Week 13: Dynamics and Endogeneity

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last week

- Individual-specific coefficients

This week

- Dynamics
  - Static models with panel data
  - Dynamic discrete choice example
  - Dynamic discrete choice models
- Endogeneity
  - Endogeneity in structural models
  - BLP estimation
  - Control function model

This week's reading

- Dynamics: Train textbook, chapter 7.7
- Endogeneity: Train textbook, chapter 13

Static Models with Panel Data

Static Vs. Dynamic Models

Static structural models

- An agent maximizes their objective function within the current time period without considering the effect on choices in future time periods
- Some simple "dynamics" can be incorporated by including:
  - Lagged or future explanatory variables
  - State dependence through lagged outcome variables
- All the models we have discussed have been static
- Panel data models can be (and usually are) static

Dynamic structural models

- An agent maximizes their objective function while explicitly considering the effect on choices in future time periods
  - A choice in one period may change the choice set in future periods and/or the utility of future choices
- Modeling this dynamic behavior requires a more complex framework

Logit Model with Panel Data

We can use the logit model to model discrete choices with panel data

$$U_{njt} = \beta' \boldsymbol{x}_{njt} + \varepsilon_{njt} \quad \Rightarrow \quad P_{nit} = \frac{e^{\beta' \boldsymbol{x}_{nit}}}{\sum_j e^{\beta' \boldsymbol{x}_{njt}}}$$

The logit assumption has to hold

$$\varepsilon_{njt} \sim \text{i.i.d. type I extreme value (Gumbel) with } Var(\varepsilon_{njt}) = \frac{\pi^2}{6}$$

- But the unobserved preferences of a decision maker that affect their choices are unlikely to be independent over time

The logit model with panel data is a sequence of static choices, not a fully dynamic model

- We assume the decision maker maximizes utility in each time period
- But we do not represent how a choice will affect future choices

Mixed Logit Model with Panel Data

We can better represent a sequence of choices over time using the mixed logit model

$$U_{njt} = \boldsymbol{\beta}_n' \boldsymbol{x}_{njt} + \varepsilon_{njt} \quad \Rightarrow \quad P_{n\boldsymbol{i}} = \int \prod_{t=1}^{T} \frac{e^{\boldsymbol{\beta}' \boldsymbol{x}_{ni_t t}}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}' \boldsymbol{x}_{njt}}} f(\boldsymbol{\beta} \mid \boldsymbol{\theta}) d\boldsymbol{\beta}$$

The individual-specific coefficients represent unobserved preferences

- We model these coefficients as random coefficients and estimate their distributions
- These individual coefficients yield unobserved correlations in choices over time periods

Although the mixed logit model better represents choices over multiple time periods, it is still inherently a sequence of static choices

- We do not represent how a choice will affect future choices

Simple "Dynamics" in Static Models

When using these static models, we can include lagged outcome variables to model

- Habit formation
- Variety-seeking behavior
- Switching costs
- Brand loyalty

These are all examples of how past choices affect the decision maker's utility in the current time period

But if past choices affect utility in the current time period, then the current choice affects future utility

- A rational decision maker will consider these effects on future utility when making the current choice
- To fully capture this discrete choice framework, we need to use a dynamic discrete choice model

Dynamic Discrete Choice Example

Dynamic Discrete Choice Example

Why do people attend college? (Or graduate school?)

- Because four years in college provides more utility than anything else the individual could have done in those four years?
- Or because college opens up a new set of jobs and higher salaries compared to not attending college?

A static model implicitly assumes that the first answer is correct

- The static model has no good way to represent that the future choice set and salaries will be different after graduating from college

To consider the second answer, we need to use a dynamic discrete choice model

- This model will explicitly represent how the decision to attend college affects the future choice set and salaries

Decision Maker with Two-Period Dynamics

A decision maker thinking about college considers two time periods

1. College (C) or work (W) for four years
   - U_{1C}: utility in period 1 from four years in college
   - U_{1W}: utility in period 1 from four years working
2. A set of J possible jobs for a career over many future years
   - U_{2j}^{C}: utility in period 2 from job j after attending college in period 1
   - U_{2j}^{W}: utility in period 2 from job j after working in period 1

The total utility obtained from attending college or working in period 1 is

$$TU_C = U_{1C} + \lambda \max_j (U_{2j}^{C})$$

$$TU_W = U_{1W} + \lambda \max_j (U_{2j}^{W})$$

where λ reflects the relative weighting of the two periods

The decision maker attends college if and only if $TU_C > TU_W$

Econometrician with Two-Period Dynamics

We decompose the utility of each alternative into an observed and an unobserved (to the econometrician) component

$$\begin{aligned} U_{1C} &= V_{1C} + \varepsilon_{1C} & \qquad U_{2j}^{C} &= V_{2j}^{C} + \varepsilon_{2j}^{C} \\ U_{1W} &= V_{1W} + \varepsilon_{1W} & \qquad U_{2j}^{W} &= V_{2j}^{W} + \varepsilon_{2j}^{W} \end{aligned}$$

and define $\boldsymbol{\varepsilon} = \{\varepsilon_{1C}, \varepsilon_{1W}, \varepsilon_{2j}^{C}, \varepsilon_{2j}^{W}\}$ with joint density $f(\boldsymbol{\varepsilon})$

The probability that the decision maker chooses to attend college is

$$\begin{aligned} P_C &= \Pr(TU_C > TU_W) \\ &= \int \mathbb{1}\Big[V_{1C} + \varepsilon_{1C} + \lambda \max_j (V_{2j}^{C} + \varepsilon_{2j}^{C}) \\ &\qquad\qquad > V_{1W} + \varepsilon_{1W} + \lambda \max_j (V_{2j}^{W} + \varepsilon_{2j}^{W})\Big] f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} \end{aligned}$$

We have to approximate this integral using simulation

Simplifications for Two-Period Dynamics

If we assume that random utility in the first period, ε_{1C} and ε_{1W}, is i.i.d. extreme value, the choice probability of attending college simplifies to

$$P_C = \int \frac{e^{V_{1C} + \lambda \max_j (V_{2j}^C + \varepsilon_{2j}^C)}}{e^{V_{1C} + \lambda \max_j (V_{2j}^C + \varepsilon_{2j}^C)} + e^{V_{1W} + \lambda \max_j (V_{2j}^W + \varepsilon_{2j}^W)}} g(\boldsymbol{\varepsilon}_2) d\boldsymbol{\varepsilon}_2$$

which is a simpler integral to simulate

If we also assume that random utility in the second period, ε_{2j}^C and ε_{2j}^W, is i.i.d. extreme value, the joint probability of attending college in period 1 and then taking job i in period 2 is

$$P_{Ci} = P_C \times \frac{e^{V_{2i}^C}}{\sum_{j=1}^J e^{V_{2j}^C}}$$

where P_C is the probability of attending college given above

Dynamic Discrete Choice Example With Three Periods

After many years working job j, the decision maker reaches retirement age and has a new choice to make

- Continue working full time
- Work part time and spend some retirement funds
- Retire and collect social security and/or pension funds

The retirement plan, social security payout, etc. differs for each of the possible jobs in period 2

- In period 2, the decision maker will consider how their job will affect utility in periods 2 and 3
- In period 1, the decision maker will consider how the decision to attend college will affect utility in all periods

Decision Maker with Three-Period Dynamics

A decision maker thinking about college considers three time periods

1. College (C) or work (W) for four years
2. A set of J possible jobs for a career over many future years
3. A set of S possible retirement plans
   - U_{3s}^{Cj}: utility in period 3 from retirement plan s after attending college in period 1 and working job j in period 2
   - U_{3s}^{Wj}: utility in period 3 from retirement plan s after working in period 1 and working job j in period 2

The total utility obtained from attending college or working in period 1 is

$$TU_C = U_{1C} + \lambda \max_j \left[U_{2j}^{C} + \theta \max_s (U_{3s}^{Cj}) \right]$$

$$TU_W = U_{1W} + \lambda \max_j \left[U_{2j}^{W} + \theta \max_s (U_{3s}^{Wj}) \right]$$

where λ and θ reflect the relative weighting of the three periods

The decision maker goes to college if and only if $TU_C > TU_W$

Econometrician with Three-Period Dynamics

We decompose the utility of each alternative into an observed and an unobserved (to the econometrician) component

$$
\begin{aligned}
U_{1C} &= V_{1C} + \varepsilon_{1C} \qquad & U_{2j}^{C} &= V_{2j}^{C} + \varepsilon_{2j}^{C} \qquad & U_{3s}^{Cj} &= V_{3s}^{Cj} + \varepsilon_{3s}^{Cj} \\
U_{1W} &= V_{1W} + \varepsilon_{1W} \qquad & U_{2j}^{W} &= V_{2j}^{W} + \varepsilon_{2j}^{W} \qquad & U_{3s}^{Wj} &= V_{3s}^{Wj} + \varepsilon_{3s}^{Wj}
\end{aligned}
$$

and define $\boldsymbol{\varepsilon} = \{\varepsilon_{1C}, \varepsilon_{1W}, \varepsilon_{2j}^{C}, \varepsilon_{2j}^{W}, \varepsilon_{3s}^{Cj}, \varepsilon_{3s}^{Wj}\}$ with joint density $f(\boldsymbol{\varepsilon})$

The probability that the decision maker chooses to attend college is

$$
\begin{aligned}
P_C &= \Pr(TU_C > TU_W) \\
&= \int \mathbb{1}\left[V_{1C} + \varepsilon_{1C} + \lambda \max_j \left[V_{2j}^{C} + \varepsilon_{2j}^{C} + \theta \max_s (V_{3s}^{Cj} + \varepsilon_{3s}^{Cj})\right]\right. \\
&\qquad\qquad \left. > V_{1W} + \varepsilon_{1W} + \lambda \max_j \left[V_{2j}^{W} + \varepsilon_{2j}^{W} + \theta \max_s (V_{3s}^{Wj} + \varepsilon_{3s}^{Wj})\right]\right] \\
&\qquad \times f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}
\end{aligned}
$$

Dynamic Discrete Choice Models

Dynamic Optimization Notation and Terminology

Some common notation and terminology for dynamic structural models

- $\{i_1, i_2, \ldots, i_t\}$: sequence of choices up to and including period t
- $U_{tj}(i_1, i_2, \ldots, i_{t-1})$: utility obtained in period t from alternative j, which depends on all previous choices
- $TU_{tj}(i_1, i_2, \ldots, i_{t-1})$: total utility (current and all future time periods) obtained from choosing alternative j in period t, assuming the optimal choice is made in all future periods
  - Known as the "conditional value function"
- $TU_t(i_1, i_2, \ldots, i_{t-1})$: total utility obtained from the optimal choice in period t, assuming the optimal choice is made in all future periods
  - $TU_t(i_1, i_2, \ldots, i_{t-1}) = \max_j TU_{tj}(i_1, i_2, \ldots, i_{t-1})$
  - Known as the "value function" or "valuation function" at time t

We need to calculate all possible values of $TU_{tj}(i_1, \ldots, i_{t-1})$ in order to express the optimal choice in each time period

Bellman Equation for Dynamic Discrete Choice

The decision maker chooses optimally (maximizes utility) in the current period knowing they will also choose optimally in every future period (and discounting the future with discount rate δ), which yields an expression for the value function at time t

$$TU_t(i_1, \ldots, i_{t-1}) = \max_j \left[U_{tj}(i_1, \ldots, i_{t-1}) + \delta \, TU_{t+1}(i_1, \ldots, i_t = j)\right]$$

This relation is the Bellman equation for dynamic discrete choice

We can also write down a Bellman equation for the conditional valuation function, $TU_{tj}(i_1, \ldots, i_{t-1})$

$$TU_{tj}(i_1, \ldots, i_{t-1}) = U_{tj}(i_1, \ldots, i_{t-1}) + \delta \max_k \left[TU_{t+1,k}(i_1, \ldots, i_t = j)\right]$$

Applying the Bellman Equation

If the number of time periods is finite, we can apply the Bellman equation through backward recursion to calculate all possible $TU_{tj}(i_1, \ldots, i_{t-1})$

1. Start in the last time period, $t = T$, with

$$TU_{Tj}(i_1, \ldots, i_{T-1}) = U_{Tj}(i_1, \ldots, i_{T-1})$$

2. Calculate total utility in period $T-1$, $TU_{T-1j}(i_1, \ldots, i_{T-2})$, as a function of the values of $TU_{Tj}(i_1, \ldots, i_{T-1})$ from step 1
3. Continue working backward until you reach period 1

We have to calculate $U_{tj}(i_1, \ldots, i_{t-1})$ for each t, each j, and each $\{i_1, i_2, \ldots, i_{t-1}\}$

- If there are J alternatives in each of T time periods, we have to calculate $J^T \times T$ utilities
- This computational burden is known as the "curse of dimensionality"

Choice Probabilities in Dynamic Discrete Choice

We will ultimately use these conditional value functions, $TU_{tj}(i_1, \dots, i_{t-1})$, to create an expression for choice probabilities

For example, the probability of choosing alternative i in period 1 is

$$\begin{aligned} P_{1i} &= \Pr(TU_{1i} > TU_{1j} \; \forall j \neq i) \\ &= \int \mathbb{1}\left[TU_{1i}(\varepsilon) > TU_{1j}(\varepsilon) \; \forall j \neq i\right] f(\varepsilon) d\varepsilon \end{aligned}$$

We have to simulate this choice probability by taking random draws from an assumed joint density of all unobserved utilities, $f(\varepsilon)$

See chapter 7.7.3 of the Train textbook for more details

Uncertainty in Dynamic Discrete Choice Models

So far we have assumed that the decision maker has perfect information about the future

- Utility of each alternative in each future time period
- How every possible sequence of choices affects this future utility
- But this is unlikely to be true!

We can model utility as a function of factors that are unknown in previous periods

- The decision maker maximizes total expected utility with the expectation taken over the density of the unknown factors
- This expectation adds another integral that has to be simulated, adding yet another layer of complexity and dimensionality to a problem that already suffers from the curse of dimensionality

See chapter 7.7.3 of the Train textbook for more details

Simplifications for Dynamic Discrete Choice Models

Use the fewest number of time periods possible

- We could model the college-job-retirement sequence of choices annually (or monthly, weekly, daily) instead of three broad time periods
- Estimation is feasible with three time periods, but it becomes (at least) an order of magnitude more difficult with 60 individual year

Assume the factors that the decision maker does not observe are the same factors that the econometrician does not observe, and these factors are i.i.d. extreme value

- Choice probabilities have closed-form expressions that are easy to calculate
- This assumption is unrealistic, but it may be the only way to make the model tractable

See chapter 7.7.3 of the Train textbook for more details

Endogeneity in Structural Models

Endogeneity in Structural Models

So far, we have (mostly) assumed that all of our explanatory variables are exogenous

- When we talked about GMM estimation, we talked how we can use it to incorporate instruments, but I did not say much about why we would want to do so

Why is exogeneity/endogeneity so important?

- We need exogenous variation in our explanatory variables in order to give our parameter estimates a "causal" interpretation
  - If the data are endogenous, our parameters can be interpreted as a kind of correlation between the data and choices, but they will not be the true structural parameters we intend to estimate
- But in most cases, exogenous variation in the explanatory variables is difficult to come by

Examples of Endogeneity

Housing choice and commute choice are correlated

- Example: people who like public transit tend to live closer to transit stations, making their transit travel time lower
- The coefficient on transit travel time will be biased upward

Price and unobserved quality are correlated

- Example: products with higher unobserved (to the econometrician) quality cost more and are preferred by consumers
- The coefficient on price will be biased downward and may even have the wrong sign

Price and unobserved marketing are correlated

- Example: large marketing campaigns may be accompanied by sales or increased prices
- The coefficient on price will be biased, but the direction is uncertain

Exogenous Variation and Causal Parameters

How do we estimate parameters with a "causal" interpretation?

- BLP estimation: use instruments to isolate exogenous variation in explanatory variables
- Control function estimation: use instruments to control for endogeneity in explanatory variables

What makes a good instrument?

- Correlated with explanatory variables
- Exogenous, or uncorrelated with random utility

Where do we get good instruments?

- Same concept as good research design for reduced-form analysis
  - Institutional knowledge, natural experiments, etc.

BLP Estimation

BLP Estimation

The context for the canonical BLP estimation approach is a mixed logit (or random coefficients) model of demand for a differentiated product using market-level data

- We want to estimate how the attributes of a product affect consumer demand
- Price is (one of) the most important attributes to consider
- But price is almost certainly correlated with the unobserved attributes (quality, etc.) of a product

Berry, Levinsohn, and Pakes (1995)—known as BLP—use instruments to isolate exogenous variation in price

- This paper developed a novel method to include instruments in a nonlinear model using market-level data

A similar procedure can be used for endogenous variables other than price

BLP Demand Model

We have data on M markets with J products in each market

- One of these products can be the "outside good" or purchase nothing

The utility that consumer n in market m obtains from product j is

$$U_{njm} = V(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \boldsymbol{\beta}_n) + \xi_{jm} + \varepsilon_{njm}$$

- p_{jm}: price of product j in market m
- $\boldsymbol{x}_{jm}$: vector of non-price attributes of product j in market m
- $\boldsymbol{s}_n$: vector of demographic characteristics of consumer n
- $\boldsymbol{\beta}_n$: vector of coefficients for consumer n
- ξ_{jm}: utility of unobserved attributes of product j in market m
- ε_{njm}: idiosyncratic unobserved utility

Endogeneity in the BLP Demand Model

We would expect the price to depend on all attributes of a product, including those that are unobserved by the econometrician

- But if consumers also get utility from those unobserved attributes, then the price is correlated with the composite error term, $\xi_{jm} + \varepsilon_{njm}$

To solve this problem, BLP use a two-step procedure

1. Estimate the average utility for product j in market m, including observable and unobservable attributes
2. Regress this average utility value on price and other observable attributes, instrumenting for price

Utility Decomposition

Decompose the utility from observed attributes, $V(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \boldsymbol{\beta}_n)$, into two components (with $\bar{\beta}$ and $\widetilde{\beta}_n$ defined analogously)

- $\bar{V}(p_{jm}, \boldsymbol{x}_{jm}, \bar{\boldsymbol{\beta}})$: component that varies over products and markets
- $\widetilde{V}(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n)$: component that varies by consumer

Then the utility that consumer n in market m obtains from product j is

$$\begin{aligned}
U_{njm} &= \bar{V}(p_{jm}, \boldsymbol{x}_{jm}, \bar{\boldsymbol{\beta}}) + \widetilde{V}(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n) + \xi_{jm} + \varepsilon_{njm} \\
&= \left[\bar{V}(p_{jm}, \boldsymbol{x}_{jm}, \bar{\boldsymbol{\beta}}) + \xi_{jm}\right] + \widetilde{V}(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n) + \varepsilon_{njm} \\
&= \delta_{jm} + \widetilde{V}(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n) + \varepsilon_{njm}
\end{aligned}$$

where

$$\delta_{jm} = \bar{V}(p_{jm}, \boldsymbol{x}_{jm}, \bar{\boldsymbol{\beta}}) + \xi_{jm}$$

This term, δ_{jm}, effectively becomes a product-market constant term that represents the average utility obtained by product j in market m

Choice Probabilities

Two distributional assumptions

- $\varepsilon_{njm} \sim$ i.i.d. type I extreme value
- $\widetilde{\boldsymbol{\beta}}_n$ has density $f(\widetilde{\boldsymbol{\beta}}_n \mid \boldsymbol{\theta})$
  - The mean of $\boldsymbol{\beta}_n$ is already modeled by $\bar{\boldsymbol{\beta}}$, so $\boldsymbol{\theta}$ will often be only a variance-covariance matrix

Then choice probabilities can be expressed as functions of δ_{jm} and $\widetilde{V}(\cdot)$

$$P_{nim} = \int \left[\frac{e^{\delta_{im} + \widetilde{V}(p_{im}, \boldsymbol{x}_{im}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n)}}{\sum_{j=1}^{J} e^{\delta_{jm} + \widetilde{V}(p_{jm}, \boldsymbol{x}_{jm}, \boldsymbol{s}_n, \widetilde{\boldsymbol{\beta}}_n)}} \right] f(\widetilde{\boldsymbol{\beta}}_n \mid \boldsymbol{\theta}) d\widetilde{\boldsymbol{\beta}}_n$$

We can use these choice probabilities to estimate the constant terms, δ_{jm}, and the $\boldsymbol{\theta}$ parameters

- But we cannot directly estimate the $\bar{\boldsymbol{\beta}}$ parameters because they are subsumed into the constant terms

Instrumenting for Price

If we assume that $\bar{V}$ is linear in parameters

$$\bar{V}(p_{jm}, \boldsymbol{x}_{jm}, \bar{\beta}) = \bar{\beta}'(p_{jm}, \boldsymbol{x}_{jm})$$

then we can express the constant terms as

$$\delta_{jm} = \bar{\beta}'(p_{jm}, \boldsymbol{x}_{jm}) + \xi_{jm}$$

Once we have estimated the constant terms, we can regress them on prices and other attributes to estimate $\bar{\beta}$

- But price is endogenous—it depends on ξ_{jm}—so we have to instrument for price in this regression
- Instrumenting in a linear model is easy if we have good instruments

Contraction Mapping

This estimation framework is theoretically feasible

- But we have to estimate δ_{jm} for each product and market, which can easily be 100s or 1000s (or more!) of terms to estimate

BLP developed an alternate approach that does not require estimating these δ_{jm} terms in the standard way

- Their insight is that these δ_{jm} terms determine predicted market shares, so we want to find the set of constant terms that equates predicted market shares with observed market shares

BLP show that

- For a given set of $\boldsymbol{\theta}$ parameters, a unique vector $\boldsymbol{\delta}$ equates predicted market shares with observed market shares
- There is an iterative "contraction mapping" algorithm that recovers this unique vector of $\boldsymbol{\delta}$

Contraction Mapping Algorithm

We want to find the vector of product-market constant terms, $\boldsymbol{\delta}$, that equates predicted market share, $\widehat{S}_{jm}(\boldsymbol{\delta})$, with observed market share, S_{jm}, for all products in all markets

1. Begin with some initial product-market constant values, $\boldsymbol{\delta}^0$
2. Predict the market share for the current constant values, $\widehat{S}_{jm}(\boldsymbol{\delta}^s)$, for each product-market
3. Adjust each product-market constant term by comparing predicted and observed market share

$$\delta_{jm}^{s+1} = \delta_{jm}^{s} + \ln\left(\frac{S_{jm}}{\widehat{S}_{jm}(\boldsymbol{\delta}^s)}\right)$$

4. Repeat steps (2) and (3) until the algorithm converges to the set of product-market constants, $\widehat{\boldsymbol{\delta}}$

Estimation

The contraction mapping is an inner algorithm loop within the larger estimation loop

Two steps to estimate this model

1. Outer loop: search over $\boldsymbol{\theta}$ to optimize the estimation objective function
    1. Inner loop: use the contraction mapping to find $\boldsymbol{\delta}(\boldsymbol{\theta})$, the vector of product-market constant terms conditional on $\boldsymbol{\theta}$
    2. Use $\boldsymbol{\theta}$ and $\boldsymbol{\delta}(\boldsymbol{\theta})$ to simulate choice probabilities, $\check{P}_{njm}(\boldsymbol{\delta}(\boldsymbol{\theta}), \boldsymbol{\theta})$
    3. Use choice probabilities to calculate the estimation objective function
2. Estimate $\bar{\beta}$ by regressing δ_{jm} on $(p_{jm}, \boldsymbol{x}_{jm})$ with price instruments, $\boldsymbol{z}_{jm}$

We can estimate this model in two different ways

- MSL for step 1 and 2SLS for step 2
- MSM for steps 1 and 2 simultaneously

BLP Estimation Using MSL and 2SLS

The MSL estimator, $\widehat{\boldsymbol{\theta}}$, is the set of parameters that maximizes the simulated log-likelihood function

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{n=1}^{N} \ln \check{P}_{ni_n m}(\boldsymbol{\theta})$$

and this estimator implies a unique vector of product-market constants

$$\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}(\widehat{\boldsymbol{\theta}})$$

Then $\bar{\boldsymbol{\beta}}$ is estimated by regressing $\widehat{\delta}_{jm}$ on $(p_{jm}, \boldsymbol{x}_{jm})$ by 2SLS

$$\widehat{\delta}_{jm} = \bar{\boldsymbol{\beta}}'(p_{jm}, \boldsymbol{x}_{jm}) + \xi_{jm}$$

with exogenous instruments $\boldsymbol{z}_{jm}$, which gives thes 2SLS estimator

$$\widehat{\bar{\boldsymbol{\beta}}} = \left(\sum_{j=1}^{J} \sum_{m=1}^{M} \boldsymbol{z}_{jm}(p_{jm}, \boldsymbol{x}_{jm}) \right)^{-1} \left(\sum_{j=1}^{J} \sum_{m=1}^{M} \boldsymbol{z}_{jm} \widehat{\delta}_{jm} \right)$$

BLP Estimation Using MSM

The population moments that correspond to the two steps of BLP estimation, respectively, are

$$E\left[\left(y_{njm} - \check{P}_{njm}(\boldsymbol{\theta})\right)\boldsymbol{z}_{njm}\right] = \mathbf{0}$$
$$E\left[\left(\delta_{jm}(\boldsymbol{\theta}) - \bar{\beta}'(p_{jm}, \boldsymbol{x}_{jm})\right)\boldsymbol{z}_{jm}\right] = \mathbf{0}$$

The MSM estimator, $(\widehat{\boldsymbol{\theta}}, \widehat{\bar{\beta}})$ is the set of parameters that solves the empirical analogs of these population moments

$$\frac{1}{NJ}\sum_{n=1}^{N}\sum_{j=1}^{J}\left(y_{njm} - \check{P}_{njm}(\widehat{\boldsymbol{\theta}})\right)\boldsymbol{z}_{njm} = \mathbf{0}$$
$$\frac{1}{JM}\sum_{j=1}^{J}\sum_{m=1}^{M}\left(\delta_{jm}(\widehat{\boldsymbol{\theta}}) - \widehat{\bar{\beta}}'(p_{jm}, \boldsymbol{x}_{jm})\right)\boldsymbol{z}_{jm} = \mathbf{0}$$

or minimizes the weighted sum of squared moments

Control Function Model

Control Function Approach

The control function approach can be thought of as the opposite of the BLP approach

- The BLP approach isolates exogenous variation
- The control function approach controls for the source of endogeneity

Why might the control function approach be better than the BLP approach?

- A control function can be used even if market shares are zero
  - The constant terms in BLP are not identified for zero market shares
- A control function can control for individual-level endogeneity, rather than market-level endogeneity
  - An individual-specific constant term is not identified in BLP
- The control function approach does not require the contraction mapping

Control Function Model

The utility that consumer n obtains from product j is

$$U_{nj} = V(y_{nj}, \boldsymbol{x}_{nj}, \boldsymbol{\beta}_n) + \varepsilon_{nj}$$

- y_{nj}: endogenous explanatory variable for consumer n and product j
- $\boldsymbol{x}_{jm}$: vector of non-price attributes for consumer n and product j
- $\boldsymbol{\beta}_n$: vector of coefficients for consumer n
- ε_{nj}: unobserved utility for consumer n and product j

The endogenous explanatory variable can be expressed as

$$y_{nj} = W(\boldsymbol{z}_{nj}, \gamma) + \mu_{nj}$$

- $\boldsymbol{z}_{nj}$: vector of exogenous instruments for y_{nj}
- γ: parameters that relate y_{nj} and $\boldsymbol{z}_{nj}$
- μ_{nj}: unobserved factors that affect y_{nj}

Endogeneity in the Control Function Model

The utility that consumer n obtains from product j is

$$U_{nj} = V(y_{nj}, \boldsymbol{x}_{nj}, \boldsymbol{\beta}_n) + \varepsilon_{nj}$$

where the endogenous variable, y_{nj}, can be expressed as

$$y_{nj} = W(\boldsymbol{z}_{nj}, \boldsymbol{\gamma}) + \mu_{nj}$$

Two assumptions about the model errors

- ε_{nj} and μ_{nj} are correlated
  - y_{nj} and ε_{nj} are correlated, so y_{nj} is endogenous
- ε_{nj} and μ_{nj} are independent of $\boldsymbol{z}_{nj}$
  - $\boldsymbol{z}_{nj}$ are good instruments for y_{nj}

Control Function

Decompose the unobserved utility, ε_{nj}, into a conditional mean and a deviation from this conditional mean

$$\varepsilon_{nj} = E[\varepsilon_{nj} \mid \mu_{nj}] + \widetilde{\varepsilon}_{nj}$$

By construction, the deviations are not correlated with μ_{nj}, so they are not correlated with y_{nj}

- If we can control for the conditional mean, then we control for the source of endogeneity

We construct a "control function" to control for the conditional mean

$$CF(\mu_{nj}, \boldsymbol{\lambda}) = E[\varepsilon_{nj} \mid \mu_{nj}]$$

- The control function is often linear, $CF(\mu_{nj}, \lambda) = \lambda\mu_{nj}$

Control Function Choice Probabilities

Substituting in the control function, the utility that consumer n obtains from product j becomes

$$U_{nj} = V(y_{nj}, \boldsymbol{x}_{nj}, \beta_n) + CF(\mu_{nj}, \boldsymbol{\lambda}) + \widetilde{\varepsilon}_{nj}$$

We make two distributional assumptions

- $\widetilde{\boldsymbol{\varepsilon}}_n$ has conditional density $g(\widetilde{\boldsymbol{\varepsilon}}_n \mid \boldsymbol{\mu}_n)$
- $\boldsymbol{\beta}_n$ has density $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta})$

Then the choice probabilities are

$$P_{ni} = \int \int \mathbb{1}\left[V_{ni} + CF_{ni} + \widetilde{\varepsilon}_{ni} > V_{nj} + CF_{nj} + \widetilde{\varepsilon}_{nj} \; \forall j \neq i\right] \\ \times g(\widetilde{\boldsymbol{\varepsilon}}_n \mid \boldsymbol{\mu}_n) f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}) d\widetilde{\boldsymbol{\varepsilon}}_n d\boldsymbol{\beta}_n$$

Control Function Estimation

Two steps to estimate this model

1. Estimate $\widehat{\mu}_{nj}$ by regressing y_{nj} on $\boldsymbol{z}_{nj}$
   - $\widehat{\mu}_{nj}$ is the residual of this regression
2. Estimate $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}})$ by MSL using simulated choice probabilities to construct a simulated log-likelihood function

An alternative approach is to estimate all parameters simultaneously

- This approach requires that we specify the joint distribution of $\boldsymbol{\varepsilon}_n$ and $\boldsymbol{\mu}_n$, whereas the sequential method requires only the conditional distribution of $\boldsymbol{\varepsilon}_n$ given $\boldsymbol{\mu}_n$
- But if we can correctly specify this joint distribution, then the simultaneous approach is more efficient