Discrete Choices

Robert A. Miller

Structural Econometrics

October 2026

Prototype Model

Choices

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

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

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

- At an abstract level assuming that choices are mutually exclusive is innocuous, because two combinations of choices sharing some features but not others can be interpreted as two different choices.

Prototype Model

Information and states

- Suppose that actions taken at time t can potentially depend on the state $z_t \in Z$.
- For Z finite denote by $f_{jt}(z_{t+1}|z_t)$, the probability of z_{t+1} occurring in period $t+1$ when action j is taken at time t.
- For example in the example above, suppose $z_t = (w_t, k_t)$ where:
  - $k_t \in \{0, 1, \ldots\}$ are the number of births before t
  - $w_t \equiv d_{1,t-1} + d_{2,t-1}$, so $w_t = 1$ if the female worked in period $t-1$, and $w_t = 0$ otherwise.
- With up to 5 offspring, 3 levels of experience, the number of states including age (say 50 years) is 750. Add in 4 levels of education (less than high school, high school, some college and college graduate) and 3 racial categories, increases this number to 9000.

Prototype Model

Large but sparse matrices

- When Z is finite there is a $Z \times Z$ transition matrix for each (j, t).
- In the example above they have $9,000^2 = 81$ million cells.
- In many applications the matrices are sparse.
- Suppose households can only increase the number of kids one at time.
- They can only change their work experience by one unit at most.
- Hence there are at most six cells they can move from (w_t, k_t):

$$\left\{ \begin{array}{l} (w_t, k_t), (w_t, k_t + 1), (w_t + 1, k_t), \\ (w_t + 1, k_t + 1), (w_t - 1, k_t), (w_t - 1, k_t + 1) \end{array} \right\}$$

- Therefore a transition matrix has at most $54,000$ nonzero elements, and all the nonzero elements are one.

Prototype Model

Preferences and expected utility

- The individual's current period payoff from choosing j at time t is determined by z_t, which is revealed to the individual at the beginning of the period t.
- The current period payoff at time t from taking action j is $u_{jt}^*(z_t)$.
- Given choices $(d_{1t}, \ldots, d_{Jt})$ in each period $t \in \{1, 2, \ldots, T\}$ and each state $z_t \in Z$ the individual's expected utility is:

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

  where $\beta \in (0,1)$ is the subjective discount factor, and at each period t the expectation is taken over $z_2, \ldots, z_T$.
- Formally β is redundant if u is subscripted by t.
- We typically include a geometric discount factor to bound infinite sums of utility so that the optimization problem is well posed.

Prototype Model

Value Function

- Write the optimal decision at period t as a decision rule denoted by $d_t^o(z_t)$ formed from its elements $d_{jt}^o(z_t)$.
- Let $V_t^*(z_t)$ denote the value function in period t, conditional on behaving according to the optimal decision rule:

$$V_t^*(z_t) \equiv E\left[\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t}d_{j\tau}^o(z_\tau)\,u_{j\tau}^*(z_\tau)\,|z_t\right]$$

- In terms of period $t+1$:

$$\beta V_{t+1}^*(z_{t+1}) \equiv \beta E\left\{\sum_{\tau=t+1}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^o(z_\tau)\,u_{j\tau}^*(z_\tau)\,|z_{t+1}\right\}$$

- Appealing to Bellman's (1958) principle we obtain, when Z is finite:

$$\begin{aligned} V_t^*(z_t) &= \sum_{j=1}^{J} d_{jt}^o u_{jt}^*(z_t) \\ &\quad + \sum_{j=1}^{J} d_{jt}^o \sum_{z \in Z} E\left[\sum_{\substack{\tau=t+1 \\ k=1}}^{T,J} \beta^{\tau-t} d_{k\tau}^o(z_\tau)\, u_{k\tau}^*(z_\tau) \,|z \right] f_{jt}(z|z_t) \\ &= \sum_{j=1}^{J} d_{jt}^o \left[u_{jt}^*(z_t) + \beta \sum_{z \in Z} V_{t+1}^*(z) f_{jt}(z|z_t) \right] \end{aligned}$$

Prototype Model

Optimization

- To compute the optimum for T finite, we first solve a static problem in the last period to obtain $d_T^o(z_T)$ for all $z_T \in Z$.
- Applying backwards induction $k \in \{1, \ldots, J\}$ is chosen to maximize:

$$u_{kt}^*(z_t) + E\left\{ \sum_{\tau=t+1}^{T} \sum_{j=1}^{J} \beta^{\tau-t-1} d_{j\tau}^o(z_\tau)\, u_{j\tau}^*(z_\tau) \,|z_t, d_{kt} = 1 \right\}$$

- In the stationary infinite horizon case we might assume $u_{jt}^*(z) \equiv u_j^*(z)$ and that $u_j^*(z) < \infty$ for all (j, z).
- Consequently expected utility each period is bounded and the contraction mapping theorem applies, proving $d_t^o(z) \longrightarrow d^o(z)$ for large T.

- Let $v_{jt}^{*}(z_t)$ denote the flow payoff of any action $j \in \{1, \ldots, J\}$ plus the expected future utility of behaving optimally from period $t+1$ on:

$$v_{jt}^{*}(z_t) \equiv u_{jt}^{*}(z_t) + \beta \sum_{z \in Z} V_{t+1}^{*}(z) f_{jt}\left(z | z_t\right)$$

- By definition:

$$d_{jt}^{o}\left(z_t\right) \equiv I\left\{v_{jt}^{*}(z_t) \geq v_{kt}^{*}(z_t) \forall\ k\right\}$$

- Suppose we observe the states z_{nt} and decisions $d_{nt} \equiv \left(d_{n1t}, \ldots, d_{nJt}\right)$ of individuals $n \in \{1, \ldots, N\}$ over time periods $t \in \{1, \ldots, T\}$.
- Could we use such data to infer the primitives of the model:
  1. A consistent estimator of $f_{jt}\left(z_{t+1} | z_t\right)$ can be obtained from the proportion of observations in the (t, j, z_t) cell transitioning to z_{t+1}.
  2. There are $(J-1)\sum_{n=1}^{N} I\left\{z_{nt} = z_t\right\}$ inequalities relating the pairs of mappings $v_{jt}^{*}(z_t)$ and $v_{kt}^{*}(z_t)$ observed at (t, z_t).
  3. Can we recursively derive the values of $u_{jt}^{*}(z_t)$ from the $v_{jt}^{*}(z_t)$ values?

Heterogeneity and Inference

Why unobserved heterogeneity is introduced into data analysis

- Note that if two people in the data set with the same (t, z_t) made different decisions, say j and k, then $v^*_{jt}(z_t) = v^*_{kt}(z_t)$.
- This raises two potential problems for modeling data this way:
  1. In a large data set it is easy to imagine that for every choice $j \in \{1, \ldots, J\}$ and every (t, z_t) at least one sampled person n sets $d_{njt} = 1$. If so, we would conclude that the population was indifferent between all the choices, and hence the model would have no empirical content because no behavior could be ruled out.
  2. This approach does not make use of the information that some choices are more likely than others; that is the proportions of the sample taking different choices at (t, z_t) might vary, some choices being observed often, others perhaps very infrequently.
- For these two reasons, treating all heterogeneity as observed, and predicting the decisions of each individual, is not a promising approach to analyzing data.

- A more modest objective is to predict the probability distribution of choices margined over factors that individuals observe, but data analysts do not.
- Predicting the behavior of a population (rather than individuals), *essentially obliterates the difference between macroeconomics and microeconomics.*
- We now assume the states can be partitioned into those which are observed, x_t, and those that are not, ϵ_t.
- Thus $z_t \equiv (x_t, \epsilon_t)$.
- Suppose the data consist of N independent and identically distributed draws from the string of random variables $(X_1, D_1, \ldots, X_T, D_T)$.
- The n^{th} observation is given by $\left\{x_1^{(n)}, d_1^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)}\right\}$ for $n \in \{1, \ldots, N\}$.

- Denote the mixed probability (density) of the pair $(x_{t+1}, \epsilon_{t+1})$, conditional on (x_t, ϵ_t) and the optimal action is j, as:

$$H_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right) \equiv d_{jt}^{o}\left(x_t, \epsilon_t\right) f_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right)$$

- The probability of $\{d_1, x_2, \ldots, d_{T-1}, x_T, d_T\}$ given x_1 is:

$$\Pr\left\{d_1, x_2, \ldots, d_{T-1}, x_T, d_T \left| x_1 \right.\right\} = \tag{1}$$
$$\int_{\epsilon_T} \cdots \int_{\epsilon_1} \left[\begin{array}{l} g\left(\epsilon_1 \left| x_1 \right.\right) \sum_{j=1}^{J} d_{jT} d_{jT}^{o}\left(x_T, \epsilon_T\right) \times \\ \prod_{t=1}^{T-1} \sum_{j=1}^{J} d_{jt} H_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

where $g\left(\epsilon_1 \left| x_1 \right.\right)$ is the density of ϵ_1 conditional on x_1.

- Let $\theta \in \Theta$ uniquely index a specification of $u_{jt}^{*}(z_t)$, $f_{jt}\left(z_{t+1}|z_t\right)$ and β under consideration.
- Conditional on $x_1^{(n)}$ suppose $\left\{d_1^{(n)}, x_2^{(n)}, \ldots, , d_T^{(n)}\right\}_{n=1}^{N}$ was generated by $\theta_0 \in \Theta$.
- The maximum likelihood (ML) estimator, θ_{ML}, selects $\theta \in \Theta$ to maximize the joint probability of observed occurrences conditional on the initial conditions:

$$\theta_{ML} \equiv \arg\max_{\theta \in \Theta} \left\{ N^{-1} \sum_{n=1}^{N} \log\left(\Pr\left\{d_1^{(n)}, x_2^{(n)}, \ldots, x_T^{(n)}, d_T^{(n)} \,\middle|\, x_1^{(n)}; \theta\right\}\right)\right\} \tag{2}$$

- The first applications followed this route:
  - **Robert Miller** (JPE 1984) on job turnover . . . *updating beliefs about nonpecuniary benefits of job match*
  - **Kenneth Wolpin** (JPE 1984) on fertility . . . *different unobserved types of females*

Heterogeneity and Inference

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 compromise 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.

- **Ariel Pakes** (Econometrica 1986) introduced simulation to substitute for numerical integration in his work on patent renewal.
- For example:
  - rather than compute $\Pr\{d_1, x_2, \ldots, d_{T-1}, x_T, d_T \,|x_1\}$
  - simulate a path starting out at x_1 for each observation
  - and choose parameters to minimize the distance between the simulated paths with the sample paths
- In this way the accumulated interpolation error of multiple integration is replaced with an exponentially growing simulation error.
- To conduct the simulation it is still necessary to solve for the optimal decision rule.

Separable Transitions in the Observed Variables

A simplification

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

$$f_{jt}\left(x_{t+1} \left| x_t, \epsilon_t \right.\right) = f_{jt}\left(x_{t+1} \left| x_t \right.\right)$$

- Since we assume x_{t+1} conveys all the information of x_t for the purposes of forming probability distributions at $t+1$:

$$\begin{aligned} f_{jt}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t \right.\right) &\equiv g_{t+1}\left(\epsilon_{t+1} \left| x_{t+1}, x_t, \epsilon_t \right.\right) f_{jt}\left(x_{t+1} \left| x_t, \epsilon_t \right.\right) \\ &= g_{t+1}\left(\epsilon_{t+1} \left| x_{t+1}, \epsilon_t \right.\right) f_{jt}\left(x_{t+1} \left| x_t \right.\right) \end{aligned}$$

- The ML estimator maximizes the same criterion function but $H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right.\right)$ simplifies to:

$$H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right.\right) =$$

$$\sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{n,t+1}, \epsilon_t \right.\right) f_{jt}\left(x_{n,t+1} \left| x_{nt} \right.\right)$$

Separable Transitions in the Observed Variables

Exploiting separability in estimation

- Note $f_{jt}\left(x_{t+1}\left|x_t\right.\right)$ is identified for each (j,t) from the transitions.
- Instead of jointly estimating the parameters, we could use a two stage estimator to reduce computation costs:
  1. Estimate $f_{jt}\left(x_{t+1}\left|x_t\right.\right)$ with a cell estimator (for x finite), a nonparametric estimator, or a parametric function, denoted by $\widehat{f}_{jt}\left(x_{t+1}\left|x_t\right.\right)$;
  2. Define:

$$\widehat{H}_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)\equiv d_{jt}^{o}\left(x_t,\epsilon_t\right)\widehat{f}_{jt}\left(x_{t+1}\left|x_t\right.\right)$$

  3. Substitute $\widehat{H}_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)$ for $H_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)$ when forming (1) and the maximand in (2).
  4. Correct standard errors from the first stage estimator to account for the loss in asymptotic efficiency.

Conditional Independence

Conditional independence defined

- Separable transitions do not, however, free us from:
  1. the curse of multiple integration (or simulation);
  2. numerically optimization to obtain the value function.
- Suppose in addition, that conditional on x_t the unobserved variable ϵ_{t+1} is independent of ϵ_t.
- **John Rust** (Econometrica 1987) introduced the conditional independence assumption, which can be stated as:

$$\begin{aligned} f_{jt}\left(x_{t+1}\left|x_t,\epsilon_t\right.\right) &= f_{jt}\left(x_{t+1}\left|x_t\right.\right) \\ g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1},\epsilon_t\right.\right) &= g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right) \end{aligned}$$

It implies:

$$f_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right) = f_{jt}\left(x_{t+1}\left|x_t\right.\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right)$$

Conditional Independence

Simplifying expressions within the likelihood

- Assuming conditional independence:

$$H_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)=d_{jt}^{o}\left(x_t,\epsilon_t\right)f_{jt}\left(x_{t+1}\left|x_t\right.\right)g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right)$$

and hence:

$$\Pr\left\{d_1,x_2,\ldots,d_{T-1},x_T,d_T\left|x_1\right.\right\}=$$

$$\int_{\epsilon_T}\cdots\int_{\epsilon_1}\left[\begin{array}{l} g\left(\epsilon_1\left|x_1\right.\right)\sum_{j=1}^{J}d_{jT}d_{jT}^{o}\left(x_T,\epsilon_T\right)\times \\ \prod_{t=1}^{T-1}\sum_{j=1}^{J}d_{jt}H_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right)\end{array}\right]d\epsilon_1\ldots d\epsilon_T$$

$$=\prod_{t=1}^{T}\int_{\epsilon_t}\left[\sum_{j=1}^{J}d_{jt}d_{jt}^{o}\left(x_t,\epsilon_t\right)g_t\left(\epsilon_t\left|x_t\right.\right)\right]d\epsilon_t$$

$$\times\prod_{t=1}^{T-1}\sum_{j=1}^{J}d_{jt}f_{jt}\left(x_{t+1}\left|x_t\right.\right)$$

Conditional Independence

Conditional choice probabilities defined

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

$$p_{jt}(x_t) \equiv \int_{\epsilon_t} d_{jt}^o(x_{nt}, \epsilon_t) g_t(\epsilon_t | x_{nt}) d\epsilon_t$$

- Using this notation, the contribution of observation n to the log likelihood can now be compactly expressed as:

$$\begin{aligned} & \sum_{t=1}^{T-1} \sum_{j=1}^{J} d_{jt} \ln f_{jt}(x_{t+1} | x_t) \\ & + \sum_{t=1}^{T} \sum_{j=1}^{J} d_{jt} \ln \int_{\epsilon_t} \left[d_{jt}^o(x_t, \epsilon_t) g_t(\epsilon_t | x_t) \right] d\epsilon_t \\ = & \sum_{t=1}^{T-1} \sum_{j=1}^{J} d_{njt} \ln \left[f_{jt}(x_{n,t+1} | x_{nt}) \right] + \sum_{t=1}^{T} \sum_{j=1}^{J} d_{njt} \ln p_{jt}(x_t) \end{aligned}$$

Conditional Independence

Reformulating the primitives

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

$$\begin{aligned} v_{jt}^{*}(x_t, \epsilon_t) &\equiv u_{jt}^{*}(x_t, \epsilon_t) \\ &\quad + \beta \int_{\epsilon} \sum_{x=1}^{X} V_{t+1}^{*}(x, \epsilon) f_{jt}\left(x \,|x_t\right) g_{t+1}\left(\epsilon \,|x\right) d\epsilon \end{aligned}$$

- Without further loss of generality we now redefine ϵ_t so that:

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

- In this way we redefine the primitives by the preferences $u_{jt}(x_t)$, 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 $\epsilon_t \equiv \left(\epsilon_{1t}, \ldots, \epsilon_{Jt}\right)$.

Conditional Independence

Conditional value functions defined

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

$$\begin{aligned} v_{jt}(x_t) &= u_{jt}(x_t) \\ &\quad + \beta \int_{\epsilon^*} \sum_{x=1}^{X} V_{t+1}^{*}(x, \epsilon) f_{jt}\left(x \left| x_t \right.\right) g_{t+1}\left(\epsilon \left| x \right.\right) d\epsilon \end{aligned}$$

- Thus $p_{jt}\left(x\right)$ is found by integrating over $\left(\epsilon_{1t}, \ldots, \epsilon_{Jt}\right)$ in the regions:

$$\epsilon_{kt} - \epsilon_{jt} \leq v_{jt}(x_t) - v_{kt}(x_t)$$

- Thus $p_{jt}(x_t)$ can be rewritten:

$$p_{jt}(x_t) = \int_{\epsilon_t} \prod_{k=1}^{J} I\left\{\epsilon_{kt} - \epsilon_{jt} \leq v_{jt}(x_{nt}) - v_{kt}(x_{nt})\right\} g_t\left(\epsilon_t \left| x_t \right.\right) d\epsilon_t$$

Bus Engines

A renewal problem (Rust, 1987)

- Replacing bus engines can be modeled as a renewal problem.
- Mr Zurcher maximizes the expected discounted sum of payoffs:

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

where:

$$x_{t+1}=\begin{cases}1 \text{ if } d_{t1}=1 \text{ (and Zurcher replaces the engine)}\\ x_t+1 \text{ if } d_{t2}=1 \text{ (and Zurcher does not replace the engine)}\end{cases}$$

  - 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).

Bus Engines

Value functions and replacement CCP

- Let $d_{jt}^{o}\left(x_t, \epsilon_t\right)$ denote the optimal decision rule for choice $j \in \{1, 2\}$.
- Let $V(x_t, s)$ 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.
- That is:

$$V(x_t, s) = E\left\{\sum_{\tau=t}^{\infty} \beta^{t-1}\left[d_{2\tau}(\theta_1 x_\tau + \theta_2 s + \epsilon_{2\tau}) + d_{1\tau}\epsilon_{1t}\right] | x_t, s\right\}$$

- Also let $v_j(x, s)$ denote the conditional value function for choice $j \in \{1, 2\}$:

$$v_j(x, s) = \begin{cases} \beta V(0, s) & \text{if } j = 1 \\ \theta_1 x + \theta_2 s + \beta V(x+1, s) & \text{if } j = 2 \end{cases}$$

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}).
- Noting $\epsilon_{j\tau}$ is T1EV, the CCP for replacing the engine given x and s, is:

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

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

Bus Engines

Exploiting the renewal property

- In 47-813 we shall see that if ϵ_{jt} is T1EV, then for all (x, s, j):

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

- Therefore the conditional value function of not replacing is:

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

- 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 (Hotz and Miller,1993)

- 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

Modifying the bus engine problem (Arcidiacono and Miller, 2011)

- 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:

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

Monte Carlo Study

Extract from Table 1 of Arcidiacono and Miller (2011)

	DGP	FIML	CCP
	(1)	(2)	(3)
θ_0 (intercept)	2	2.0100	1.9911
		(0.0405)	(0.0399)
θ_1 (mileage)	−0.15	−0.1488	−0.1441
		(0.0074)	(0.0098)
θ_2 (unobs. state)	1	0.9945	0.9726
		(0.0611)	(0.0668)
β (discount factor)	0.9	0.9102	0.9099
		(0.0411)	(0.0554)
Time (minutes)		130.29	0.078
		(19.73)	(0.0041)

Why Structural Econometrics?

Statistical summaries

- **Every statistical model has structure**:
  - typically comprising equations and/or inequalities.
  - For example the linear regression model has the following structure:

$$\epsilon_n = y_n - x_n\beta \tag{3}$$

- **The key components of this "statistical" structure are the:**
  - observed variables . . . (y_n, x_n)
  - unobserved variables . . . ϵ_n
  - parameters to be estimated . . . β
- **A data generating process supports the structure. For example:**
  - the observed variables . . . (y_n, x_n)
  - are generated by the unobserved variables . . . ϵ_n
  - by linking (y_n, x_n) through (3) via the unknown parameter β.
- **Estimators of** β are:
  - justified assumptions about the unobserved variables . . . ϵ_n
  - often founded on Laws of Large Numbers and Central Limit Theorems

Why Structural Econometrics?

"Structural econometrics" is based on an economics model

- **The data generating process (DGP) is founded upon:**
  - optimization problems individuals solve
  - equilibrium
- **Demands internal consistency . . .**
  - rational individuals facing constraints
  - uncertainty is treated as a probability distribution
  - equilibrium (competitive, Nash refinement, optimal contract)
  - identification based on data generating process
  - estimation (founded on Laws of Large Numbers and Central Limit Theorems)
- **Causality . . .**
  - is an abstraction (nothing more)
  - based on explicit assumptions of economic framework
  - leads to counterfactual predictions derived from the model

Why Structural Econometrics?

Estimated stuctural econometrics do not replicate reality

- Suppose reality is the best possible data set anyone can imagine.
- Essentially every applied econometric analysis is a **data reduction exercise**:
  - a mapping from all the points in the data to a much smaller number of parameters
- Thus "structural econometrics" is a nonlinear projection of the data.
- Therefore results of the econometric analysis contain **less information** than the actual data set.

Why Structural Econometrics?

Advantages and challenges

- **Elegance and transparency** . . .
  - steps can be independently verified
  - partly depends how easily the estimator is computed
- **Counterfactual predictions** . . .
  - derived from the model
  - strictly applies only to the model
- **Nonlinearities** . . .
  - choices and outcomes typically nonlinear in model parameters
- **Unobserved heterogeneity** . . .
  - data is an incomplete description of what is relevant