1. Introduction to Dynamic Structural Econometrics

Robert A. Miller

DSE at UCL

June 2025

Some Context

Why structural econometrics?

- **Internal consistency** . . .
  - rational individuals facing constraints
  - uncertainty is treated as a probability distribution
  - equilibrium (competitive, Nash refinement, optimal contract)
  - data generating process (as if sample comes from model population)
  - estimation (founded on LLN and CLT)
- **Elegance and transparency** . . .
  - steps can be independently verified
  - less discretion (*but what is numerical zero?*)
- **Causality** . . .
  - a model based concept
  - economic framework based on explicit assumptions
  - causal econometrics: open ended question about valid instruments
- **Counterfactual predictions** . . .
  - derived from the model
  - strictly applies only to the model

- Heterogeneity inspires, enriches and complicates **theory** . . .
  - specialization and trade
  - social interactions within a homogeneous population seem limited
- Heterogeneity in **dynamic** environments . . .
  - physical investment . . . and consumption/saving decision
  - investment in human capital
  - atrophy and death
  - sequential revelation of information
- **Inference** with heterogeneous populations . . .
  - complicates interpretation of aggregated data
  - aids identification if observed
  - complicates estimation if unobserved

Some Context

Policy evaluation . . .

- How can we conduct **policy evaluation** without a model?
  - (*I don't know.*)
- Should the model's **parameters be determined by the population** under consideration?
  - (*At least wouldn't that be the ideal?*)
- Can a model be useful **without being realistic**?
  - (*Are lab rats and mice really human?*)
- What is realism . . . . accepting received **orthodoxy**?
  - (*Who decides what is realistic?*)
- What is research . . **challenging** orthodoxy?
  - (*in order to create value . . perhaps?*)

Where to Start
Data

- The data typically comprise a sample of individuals for which there are records on some of their:
  - background characteristics
  - choices
  - outcomes from those choices.
- What are the challenges to making predictions and testing hypotheses when we take this approach?
  1. The choices and outcomes of economic models are typically nonlinear in the underlying parameters of the model we wish to estimate.
  2. The data variables on background, choices and outcomes might be an incomplete description about what is relevant to the model.

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.
- *Modeling the state space is an art . . . or a task for machine learning?*

- 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 $d_t^o\left(z_t\right)$ 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:

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

- In terms of period $t+1$:

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

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

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

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 $i \in \{1, \ldots, J\}$ is chosen to maximize:

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

- 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.

Heterogeneity and Inference

Estimating a model when all heterogeneity is observed

- 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:

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

- By definition:

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

- Suppose we observe the states z_{nt} and decisions $d_{nt} \equiv (d_{n1t}, \ldots, d_{nJt})$ 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 $f_{jt}(z_{t+1}|z_t)$ can be obtained from the proportion of observations in the (t, j, z_t) cell transitioning to z_{t+1}.
  2. There are $(J-1)\sum_{n=1}^{N} I\{z_{nt} = z_t\}$ inequalities relating the pairs of mappings $v_{jt}(z_t)$ and $v_{kt}(z_t)$ for each observation on d_{nt} 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 trying to predict the decisions of individuals, is not a very 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:

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

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

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

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 $u_{jt}(z_t)$, $f_{jt}\left(z_{t+1} \middle| z_t\right)$ 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:

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

- 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*

- **Ariel Pakes** (Econometrica 1986) introduced simulation to substitute for numerical integration in his work on patent renewal.
- There has been considerable amount of work devoted to handling multiple integration, some of which I will discuss tomorrow.
- **Victor Aguirregaberia**'s lecture on fixed effects tomorrow is a new approach to this challenge.

- **John Rust** (Econometrica 1987) dispensed with the integration altogether by introducing the conditional independence assumption in Harold Zurcher paper.
- The joint mixed density function for the state in period $t+1$ conditional on (x_t, ϵ_t), denoted by $g_{t,x,\epsilon}\left(x_{t+1}, \epsilon_{t+1} \left| x_t, \epsilon_t\right.\right)$, satisfies the *conditional independence assumption*:

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

where:

  - $g_t\left(\epsilon_t | x_t\right)$ is a conditional density for the disturbances
  - $f_{jt}(x_{t+1}|x)$ is a transition probability for x conditional on (j, t).
- This assumption is widely used in the estimation of dynamic discrete choice models.

A Framework with Conditional Independence

Bounded additively separable preferences

- Denote the discount factor by $\beta \in (0,1)$ and the current payoff from taking action j at t given (x_t, ϵ_t) by $u_{jt}(x_t) + \epsilon_{jt}$.
- To ensure a transversality condition is satisfied, assume $\{u_{jt}(x)\}_{t=1}^{T}$ is a bounded sequence for each $(j,x) \in \{1,\ldots,J\} \times \{1,\ldots,X\}$, and so is:
$$\left\{ \int \max\left\{ |\epsilon_{1t}|, \ldots, |\epsilon_{Jt}| \right\} g_t\left(\epsilon_t | x_t\right) d\epsilon_t \right\}_{t=1}^{T}$$
- At the beginning of each period t the agent observes the realization (x_t, ϵ_t) chooses d_t to sequentially maximize:
$$E\left\{ \sum_{\tau=t}^{T} \sum_{j=1}^{J} \beta^{\tau-1} d_{j\tau} \left[u_{j\tau}(x_\tau) + \epsilon_{j\tau} \right] | x_t, \epsilon_t \right\} \qquad (1)$$
where the expectation is taken over future realized values $x_{t+1}, \ldots, x_T$ and $\epsilon_{t+1}, \ldots, \epsilon_T$ conditional on (x_t, ϵ_t).

A Framework with Conditional Independence

Optimization

- Denote the optimal decision rule at t as $d_t^o(x_t, \epsilon_t)$, with j^{th} element $d_{jt}^o(x_t, \epsilon_t)$, and define the *social surplus function* as:

$$V_t(x_t) \equiv E\left\{\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^o\left(x_\tau,\epsilon_\tau\right)\left(u_{j\tau}(x_\tau)+\epsilon_{j\tau}\right)\right\}$$

- The *conditional value function*, $v_{jt}(x_t)$, is defined as:

$$v_{jt}(x_t) \equiv u_{jt}(x_t) + \beta\sum_{x=1}^{X}V_{t+1}(x)f_{jt}\left(x|x_t\right)$$

- Integrating $d_{jt}^o(x_t, \epsilon)$ over $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_J)$ define the *conditional choice probabilities* CCPs by:

$$p_{jt}(x_t) \equiv E\left[d_{jt}^o\left(x_t,\epsilon\right)|x_t\right] = \int d_{jt}^o\left(x_t,\epsilon\right)g_t\left(\epsilon\,|x_t\right)d\epsilon$$

Extension to Dynamic Markov Games

Players, choices and state variables

- Consider a dynamic game for I countable players:
  1. $d_t^{(i)} \equiv \left(d_{t1}^{(i)}, \ldots, d_{tJ}^{(i)} \right)$ choice of player i in period t.
  2. $d_t \equiv \left(d_t^{(1)}, \ldots, d_t^{(I)} \right)$ choices of all the players in period t.
  3. $d_t^{(-i)} \equiv \left(d_t^{(1)}, \ldots, d_t^{(i-1)}, d_t^{(i+1)}, \ldots, d_t^{(I)} \right)$ choices of all but i in t.
  4. x_t value of state variables of the game in period t.
  5. $F\left(x_{t+1} \left| x_t, d_t \right.\right)$ transition probability for x_{t+1} given (x_t, d_t).
  6. $F_j\left(x_{t+1} \left| x_t, d_t^{(-i)} \right.\right) \equiv F\left(x_{t+1} \left| x_t, d_t^{(-i)}, d_{jt}^{(i)} = 1 \right.\right)$ transition probability for x_{t+1} given x_t, i choosing j, and everyone else $d_t^{(-i)}$.

Extension to Dynamic Markov Games

Payoffs, information and CCPs

- The summed discounted payoff to i from playing the game is:

$$\sum_{t=1}^{T} \sum_{j=1}^{J} \beta^{t-1} d_{jt}^{(i)} \left[U_j^{(i)} \left(x_t, d_t^{(-i)} \right) + \epsilon_{jt}^{(i)} \right]$$

  where:

  1. $U_j^{(i)} \left(x_t, d_t^{(-i)} \right)$ depends on the choices of all the players.
  2. $\epsilon_t^{(i)} \equiv \left(\epsilon_{1t}^{(i)}, \dots, \epsilon_{Jt}^{(i)} \right)$ is *iid* across i with density $g \left(\epsilon_t^{(i)} | x_t \right)$.
  3. neither $d_t^{(-i)}$ nor $\epsilon_t^{(-i)}$ are observed by i.

- Analogous to the single agent setup define:

  1. $p_j^{(i)}(x_t) = \int d_j^{(i)} \left(x_t, \epsilon_t^{(i)} \right) g \left(\epsilon_t^{(i)} \right) d\epsilon_t^{(i)}$ as the CCP for the i choosing j in period t.
  2. $P \left(d_t^{(-i)} | x_t \right) = \prod_{i'=1, i' \neq i}^{I} \left(\sum_{j=1}^{J} d_{jt}^{(i')} p_j^{(i')}(x_t) \right)$ as the CCP for all the other players choosing $d_t^{(-i)}$ in period t.

Extension to Dynamic Markov Games

Equilibrium defined

- Then $\left(p_1^{(i)}(x_t), \ldots, p_J^{(i)}(x_t)\right)$ is an equilibrium if $d_j^{(i)}\left(x_t, \epsilon_t^{(i)}\right)$ solves the individual optimization problem (1) for each $\left(i, x_t, \epsilon_t^{(i)}\right)$ when:

$$u_j^{(i)}\left(x_t\right) = \sum_{d_t^{(-i)}} P\left(d_t^{(-i)} \left| x_t\right.\right) U_j^{(i)}\left(x_t, d_t^{(-i)}\right) \qquad (2)$$

  and:

$$f_j^{(i)}\left(x_{t+1} \left| x_t^{\cancel{(i)}}\right.\right) = \sum_{d_t^{(-i)}} P\left(d_t^{(-i)} \left| x_t^{(i)}\right.\right) F_j\left(x_{t+1} \left| x_t, d_t^{(-i)}\right.\right) \qquad (3)$$

- To analyze dynamic games taking this form:
  1. interpret $u_j^{(i)}\left(x_t\right)$ with (2) and $f_j^{(i)}\left(x_{t+1} \left| x_t^{(i)}\right.\right)$ with (3)
  2. in estimation treat the *best reply function* as the solution to a dynamic discrete choice optimization problem within the equilibrium played out by the *data generating process* DGP.

"1000 flowers bloom . . . 100 schools of thought contend"

- **Nesting the equilibrium solution within the estimation algorithm:**
  - integrate the model solution into the estimation routine with a nested fixed point algorithm, for example NFXP
  - yields the maximum likelihood estimator.
  - is a way to achieve asymptotic efficiency.
  - and the fixed point algorithm doubles as the solution to counterfactuals.
- **Bertel Schjerning** and later **Fedor Iskhakov** will lecture on this approach later today.

- **Separating inference from the model solution:**
    - exploit model data generating process (without solving it) to determine identification and obtain estimates
    - gives the identification conditions.
    - yields less efficient but much faster estimates.
    - requires the model solution to compute counterfactuals.
- I take this approach in the next lecture.

- **Calibration methods:**
  - typically disconnects sample variation from population probabilities.
  - can dispense with the estimation step altogether.
  - use numerical values drawn from published empirical work to quantify model solution, sometimes called calibration.
  - do not typically gives estimates of precision.
  - focuses on key restrictions and model moments.

"1000 flowers bloom . . . 100 schools of thought contend"

How should we solve and estimate dynamic models?

- Academics squabble . . .
- Relevant factors for this debate might be:
  - the kind of data including how much
  - the complexity of the model
  - the sensitivity of the estimates to the underlying assumptions
  - is sample variation an important factor in assessing precision
  - what is the specific policy question
- Let's postpone that discussion until we see more clearly what each approach entails.

Bus Engines (Rust,1987)

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:

  - $d_{t1}=1$ and $x_{t+1}=1$ if Zurcher replaces the engine
  - $d_{t2}=1$ and bus mileage advances to $x_{t+1}=x_t+1$ if he keeps the engine
  - 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).

- Define the conditional value function for each choice as:

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

  where $V(x,s)$ denotes the social surplus function.

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

- ~~The previous lecture implies that~~ if ϵ_{jt} is T1EV, then for all (x, s, j):

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

- Therefore the conditional value function of not replacing is:

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

- Similarly:

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

- 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:

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

- Therefore:

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

- 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.

- 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:

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

Monte Carlo Study

Extract from Table 1 of Arcidiacono and Miller (2011)

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