Lecture 8. Conditional Choice Probability (CCP) estimators

Ciprian Domnisoru

Aalto University

A Review of the Dynamic Discrete Choice Model: Choices

- Each period $t \in \{1, 2, \ldots, T\}$ for $T \leq \infty$, an individual chooses among J mutually exclusive actions.
- Let d_{tj} equal one if action $j \in \{1, \ldots, J\}$ is taken at time t and zero otherwise:

$$d_{tj} \in \{0, 1\}$$

$$\sum_{j=1}^{J} d_{tj} = 1$$

- Suppose that actions taken at time t can potentially depend on the state $z_t \in \mathbb{Z}$.
- A transition probability $F_{tj}\left(z_{t+1} \left| z_t\right.\right)$, with density $f_{tj}\left(z_{t+1} \left| z_t\right.\right)$ when z_t is continuous, determines how z_t evolves stochastically over time with actions j.

A Review of the Dynamic Discrete Choice Model: Utility

- The current period payoff at time t from taking action j is $u_{tj}(z_t)$.
- Given choices $(d_{t1}, \ldots, d_{tJ})$ in each period $t \in \{1, 2, \ldots, T\}$ the individual's lifetime expected utility is:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1}d_{tj}u_{tj}(z_t)\,|z_1\right\}$$

where $\beta \in (0, 1)$ is the discount factor, and the expectation is taken over $z_{t+1}, \ldots, z_T$ given z_1.

A Review of the Dynamic Discrete Choice Model

Value function and optimization

- Denote the optimal decision rule by $d_t^o(z_t) \equiv (d_{t1}^o(z_t), \ldots, d_{tJ}^o(z_t))$.
- The current value function $V_t(z_t)$ is then defined as:

$$\begin{aligned} V_t(z_t) &= E\left\{\sum_{s=t}^{T}\sum_{j=1}^{J}\beta^{s-t} d_{sj}^o(z_s)\, u_{sj}(z_s)\,|z_t\right\} \\ &= \sum_{j=1}^{J} d_{tj}^o(z_t)\left[u_{tj}(z_t) + \beta\int V_{t+1}(z_{t+1}) f_{tj}(z_{t+1}\,|z_t)\,dz_{t+1}\right] \end{aligned}$$

- Let $v_{tj}(z_t)$ denote the flow payoff of action j plus the expected future utility of behaving optimally from period $t+1$ on:

$$v_{tj}(z_t) \equiv u_{tj}(z_t) + \beta\int V_{t+1}(z_{t+1}) f_{tj}(z_{t+1}\,|z_t)\,dz_{t+1}$$

- Bellman's principle implies:

$$d_{tj}^o(z_t) \equiv \prod_{k=1}^{K} \mathbf{1}\{v_{tj}(z_t) \geq v_{tk}(z_t)\}$$

A Review of the Dynamic Discrete Choice Model

Parameterizing the data generating process

- Typically we acknowledge that some of the factors affecting individual decision making are unobserved.
- This could explain why we:
  - cannot predict individual behavior exactly
  - estimate a probability distribution to stochastically characterize individual behavior.
- Accordingly partition $z_t \equiv (x_t, \epsilon_t)$ where x_t is observed, but ϵ_t is not.
- We define the data generating process, the DGP, as the probability distribution of the data, that is margined over the unobserved variables.
- The data comprise $\{d_{nt1}, \ldots, d_{ntJ}, x_{nt}\}$ for observations $(n, t) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$.

A Review of the Dynamic Discrete Choice Model: Estimation

- We assume $u_{tj}(z_t)$, $F_{tj}\left(z_{t+1} \middle| z_t\right)$ and β are fully characterized by $\theta \in \Theta$, where for example $\Theta \subseteq \mathbb{R}^p$, and p is a counting number.
- Thus the DGP is characterized by some unknown $\theta_0 \in \Theta$.
- Denote the *pdf* of $(x_{t+1}, \epsilon_{t+1})$ conditional on $(d_{t1}, \ldots, d_{tJ}, x_t, \epsilon_t)$ by:

$$
\begin{aligned}
& H_t\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right) \\
\equiv \quad & \sum_{j=1}^{J} d_{tj} d_{tj}^{o}\left(x_t, \epsilon_t; \theta\right) f_{tj}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right)
\end{aligned}
$$

- The ML estimator chooses θ to maximize:

$$
\prod_{n=1}^{N} \int_{\epsilon_T \ldots \epsilon_1} \left[\begin{array}{l} \sum_{j=1}^{J} d_{nTj} d_{Tj}^{o}\left(x_{nT}, \epsilon_T; \theta\right) \times \\ f_1\left(\epsilon_1 \left| x_{n1} \right. ; \theta\right) \prod_{t=1}^{T-1} H_t\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right. ; \theta\right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T
$$

A Review of the Dynamic Discrete Choice Model

A computational challenge

- What are the computational challenges to enlarging the state space?
  1. Computing the value function;
  2. Solving for equilibrium in a multiplayer setting;
  3. Integrating over unobserved heterogeneity.
- These challenges have led researchers to compromises on several dimensions:
  1. Keep the dimension of the state space small;
  2. Assume all choices and outcomes are observed;
  3. Model unobserved states as a matter of computational convenience;
  4. Consider only one side of market to finesse equilibrium issues;
  5. Adopt parameterizations based on convenient functional forms.

Separable Transitions in the Observed Variables

A simplification

- We could assume that for all (t, j, x_t, ϵ_t) the transition of the observed variables does not depend on the unobserved variables:

$$F_{tj}\left(x_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right) = F_{tj}\left(x_{t+1} \left| x_t \right. ; \theta\right)$$

- Note $F_{tj}\left(x_{t+1} \left| x_t \right.\right)$ is identified for each (t, j) from the transitions, so there is no conceptual reason for parameterizing this distribution.
- The ML estimator maximizes the same criterion function but $H_t\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right. ; \theta\right)$ simplifies to:

$$H_t\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right. ; \theta\right) \equiv$$
$$\sum_{j=1}^{J} d_{tj} d_{tj}^{o}\left(x_t, \epsilon_t ; \theta\right) f_{tj}\left(x_{t+1} \left| x_t \right. ; \theta\right) f_{t+1}\left(\epsilon_{t+1} \left| x_{t+1}, x_t, \epsilon_t \right. ; \theta\right)$$

Separable Transitions in the Observed Variables

Exploiting separability in estimation

- Instead of jointly estimating the parameters, we could use a two stage estimator to reduce computation costs:
  1. Estimate $F_{tj}\left(x_{t+1}\left|x_t\right.;\theta\right)$ with a cell estimator, a parametric function, or a nonparametric estimator, with $\widehat{F}_{tj}\left(x_{t+1}\left|x_t\right.;\theta\right)$.
  2. Define:

$$\widehat{H}_t\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.;\theta\right)\equiv \sum_{j=1}^{J} d_{tj}d_{tj}^{o}\left(x_t,\epsilon_t;\theta\right)\widehat{f}_{tj}\left(x_{t+1}\left|x_t\right.;\theta\right)f_{t+1}\left(\epsilon_{t+1}\left|x_{t+1},x_t,\epsilon_t\right.;\theta\right)$$

  3. Choose θ to maximize:

$$\prod_{n=1}^{N}\int_{\epsilon_T\ldots\epsilon_1}\left[\begin{array}{l}\sum_{j=1}^{J} d_{nTj}d_{Tj}^{o}\left(x_{nT},\epsilon_T;\theta\right)\times \\ f_1\left(\epsilon_1\left|x_{n1}\right.;\theta\right)\prod_{t=1}^{T-1}\widehat{H}_t\left(x_{n,t+1},\epsilon_{t+1}\left|x_{nt},\epsilon_t\right.;\theta\right)\end{array}\right]d\epsilon_1\ldots d\epsilon_T$$

  4. Correct standard errors induced at the first stage of estimation.

Conditional independence

- Separable transitions do not, however, free us from:
  1. the curse of multiple integration.
  2. numerical optimization to obtain the value function.
- Suppose we assume in addition that ϵ_{t+1}, conditional on x_{t+1}, is independent of x_t (plausible) and ϵ_t (questionable).
- Conditional independence embodies both assumptions:

$$\begin{aligned} F_{tj}\left(x_{t+1}\left|x_t,\epsilon_t\right.\right) &= F_{tj}\left(x_{t+1}\left|x_t\right.;\theta\right) \\ F_{t+1}\left(\epsilon_{t+1}\left|x_{t+1},x_t,\epsilon_t\right.\right) &= G_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.;\theta\right) \end{aligned}$$

- Conditional independence implies:

$$F_{tj}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right) = F_{tj}\left(x_{t+1}\left|x_t\right.;\theta\right) G_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.;\theta\right)$$

Conditional Independence

Simplifying expressions within the likelihood

- Conditional independence implies:

$$
\begin{aligned}
& \sum_{j=1}^{J} d_{nTj} d_{Tj}^{o} \left(x_{nT}, \epsilon_{T}; \theta\right) g_{1}\left(\epsilon_{1} \left| x_{n1} \right. ; \theta\right) \\
& \times \prod_{t=1}^{T-1} H_{t}\left(x_{t+1}, \epsilon_{t+1} \left| x_{t}, \epsilon_{t} \right. ; \theta\right) \\
= \; & \sum_{j=1}^{J} d_{nTj} d_{Tj}^{o} \left(x_{nT}, \epsilon_{T}; \theta\right) g_{1}\left(\epsilon_{1} \left| x_{n1} \right. ; \theta\right) \\
& \times \prod_{t=1}^{T-1} \sum_{j=1}^{J} \left[d_{tj} d_{tj}^{o} \left(x_{t}, \epsilon_{t}; \theta\right) f_{tj}\left(x_{t+1} \left| x_{t} \right. ; \theta\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{t+1} \right. ; \theta\right) \right] \\
= \; & \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{tj} f_{tj}\left(x_{t+1} \left| x_{t} \right. ; \theta\right) \\
& \times \prod_{t=1}^{T} \sum_{j=1}^{J} d_{tj} d_{tj}^{o} \left(x_{t}, \epsilon_{t}; \theta\right) g_{t}\left(\epsilon_{t} \left| x_{t} \right. ; \theta\right)
\end{aligned}
$$

ML under conditional independence

- Hence the contribution of $n \in \{1, \ldots, N\}$ to the likelihood is:

$$\int_{\epsilon_T \ldots \epsilon_1} \left[\begin{array}{l} \sum_{j=1}^{J} d_{nTj} d^o_{Tj} \left(x_{nT}, \epsilon_T; \theta \right) \times \\ g_1 \left(\epsilon_1 \left| x_{n1} \right. ; \theta \right) \prod_{t=1}^{T-1} H_t \left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right. ; \theta \right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

$$= \int_{\epsilon_T \ldots \epsilon_1} \left[\begin{array}{l} \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{tj} f_{tj} \left(x_{t+1} \left| x_t \right. \right) \times \\ \prod_{t=1}^{T} \sum_{j=1}^{J} d_{tj} d^o_{tj} \left(x_t, \epsilon_t; \theta \right) g_t \left(\epsilon_t \left| x_t \right. ; \theta \right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

$$= \prod\nolimits_{t=1}^{T-1} \sum\nolimits_{j=1}^{J} d_{tj} f_{tj} \left(x_{t+1} \left| x_t \right. \right)$$

$$\times \prod\nolimits_{t=1}^{T} \int_{\epsilon_t} \sum\nolimits_{j=1}^{J} d_{tj} d^o_{tj} \left(x_t, \epsilon_t \right) g_t \left(\epsilon_t \left| x_t \right. ; \theta \right) d\epsilon_t$$

Conditional choice probabilities defined

- Under conditional independence, we define for each (t, x_t) the conditional choice probability (CCP) for action j as:

$$\begin{aligned} p_{tj}(x_t) &\equiv \int_{\epsilon_t} d^o_{tj}(x_t, \epsilon_t)\, g_t(\epsilon_t | x_t)\, d\epsilon_t \\ &= E\left[d^o_{tj}(x_t, \epsilon_t) | x_t \right] \\ &= \int_{\epsilon_t} \prod_{k=1}^{J} I\left\{ v_{tk}(x_t, \epsilon_t) \le v_{tj}(x_t, \epsilon_t) \right\} g_t(\epsilon_t | x_t)\, d\epsilon_t \end{aligned}$$

- Using this notation, the log likelihood can now be compactly expressed as:

$$\sum_{n=1}^{N} \sum_{t=1}^{T-1} \sum_{j=1}^{J} d_{ntj} \ln\left[f_{tj}(x_{n,t+1} | x_{nt}\, ; \theta) \right] + \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} d_{ntj} \ln p_{tj}(x_t; \theta)$$

Reformulating the primitives

- Conditional independence implies that $v_{tj}(x_t, \epsilon_t)$ only depends on ϵ_t through $u_{tj}(x_t, \epsilon_t)$ because:

$$
\begin{aligned}
v_{tj}(x_t, \epsilon_t) \quad \equiv \quad & u_{tj}(x_t, \epsilon_t) \\
& + \beta \int_{\epsilon} \int_{x_{t+1}} \left\{ \begin{array}{l} V_{t+1}(x_{t+1}, \epsilon) \times \\ f_{tj}\left(x_{t+1} \left| x_t \right.\right) g_{t+1}\left(\epsilon \left| x_{t+1} \right.\right) dx_{t+1} d\epsilon \end{array} \right\}
\end{aligned}
$$

- Without further loss of generality we now redefine the primitives by:
  - the preferences $u_{tj}^{*}(x_t) \equiv E\left[u_{tj}(x_t, \epsilon_t) \left| x_t \right.\right]$
  - the observed variables transitions $f_{jt}\left(x_{t+1} | x_t\right)$
  - and the distribution of unobserved variables $g_t^{*}\left(\epsilon_t^{*} | x_t\right)$ where $\epsilon_t^{*} \equiv (\epsilon_{1t}^{*}, \ldots, \epsilon_{Jt}^{*})$ and $\epsilon_{jt}^{*} \in \mathbb{R}$ for all (j, t), and:

$$
\epsilon_{tj}^{*} \equiv u_{tj}(x_t, \epsilon_t) - E\left[u_{tj}(x_t, \epsilon_t) \left| x_t \right.\right]
$$

Conditional value functions defined

- Given conditional independence, define the conditional value function as:

$$v_{tj}^{*}(x_t) \equiv u_{tj}^{*}(x_t) + \beta \int_{\epsilon} \int_{x_{t+1}} \left\{ \begin{array}{l} V_{t+1}(x_{t+1}, \epsilon^{*}) \times \\ f_{tj}\left(x_{t+1} \left| x_t \right.\right) g_{t+1}^{*}\left(\epsilon^{*} \left| x_{t+1} \right.\right) dx_{t+1} d\epsilon^{*} \end{array} \right\}$$

- Thus $p_{tj}(x)$ is found by integrating over $(\epsilon_{t1}^{*}, \ldots, \epsilon_{tJ}^{*})$ in the regions:

$$\epsilon_{tk}^{*} - \epsilon_{tj}^{*} \leq v_{tj}^{*}(x_t) - v_{tk}^{*}(x_t)$$

hold for all $k \in \{1, \ldots, J\}$. That is $p_{tj}(x_t)$ can be rewritten:

$$\begin{aligned} & \int_{\epsilon_t} \prod_{k=1}^{J} \mathbf{1} \left\{ v_{tk}(x_{nt}, \epsilon_t) \leq v_{tj}(x_{nt}, \epsilon_t) \right\} g_t\left(\epsilon_t \left| x_t \right.\right) d\epsilon_t \\ = \quad & \int_{\epsilon_t} \prod_{k=1}^{J} I \left\{ \epsilon_{tk}^{*} - \epsilon_{tj}^{*} \leq v_{tj}^{*}(x_{nt}) - v_{tk}^{*}(x_{nt}) \right\} g_t^{*}\left(\epsilon_t^{*} \left| x_t \right.\right) d\epsilon_t^{*} \end{aligned}$$

Connection with static models

- Suppose we only had data on the last period T, and wished to estimate the preferences determining choices in T.
- By definition this is a static problem in which $v^*_{Tj}(x_T) \equiv u^*_{Tj}(x_T)$.
- For example to the probability of observing the J^{th} choice is:

$$p_{TJ}\left(x_T\right) \equiv \int_{-\infty}^{\epsilon^*_{TJ}+u^*_{TJ}(x_T) - u^*_{T1}(x_T)} \dots \int_{-\infty}^{\epsilon^*_{TJ}+u^*_{TJ}(x_T) - u^*_{T,J-1}(x_T)} \int_{-\infty}^{\infty} g^*_T\left(\epsilon^*_T \left| x_T\right.\right) d\epsilon^*_T$$

- The only essential difference between a estimating a static discrete choice model using ML and a estimating a dynamic model satisfying conditional independence using ML is that parameterizations of $v^*_{tj}(x_t)$ based on $u^*_{tj}(x_t)$ do not have a closed form, but must be computed numerically.

Bus Engines (Rust,1987)

Another renewal problem

- The job matching model (JPE 1984) is a renewal problem: with only one occupation and an infinite number of jobs, every new job match restarts life.
- However the model does not satisfy conditional independence, because posterior beliefs are unobserved state variables.
- Replacing bus engines is also a renewal problem.
- Mr. Zurcher decides whether to replace the existing engine ($d_{t1} = 1$), or keep it for at least one more period ($d_{t2} = 1$).
- If Zurcher keeps the engine ($d_{t2} = 1$) bus mileage advances to $x_{t+1} = x_t + 1$; alternatively $d_{t1} = 1$ and $x_{t+1} = 1$.
- Buses are also differentiated by a fixed characteristic $s \in \{0, 1\}$.
- The choice-specific shocks ϵ_{tj} are *iid* Type 1 extreme value (T1EV).

The value function and optimal decision rule

- Zurcher maximizes the expected discounted sum of payoffs:

$$E\left\{\sum_{t=1}^{\infty}\beta^{t-1}\left[d_{t2}(\theta_1 x_t+\theta_2 s+\epsilon_{t2})+d_{t1}\epsilon_{t1}\right]\right\}$$

- Because this is a stationary infinite horizon problem, age and time have no role.
- Let $V(x,s)$ denote the ex-ante value function at the beginning of period t, the discounted sum of current and future payoffs just before ϵ_t is realized and before the decision at t is made.
- We also define the conditional value function for each choice as:

$$v_j(x,s)=\begin{cases}\beta V(1,s) & \text{if } j=1\\ \theta_1 x+\theta_2 s+\beta V(x+1,s) & \text{if } j=2\end{cases}$$

- Optimizing behavior implies:

$$d_1^o\left(x,s,\epsilon_t\right)=\mathbf{1}\left\{\epsilon_{t2}-\epsilon_{t1}\leq v_1(x,s)-v_2(x,s)\right\}=1-d_2^o\left(x,s,\epsilon_t\right)$$

Bus Engines: The DGP and the CCPs

- We suppose the data comprises a cross section of N observations of buses $n \in \{1, \ldots, N\}$ reporting their:
  - fixed characteristics s_n,
  - engine miles x_n,
  - and maintenance decision (d_{n1}, d_{n2}).
- Let $p_1(x, s)$ denote the conditional choice probability (CCP) of replacing the engine given x and s.
- Stationarity and T1EV imply that for all t :

$$
\begin{aligned}
p_1(x, s) &\equiv \int_{\epsilon_t} d_1^o(x, s, \epsilon_t)\, g(\epsilon_t)\, d\epsilon_t \\
&= \int_{\epsilon_t} \mathbf{1}\{\epsilon_{t2} - \epsilon_{t1} \leq v_1(x, s) - v_2(x, s)\}\, g(\epsilon_t | x_t)\, d\epsilon_t \\
&= \{1 + \exp[v_2(x, s) - v_1(x, s)]\}^{-1}
\end{aligned}
$$

- An ML estimator could be formed off this equation following the steps described above.

Bus Engines: Exploiting the renewal property

- In future lectures we show that if ϵ_{jt} is T1EV, then for all (x, s, j):

$$V(x,s) = v_j(x,s) - \beta \log\left[p_j(x,s)\right] + 0.57\ldots$$

- Therefore the conditional value function of not replacing is:

$$\begin{aligned} v_2(x,s) &= \theta_1 x + \theta_2 s + \beta V(x, s+1) \\ &= \theta_1 x + \theta_2 s + \beta \left\{ v_1\left(x+1, s\right) - p_1(x+1, s) + 0.57\ldots \right\} \end{aligned}$$

- Similarly:

$$v_1(x,s) = \beta V(1,s) = \beta \left\{ v_1(1,s) - \ln\left[p_1(1,s)\right] + 0.57 \right\}\ldots$$

- Because bus engine miles is the only factor affecting bus value given s:

$$v_1(x+1, s) = v_1(1, s)$$

Bus Engines: Using CCPs to represent differences in continuation values

- Hence:

$$v_2(x,s) - v_1(x,s) = \theta_1 x + \theta_2 s + \beta \ln\left[p_1(1,s)\right] - \beta \ln\left[p_1(x+1,s)\right]$$

- Therefore:

$$\begin{aligned} p_1(x,s) &= \frac{1}{1+\exp\left[v_2(x,s) - v_1(x,s)\right]} \\ &= \frac{1}{1+\exp\left\{\theta_1 x + \theta_2 s + \beta \ln\left[\frac{p_1(1,s)}{p_1(x+1,s)}\right]\right\}} \end{aligned}$$

- Intuitively the CCP for current replacement is the CCP for a static model with an offset term.
- The offset term accounts for differences in continuation values using future CCPs that characterize optimal future replacements.

Bus Engines: CCP estimation

- Consider the following CCP estimator:
  1. Form a first stage estimator for $p_1(x, s)$ from the relative frequencies:

$$\hat{p}_1(x, s) \equiv \frac{\sum_{n=1}^{N} d_{n1} I(x_n = x) I(s_n = s)}{\sum_{n=1}^{N} I(x_n = x) I(s_n = s)}$$

  2. Substitute $\hat{p}_1(x, s)$ into the likelihood as incidental parameters to estimate $(\theta_1, \theta_2, \beta)$ with a logit:

$$\frac{d_{n1} + d_{n2} \exp(\theta_1 x_n + \theta_2 s_n + \beta \ln\left[\frac{\hat{p}_1(1, s_n)}{\hat{p}_1(x_n + 1, s_n)}\right]}{1 + \exp(\theta_1 x_n + \theta_2 s_n + \beta \ln\left[\frac{\hat{p}_1(1, s_n)}{\hat{p}_1(x_n + 1, s_n)}\right]}$$

  3. Correct the standard errors for $(\theta_1, \theta_2, \beta)$ induced by the first stage estimates of $p_1(x, s)$.
- Note that in the second stage $\ln\left[\frac{\hat{p}_1(1, s_n)}{\hat{p}_1(x_n + 1, s_n)}\right]$ enters the logit as an individual specific component of the data, the β coefficient entering in the same way as θ_1 and θ_2.

Monte Carlo Study (Arcidiacono and Miller, 2011)

Modifying the bus engine problem

- Suppose bus type $s \in \{0, 1\}$ is equally weighted.
- Two state variables affect wear and tear on the engine:
  1. total accumulated mileage:

$$x_{1,t+1} = \begin{cases} \Delta_t \text{ if } d_{1t} = 1 \\ x_{1t} + \Delta_t \text{ if } d_{2t} = 1 \end{cases}$$

  2. a permanent route characteristic for the bus, x_2, that systematically affects miles added each period.
- More specifically we assume:
  - $\Delta_t \in \{0, 0.125, \ldots, 24.875, 25\}$ is drawn from a discretized truncated exponential distribution, with:

$$f(\Delta_t | x_2) = \exp\left[-x_2(\Delta_t - 25)\right] - \exp\left[-x_2(\Delta_t - 24.875)\right]$$

  - x_2 is a multiple 0.01 drawn from a discrete equi-probability distribution between 0.25 and 1.25.

Monte Carlo Study

Including the age of the bus in panel estimation

- Let θ_{0t} denote other bus maintenance costs tied to its vintage.
- This modification renders the optimization problem nonstationary.
- The payoff difference from retaining versus replacing the engine is:

$$u_{t2}(x_{t1}, s) - u_{t1}(x_{t1}, s) \equiv \theta_{0t} + \theta_1 \min\{x_{t1}, 25\} + \theta_2 s$$

- Denoting $x_t \equiv (x_{1t}, x_2)$, this implies:

$$\begin{aligned} v_{t2}(x_t, s) - v_{t1}(x_t, s) &= \theta_{0t} + \theta_1 \min\{x_{t1}, 25\} + \theta_2 s \\ &\quad + \beta \sum_{\Delta_t \in \Lambda} \left\{ \ln \left[\frac{p_{1t}(\Delta_t, s)}{p_{1t}(x_{1t} + \Delta_t, s)} \right] \right\} f(\Delta_t | x_2) \end{aligned}$$

- In the first three columns of the next table each sample simulation has 1000 buses observed for 20 periods.
- In the fourth column 2000 buses are observed for 10 periods.
- The mean and standard deviations are compiled from 50 simulations.

Monte Carlo Study: Extract from Table 1 of Arcidiacono and Miller (2011)

TABLE I

MONTE CARLO FOR THE OPTIMAL STOPPING PROBLEM[a]

		s Observed		Ignoring s	s Unobserved		Time Effects	
							s Observed	s Unobserved
	DGP	FIML	CCP	CCP	FIML	CCP	CCP	CCP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
θ_0 (intercept)	2	2.0100	1.9911	2.4330	2.0186	2.0280		
		(0.0405)	(0.0399)	(0.0363)	(0.1185)	(0.1374)		
θ_1 (mileage)	−0.15	−0.1488	−0.1441	−0.1339	−0.1504	−0.1484	−0.1440	−0.1514
		(0.0074)	(0.0098)	(0.0102)	(0.0091)	(0.0111)	(0.0121)	(0.0136)
θ_2 (unobs. state)	1	0.9945	0.9726		1.0073	0.9953	0.9683	1.0067
		(0.0611)	(0.0668)		(0.0919)	(0.0985)	(0.0636)	(0.1417)
β (discount factor)	0.9	0.9102	0.9099	0.9115	0.9004	0.8979	0.9172	0.8870
		(0.0411)	(0.0554)	(0.0591)	(0.0473)	(0.0585)	(0.0639)	(0.0752)
Time (minutes)		130.29	0.078	0.033	275.01	6.59	0.079	11.31
		(19.73)	(0.0041)	(0.0020)	(15.23)	(2.52)	(0.0047)	(5.71)

[a]Mean and standard deviations for 50 simulations. For columns 1–6, the observed data consist of 1000 buses for 20 periods. For columns 7 and 8, the intercept (θ_0) is allowed to vary over time and the data consist of 2000 buses for 10 periods. See the text and the Supplemental Material for additional details.