Demand Estimation Using Aggregate Data: Static Discrete Choice Models

K. Sudhir

Yale School of Management

Quantitative Marketing and Structural Econometrics Workshop

Duke University

July 30 - August 1 2013

Agenda

- Session 1
  - Why is BLP popular in marketing?
  - The BLP algorithm to estimate discrete choice random coefficients logit model with aggregate data
  - Reviewing BLP code for a simple example
- Session 2
  - Compare BLP algorithm with MPEC algorithm
  - Instruments
  - Identification
  - Adding micro moments
  - Adding supply moments

Why has BLP demand estimation become popular in marketing?

Motivation: Demand estimation using aggregate data

- Demand estimation is critical element of marketing analysis
  - Value of demand estimation using aggregate data
    - Marketers often only have access to **aggregate data**
    - Even if HH data available, these are not fully representative
  - Two main challenges in using aggregate data
    - Heterogeneity: Marketers seek to *differentiate products* that appeal *differentially to different segments* to reduce competition and increase margins; need to account for this
    - Endogeneity: Researchers typically do not know (or have data) all factors that firms offer and consumers value in a product at a given time or market, firms account for this in setting marketing mix—this creates a potential **endogeneity problem**

Why is BLP demand estimation popular in marketing?

- Berry Levinsohn and Pakes (1995) addresses all three issues
  - estimates differentiated product demand systems with **aggregate data**
  - uses discrete choice models with random coefficients (**heterogeneity**)
  - accounts for researcher unobservables that affect consumer choice, and firm's marketing mix choices, (**endogeneity**)
- BLP gained quick acceptance because demand modelers using household scanner data
  - Immediately understood the importance of accounting for heterogeneity with aggregate data soon after the first papers in marketing were published (Sudhir 2001; Chintagunta 2001)
  - But it took a bit longer to accept the endogeneity issue

What data do we need for estimation?

Canonical aggregate market level data

- Aggregate “Market” Data
  - Longitudinal: one market/store across time (e.g., BLP Ecta 1995/Sudhir Mkt Sci 2001/Chintagunta Mkt Sci 2001)
  - Cross-sections: multiple markets/stores (e.g., Datta and Sudhir 2011)
  - Panel: multiple markets/stores across time (Nevo Ecta 2001; Chintagunta, Singh and Dube QME 2003)
- Typical variables used in estimation
  - Aggregate quantity
  - Prices/attributes/instruments
  - Definition of market size
  - Distribution of demographics (sometimes)

Typical Data Structure

Market/Time 1

Product	Quantity	Price	Attributes	Instruments

Market/Time 2

Product	Quantity	Price	Attributes	Instruments

Market size (M) assumption needed to get shares

$$s_{jt} = \frac{q_j}{M_t} \qquad s_{0t} = 1 - \sum_{j=1}^{J} s_j$$

Notation

- Markets or Periods: $t = 1,\ldots,T$
- Product $j = 0,1,\ldots,J,$ with $j = 0$ the outside good
- Consumer i makes choice $j \in \{0,1,\ldots,J\}$ in market t
- Indirect Utility Function: $U(x_{jt}, \xi_{jt}, p_{jt}, D_{it}, \nu_{it}; \theta)$
  - x_{jt} – observed product characteristics
  - ξ_{jt} – unobserved (to researcher) product characteristics
  - p_{jt} – price
  - D_{it} – "observable" consumer demographics
  - ν_{it} – unobserved consumer attributes
- Indirect Utility Function: $u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$

Notation

- Indirect Utility Function:
  - $u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$
  - where $\begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} = \underbrace{\begin{pmatrix}\alpha \\ \beta\end{pmatrix}}_{\theta_1} + \underbrace{\Pi D_i + \Sigma_{\nu_i}}_{\theta_2 = (\Pi,\ \Sigma_{\nu_i})}$

    Mean Heterogeneity

- Convenient to split the indirect utility into two parts
  - $u_{ijt} = \underbrace{\delta(x_{jt}, p_{jt}, \xi_{jt}; \theta_1)}_{\text{Mean}} + \underbrace{\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)}_{\text{HH Deviations from Mean}} + \varepsilon_{ijt}$
  - $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$, and $\mu_{ijt} = (p_{jt}, x_{jt})(\Pi D_i + \Sigma_{\nu_i})$

Key Challenges for Estimation with Market Level Data

- Heterogeneity: (β_i, α_i)
  - Recovering heterogeneity parameters without consumer data
  - Not an issue with consumer level data, because we have heterogeneity in choices across consumers
- Endogeneity: $Corr(p_{jt}, \xi_{jt})$
  - ξ_{jt} is potentially correlated with price (and other marketing mix variables)
  - Contrary to earlier conventional wisdom this can be an issue with even consumer level data

Motivation for Addressing Endogeneity

- Trajtenberg (JPE 1989) famous analysis of Body CT scanners using nested logit model
- Positive coefficient on price—upward sloping demand curve
  - Attributed to omitted quality variables

	1976	1977
RPRICE	11.252	.993
	(6.4)	(4.8)
SPEED	−2.292	2.138
	(−7.3)	(2.8)
SPEED^2	.236	−1.264
	(4.0)	(−3.4)
RESOL	69.107	9.113
	(7.3)	(2.4)
RESOL^2	−23.360	−2.533
	(−7.6)	(−1.5)
RTIME	−3.931	5.082
	(−5.3)	(7.0)
RTIME^2	1.054	−2.370
	(4.5)	(−6.7)
$\rho^2 = 1 - [L(\beta^*)/L(\beta^0)]$	.29	.12
Corr(π^*, π)	.999	.877
	(.0001)	(.0001)
Number of scanners	8	15
Number of observations	285	324

NOTE.—Asymptotic t-values are in parentheses.

Special Cases: Homogeneous Logit and Nested Logit

Homogeneous Logit Notation

- Indirect Utility Function:
  - $u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$
  - where $\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{\theta_1} + \xcancel{\underbrace{\Pi D_i + \Sigma_{\nu_i}}_{\theta_2 = (\Pi,\ \Sigma_{\nu_i})}}$

    Average ~~Heterogeneity~~

- Convenient to split the indirect utility into two parts
  - $u_{ijt} = \underbrace{\delta(x_{jt}, p_{jt}, \xi_{jt}; \theta_1)}_{\text{Mean}} + \xcancel{\underbrace{\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)}_{\text{HH Deviations from Mean}}} + \varepsilon_{ijt}$
  - $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$, and $\mu_{ijt} = \xcancel{(p_{jt}, x_{jt})(\Pi D_i + \Sigma_{\nu_i})}$

Homogenous Logit with Aggregate Data

- If we have market share data, s_{jt}
- $s_{jt} = \dfrac{\exp\{\delta_{jt}\}}{\sum_{k=0}^{J} \exp\{\delta_{kt}\}}$ $\quad s_{0t} = \dfrac{1}{\sum_{k=0}^{J} \exp\{\delta_{kt}\}}$ $\quad$ Normalize $\delta_{0t} = 0$
- $\dfrac{s_{jt}}{s_{0t}} = \exp\{\delta_{jt}\}$
- $\underbrace{\ln(s_{jt}) - \ln(s_{0t})}_{y_{jt}:\ \text{Data}} = \delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$
- With homogeneous logit, we can invert shares to get mean utility (δ_{jt})

Homogenous Logit with Aggregate Data: 2SLS

- $\underbrace{\ln(s_{jt}) - \ln(s_{0t})}_{y_{jt}\text{: Data}} = \delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$
- If no endogeneity, we can use OLS
- Given endogeneity of price, one can instrument for price and use 2SLS

Homogeneous Logit with aggregate data: GMM

- Alternatively, Berry (1994) suggests a GMM approach with a set of instruments Z
- Step 1: Compute $\hat{\delta}_{jt} = \ln(s_{jt}) - \ln(s_{0t})$
  - Let $\xi_{jt}(\theta) = \hat{\delta}_{jt} - x_{jt}\beta - \alpha p_{jt}$ where $\theta = (\beta, \alpha)$
- Step 2: GMM with moment conditions: $E(\xi(\theta)' Z) = 0$
  - $\min_{\theta} \ \xi(\theta)' Z W Z' \xi(\theta)$ where $W = (E(Z'\xi\xi' Z)\}^{-1}$
  - We have a nice analytical solution $\theta = (X'ZWZ'X)^{-1}(X'ZWZ'\delta)$
  - where $X = [x \ \ p]$
  - Start with $W = I$ or $W = (Z'Z)^{-1}$ to get initial θ estimate
    - Re-compute $W = (E(Z'\xi\xi' Z)\}^{-1}$ for new estimate of θ

Why homogeneous logit not great as a demand system

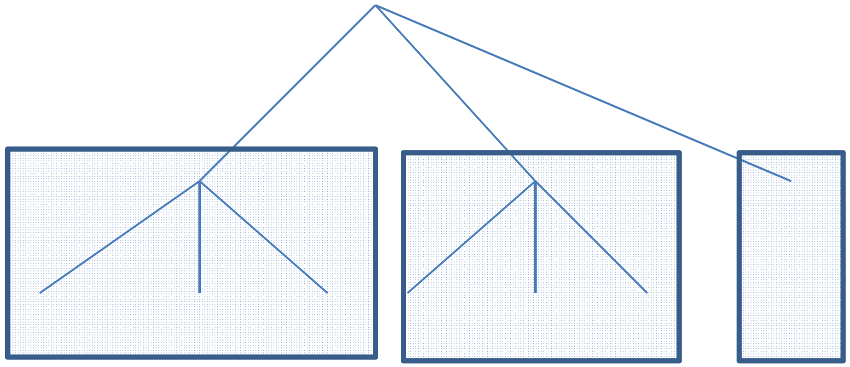
- Own Elasticity: $\eta_j = \left(\dfrac{\partial s_j}{\partial p_j}\right) \Big/ \dfrac{s_j}{p_j} = \alpha p_j (1 - s_j)$
- Cross Elasticity: $\eta_{jk} = \left(\dfrac{\partial s_j}{\partial p_k}\right) \Big/ \dfrac{s_j}{p_k} = \alpha p_k s_k$
- Bad Properties:
  - Own elasticities proportional to price, so conditional on share more expensive products tend to be more price elastic!!
    - BMW328 will be more price elastic than Ford Mustang.
  - Cross-elasticity of product *j*, w.r.t. price of product *k*, is dependent only on product *k*'s price and share
    - BMW328 and Toyota Corolla will have same cross price elasticity with respect to Honda Civic!!

Nested Logit with Aggregate Data: Applying GMM

- Nested logit provides more flexible elasticity patterns

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \rho \ln(s_{jt} / s_{Gt}) + \xi_{jt}$$

  - Where $0 < \rho < 1$ proxies for intra-group correlation in preferences
- Even if no price endogeneity, we cannot avoid instruments
  - Additional moments are needed to estimate ρ



- The GMM approach will still work, as long as we have two or more instruments to create enough identifying restrictions—one each for α and ρ

The Canonical BLP Random Coefficients model

The Canonical BLP Random Coefficients Logit Model

- Indirect Utility Function:
  - $u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$
  - where $\begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} = \underbrace{\begin{pmatrix}\alpha \\ \beta\end{pmatrix}}_{\theta_1} + \underbrace{\Pi D_i + \Sigma_{\nu_i}}_{\theta_2 = (\Pi,\ \Sigma_{\nu_i})}$

    Average Heterogeneity

- Split the indirect utility into two parts
  - $u_{ijt} = \underbrace{\delta(x_{jt}, p_{jt}, \xi_{jt}; \theta_1)}_{\text{Mean}} + \underbrace{\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)}_{\text{HH Deviations from Mean}} + \varepsilon_{ijt}$
  - $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$, and $\mu_{ijt} = (p_{jt}, x_{jt})(\Pi D_i + \Sigma_{\nu_i})$
  - Analytical inversion of δ_{jt} no longer feasible

Elasticities with heterogeneity— better demand system

- $$s_{jt}(\delta_t, \theta_2) = \int_{D_i, \nu_i} \frac{\exp\{\delta_{jt} + \mu_{ijt}(D_i, \nu_i; \theta_2)\}}{\sum_{k=0}^{J} \exp\{\delta_{kt} + \mu_{ikt}(D_i, \nu_i; \theta_2)\}} dF(D_i, \nu_i; \theta_2)$$
- Own Elasticity: $\eta_j = \left(\frac{\partial s_j}{\partial p_j}\right) \Big/ \frac{s_j}{p_j} = \frac{p_j}{s_j} \int \alpha_i s_{ij}(1 - s_{ij}) dF(D, \nu)$
- Cross Elasticity: $\eta_{jk} = \left(\frac{\partial s_j}{\partial p_k}\right) \Big/ \frac{s_j}{p_k} = -\frac{p_k}{s_j} \int \alpha_i s_{ij} s_{ik} dF(D, \nu)$
- Good Properties
  - Higher priced products more likely purchased by low α_i customers—can have lower elasticities
  - Cross elasticities vary across products—price cut on Honda Civic induces more switching from Toyota Corolla than from BMW 328

Logit vs RC logit: the value of heterogeneity (BLP)

- With logit, outside good captures all effects of price increase due to IIA
- With RC logit, IIA problem reduced
  - Expensive cars have less substitution to outside good

TABLE VII

SUBSTITUTION TO THE OUTSIDE GOOD

Model	Given a price increase, the percentage who substitute to the outside good (as a percentage of all who substitute away.) Logit	BLP
Mazda 323	90.870	27.123
Nissan Sentra	90.843	26.133
Ford Escort	90.592	27.996
Chevy Cavalier	90.585	26.389
Honda Accord	90.458	21.839
Ford Taurus	90.566	25.214
Buick Century	90.777	25.402
Nissan Maxima	90.790	21.738
Acura Legend	90.838	20.786
Lincoln Town Car	90.739	20.309
Cadillac Seville	90.860	16.734
Lexus LS400	90.851	10.090
BMW 735*i*	90.883	10.101

Estimation using market level data: BLP algorithm

1. Draw ν_i (and D_i if needed) for a set of (say NS=50) consumers. Compute initial δ_{jt} based on homogeneous logit.

2. Predicted shares
   - For given θ_2 compute the HH deviations from mean utility

     $\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)$
   - For given mean utility (δ_t) & θ_2,compute predicted shares,

     $\sigma_j(\boldsymbol{\delta}_{\mathbf{t}}, \mathbf{x}_t, \mathbf{p}_t; \theta_2)$

3. Contraction Mapping : Given nonlinear parameters θ_2, search for δ_t such that $s_{jt} = \sigma_j(\boldsymbol{\delta}_{\mathbf{t}}, \mathbf{x}_t, \mathbf{p}_t; \theta_2)$

4. From $\delta_{t'}$, estimate the linear parameters θ_1 using an analytical formula. Then form the GMM obj function $Q(\theta_2)$

5. Minimize $Q(\theta_2)$ over θ_2 with Steps 2-4 nested for every θ_2 trial

An illustrative problem

- Code and Data
  - Data provided in data.csv
  - Matlab code: Agglogit.m (main program) calls AgglogitGMM.m (the function to be minimized)
- Problem Definition
  - J=2 (brands), T=76 (periods)
  - Variables: price, advertising, 3 quarterly dummies
  - Cost instruments: 3 for each brand
  - Heterogeneity only on the 2 brand intercepts and price
    - Covariance:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \qquad L = \begin{pmatrix} b_1 & 0 & 0 \\ b_2 & b_3 & 0 \\ 0 & 0 & b_4 \end{pmatrix}$$

Step 1: Simulating HH draws

```
%Fix these draws outside
the estimation
w1=randn(NObs1,NCons);
w2=randn(NObs1,NCons);
wp=randn(NObs1,NCons);
%Convert to multivariate
draws within nonlinear
parameter estimation
aw1=b(1)*w1+b(2)*w2;
aw2=b(3)*w1;
aw2=b(3)*w1;
```

Cholesky Decomposition

$$\Sigma_{\nu_i} = LL'$$

$$L = \begin{pmatrix} b_1 & 0 & 0 \\ b_2 & b_3 & 0 \\ 0 & 0 & b_4 \end{pmatrix}$$

Step 2: Computing market shares based on model

- For logit and nested logit, can use analytic formulas
- For random coefficients logit, integrate over the heterogeneity by simulation
- $$\tilde{\sigma}_{jt} = \frac{1}{NS}\sum_{i=1}^{NS}\frac{\exp\{\delta_{jt} + (p_{jt}\ \ x_{jt})(\Pi D_i + \nu_i)\}}{\sum_{k=0}^{J}\exp\{\delta_{kt} + (p_{kt}\ \ x_{kt})(\Pi D_i + \nu_i)\}}$$
  - Where ν_i and D_i, $i = 1, \ldots, NS$ are draws from $F_v^*(v)$ and $F_D^*(D)$ that are drawn and fixed over optimization
- Simulation variance reduction (see Train Ch. 9)
  - Importance sampling: BLP oversamples on draws leading to auto purchases, relative to no purchase
  - Halton draws (100 Halton draws found better than 1000 random draws for mixed logit estimation; Train 2002)

Step 3: Contraction Mapping to get mean utility (δ_t)

- For logit and nested logit, you can get mean utility analytically

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} - \delta_{0t} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$$

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} - \delta_{0t} = x_{jt}\beta + \alpha p_{jt} + \rho \ln(s_{jt} / s_{Gt}) + \xi_{jt}$$

- For random coefficients, logit, you need a contraction mapping, where you iterate till convergence (i.e., $\delta_t^{h+1} = \delta_t^h <$ Tolerance)

$$\delta_t^{h+1} = \delta_t^h + \ln(s_t) - \ln(\tilde{\sigma}_j(\delta_t^h, \mathbf{x}_t, \mathbf{p}_t, F_{NS}; \theta_2)), h = 1, \ldots, H$$

Steps 2&3:Market Shares and Contraction Mapping

```
while (Err >= Tol)
    de=de1;
    sh=zeros(NObs1,1);
    psh=zeros(NObs1,1);
    %Integrating over consumer heterogeneity
    for i=1:1:NCons;
        psh=exp(aw(:,i)+awp(:,i)+de); psh=reshape(psh',2,NObs)';
        spsh=sum(psh')';
        psh(:,1)=psh(:,1)./(1+spsh); psh(:,2)=psh(:,2)./(1+spsh);
        sh=sh+reshape(psh',NObs1,1);
    end;
    %Predicted Share
    sh=sh/NCons;
    %Adjust delta_jt
    de1=de+log(s)-log(sh);
    Err=max(abs(de1-de));
end;
delta=de1;
```

Step 4: Estimate parameters in two steps

- A. Given $\delta_t(\theta_2)$, estimate θ_1 and compute ξ_{jt}

$$\theta_1 = (X'ZWZ'X)^{-1}(X'ZWZ'\delta)$$

$$\xi_{jt}(\theta_2) = \delta_{jt} - x_{jt}\beta - \alpha p_{jt}$$

- B. Minimize the GMM Objective: $\underset{\theta_2}{\text{Min}}\ \xi(\theta)'ZWZ'\xi(\theta)$
  - Where W is the GMM weight matrix $\left(E((\xi(\theta)'Z)'(Z'\xi(\theta)))\right)^{-1}$
  - But W is based on θ. Start with weighting matrix $W = I$ and iterate on W, based on new estimates. In practice, I start with W based on homogeneous logit, when estimating random coefficients logit

Step 4: Linear parameters and GMM objective function

```
% Analytically estimating linear parameters
blin=inv(xlin'*z*W*z'*xlin)*(xlin'*z*W*z'*delta);

% GMM Objective function over nonlinear parameters
er=delta-xlin*blin;
f=er'*z*W*z'*er;
```

Step 5: Optimizing over θ_2

- Use a nonlinear optimizer to minimize the GMM objective function in Step 4
  - The main file with data setup, homogeneous logit OLS, homogeneous logit IV, and calling the nonlinear optimizer are in file AggLogit.m
  - The GMM objective function nesting steps 2-4 are in the file AggLogitGMM.m

```
% Calling the optimizer with appropriate options
[b, fval,exitflag,output,grad,hessian] = fminunc('AgglogitGMM',b0,options);
```

- Standard errors should be computed by standard GMM formula (Hansen 1982)

Summary

- Why is BLP popular?
  - Handles aggregate data, heterogeneity and endogeneity
- Reviewed estimation algorithms
  - Homogenous and nested logit reduces to a linear model and can be estimated using an analytical formula
  - Random coefficients logit requires a nested algorithm
- Reviewed an illustrative coding example

Summary: Key elements in programming BLP

- BLP illustrative example code:
  - Simulation to integrate over random coefficients distribution
    - Drawing from a multivariate distribution
  - Contraction mapping
    - Linearization of the mean utility to facilitate IV
  - Generalized Method of Moments
    - We numerically optimize only over the nonlinear parameters, while estimating the linear parameters affecting mean utility through an analytical formula (as with homogenous logit)

Session 2

Session 2: Agenda

- Contrasting the contraction mapping algorithm with the MPEC approach
- Instruments
- Identification
- Improving identification and precision
  - Adding micro moments
  - Adding supply moments

Recall: The BLP Random Coefficients Logit Model

- Indirect Utility Function: (Ignore income effects)
  - $u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}$
  - where $$\begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} = \underbrace{\begin{pmatrix}\alpha \\ \beta\end{pmatrix}}_{\theta_1} + \underbrace{\Pi D_i + \Sigma_{\nu_i}}_{\theta_2 = (\Pi,\ \Sigma_{\nu_i})}$$

    Average Heterogeneity

- Split the indirect utility into two parts
  - $$u_{ijt} = \underbrace{\delta(x_{jt}, p_{jt}, \xi_{jt}; \theta_1)}_{\text{Mean}} + \underbrace{\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)}_{\text{HH Deviations from Mean}} + \varepsilon_{ijt}$$
  - $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt},\ \text{and}\ \mu_{ijt} = (p_{jt}, x_{jt})(\Pi D_i + \Sigma_{\nu_i})$

Estimation using market level data: BLP algorithm

1. Draw ν_i (and D_i if needed) for a set of (say NS=50) consumers. Compute initial δ_{jt} based on homogeneous logit.

2. Predicted shares

- For given θ_2 compute the HH deviations from mean utility

  $\mu(x_{jt}, p_{jt}, D_i, \nu_i; \theta_2)$

- For given mean utility (δ_t) & θ_2, compute predicted shares,

  $\sigma_j(\boldsymbol{\delta}_\mathbf{t}, \mathbf{x}_t, \mathbf{p}_t; \theta_2)$

3. Contraction Mapping : Given nonlinear parameters θ_2, search for δ_t such that $s_{jt} = \sigma_j(\boldsymbol{\delta}_\mathbf{t}, \mathbf{x}_t, \mathbf{p}_t; \theta_2)$

4. From δ_t, estimate the linear parameters θ_1 using an analytical formula. Then form the GMM obj function $Q(\theta_2)$

5. Minimize $Q(\theta_2)$ over θ_2 with Steps 2-4 nested for every θ_2 trial

Problem with BLP Contraction Mapping

- BLP: Nesting a Contraction mapping for each trial of θ

$$\underset{\theta}{\text{Min}} \ \xi^{'}(\theta) Z W Z^{'} \xi(\theta)$$

- Problems:
  - Can be slow: For each trial of θ, we have to do a contraction mapping to obtain δ_t. This can be really slow if we have poor trial values of θ
  - Unstable if the tolerance levels used in the nesting is too high (suggested 10^{-12})
- An alternative approach is to use the MPEC approach that avoids the contraction mapping

MPEC Approach (Dube, Fox and Su, Ecta 2011)

- MPEC: Mathematical Programming with Equilibrium Constraints
  - Reduce to a constrained nonlinear programming problem

$$\underset{\theta,\xi}{\text{Min}}\ \xi^{'} Z W Z^{'} \xi$$

$$\text{subject to: } \tilde{\sigma}_{j}(\xi, \mathbf{x}, \mathbf{p}, F_{ns}, \theta) = S$$

- You have to search over both θ, ξ
  - With J brands and T periods(markets), JT ξ parameters
    - But no contraction mapping for each trial of θ
    - Nonlinear optimizers can do this effectively for small JT, but convergence can be tricky as JT becomes large

Contrast MPEC with BLP Contraction Mapping

- MPEC (Dube, Fox and Su 2011)

$$\underset{\theta,\xi}{\text{Min}}\ \xi' ZWZ' \xi$$

subject to: $\tilde{\sigma}_j(\xi, \mathbf{x}, \mathbf{p}, F_{ns}) = S_j$

- BLP: Nesting a contraction mapping for each trial of θ

$$\underset{\theta}{\text{Min}}\ \xi'(\theta) ZWZ' \xi(\theta)$$

Choosing Instruments

Choosing instruments

- The BLP (and MPEC) estimation procedure is based on instruments Z that satisfy the moment condition
  $E(\xi_{jt} \mid Z) = 0$
- IV's are needed for:
  - Moment conditions to identify θ_2 (heterogeneity)
    - Recall nested logit needed instruments even if price were not endogenous
  - Correcting for price (and other marketing mix) endogeneity
    - IV should be correlated with price but not with ξ_{jt}

Common Instruments: (1) Product Characteristics

- Own product characteristics (Almost all papers)
  - These can just identify the linear parameters associated with these characteristics in the mean utility
- Other product characteristics (BLP)
  - Sum of characteristics of other products produced by firm
  - Sum of characteristics of competitor products
  - Sudhir (2001) use sums by product group
- Intuition for instrument validity: other product characteristics have no direct impact on consumer utility for product, but through competition impacts prices
- Key assumption: Characteristics are chosen before ξ_{jt} known
- Widely used because it is generally available

Common Instruments (2): Cost Shifters

- Characteristics entering cost, but not demand
  - Generally hard to find
  - BLP use scale economies argument to use total production as a cost instrument
- Input factor prices
  - Affects costs and thus price, but not directly demand
  - Often used to explain price differentials across time, but often does not vary across brands (e.g., market wages)
    - If we know production is in different states or countries, we can get brand specific variation in factor costs (e.g., Sudhir, Chintagunta and Kadiyali (Mkt Sci 2005) use US and Japanese factor costs for Kodak and Fuji respectively.

Common Instruments (3): Prices in other markets

- Prices of products in other markets (Nevo 2001; Hausman 1996)
  - If there are common cost shocks across markets, then price in other markets can be a valid instrument
  - But how to justify no common demand shocks? (e.g., national advertising; seasonality)

Common Instruments (4): Lagged Characteristics

- When current characteristics are simultaneously related to the unobservables, one may motivate use of lagged characteristics similar to the dynamic panel data literature
- Example: Sweeting (Ecta, 2012) assumes an AR(1) process on the unobservables $\xi_{jt} = \rho\xi_{jt-1} + \eta_{jt}$, where $E(\eta_{jt} \mid x_{t-1}) = 0$ to justify the moment condition $E(\xi_{jt} - \rho\xi_{jt-1} \mid x_{t-1}) = 0$
- Can lagged prices be a valid instrument?
  - Not if last week's promotions drives this week's unobservable (e.g., due to stockpiling, which is unobserved)!!

Importance of IV Correction: BLP

TABLE III

RESULTS WITH LOGIT DEMAND AND MARGINAL COST PRICING
(2217 OBSERVATIONS)

Variable	OLS Logit Demand	IV Logit Demand	OLS ln (*price*) on *w*
Constant	−10.068	−9.273	1.882
	(0.253)	(0.493)	(0.119)
*HP/Weight**	−0.121	1.965	0.520
	(0.277)	(0.909)	(0.035)
Air	−0.035	1.289	0.680
	(0.073)	(0.248)	(0.019)
MP$	0.263	0.052	—
	(0.043)	(0.086)	
*MPG**	—	—	−0.471
			(0.049)
*Size**	2.341	2.355	0.125
	(0.125)	(0.247)	(0.063)
Trend	—	—	0.013
			(0.002)
Price	−0.089	−0.216	—
	(0.004)	(0.123)	
No. Inelastic Demands	1494	22	*n.a.*
(+/− 2 *s.e.'s*)	(1429–1617)	(7–101)	
R^2	0.387	*n.a.*	.656

Notes: The standard errors are reported in parentheses.
* The continuous product characteristics—hp/weight, size, and fuel efficiency (*MP$* or *MPG*)—enter the demand equations in levels, but enter the column 3 price regression in natural logs.

Identification

Step back: suppose we have consumer choice data

- Step 1: Estimate (δ, θ_2) by Simulated ML
  - Assuming iid double exponential for ε_{ijt}

$$P_{ijt} = \int_{\nu_i} \frac{\exp\{\delta_{jt} + (p_{jt} \ \ x_{jt})\Pi D_i + \nu_i\}}{\sum_{k=0}^{J} \exp\{\delta_{kt} + (p_{kt} \ \ x_{kt})\Pi D_i + \nu_i\}} dF(\nu_i)$$

  - Note, $\theta_2 = (\Pi, \Sigma_{\nu_i})$, i.e., heterogeneity is identified off differences in household choices for same $(p_{jt} \ \ x_{jt})$- not available in market data
- Step 2: Estimate θ_1
  - θ_1 identified based on cross market/time variations
  - **Correction for endogeneity needed even with consumer data**

$$\hat{\delta}_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$$

Variation in aggregate data that allows identification

- Some of the identification is due to functional form
  - But we also need enough variation in product characteristics and prices and see shares changing in response across different demographics for identification
- Across markets, variation in
  - demographics , choice sets (possibly)
- Across time, variation in
  - choice sets (possibly), demographics (possibly)
- To help further in identification
  - add micro data
  - add supply model

Variation in some familiar papers

- BLP (1995)
  - National market over time (10 years)
  - Demographics hardly changes, but choice sets (characteristics and prices) change
  - Identification due to changes in shares due to choice sets
- Nevo (2001)
  - Many local markets, over many weeks
  - Demographics different across markets, product characteristics virtually identical, except prices
  - Identification comes from changes in shares across choice sets and across demographics

Caveats about cross-sectional variation across markets

- Selection Problem: Is ξ_{jm} affected by market characteristics?
  - E.g., less fresh vegetables sold at lower prices in poor neighborhoods
  - May need to model selection
- Can distribution of ν_i be systematically different across markets?
  - If richer cities have better cycling paths for bikes (a market unobservable), distribution of random coefficients for bike characteristics may differ across markets (by income)
  - allowing for heterogeneity in distributions across markets may be necessary.
- Caution: identification of heterogeneity is tough with aggregate data, so we should not get carried away in demanding more complexity in modeling

Two other ways to improve identification and precision

- Add micro data based moments
- Add supply moments

Identification: Adding Micro Moments

Aggregate data may be augmented with micro data

- Consumer Level Data
  - panel of consumer choice data (Chintagunta and Dube JMR 2005)
  - cross-sections of consumer choice
    - second choice data on cars (BLP, JPE 2004)
    - Survey of consideration sets (theater, dvd) (Luan and Sudhir, WP 2006)
- Segment Summaries
  - Quantity/share by demographic groups (Petrin, JPE 2002)
  - Average demographics of purchasers of good j (Petrin JPE 2002)
  - Consideration set size distributions (Albuquerque et al. Mkt Sci 2009)

Augmenting market data with micro data (e.g., Petrin)

- Examples of micro-moments added by Petrin (2002)
  - Set 1: Price Sensitivity

    $$E[\{i \text{ purchases new vehicle} | \{y_i < \overline{y}_1\}]$$

    $$E[\{i \text{ purchases new vehicle} | \{\overline{y}_1 < y_i < \overline{y}_2\}]$$

    $$E[\{i \text{ purchases new vehicle} | \{y_i > \overline{y}_2\}]$$

  - Set 2: Choice and demographics (Family Size, Age)

    $$E[\{fs_i | i \text{ purchases a minivan}\}]$$

    $$E[\{fs_i | i \text{ purchases a station wagon}\}]$$

    $$E[\{fs_i | i \text{ purchases a SUV}\}]$$

    $$E[\{fs_i | i \text{ purchases a fullsize van}\}]$$

Importance of micro data (Petrin 2002)

TABLE 4
PARAMETER ESTIMATES FOR THE DEMAND-SIDE EQUATION

Variable	OLS Logit (1)	Instrumental Variable Logit (2)	Random Coefficients (3)	Random Coefficients and Microdata (4)
			A. Price Coefficients (α's)	
α_1	.07	.13	4.92	7.52
	(.01)**	(.01)**	(9.78)	(1.24)**
α_2			11.89	31.13
			(21.41)	(4.07)**
α_3			37.92	34.49
			(18.64)**	(2.56)**

Importance of micro data (Petrin 2002)

	B. Base Coefficients (β's)			
Constant	−10.03	−10.04	−12.74	−15.67
	(.32)**	(.34)**	(5.65)**	(4.39)**
Horsepower/weight	1.48	3.78	3.40	−2.83
	(.34)**	(.44)**	(39.79)	(8.16)
Size	3.17	3.25	4.60	4.80
	(.26)**	(.27)**	(24.64)	(3.57)*
Air conditioning standard	−.20	.21	−1.97	3.88
	(.06)**	(.08)**	(2.23)	(2.21)*
Miles/dollar	.18	.05	−.54	−15.79
	(.06)**	(.07)	(3.40)	(.87)**
Front wheel drive	.32	.15	−5.24	−12.32
	(.05)**	(.06)**	(3.09)	(2.36)**
Minivan	.09	−.10	−4.34	−5.65
	(.14)	(.15)	(13.16)	(.68)**
Station wagon	−1.12	−1.12	−20.52	−1.31
	(.06)**	(.07)**	(36.17)	(.36)**
Sport-utility	−.41	−.61	−3.10	−4.38
	(.09)**	(.10)**	(10.76)	(.41)**
Full-size van	−1.73	−1.89	−28.54	−5.26
	(.16)**	(.17)**	(235.51)	(1.30)**
% change GNP	.03	.03	.08	.24
	(.01)**	(.01)**	(.02)**	(.02)**

Identification:
Adding Supply Moments

Adding a Supply Equation

- Adding a supply equation can increase precision of demand side estimates
  - A Bertrand pricing assumption is usually used (BLP); using FOC

$$\begin{pmatrix} p_1 \\ \dots \\ p_J \end{pmatrix} = \begin{pmatrix} c_1 \\ \dots \\ c_J \end{pmatrix} + \left[\begin{pmatrix} \frac{ds_1}{dp_1} & \dots & \frac{ds_1}{dp_J} \\ \dots & \dots & \dots \\ \frac{ds_J}{dp_1} & \dots & \frac{ds_J}{dp_J} \end{pmatrix} .* O \right]^{-1} \begin{pmatrix} s_1 \\ \dots \\ s_J \end{pmatrix}$$

  - Where O is the ownership matrix where $O_{ij}=1$, if i and j are owned by the same firm
  - Even if Bertrand assumption is incorrect (supply equation is mis-specified, say due to dynamics), demand estimates tend to be consistent in characteristics based models as
    - (1) characteristics based price regressions (hedonics) tend to have good R^2
    - (2) link between margins and elasticity (own and cross) is generally right.

Adding a Supply Equation (contd...)

- $$\begin{pmatrix} c_{1t} \\ \cdots \\ c_{Jt} \end{pmatrix} = \begin{pmatrix} p_{1t} \\ \cdots \\ p_{Jt} \end{pmatrix} - \underbrace{\left(\left[\begin{pmatrix} \frac{ds_{1t}}{dp_{1t}} & \cdots & \frac{ds_{1t}}{dp_{Jt}} \\ \cdots & \cdots & \cdots \\ \frac{ds_{Jt}}{dp_{1t}} & \cdots & \frac{ds_{Jt}}{dp_{Jt}} \end{pmatrix} .* O_t \right]^{-1} \begin{pmatrix} s_{1t} \\ \cdots \\ s_{Jt} \end{pmatrix} \right)}_{\text{Margin } (m_{jt})}$$
- $$p_{jt} - m_{jt} = \underbrace{wW_{jt} + \omega_{jt}}_{c_{jt}}$$
- Construct the supply errors as

$$\omega_{jt} = p_{jt} - m_{jt}(\theta) - \omega W_{jt}$$

- One can create and stack supply moments using these supply errors with cost based instruments (see BLP 1995; Sudhir 2001)

Adding a supply equation (contd…)

- Supply errors: $\omega_{jt} = p_{jt} - m_{jt}(\theta) - \omega W_{jt}$
- Supply moments based on cost side instruments (Z_c):

$$E(\omega_{jt} \mid Z_c) = 0$$

- Stack the supply moments over the demand moments:

$$E\begin{bmatrix} \xi_{jt}(\theta) \mid Z \\ \omega_{jt}(\theta,\omega) \mid Z_c \end{bmatrix} = 0$$

- Since there is a pricing equation for each product
  - in effect, we double the number of observations (of course correlation between equations),
  - at the expense of estimating few more cost parameters
  - This helps improve the precision of estimates

Where to modify earlier code for supply equation…

```
while (Err >= Tol)
    de=de1;
    sh=zeros(NObs1,1);
    psh=zeros(NObs1,1);
    %Integrating over consumer heterogeneity
    for i=1:1:NCons;
        psh=exp(aw(:,i)+awp(:,i)+de); psh=reshape(psh',2,NObs)';
        spsh=sum(psh')';
        psh(:,1)=psh(:,1)./(1+spsh); psh(:,2)=psh(:,2)./(1+spsh);
        sh=sh+reshape(psh',NObs1,1);
    end;
    %Predicted Share
    sh=sh/NCons;
    %Adjust delta_jt
    de1=de+log(s)-log(sh);
    Err=max(abs(de1-de));
end;
delta=de1;
```

1. Compute own and cross price elasticity (see next slide for formula) for each household along with shares
2. Use these to construct margins along with delta

Recall: Elasticities with heterogeneity

- Own Elasticity: $$\eta_j = \left(\frac{\partial s_j}{\partial p_j}\right) \Big/ \frac{s_j}{p_j} = \frac{p_j}{s_j} \int \alpha_i s_{ij} (1 - s_{ij}) dF(D, \nu)$$
- Cross Elasticity: $$\eta_{jk} = \left(\frac{\partial s_j}{\partial p_k}\right) \Big/ \frac{s_j}{p_k} = -\frac{p_k}{s_j} \int \alpha_i s_{ij} s_{ik} dF(D, \nu)$$

Where to modify earlier code for supply equation…

```
% Analytically estimating linear parameters
blin=inv(xlin'*z*W*z'*xlin)*(xlin'*z*W*z'*delta);

% GMM Objective function over nonlinear parameters
er=delta-xlin*blin;
f=er'*z*W*z'*er;
```

3. Estimate the linear cost parameters
4. Construct the supply error
5. Stack the supply moments with appropriate weighting matrix in constructing the GMM objective function

Exercises

- Estimate the model with supply side moments added
- Compute the standard errors for the estimated demand side parameters
  - See Appendix to these slides on computing standard errors.

Summary

- Session 1:
  - Why is BLP so popular in marketing?
    - Handles heterogeneity and endogeneity in estimating demand systems with easily available aggregate data
  - The BLP algorithm and illustrative code
    - Simulation based integration, contraction mapping for mean utility, analytical formula for linear parameters, numerical optimization for the nonlinear parameters
- Session 2
  - MPEC versus BLP Contraction mapping (Nested Fixed Point)
  - Instruments
  - Identification
  - Improving precision through
    - Micro data based moments
    - Supply moments

```
% Sample code to illustrate estimation of BLP random coeffients model
% with aggregate data
% Written by K. Sudhir, Yale SOM
% For the Quantitative Marketing and Structural Econometrics Workshop
% at Duke University-2013

%Variables for demand equation (y, x) and the instruments (z)
y=log_s_s0;
x=[int p ad qtr];
z=[int cost lad qtr];

%Homogeneous logit without endogeneity fix
bOLS=inv(x'*x)*x'*y;
bOLS

%Homogeneous logit with endogeneity fix (W=I)
W=eye(size(z,2),size(z,2));
bIV1=inv(x'*z*W*z'*x)*x'*z*W*z'*y;
bIV1

%Homogeneous logit with endogeneity fix (W=inv(z'*z)) better than W=I, when
%different instruments have very different numerical  magnitudes; this
%equalizes the relative weights of the instruments in GMM
W=inv(z'*z);
bIV2=inv(x'*z*W*z'*x)*x'*z*W*z'*y;
bIV2

%Homogeneous logit with endogeneity fix (W=inv(E(z'ksi*ksi'*z)))
derr=y-x*bIV2;
zderr=z;
for i=1:1:size(z,2);
     zderr(:,i)=z(:,i).*derr;
 end;
W=inv((zderr)'*(zderr));
bIV3=inv(x'*z*W*z'*x)*x'*z*W*z'*y;
bIV3

%Heterogeneous logit with endogeneity fix
%Note I draw different 50 individuals for the different markets (time periods) from the
%same distribution. With markets, this is perfectly logical.
%With time, one could argue we need the same 50 individuals across
%markets. However, since there is no panel structure in  choices across time,
%different individuals across time is also correct.
%The advantage I see with different individuals is that you sample across
% a wider set of households from the distribution

NCons=50;
xlin=[int p ad qtr];
w1=randn(NObs1,NCons);
w2=randn(NObs1,NCons);
wp=randn(NObs,NCons);
```

```
wp=(reshape([wp';wp'],NCons,NObs1))';
b0=ones(4,1);

blin=bIV3;

options=optimset('Display','iter','TolFun',1e-12,'TolX',1e-12,'MaxIter',2500,↙
'MaxFunEvals',5000, 'LargeScale','off', 'HessUpdate', 'dfp');
[b, fval,exitflag,output,grad,hessian] = fminunc('AgglogitGMM',b0,options);


%Comparing the linear parameters across the different methods
bResults=[bOLS bIV1 bIV2 bIV3 blin]
%Formatted Reporting of the same parameters
horz= ['bOLS' '  bIV-W=I   ' '  bIV-W=zz '  ' bIV-W-ksi*z'     'blin-hetero'];
vert=['Int1 '; 'Int2 ';  'Price';  'Ad   '; 'Q1   ';  'Q2   ';  'Q3   '];
disp(horz)
for i = 1:1: size(vert);
    disp(vert(i,:))
    disp(bResults(i,:))
end;
% The RC Logit Model estimates with both linear (blin) and nonlinear (b) parameters
bResults1=[blin; b];

%Formatted Reporting of the same parameters
horz= ['bFull-hetero ' ];
vert=['Int1 '; 'Int2 ';  'Price';  'Ad   '; 'Q1   ';  'Q2   ';  'Q3   '; 'L11  '; 'L12↙
'; 'L22  '; 'Sigp ';];
disp(horz)
for i = 1:1: size(vert);
    disp(vert(i,:))
    disp(bResults1(i,:))
end;
```

```
% Sample code to illustrate estimation of BLP random coeffients model
% Written by K. Sudhir, Yale SOM
% For  Quantitative Marketing & Structural Econometrics Workshop @Duke-2013
% This is code for the GMM objective function that needs to minimized
function f=AgglogitGMM(b)
global w1 w2 wp x y z NCons NObs NObs1 xlin blin W z s blin log_s_s0;
%Step 1: Multiplying fixed Standard Normal draws  by lower triangular Cholesky matrix
% parameters to get the multivariate heterogeneity draws on intercepts and
% prices
aw=w1;
awp=wp;
aw1=b(1)*w1+b(2)*w2;
aw2=b(3)*w1;
%Step 2: Constructing the nonlinear part of the share equation (mu)
% using the heterogeneity draws on intercepts and price coeff
for i=1:1:size(w1,2);
    aw(:,i)=aw1(:,i).*x(:,1)+aw2(:,i).*x(:,2);
    awp(:,i)=b(4)*x(:,3).*wp(:,i);
end;
delta=log_s_s0;
Err=100;
Tol=1e-12;
de1=delta;
% Step 3: Contraction Mapping to get the delta_jt until Tolerance level met
while (Err >= Tol)
    de=de1;
    sh=zeros(NObs1,1);
    psh=zeros(NObs1,1);
    %Obtaining the predicted shares based on model
    for i=1:1:NCons;
        psh=exp(aw(:,i)+awp(:,i)+de);
        psh=reshape(psh',2,NObs)';
        spsh=sum(psh')';
        psh(:,1)=psh(:,1)./(1+spsh);
        psh(:,2)=psh(:,2)./(1+spsh);
        sh=sh+reshape(psh',NObs1,1);
    end;
    sh=sh/NCons;
    %Updating delta_jt based on difference between actual share and
    %predicted shares
    de1=de+log(s)-log(sh);
    Err=max(abs(de1-de));
end;
delta=de1;
% Step 4: Getting the linear parameters and setting up the objective fn
blin=inv(xlin'*z*W*z'*xlin)*(xlin'*z*W*z'*delta);
ksi=delta-xlin*blin;
% The GMM objective function that will be optimized over to get
% nonlinear parameters
f=ksi'*z*W*z'*ksi;
```

Estimating Standard Errors for a BLP Model

K. Sudhir, Yale School of Management

Prepared for the Quantitative Marketing and Structural Econometrics Conference

Duke University-2013

The asymptotic covariance matrix for the GMM estimator is given by (See Cameron and Trivedi, 2005; or any standard textbook which discusses GMM).

$$\hat{V}(\hat{\beta}_{GMM}) = N[D'ZWZ'D]^{-1}[D'ZW\hat{S}Z'D][D'ZWZ'D]^{-1}$$

Where '=

1. $D = [X \ \nabla_{\theta_2}(\hat{\delta}(\theta_2)]$, where D is the Jacobian of the moment conditions with respect to θ. For the linear parameters θ_1, this reduces to the corresponding variables *X*, while we need to take the derivatives with respect to θ_2
2. $\hat{S} = \frac{1}{N}\sum_{i=1}^{N} Z_i'\hat{\xi}_i\hat{\xi}_i'Z_i$
3. W is the weighting matrix used in the GMM estimation

$\nabla_{\theta_2}(\hat{\delta}(\theta_2)$ needs to be computed as an integral over the consumer heterogeneity, and therefore needs to be embedded in the code where the simulated shares are calculated. Since these gradients are not needed in the estimation, it should be computed outside the estimation loops.

$$\nabla_{\theta_2}(\hat{\delta}(\theta_2) = D\delta_{.t} = \begin{bmatrix} \frac{\partial\delta_{1t}}{\partial\theta_{21}} & \cdots & \frac{\partial\delta_{1t}}{\partial\theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\delta_{Jt}}{\partial\theta_{21}} & \cdots & \frac{\partial\delta_{Jt}}{\partial\theta_{2L}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_{1t}}{\partial\delta_{21}} & \cdots & \frac{\partial s_{1t}}{\partial\delta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial\delta_{21}} & \cdots & \frac{\partial s_{Jt}}{\partial\delta_{2L}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial s_{1t}}{\partial\theta_{21}} & \cdots & \frac{\partial s_{1t}}{\partial\theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial\theta_{21}} & \cdots & \frac{\partial s_{Jt}}{\partial\theta_{2L}} \end{bmatrix}$$

$$\frac{\partial s_{jt}}{\partial\delta_{jt}} = \frac{1}{NS}\sum_{i=1}^{NS} s_{jti}(1 - s_{jti})$$

$$\frac{\partial s_{jt}}{\partial\delta_{kt}} = \frac{1}{NS}\sum_{i=1}^{NS} s_{jti}s_{kti}$$

$$\frac{\partial s_{jt}}{\partial\theta_{2l}} = \frac{1}{NS}\sum_{i=1}^{NS} \nu_i^l s_{jti}\left(x_{jt}^l - \sum_{k=1}^{J} x_{kt}^l s_{kti}\right)$$

Standard errors are obtained from the square root of the diagonal of $\hat{V}(\hat{\beta}_{GMM})$.