ECO 310: Empirical Industrial Organization

Lecture 7: Demand Systems: Discrete Choice Models [2]

Victor Aguirregabiria (University of Toronto)

October 29, 2018

Outline on today's lecture

1. **Estimation of the Standard Logit Model**
   - 1.1. **Endogeneity problem & bias of OLS estimator**
   - 1.2. **Instrumental Variables estimation**
2. **Logit model with heterogeneous coefficients**

1. Estimation of the Standard Logit Model

Estimation Standard Logit Model: Data

- Suppose that we have data on quantities (sold), prices, and characteristics of all the J products in a market:

$$\text{Data} = \{q_j, \; p_j, \; X_{1j}, \; ..., X_{Kj} \text{: for } j = 1, 2, ..., J\}$$

- Suppose that we also observe the consumers who have not purchased any of the J products, q_0.

- For instance, the Stata dataset `verboven_cars.dta` contains the following variables for $J = 356$ car models in the markets of five different European countries.

  price; quantity; brand; displacement (in cc); horsepower (in kW); weight (in kg); seats; doors; length; (in cm); width (in cm); height (in cm); fuel efficiency (liter per km); maximum speed (km/hour); time to acceleration (secs from 0 to 100 km/h).

Estimation Logit Model

- Given quantities, we can construct market shares. Market size (number of consumers) is: $H = q_0 + q_1 + ... + q_J$. And the market share of product j is $s_j = q_j / H$.

- The logit model implies the regression model:

$$y_j = \beta_p \; p_j + \beta_1 \; X_{1j} + ... + \beta_K \; X_{Kj} + \xi_j$$

  where $y_j \equiv \ln(s_j) - \ln(s_0)$, $\beta_p = -\alpha$.

- The error term ξ_j represents characteristics of product j valuable to the consumers but unobservable to us as researchers.

- Given these data, we can estimate parameters $(\beta_p, \beta_1, ..., \beta_K)$.

OLS Estimation: Endogeneity problem

- Unfortunately, the OLS estimator does not provide unbiased (consistent) estimates of the parameters of the model.
- Products with higher unobserved quality ξ_j tend to have higher prices [See next slide]:

$$cov(p_j, \xi_j) > 0$$

- This endogeneity problem implies that the OLS estimate $\widehat{\beta}_p^{OLS}$ estimates the combination of two effects:

  (a) the causal effect of price on y_j: i.e., $\beta_p < 0$;

  (b) an indirect positive effect (not causal) that comes from the correlation between price and unobserved product quality.

$$\widehat{\beta}_p^{OLS} \rightarrow \frac{cov(y_j, p_j)}{var(p_j)} = \beta_p + \frac{cov(p_j, \xi_j)}{var(p_j)}$$

- We could even get $\widehat{\beta}_p^{OLS} > 0$.

Endogeneity problem: Example

- Suppose that the profit maximization condition, Marginal Revenue = Marginal Cost, implies the following optimal price for the firm selling product j:

  $$p_j = \gamma_1 \ X_{1j} + ... + \gamma_K \ X_{Kj} + \gamma_\xi \ \xi_j$$

  where $\gamma' s$ are parameters.

- Product characteristics affect price because: (1) they affect MCs, i.e., higher quality products are more costly to produce; and (2) they enter in demand and affect marginal revenue.

- The model consist of the logit demand equation and the pricing equation. For simplicity, let's omit the X variables:

  $$\begin{aligned} y_j &= \beta_p \ p_j + \xi_j \\ p_j &= \gamma_\xi \ \xi_j \end{aligned}$$

Endogeneity problem: Example [2]

- These are the structural equations of the model (I have omitted the constant terms; the variables are in deviations with respect to their respective means).

$$
\begin{aligned}
y_j &= \beta_p \, p_j + \xi_j \\
p_j &= \gamma_\xi \, \xi_j
\end{aligned}
$$

- Solving the price equation into the demand equation, we have that:

$$
y_j = (\beta_p \gamma_\xi + 1) \; \xi_j
$$

- Therefore:

$$
\begin{aligned}
cov(y_j, p_j) &= (\beta_p \gamma_\xi + 1) \; \gamma_\xi \; var(\xi_j) \\
var(p_j) &= (\gamma_\xi)^2 \; var(\xi_j)
\end{aligned}
$$

Endogeneity problem: Example [3]

- Then, in this model, the OLS estimator is such that:

$$\begin{aligned}\widehat{\beta}_p^{OLS} &\rightarrow \frac{cov(y_j, p_j)}{var(p_j)} = \frac{(\beta_p \gamma_\xi + 1)\gamma_\xi \ var(\xi_j)}{(\gamma_\xi)^2 \ var(\xi_j)} \\ &= \beta_p + \frac{1}{\gamma_\xi}\end{aligned}$$

- Since $\dfrac{1}{\gamma_\xi} > 0$, the OLS estimator is an upward biased estimate of the true β_p.
- Since $\beta_p < 0$, we have that the estimate is biased towards zero, or it could be even positive.

Instrumental Variables (IV) Estimation

- To deal with this endogeneity problem, we can use IV estimation.
- We need a variable (or multiple variables), Z_j, that satisfies the following conditions.
- **[1] Exclusion.** Z_j is NOT an explanatory variable in the demand equation of product j, i.e., Z_j is not part of vector $\mathbf{X}_j$.
- **[2] No correlation with error.** Z_j is NOT correlated with product j unobserved quality ξ_j.
- **[3] Relevance.** In the regression of price, p_j, on the vector $\mathbf{X}_j$ and on Z_j, variable Z_j has a significant (partial) correlation with p_j.

IV Estimation in Two stages (2SLS)

- To implement the IV estimator we can use a two stage least squares (2SLS) method.
- **[Stage 1]** We run an OLS regression for price on the exogenous variables of the model (vector $\mathbf{X}_j$) and the instrument (Z_j):

$$p_j = \gamma_z \; Z_j + \gamma_1 \; X_{1j} + ... + \gamma_K \; X_{Kj} + e_j$$

  And obtain the fitted values: $\widehat{p}_j = \widehat{\gamma}_z \; Z_j + \widehat{\gamma}_1 \; X_{1j} + ... + \widehat{\gamma}_K \; X_{Kj}$.

- **[Stage 2]** We run an OLS regression of the demand equation but using the fitted values from stage 1 ($\widehat{p}_j$) instead of price (p_j) as explanatory variable:

$$y_j = \beta_p \; \widehat{p}_j + \beta_1 \; X_{1j} + ... + \beta_K \; X_{Kj} + \xi_j^*$$

- The estimator in this second stage is the IV estimator. Standard errors should be corrected.

How to get instruments?

- Under the assumption that the observable characteristics (other than price) X_{kj} are not correlated with the unobserved quality ξ_j, the model of demand and price competition of differentiated products provides IVs.
- This model implies that the profit-maximizing price for product j depends not only on its own characteristics ($\mathbf{X}_j$ and ξ_j) but also on the characteristics of other products competing with product j ($\mathbf{X}_i$ and ξ_i).
- Intuitively, if the values of $\mathbf{X}$ are such that there are other products with similar characteristics as product j, price competition is intense and price p_j is low:

$$p_j \text{ depends positively on distance}(\mathbf{X}_j, \mathbf{X}_i)$$

How to get instruments? [2]

- Under this argument, we can use the characteristics of products other than j (i.e., X_{ki} for $i \neq j$) as an instrument for product p_j.
- These instruments are called the **Berry-Levinsohn-Pakes (BLP) instruments** in the demand of differentiated products.
- For instance, we can use:

$$Z_j = \min_{i \neq j} \|\mathbf{X}_j - \mathbf{X}_i\|$$

where $\|\mathbf{a} - \mathbf{b}\|$ is the Euclidean distance between vectors **a** and **b**:

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(a_1 - b_1)^2 + ... + (a_K - b_K)^2}$$

- Or we can use as instrument Z_j other functions of $\mathbf{X}_j$ and $\mathbf{X}_i$:

$$Z_j = \frac{\sum_{i \neq j} \|\mathbf{X}_j - \mathbf{X}_i\|}{J-1} \quad \text{or} \quad Z_j = \frac{\sum_{i \neq j} \mathbf{X}_i}{J-1}$$

2. Logit Model with Heterogeneous Coefficients

Logit Demand Model: Some limitations

- The standard Logit model imposes some strong restrictions on price elasticities.
- For any, products i and j:

$$\frac{\partial s_j}{\partial p_i} \frac{p_i}{s_j} = \alpha \ s_i \ p_i$$

- And increase in the price of product i (e.g., a luxury car) implies the same proportional increase in the demand of product j (other luxury car) as in the demand of product b (a very low quality car).
- We can introduce two extensions in the model that relax this restriction:

  [A] Consumer heterogeneous coefficients β_h

  [B] Nested Logit model for ε_h

Logit with heterogeneous coefficients: Micro data

- Suppose that our dataset is such that we observe consumer purchasing decisions for G groups of consumers, indexed by $g = 1, 2, ..., G$.
- These G groups are based on consumer demographic characteristics such as age, gender, income, geographic location, etc.
- For instance, group 1 could be defined as: "Consumers in age group 20-to-30; Female; income group [$70K-$80K]; in city A".
- For each group g, we observe quantities q_{gj} and the number of consumers H_g, such that we can construct the market shares $s_{gj} = q_{gj} / H_g$.

Logit with heterogeneous coefficients

- Suppose that consumer groups are heterogeneous in the preferences: in the utility parameters α and β.
- The logit model for group g is:

$$\ln(s_{gj}) - \ln(s_{g0}) = -\alpha_g \; p_j + \beta_{1g} \; X_{1j} + ... + \beta_{Kg} \; X_{Kj} + \xi_{gj}$$

- Note that the explanatory variables (p_j and X_j) are the same for each group, but the dependent variable and the parameters are different.
- We have G different regression equations, one for each group. We can estimate the model parameters separately for each group using the IV method described above.

Heterogeneous coeff. deal with limitations of standard Logit

- For each group g, the model has the same structure as the standard logit. However, now the aggregate demand of product j has a different structure.

- The aggregate demand of product j is:

$$q_j = \sum_{g=1}^{G} q_{gj} = \sum_{g=1}^{G} H_g \ s_{gj} = \sum_{g=1}^{G} H_g \left[\frac{\exp\left\{\delta_{gj}\right\}}{\sum_{i=0}^{J} \exp\left\{\delta_{gi}\right\}} \right]$$

with $\delta_{gj} = -\alpha_g \ p_j + \beta_{1g} \ X_{1j} + ... + \beta_{Kg} \ X_{Kj} + \xi_{gj}$.

- Now, we have:

$$\frac{\partial q_j}{\partial p_i} \frac{p_i}{q_j} = \left[\sum_{g=1}^{G} H_g \ \frac{\partial s_{gj}}{\partial p_i} \right] \frac{p_i}{q_j} = \left[\sum_{g=1}^{G} H_g \ \alpha_g \ s_{gj} \ s_{gi} \right] \frac{p_i}{q_j}$$

Heterogeneous coeff. Logit: Price elasticities

$$\frac{\partial q_j}{\partial p_i}\frac{p_i}{q_j} = \left[\sum_{g=1}^{G} H_g \ \alpha_g \ s_{gj} \ s_{gi}\right]\frac{p_i}{q_j}$$

- Note that $\sum_{g=1}^{G} H_g \ \alpha_g \ s_{gj} \ s_{gi}$ is no longer equal to $\alpha \ s_j \ s_i$.
- The term $\sum_{g=1}^{G} H_g \ \alpha_g \ s_{gj} \ s_{gi}$ measures the covariation of the market shares of products j and i across groups.
- This covariation depends on the characteristics of these products.
- If the two products have similar characteristics, then $\sum_{g=1}^{G} H_g \ \alpha_g \ s_{gj}$ s_{gi} is large.
- If the two products have very different characteristics, then $\sum_{g=1}^{G} H_g$ $\alpha_g \ s_{gj} \ s_{gi}$ is small.

Heterogeneous coeff. Logit: Price elasticities

- To see the math of this result, consider the case on only two products, 1 and 2, without outside alternative 0, and $H_g = H$ for all the groups, and $\alpha_g = \alpha = 1$.
- The value $s_{g1}\ s_{g2}$ is maximized (given $s_{g1} + s_{g2} = 1$) when $s_{g1} = s_{g2} = 1/2$, and it declines when the distance between s_{g1} and s_{g2} increases.
- Then, $\sum_{g=1}^{G} s_{g1}\ s_{g2}$ reaches its maximum value if $s_{g1} = s_{g2} = 1/2$ for every group g.
- If the characteristics of the two products are similar, then $\delta_{g1} \simeq \delta_{g2}$ for every group g and $\sum_{g=1}^{G} s_{g1}\ s_{g2}$ is close to its maximum.
- If the characteristics of the two products are very different then, for some groups $s_{g1} \simeq 1$ and $s_{g2} \simeq 0$, and viceversa for other groups, such that $\sum_{g=1}^{G} s_{g1}\ s_{g2}$ is close to zero.