model in terms of $h(\tau)$ rather than $f(\tau)$.
- The log-likelihood for the n^{th} failure time, τ_n, is:

$$\log f(\tau_n) = \log\left[h(\tau_n)\exp\left(-\int_0^{\tau_n} h(s)\,ds\right)\right] = \log h(\tau_n) - H(\tau_n)$$

where $H(\tau)$ is the *integrated hazard*. (Thus $h(s) = H'(\tau)$.)

Modeling in Continuous time

Competing risk

- Extending the model now suppose:
  - There are J possible failure types for a system
  - T_j denotes the failure time for a failure of type $j \in \{1, \ldots, J\}$
  - Let $T \equiv \min \{T_1, \ldots, T_J\}$ denote the failure time of the system
- Define the marginal hazard for the latent event T_j by $h_j(t)$
- Since the probabiity of two or more failures occurring simultaneous is zero, the hazard rate for T is:

$$h(\tau) = \sum_{j=1}^{J} h_j(\tau)$$

- Hence the contribution of the n^{th} failure time, τ_n, is:

$$\log h(\tau_n) - H(\tau_n) = \sum_{j=1}^{J} h_j(\tau)$$

- endogenously determined by the choices individuals make.

- Extending the model still further suppose:
  - there are $j \in \{0, \ldots, J\}$ types of failures (events)
  - they occur at random times $\{t_1, t_2, \ldots\}$ within a competing risk framework.
  - conditional on a failure ocurring, let θ_j denote the probability it is a type j failure
- In a economic scenario, imagine:
  - at sporadic intervals agent(s) in the model moves (move)
  - j is the outcome of a particular individual solving an optimization problem
  - or j is an equilibrium, simultaneaous move of several agents in a noncooperative game
- The data:
  - are, however, truncated, never reporting $j = 0$ type events whenever they occur
  - comprise a sequence of N observations of the form $\{t_n, j_n\}_{n=1}^{N}$

Modeling in Continuous time

Estimating models of competing risk with selection on unobserved variables

- Then the joint probability must account for the possibility that in intervals $[t_n, t_{n+1})$:
  1. either the system did not fail
  2. or a zero type event occurred at least once
- Hence the probability of not observing the system fail in the time interval $[t, t + dt)$ is:

$$1 - \sum_{j=1}^{J} h_j(t)\, dt + h_0(t)\, \theta_0 dt \equiv 1 - h^*(t)\, dt + h_0(t)\, \theta_0 dt$$

- Therefore log likelihood for the sequence is:

$$\sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \left\{ \log h(t_n - t_{n-1}) - H(t_n - t_{n-1}) + \ln \theta_j + \ln h_0(t)\, \theta_0 \right\}$$

- For the most part these frameworks generate nonlinear equations

Modeling in Continuous time

Competing risk with selection

$$
\begin{aligned}
& \Pr\left\{t+\tau_{execute} \leq t+\tau \,|z_t\,, d_{kt}^{(b)}=1\right\} \\
= \quad & F_{execute}\left(\tau\,|z_t\,, d_{kt}^{(b)}=1\right) \\
= \quad & 1-\exp\left[-\exp\left(z_t'\kappa_k^{(b)}\right)\tau^{\beta_k^{(b)}}\right] \\
& \Pr\left\{t+\tau_{cancel} \leq t+\tau \,|z_t\,, d_{kt}^{(b)}=1\right\} \\
= \quad & F_{cancel}\left(\tau\,|z_t\,, d_{kt}^{(b)}=1\right) \\
= \quad & 1-\exp\left[-\exp\left(z_t'\gamma_k^{(b)}\right)\tau^{\alpha_k^{(b)}}\right]
\end{aligned}
$$

- The hazard rate is:

$$
\begin{aligned}
& \Pr\left\{t+\tau_{execute} \leq t+\tau \,|z_t\,, d_{kt}^{(b)}=1\right\} \\
= \quad & F_{execute}\left(\tau\,|z_t\,, d_{kt}^{(b)}=1\right)
\end{aligned}
$$