Estimation with Micro Data or Panel Data
Petrin (2002)

Phil Haile
Yale University

Fall 2018

Micro Data

micro data (in discrete choice setting) = observation of individual consumer characteristics D_i *matched* to individual consumer choices

Note:

- individual consumer choices without matched consumer characteteristics has *no new information* relative to market level data
- consumer characteristics not matched to individual choices is just a version of market level data, as in BLP95, Nevo, etc.

Gains from Micro Data?

Micro data links household choices directly to household characteristics, e.g.,

- large family$\longrightarrow$like station wagon or minivan
- rich$\longrightarrow$less sensitive to price

Gains from Micro Data?

Micro data links household choices directly to household characteristics, e.g.,

- large family$\longrightarrow$like station wagon or minivan
- rich$\longrightarrow$less sensitive to price

With market level data, we learn about the *marginal* distributions $F_D\left(D_i\right), F_Y\left(Y_i\right)$ of demographics and choices.

With micro data, we learn about their *joint* distribution $F_{DY}\left(D_i, Y_i\right)$.

The Value of Micro Data

Why is this additional information useful?

- we could put every value of D_i in its own market (so model completely flexible wrt D_i)
- or, with some structure on how D_i enters (e.g., distance to retailer j affects utility of buying from j, but not that of buying from k) we can exploit this observable variation to help pin down substitution patterns—i.e., to replace/supplement exogenous variation in quantities via instruments

The main challenge in estimating demand with market level data is finding enough good instruments. Micro data can relax the IV requirements by providing exogenous variation in teh choice problem *within a single market* (fixed ξ_{jt}). We'll discuss this more formally later.

BLP Style Preferences

$u_{ijmt} = x_{jmt}\beta_{im} + \xi_{jmt} + \epsilon_{ijmt}$, where

- $t \in$ indexes time
- $m \in \{1, \ldots, M\}$ indexes geographic markets
- $j \in \{0, 1, \ldots, J\}$ indexes goods
- $i \in \{1, \ldots, N\}$ indexes consumers
- observables $x_{jmt} \in \mathbb{R}^G$ include price
- consumer heterogeneity
    - $\epsilon_{imt} = (\epsilon_{i0mt}, \ldots, \epsilon_{iJmt})$, i.i.d. type 1 EV
    - $\beta_{im} = \beta_0 + \gamma d_{im} + \sigma\zeta_{im}$, where d_{im} are demographics, ζ_{im} is random vector $\sim \Phi(\cdot)$ iid across consumers

Note: consumer tastes β_{im} labeled as constant across time. This is not essential, and is only an assumption if one has a consumer panel (otherwise, i does not define a fixed consumer across time anyway).

Discrete Choice

Rewrite

$$u_{ijmt} = \delta_{jmt} + \mu_{ijmt}(\sigma,\gamma) + \epsilon_{ijmt}$$

where

$$\begin{aligned}
\delta_{jmt} &= x_{jmt}\beta_0 + \xi_{jmt} \\
\mu_{ijmt}(\sigma,\gamma) &= x_{jmt}\left(\gamma d_{imt} + \sigma\zeta_{im}\right).
\end{aligned}$$

Discrete Choice

Rewrite

$$u_{ijmt} = \delta_{jmt} + \mu_{ijmt}\left(\sigma, \gamma\right) + \epsilon_{ijmt}$$

where

$$\begin{aligned} \delta_{jmt} &= x_{jmt}\beta_0 + \xi_{jmt} \\ \mu_{ijmt}\left(\sigma, \gamma\right) &= x_{jmt}\left(\gamma d_{imt} + \sigma \zeta_{im}\right). \end{aligned}$$

Probability that consumer i in market m chooses good j in period t takes the form

$$\begin{aligned} & \int_{\mathbb{R}^G} \Pr\left\{ j \in \arg\max_k \ \delta_{kmt} + \mu_{ikmt}\left(\sigma, \gamma\right) + \epsilon_{ikmt} \right\} d\Phi\left(\zeta_{im}\right) \\ = \ & \int_{\mathbb{R}^G} \frac{\exp(\delta_{jmt} + \mu_{ijmt}\left(\sigma, \gamma\right))}{1 + \sum \exp(\delta_{kmt} + \mu_{ikmt}\left(\sigma, \gamma\right))} d\Phi\left(\zeta_{im}\right). \end{aligned}$$

An Estimation Approach

$$\int_{\mathbb{R}^G} \frac{\exp(\delta_{jmt} + \mu_{ijmt}(\sigma, \gamma))}{1 + \sum \exp(\delta_{kmt} + \mu_{ikmt}(\sigma, \gamma))} d\Phi(\zeta_{im})$$

If there is no variation in t (micro data, but no panel) this choice probability (for the good j actually selected) is the likelihood contribution of each observation (consumer) as a function of the parameters (δ, γ, σ).

One could estimate all parameters by MSM, with moments.

- score of likelihoood wrt (δ, γ, σ) (can use contraction for $\delta | \gamma, \sigma$)
- orthogonality conditions $E\left[Z'\left(\delta - x_{jmt}\beta_0\right)\right] = 0$

Note: micro data reduces the role of orthogonality conditions—if we dropped those moments, we might still be able to estimate (δ, γ, σ), which includes the "nonlinear parameters" governing substitution patterns. Formal results later.

Panel Data

Micro data is already a form of panel data: we see many consumers within a given market. This allows us to hold the market fixed (including all ξ_{jt}) and see how changes in consumer characteristics alter choices. This "within market variation" has no endogeneity problem because the structural error responsible for endogeneity is fixed within a market.

But in a discrete choice setting, "panel data" often refers to a situation where we have multiple observations per consumer.

Consumer Panel

With a consumer panel, we can exploit the fact that we see the same consumer on different choice occasions, ideally facing different choice sets. This will provide even more information about the role of individual characteristics in determining substitution patterns. . .

2-Period Consumer Panel

Suppose $t \in \{0,1\}$. Then $s_{ijkm}(\delta,\sigma) \equiv \Pr(y_{i0}=j, y_{i1}=k) =$

$$\int \left[\frac{\exp\left(\delta_{jm0}+\mu_{ijm0}(\sigma,\gamma)\right)}{1+\sum_\ell \exp\left(\delta_{\ell m0}+\mu_{i\ell m0}(\sigma,\gamma)\right)}\right]\left[\frac{\exp\left(\delta_{km1}+\mu_{ikm1}(\sigma,\gamma)\right)}{1+\sum_\ell \exp\left(\delta_{\ell m1}+\mu_{i\ell m1}(\sigma,\gamma)\right)}\right] d\Phi\left(\zeta_{im}\right).$$

Replacing j and k with y_{0i} and y_{1i} yields the likelihood contribution for consumer i, as a function of the parameters (δ,γ,σ). In practice we'd have to simulate the integral in the likelihood (or its score).

(Assuming here no state dependence—e.g., brand "inertia" or inattention. That would require a change to the model and, often, dealing with an "initial conditions" problem).

Consumer Panel: A Caution

More details on simulation-based estimation in Ken Train's book *Discrete Choice Methods with Simulation.*

But reliance on the likelihood often implies that we will need especially many simulation draws and that importance sampling and/or other tricks may be critical. For a moderate number of products this may work easily—no tiny choice probabilities to simulate.

But with two periods of data the choice probabilities in the likelihood are probabilities over $(J+1)^2$ combinations of choices, and *some jk combinations may be very rare*. Score of (log-) likelihood requires that we simulate well the *derivatives* of all choice probabilities. Often it will make sense to estimate instead by GMM, choosing moments carefully (e.g., aggregate).

Petrin (2002)

Some Big Economic Questions

- how big are the welfare gains from innovation?
- what share of these gains are captured by the innovator?

Some Big Economic Questions

- how big are the welfare gains from innovation?
- what share of these gains are captured by the innovator?
- how big are the negative externalities ("business stealing") on other producers?

Some Big Economic Questions

- how big are the welfare gains from innovation?
- what share of these gains are captured by the innovator?
- how big are the negative externalities ("business stealing") on other producers?
- in the case of innovation by an incumbant, how big is the negative effect of innovation on itself ("cannibalization")?

Some Big Economic Questions

- how big are the welfare gains from innovation?
- what share of these gains are captured by the innovator?
- how big are the negative externalities ("business stealing") on other producers?
- in the case of innovation by an incumbant, how big is the negative effect of innovation on itself ("cannibalization")?
- how does innovation alter competition/market power?

Some Big Economic Questions

- how big are the welfare gains from innovation?
- what share of these gains are captured by the innovator?
- how big are the negative externalities ("business stealing") on other producers?
- in the case of innovation by an incumbant, how big is the negative effect of innovation on itself ("cannibalization")?
- how does innovation alter competition/market power?

Petrin: A case study of the minivan, combining market level data and micro data. Methods: combine micro data and market level data (no panel).

1983 Chrysler Town & Country Station Wagon

1984 Chrysler Caravan Minivan

Combining Market and Micro Data

1. Market level data (like BLP, Nevo)
    - car characteristics and market shares (same as BLP)
    - all of U.S., 1982-1993
    - household demographics for representative sample of U.S. population (CES)

Combining Market and Micro Data

1. Market level data (like BLP, Nevo)
   - car characteristics and market shares (same as BLP)
   - all of U.S., 1982-1993
   - household demographics for representative sample of U.S. population (CES)

2. "Micro data"
   - "CEX" : Extended Consumer Expenditure Study
   - demographics and new car purchases for smaller sample: 30,000 households
     - too few households to use these data alone
     - for example, only about 2700 purchases; many cars never purchased in this sample.

TABLE 2
AVERAGE CONSUMER CHARACTERISTICS FOR THE UNITED STATES AND SELECTED SUBPOPULATIONS, 1987–92

	United States		Purchasers of				
	Mean	Standard Deviation	New Vehicles	Minivans	Station Wagons	Sport-Utilities	Full-Size Vans
Income	23,728	21,255	36,113	39,476	40,196	41,569	31,164
Family size	2.58	1.53	2.87	3.86	3.17	2.97	3.47
Midage	.55	.49	.64	.78	.73	.74	.65

SOURCE.—Consumer Expenditure Survey.

NOTE.—Income is measured in 1982–84 CPI-adjusted dollars. Family size is the number of household members. Midage is a binary variable for the age of the head of household between 30 and 60 inclusive.

Discrete Choice Model

$$u_{ijt} = \alpha_i \ln\left(y_i - p_{jt}\right) + x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt}$$

Discrete Choice Model

$$u_{ijt} = \alpha_i \ln\left(y_i - p_{jt}\right) + x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt}$$

where

$$\alpha_i = \begin{cases} \alpha_1 & y_i < Y_1 \\ \alpha_2 & Y_1 \leq y_i \leq Y_2 \\ \alpha_3 & Y_2 \leq y_i \end{cases}$$

Discrete Choice Model

$$u_{ijt} = \alpha_i \ln\left(y_i - p_{jt}\right) + x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt}$$

where

$$\alpha_i = \begin{cases} \alpha_1 & y_i < Y_1 \\ \alpha_2 & Y_1 \leq y_i \leq Y_2 \\ \alpha_3 & Y_2 \leq y_i \end{cases}$$

$$\beta_{it}^k = \begin{cases} \beta_0^k + \gamma^k \ln\left(fs_i\right) \nu_{it}^k & \text{minivan, sw dummies} \\ \beta_0^k + \gamma^k \nu_{it}^k & \text{all other } k \end{cases}$$

Discrete Choice Model

$$u_{ijt} = \alpha_i \ln \left(y_i - p_{jt}\right) + x_{jt}\beta_{it} + \xi_{jt} + \epsilon_{ijt}$$

where

$$\begin{aligned}
\alpha_i &= \begin{cases} \alpha_1 & y_i < Y_1 \\ \alpha_2 & Y_1 \leq y_i \leq Y_2 \\ \alpha_3 & Y_2 \leq y_i \end{cases} \\
\beta^k_{it} &= \begin{cases} \beta^k_0 + \gamma^k \ln\left(fs_i\right) \nu^k_{it} & \text{minivan, sw dummies} \\ \beta^k_0 + \gamma^k \nu^k_{it} & \text{all other } k \end{cases} \\
\epsilon_{ijt} &\sim \text{i.i.d. EV, } \nu_{it} \text{ normal or } \chi^2.
\end{aligned}$$

Note: this specification possible with market level data too.

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- + "micro moments" .

$$E\left[i \text{ buys minivan} \middle| y_i \in \text{``bin'' } b\right] \qquad b = 1, 2, 3$$

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- $+$ "micro moments" .

$$E\left[i \text{ buys minivan} | y_i \in \text{“bin” } b\right] \qquad b = 1, 2, 3$$

$$E\left[fs_i | \text{buy minivan}\right]$$

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- + “micro moments” .

$$E\left[i \text{ buys minivan} \middle| y_i \in \text{“bin” } b\right] \qquad b = 1, 2, 3$$

$$E\left[fs_i \middle| \text{buy minivan}\right]$$

$$E\left[fs_i \middle| \text{buy station wagon}\right]$$

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- + "micro moments" .

$$E\left[i \text{ buys minivan} | y_i \in \text{“bin” } b\right] \qquad b = 1, 2, 3$$

$$E\left[fs_i | \text{buy minivan}\right]$$

$$E\left[fs_i | \text{buy station wagon}\right]$$

$$E\left[fs_i | \text{S.U.V.}\right]$$

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- + "micro moments" .

$$E\left[i \text{ buys minivan} | y_i \in \text{“bin” } b\right] \qquad b = 1, 2, 3$$
$$E\left[fs_i | \text{buy minivan}\right]$$
$$E\left[fs_i | \text{buy station wagon}\right]$$
$$E\left[fs_i | \text{S.U.V.}\right]$$
$$E\left[fs_i | \text{full size passenger van}\right]$$

Estimation

BLP-type MSM estimator

- BLP moments (including supply side)
- + "micro moments" .

$$
\begin{aligned}
&E\left[i \text{ buys minivan} \middle| y_i \in \text{“bin” } b\right] \qquad b = 1, 2, 3 \\
&E\left[fs_i \middle| \text{buy minivan}\right] \\
&E\left[fs_i \middle| \text{buy station wagon}\right] \\
&E\left[fs_i \middle| \text{S.U.V.}\right] \\
&E\left[fs_i \middle| \text{full size passenger van}\right]
\end{aligned}
$$

- model predicts these (using Bayes' rule)
- expected difference between prediction and sample mean is zero

Note: not the only possible type of micro moments; but averages conditional on coarse partition are good choices when survey is small (e.g., here, 120 purchasers of minivans) or when individual market shares very small.

Benefits (or Harm) of the Minivan Introduction

- to innovating firm (Chrysler)
- to imitating firms (almost all others eventually imitate)

Benefits (or Harm) of the Minivan Introduction

- to innovating firm (Chrysler)
- to imitating firms (almost all others eventually imitate)
- to firms with competing products (e.g., station wagon)

Benefits (or Harm) of the Minivan Introduction

- to innovating firm (Chrysler)
- to imitating firms (almost all others eventually imitate)
- to firms with competing products (e.g., station wagon)
- to consumers who buy minivan

Benefits (or Harm) of the Minivan Introduction

- to innovating firm (Chrysler)
- to imitating firms (almost all others eventually imitate)
- to firms with competing products (e.g., station wagon)
- to consumers who buy minivan
- to consumers who buy other cars at reduced price.

TABLE 3
FAMILY VEHICLE SALES AS A PERCENTAGE OF TOTAL VEHICLE SALES:
U.S. AUTOMOBILE MARKET, 1981–93

Year	Minivans (1)	Station Wagons (2)	Sport-Utilities (3)	Full-Size Vans (4)	Minivans and Station Wagons (5)	U.S. Auto Sales (Millions) (6)
1981	.00	10.51	.58	.82	10.51	7.58
1982	.00	10.27	.79	1.17	10.27	7.05
1983	.00	10.32	3.51	1.04	10.32	8.48
1984	1.58	8.90	5.51	1.20	10.48	10.66
1985	2.32	7.33	6.11	1.05	9.65	11.87
1986	3.63	6.70	5.73	.85	10.43	12.21
1987	4.86	6.47	6.44	.73	11.33	11.21
1988	5.97	5.14	7.18	.69	11.11	11.76
1989	6.45	4.13	7.47	.61	10.58	11.06
1990	7.95	3.59	7.78	.27	11.54	10.51
1991	8.29	3.05	7.80	.29	11.34	9.75
1992	8.77	3.07	9.33	.39	11.84	10.12
1993	9.93	3.02	11.66	.29	12.95	10.71

Counterfactual Simulation

Question: what would market look like (consumer welfare, firm profits) without minivan?

Counterfactual Simulation

Question: what would market look like (consumer welfare, firm profits) without minivan?

Approach

- take minivan out of the market
- construct new pricing equilibrium using the estimated marginal costs

Counterfactual Simulation

Question: what would market look like (consumer welfare, firm profits) without minivan?

Approach

- take minivan out of the market
- construct new pricing equilibrium using the estimated marginal costs
- simulate new market shares, profits, and welfare:
  - draw "a consumer" (D_i, ν_i, ϵ_i) at random, calculate
    - choice and utility under full choice set
    - choice and utility under reduced choice set
    - compensating variation CV_i for this consumer

Counterfactual Simulation

Question: what would market look like (consumer welfare, firm profits) without minivan?

Approach

- take minivan out of the market
- construct new pricing equilibrium using the estimated marginal costs
- simulate new market shares, profits, and welfare:
  - draw "a consumer" (D_i, ν_i, ϵ_i) at random, calculate
    - choice and utility under full choice set
    - choice and utility under reduced choice set
    - compensating variation CV_i for this consumer
  - repeat

[actually, with logit, analytic formula for expected CV (Small and Rosen, 1981), so don't need to include ϵ in the simulation].

A Reminder

\

To compare counterfactual world to real world, we almost always want to compare the model predictions for these two scenarios. Here, we wnat to calculate the model's predicted welfare and profits with and without the minivan. (This is what Petrin does).

Another Caution

Welfare analysis is a tricky business, and sloppiness about welfare is easy to find in the empirical literature. When we specify consumer utilities, it is tempting to add them up to construct "total consumer welfare." In general this is nonsense—as we tell freshman, one can't compare utilities across people.

Welfare Analysis in Petrin

- Petrin constructs compensating variation (CV), which is a valid notion of aggregate welfare (if one assumes optimal redistribution!).
- But note that he does this by calculating the CV for each consumer—he is learning about changes in welfare of each consumer even though he sees each consumer only once! How is this possible?
- In fact, exactly the same procedure could be followed in a market-data setting. How can it be that we learn about individuals' welfare from only market level data?
- The answer is that this reflects functional form assumptions—these allow us to know individual demands because we assume an individual = a vector (y_i, ϵ_i, ν_i).

Counterfactual Simulation

Important maintained assumption: only prices would have been different in the counterfactual world; e.g.,

- no entry of another type of vehicle
- no acceleration of SUV entry
- products that exited in the sample (a lot of station wagons) still exit in counterfactual

This is not really a critique of the paper, as almost any empirical work holds many things fixed—explicitly or implicitly—that may eventually react to the intervention/counterfactual being evaluated. Being able to say what is/isn't accounted for is a virtue. Here the model makes this clear. Nonetheless, clearly room for embedding this kind of model in one that treats, e.g., entry-exit.

Chrysler Big Winner, Ford and GM Losers

TABLE 11
CHANGE IN INDUSTRY AND BIG THREE TOTAL VARIABLE PROFITS WITH THE ADVENT OF MINIVANS

YEAR	INDUSTRY	CHRYSLER		FORD		GM	
1984	−.21%	\$202.5	14.38%	−\$31.8	−1.16%	−\$155.8	−1.50%
1985	−.13%	\$259.1	13.99%	−\$37.4	−1.29%	−\$171.0	−1.63%
1986	.14%	\$201.1	12.42%	\$54.7	1.84%	−\$119.9	−1.09%
1987	.17%	\$346.1	23.27%	−\$22.8	−.66%	−\$174.5	−2.14%
1988	.65%	\$504.1	32.50%	−\$24.7	−.70%	−\$235.4	−2.90%

NOTE.—Dollar figures are given in millions. The numbers are computed using the model to estimate profits both with minivans in the market and with minivans removed from the market (see Sec. V).

Big Gains to Consumers, Even Non-Buyers of MV

TABLE 13
CHANGE IN U.S. WELFARE FROM THE MINIVAN INNOVATION, 1984–88 ($ Millions)

Year	Compensating Variation	Change in Producer Profits	Welfare Change
1984	367.29	−36.68	330.61
1985	625.04	−25.07	599.97
1986	439.93	27.30	467.23
1987	596.59	29.75	626.34
1988	775.70	110.24	885.94
Total	2,804.55	105.54	2,910.09

NOTE.—Computations were done using 1982–84 CPI-adjusted dollars.

Standard Errors?

Missing from the last two tables are standard errors.

Standard errors on counterfactual predictions often not easy, because the map from parameters to counterfactuals is complex (so delta method hard). Bootstrap often the best available approach, and more feasible now than 15-20 years ago.