Inverse Problems in Economics

Suzhou Huang
Outline

- Motivation

- Inverse problems in economics
  - The map: consumer preference ⇔ market behavior
  - Practice in typical econometric applications
  - Sources of uncertainty
  - Discrete-choice models
  - Endogeneity problem & model misspecification problem

- An illustrative example in a durable-goods market

- Conclusion
Motivation: **Importance of consumer preference**

- Consumer preference (*utility*) is fundamental
- Quantification of business
  - Market shares, costs, prices, profits, consumer surplus
  - Substitution patterns, market response to price changes
- Guidance to planning and product development
  - Capacity, product mix, R&D investment
  - New products, product improvements, cost reductions
- Public policy implications
  - Social welfare
  - Various regulations
Motivation: **Data analysis in different contexts**

- **Natural sciences**
  - Data from designed experiments
  - Well-controlled environments
  - Well understood theories
  - Small errors in regression equations

- **Social sciences**
  - Passive observations
  - Less-controlled environments
  - Poorly understood theories
  - Large errors in regression equations
  - Analysis technique matters a lot

- **Geosciences: somewhere in between?**
Forward Map: **Theory of Rational Choice**

- **Consumer preference**
- **Product information**
- **Market behavior**

**Decision process** --- utility maximization

**Market structure**
Inverse Map: Theory of Revealed Preference

- Consumer preference
- Decision process (utility maximization)
- Market structure
- Product information
- Market behavior

April 24, 2002
Inverse Problems in Economics
Degree of Inversion: What can be extracted?

- Depending on the available data
- Data types
  - Aggregate data: *prices, quantities, attributes*
  - Micro data: *+ demographics; one observation/consumer*
  - Scanner data: *multiple observations/consumer*
  - Limit case: *lots of observations under various conditions*
- Only in the limit case is a full inversion possible
- In realistic applications, scanner data are the best hope for information content
Practice in Econometric Applications

- Typically, aggregate consumer preference is inferred, based on aggregate market observations
- Parametric approaches
  - Focusing on parameter extraction
  - Methods of least squares, instrumental variables
  - Generalized method of moments
- Non-parametric approaches
  - Focusing on numerical distribution
  - Bayesian inference: Markov Chain Monte Carlo
  - Typically requiring a crowded product space
- Dynamic applications: extremely rare
  - requiring the setting of dynamic games
Sources of Uncertainty

(1) Measurement errors
- Imperfect market observations
- Standard filtering problem

(2) Endogeneity problem
- Unobserved product information
- Circumventable in “lucky” cases

(3) Model misspecification problem
- Inadequate decision models
- Model selection techniques may offer some help
Discrete-choice Models: Decision making

- In a perishable-goods market:

  Choosing a brand

  \[ R_0 = \max_{a \in A} \{ u_0[a] \}, \quad A = \{\text{all brands}\} \]

  where \( R_0 \in \{B_1, B_2, \cdots\} \)

- In a durable-goods market with transaction costs:

  Choosing a consumption pattern (over time) of a brand

  \[ R_0[s] = \max_{a \in A} \{ u_0[s,a] + \rho V_0[a] \}, \quad \forall s \in A = \{\text{allowed actions}\} \]

  where \( R_0[s] \in \{L_1L_1, N_1U_1, U_1U_1, L_2L_2, N_2U_2, U_2U_2, \cdots\} \)
Discrete-choice Models: Consumer preference

- Consumer’s utility: \( u_\theta[j] = \delta_j(x, \beta) + \nu_j(x, \theta) \)

- Mean utility: \( \delta_j(x, \beta) = x_j \cdot \beta + \xi_j \)
  where \( x_j \) and \( \xi_j \) are observable and unobservable attributes

- Pure hedonic: \( \nu_j(x, \theta) = x_j \cdot \theta \) with \( E[\theta] = 0 \)

- Other specifications, such as Logit models
  - No interaction between \( x \) and \( \theta \)
  - Suffers from Independence of Irrelevant Alternative

- Tradeoff: ease of aggregation and descriptive power
Endogeneity problem

- $L$ independent market observations

- Directly using market shares: *nonlinear regression*

- An easier regression: $\delta_j^l = x_j^l \cdot \beta - \alpha p_j^l + \xi_j^l, \quad l \in \{L\}$

- Statistical dependence: $\frac{1}{L} \sum_{l=1}^{L} p_j^l \xi_j^l \neq 0 \quad \Rightarrow \text{endogenous } p$

- Instrumental variable $z$ defined as

\[
\frac{1}{L} \sum_{l=1}^{L} z_j^l \xi_j^l = 0 \quad \text{and} \quad \frac{1}{L} \sum_{l=1}^{L} z_j^l p_j^l \neq 0
\]
Model misspecification problem

- Mean utilities are not directly observable

- Market shares are determined by mean utilities:
  \[ S_{\text{observed}} \rightarrow S(\delta) \]

- For a given discrete-choice model, they can be solved from market shares:
  \[ \delta_j \rightarrow \Delta_j(S_{\text{observed}}) \]

- A model misspecification implies a wrong functional form for \( \Delta_j(S_{\text{observed}}) \)
A stylized model

- A duopoly market: firm $A$ and firm $B$, each makes one product
- Consumer preference: pure hedonic
- Local markets are perturbed by unobserved attributes

Case 1: close substitutes

Case 2: well differentiated
Modeling procedure

- **Forward problem (data generation):** for local market $l$
  - Generate an independent sample: $\{x_j^l, \xi_j^l, c_j^l\}$
  - Solve a Nash equilibrium: $\{p_j^l, S_j^l, \cdots\}$

- **Inverse problem (regression):** using all $L$ markets
  - Treat $\{x_j^l, c_j^l, p_j^l, S_j^l\}$ as observable, and $\xi_j^l$ unobservable
  - Uncover the parameter vector: $\{\beta_0, \beta_1, \beta_2, \alpha\}$

- **Assess the severity of model misspecification**
Segmentation pattern

☐ In durable-goods market: \( \{ L_A, N_A, U_A, L_B, N_B, U_B, I \} \)

Figure 1: Consumer segmentation patterns.

☐ In perishable-goods market: \( \{ N_A, N_B, I \} \)
Regression Result: **the Correct Model**

Table 5: Regression results of the correctly specified model with $L = 500$.

<table>
<thead>
<tr>
<th>$\sigma_\varepsilon$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
<th>Estimation</th>
<th>$\sigma^2_\varepsilon/\sigma^2_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.1012(3)</td>
<td>0.988(3)</td>
<td>0.986(3)</td>
<td>0.988(3)</td>
<td>OLS</td>
<td>$\sim 2%$</td>
</tr>
<tr>
<td></td>
<td>0.101(1)</td>
<td>0.99(1)</td>
<td>0.99(1)</td>
<td>0.99(1)</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>0.003</td>
<td>0.110(1)</td>
<td>0.88(1)</td>
<td>0.88(1)</td>
<td>0.89(1)</td>
<td>OLS</td>
<td>$\sim 15%$</td>
</tr>
<tr>
<td></td>
<td>0.096(4)</td>
<td>1.03(4)</td>
<td>1.04(4)</td>
<td>1.03(4)</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>0.006</td>
<td>0.134(2)</td>
<td>0.64(2)</td>
<td>0.62(2)</td>
<td>0.65(2)</td>
<td>OLS</td>
<td>$\sim 50%$</td>
</tr>
<tr>
<td></td>
<td>0.101(7)</td>
<td>0.99(8)</td>
<td>0.98(8)</td>
<td>0.98(7)</td>
<td>IV</td>
<td></td>
</tr>
</tbody>
</table>

**Case 1:** when $\lambda$ and $\delta$ are close substitutes

- Ordinary Least Squares method: *quantitatively inconsistent*
- Instrumental Variable method: *quantitatively consistent*

- $\sigma^2_\varepsilon/\sigma^2_\delta$
## Regression Result: the ND model

### Table 6: Regression results of the ND model with $L = 2000.$

<table>
<thead>
<tr>
<th>$\sigma_\xi$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.001</strong></td>
<td>0.108(4)</td>
<td>0.023(8)</td>
<td>0.29(1)</td>
<td>-0.024(7)</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>0.123(4)</td>
<td>0.085(9)</td>
<td>0.31(1)</td>
<td>0.090(9)</td>
<td>IV</td>
</tr>
<tr>
<td><strong>0.003</strong></td>
<td>0.10(1)</td>
<td>0.01(3)</td>
<td>0.29(4)</td>
<td>-0.06(2)</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>0.12(1)</td>
<td>0.09(3)</td>
<td>0.32(3)</td>
<td>0.10(3)</td>
<td>IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sigma_\xi$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.001</strong></td>
<td>0.203(4)</td>
<td>-0.04(1)</td>
<td>0.02(1)</td>
<td>-0.03(1)</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>0.229(5)</td>
<td>0.01(1)</td>
<td>0.06(2)</td>
<td>0.07(1)</td>
<td>IV</td>
</tr>
<tr>
<td><strong>0.003</strong></td>
<td>0.20(1)</td>
<td>-0.05(4)</td>
<td>0.01(4)</td>
<td>-0.04(2)</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>0.24(2)</td>
<td>0.01(4)</td>
<td>0.06(4)</td>
<td>0.11(4)</td>
<td>IV</td>
</tr>
</tbody>
</table>

- Model misspecification: goods are treated as non-durable
- Ordinary Least Squares method: *qualitatively inconsistent*
- Instrumental Variable method: *quantitatively inconsistent*
Regression result: **the Logit Model**

Table 7: Regression results of the Logit model with $L = 5000$.

<table>
<thead>
<tr>
<th>$\sigma_\xi$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>$-2.811(3)$</td>
<td>$3.868(3)$</td>
<td>$-0.184(9)$</td>
<td>$-3.860(5)$</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>$-1.670(3)$</td>
<td>$8.288(6)$</td>
<td>$1.244(9)$</td>
<td>$4.445(6)$</td>
<td>IV</td>
</tr>
<tr>
<td>0.003</td>
<td>$-3.132(7)$</td>
<td>$2.77(2)$</td>
<td>$-0.54(3)$</td>
<td>$-6.07(2)$</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>$-1.67(7)$</td>
<td>$8.40(2)$</td>
<td>$1.26(3)$</td>
<td>$4.52(2)$</td>
<td>IV</td>
</tr>
</tbody>
</table>

Case 2: when $A$ and $B$ are well differentiated

<table>
<thead>
<tr>
<th>$\sigma_\xi$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
<th>Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>$-1.378(3)$</td>
<td>$-0.237(8)$</td>
<td>$-0.279(8)$</td>
<td>$-0.615(6)$</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>$-1.114(3)$</td>
<td>$0.253(8)$</td>
<td>$0.151(8)$</td>
<td>$0.410(7)$</td>
<td>IV</td>
</tr>
<tr>
<td>0.003</td>
<td>$-1.461(7)$</td>
<td>$-0.34(2)$</td>
<td>$-0.36(3)$</td>
<td>$-0.90(2)$</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>$-1.07(1)$</td>
<td>$0.41(3)$</td>
<td>$0.29(3)$</td>
<td>$0.64(2)$</td>
<td>IV</td>
</tr>
</tbody>
</table>

- Model misspecification: + *introducing product specific tastes*
- Ordinary Least Squares method: *fails miserably*
- Instrumental Variable method: *barely consistent qualitatively*
Inverse problems in economics are often ill-posed, due to endogeneity problem (unobservability) and model misspecification problem (poor modeling).

It is necessary to go beyond the standard statistical description of uncertainties and model selection.

Seeking techniques that can assess uncertainties when the true underlying model is unknown.

Methodology developed in geosciences is likely to be useful in economics.