Industrial Organization II (ECO 2901)
Winter 2013. Victor Aguirregabiria

Problem Set #1
Due of Friday, March 22, 2013

TOTAL NUMBER OF POINTS: 200

PROBLEM 1 [30 points]. Consider the estimation of a model of demand of differentiated products using aggregate market data as in Berry (1994) and Berry, Levinsohn, and Pakes (1995). The dataset includes information on prices (p), quantities (q), and product characteristics other than price (X) for J products over M separate markets:

$$\text{Data} = \{p_{jm},\, q_{jm},\, X_{jm} : j = 1, 2, ..., J;\, m = 1, 2, ..., M\}$$

where we index markets by m and products by j. In this dataset the number of markets M is large relative to the number of product varieties, e.g., $J = 50$ and $M = 1,000$. The specification of the model is the one in the random-coefficients 'BLP' model, where the utility of buying product j for consumer i in market m is:

$$V_{ijm} = X_{jm} \left[\beta + \upsilon_{im}^{\beta}\right] - p_{jm} \left[\alpha + \upsilon_{im}^{\alpha}\right] + \xi_{jm} + \varepsilon_{ijm}$$

where υ_{im}^{β} and υ_{im}^{α} are zero mean normal random variables that capture consumer heterogeneity in the marginal utility of product characteristics, and ε_{ijm} is a type 1 extreme value distributed variable that also captures consumer heterogeneity in preferences. For the outside alternative, $j = 0$, we have that $V_{i0m} = 0$. H_m is the number of consumers in market m, i.e., market size.

Question 1.1 [10 points] Suppose that any observable measure of market size H_m available to the researcher includes substantial measurement error. Propose a simple approach to deal with this problem. Explain in detail your proposed method.

Question 1.2 [10 points] Suppose that a substantial proportion of products are not available in all the M markets. For instance, the top-5 products (according to their market shares at the national level) are available in 95% of the local markets, while products below the top-20 are available only in 60% of the local markets. There are multiple factors that contribute to explain why a product is available or not in a local market, e.g., market size, competition, local consumer preferences, distance to production size, economies of density, etc. We believe that in the industry under study an important factor to explain these differences in product availability across markets has to do with heterogeneity among local markets in the preferences of the average local consumer, as represented by the unobserved variables $\{\xi_{jm}\}$.

(a) Discuss the implications of this issue on the properties of the standard GMM estimator using BLP moment conditions.

(b) Propose an approach to deal with this problem. Explain in detail your proposed method.

Question 1.3 [10 points] Recently, Petrin and Train (Journal of Marketing Research, 2010) and Kim and Petrin (WP, 2011) have proposed Control Function (CF) approaches to estimate the 'BLP' model and extensions of this model that allow for interactions between market level unobservables ξ and price p in the utility function. This CF is in the spirit of Rivers and Vuong (JE, 1988) at it operates in two-steps. The first step is an OLS estimation of a linear regression for the reduced form equation of prices. In the second step, the residuals from the first-step regression are plugged-in utility function to control the unobservables $\{\xi_{jm}\}$ and the the parameters of the model are estimated by Maximum Likelihood.

(a) Discuss in more detail the CF approach providing specific equations and formulas.

(b) Discuss the relative advantages and limitations of the CF approach versus the GMM-BLP approach.

PROBLEM 2 [10 points]. Describe in detail Ackerberg-Frazer-Caves (2006) criticism to the identification of the parameters in the Cobb-Douglas Production Function using Olley-Pakes Control Function approach.

PROBLEM 3 [30 points]. Consider the Two-Players Binary Choice Probit Game of complete information in Tamer (REStud, 2003). The structural equations of the model are the following best response functions:

$$\begin{aligned} Y_1 &= 1\{\ \alpha_1' X + \beta_1\ Z_1 - \delta_1\ Y_2 - \varepsilon_1 \geq 0\ \} \\ Y_2 &= 1\{\ \alpha_2' X + \beta_2\ Z_2 - \delta_2\ Y_1 - \varepsilon_2 \geq 0\ \} \end{aligned}$$

where: $Y_1 \in \{0,1\}$ and $Y_2 \in \{0,1\}$ represent players' decisions; α_1, α_2, β_1, β_2, δ_1, and δ_2 are parameters, and we assume that $\delta_1 \geq 0$ and $\delta_2 \geq 0$; X, Z_1, and Z_2 are exogenous observable variables; and ε_1 and ε_2 are Normal random variables independent of (X, Z_1, Z_2) with zero mean, unit variances, and correlation parameter ρ. We use the $\Phi^{(2)}(\varepsilon_1, \varepsilon_2; \rho)$ to represent the CDF of $(\varepsilon_1, \varepsilon_2)$. The researcher observes a random sample of M markets with information on $\{Y_{1m}, Y_{2m}, X_m, Z_{1m}, Z_{2m} : m = 1, 2, ..., M\}$. We are interested in using this sample to estimate the vector of structural parameters $\theta = (\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2,\ \delta_1,\ \delta_2,\ \rho)'$. For $(y_1, y_2) \in \{(0,0),\ (0,1),\ (1,0),\ (1,1)\}$, define the *Conditional Choice Probability* (CCP) function

$$P(y_1, y_2 \mid x, z_1, z_2; \theta) = \Pr\left(Y_1 = y_1,\ Y_2 = y_2 \mid X = z,\ Z_1 = z_1,\ Z_2 = z_2,\ \theta\right)$$

Question 3.1 [10 points]. Obtain the reduced form equations of the model, i.e., the relationship between the four possible values of the endogenous variables (Y_1,Y_2) and the exogenous variables and parameters.

Question 3.2 [5 points]. Using the reduced form equations, obtain the expressions for the CCPs $P(0,0 \mid x, z_1, z_2; \theta)$ and $P(1,1 \mid x, z_1, z_2; \theta)$. Use $\Phi^{(2)}(\varepsilon_1, \varepsilon_2; \rho)$ to represent the join CDF of $(\varepsilon_1, \varepsilon_2)$.

Question 3.3 [15 points]. Suppose that: (i) $\beta_1 \neq 0$ and $\beta_2 \neq 0$; and (ii) the distribution of (Z_1,Z_2) is such that the support set is $\mathbb{R}^2$, in the limit as $Z_1 \to \infty$ we have that the distribution of (X, Z_2) is non-degenerate, and similarly in the limit as $Z_2 \to \infty$ the distribution of (X, Z_1) is non-degenerate. For instance, condition (ii) is satisfied if (Z_1,Z_2) are jointly normally distributed conditional on Z. Prove formally that under conditions (i) and (ii) θ is identified using the data and CCP functions $P(0,0 \mid x, z_1, z_2; \theta)$ and $P(1,1 \mid x, z_1, z_2; \theta)$. [Hint: Read the proof in the Appendix of Tamer (2003)).

PROBLEM 4 [30 points]. Consider a Two-Player Game of Market Entry with Incomplete Information. The players' payoff functions are:

$$\begin{aligned} \Pi_1 &= \pi_1^M(X, Z_1) + Y_2\ \left[\pi_1^D(X, Z_1) - \pi_1^M(X, Z_1)\right] - \varepsilon_1 \\ \Pi_2 &= \pi_2^M(X, Z_2) + Y_1\ \left[\pi_2^D(X, Z_2) - \pi_2^M(X, Z_2)\right] - \varepsilon_2 \end{aligned}$$

For every firm i, $Y_i \in \{0,1\}$ represents the market entry decision of firm i. $\pi_i^M(.)$ and $\pi_i^D(.)$ are functions that represent the profit of firm i under monopoly and under duopoly, respectively. X, Z_1, and Z_2 are exogenous variables which are observable to the researcher and common knowledge to the players. For each player, ε_i is a random variable that represents a component of the fixed cost of player i that is private information of this player. We assume that ε_1 and ε_2 are independent of (X, Z_1, Z_2) and independently distributed between them with standard Normal distributions. Define the CCP functions $P_i(x, z_1, z_2) \equiv \Pr\left(Y_i = 1 \mid X = z,\ Z_1 = z_1,\ Z_2 = z_2\right)$ for $i = 1, 2$.

Question 4.1 [10 points]. Describe the equilibrium mapping in the space of CCPs such that a pair of equilibrium probabilities $(P_1(x, z_1, z_2)\ ,\ P_2(x, z_1, z_2))$ is a fixed point of that mapping.

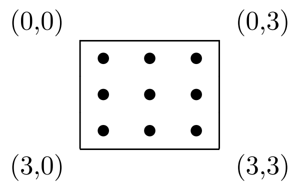
Question 4.2 [15 points]. Suppose that: (i) for any value of X, the function $\pi_i^M(X, Z_i)$ depends on Z_i; and (ii) the distribution of (X,Z_1,Z_2) is such that for any value of (X, Z_1) the distribution of Z_2 is non-degenerate, and similarly for any value of (X, Z_2) the distribution of Z_1 is non-degenerate.

Prove formally that under conditions (i) and (ii) the payoff functions π_1^M, π_1^D, π_2^M, and π_2^D are nonparametrically identified.

Question 4.3 [5 points]. Discuss the implications of multiple equilibria in the model on the identification and estimation of the payoff function.

PROBLEM 5 [100 points]. Here we consider a **Game of Market Entry and Spatial Location**. The market is a square city where the measure of a side of this square is 3 Km. We represent this city in the two-dimension Euclidean space with vertices at points (0,0), (0,3), (3,0), and (3,3). There are $L = 9$ locations where firms can open stores. The following figure represents the city and the feasible business locations.

Market and feasible business locations (represented with •)



We index locations by ℓ that belongs to the set $\{1, 2, ..., L\}$. There are two potential entrants in the market that we represent as firm i and firm j. Each potential entrant decides whether to operate a store in the market and the location of the store. Let a_i represent the decision of firm/potential entrant i, such that $a_i \in \{0, 1, ..., L\}$ and $a_i = 0$ represents "no entry", and $a_i = \ell > 0$ represents entry in location ℓ. The profit of not being active in the market is normalized to zero. The profit of a store in location ℓ is:

$$\Pi_{i\ell} = POP_\ell \left[\delta_{0i} - \delta_{1i}\ 1\{a_j = \ell\} - \delta_{2i} \left(\sum_{\ell' \in b(\ell)} 1\{a_j = \ell'\} \right) \right] - \alpha_i\ RENT_\ell - \varepsilon_{i\ell}$$

POP_ℓ and $RENT_\ell$ are exogenous variables that represent the population and the average rental price in location ℓ, respectively. a_j represents the entry decision of the competing firm $j \neq i$. $1\{.\}$ is the indicator function such that $1\{a_j = \ell\}$ is the indicator of the event "firm j has decided to have a store in location ℓ. $b(\ell)$ represents the set of locations sharing a boundary with location ℓ. The term $\delta_{0i} - \delta_{1i}\ 1\{a_j = \ell\} - \delta_{2i} \sum_{\ell' \in b(\ell)} 1\{a_j = \ell'\}$ is the variable profit per-potential-customer for firm i, where δ_{0i}, δ_{1i}, and δ_{2i} are parameters that capture the effect of competition. α_i is also a parameter. Finally, $\varepsilon_i = \{\varepsilon_{i\ell} : \ell = 0, 1, ..., L\}$ is a vector of private information variables of firm i at every possible location and it is i.i.d. over firms and locations with a type 1 extreme value distribution.

Given the vector of structural parameters of the model, $\theta \equiv (\alpha_i,\ \delta_{0i},\ \delta_{1i},\ \delta_{2i},\ \alpha_j,\ \delta_{0j},\ \delta_{1j},\ \delta_{2j})'$, and the "landscape" of the exogenous variables over the city locations, $\mathbf{X} \equiv \{POP_\ell,\ RENT_\ell : \ell =$

$1, 2, ..., L\}$, let $P_{i\ell}(\mathbf{X}, \theta)$ be the probability that firm i enters in location ℓ, i.e., $P_{i\ell}(\mathbf{X}, \theta) = \Pr(a_{i\ell} = 1 \mid \mathbf{X}, \theta)$. And let $\mathbf{P}_i(\mathbf{X}, \theta)$ be the "landscape" of entry probabilities over the L city locations for firm i, i.e., $\mathbf{P}(\mathbf{X}, \theta) \equiv \{P_{i\ell}(\mathbf{X}, \theta) : \ell = 1, 2, ..., L\}$. Given $(\mathbf{X}, \theta)$, the pair of vectors of probabilities $\mathbf{P}_i(\mathbf{X}, \theta)$ and $\mathbf{P}_j(\mathbf{X}, \theta)$ can be defined as a Bayesian Nash Equilibrium of this model.

Question 5.1 [15 points]. Obtain the expression for the expected profit of a potential entrant in location ℓ, obtain the best response probability for entry in location ℓ, and the equilibrium mapping in probability space. Describe $\{\mathbf{P}_i(\mathbf{X}, \theta)$,$\mathbf{P}_j(\mathbf{X}, \theta)\}$ as a fixed point of this equilibrium mapping.

Question 5.2 [10 points]. Suppose that we have cross-sectional data from M cities. For each city, we distinguish 9 geographic regions as in the figure above. Suppose that we observe the entry and location decisions of firms i and j in these M cities:

$$\text{Data } = \{\ POP_{m\ell},\ RENT_{m\ell},\ a_{im},\ a_{jm} : \ell = 0, 1, ..., L;\ m = 1, 2, ..., M\}$$

where we index cities with m. Obtain the expression of the likelihood function (or correspondence) for this model and data.

Question 5.3 [15 points]. Suppose that we treat firms' beliefs about the probabilities of entry of the other firm as incidental parameters. Let the vector of probabilities $\mathbf{B}_i(\mathbf{X}_m) \equiv \{B_{i\ell}(\mathbf{X}_m) : \ell = 0, 1, ..., L\}$ represent firm i' beliefs about the probability of entry of firm j at the different locations of city m. Treating $\mathbf{B}_i(\mathbf{X}_m)$ and $\mathbf{B}_j(\mathbf{X}_m)$ as a vectors of parameters, obtain the expression for the (pseudo) likelihood function $Q(\theta, \mathbf{B}_i, \mathbf{B}_j)$ for the data and model where the choice probabilities in this likelihood are best responses to the beliefs $(\mathbf{B}_i, \mathbf{B}_j)$.

Question 5.4 [10 points]. Show that under the assumption of rational beliefs, we can obtain Nonparametric Reduced Form estimates of firms' beliefs $\mathbf{B}$. Given this consistent estimator of beliefs, propose a two-step consistent estimator of the vector of structural parameters θ.

Question 5.5 [50 points]. The STATA datafile `eco2901_problemset_01_2013.dta` contains a cross-sectional dataset as the one described in Question 5.2 for $M = 1,000$ cities or metropolitan areas.

(a) Use these data to obtain a reduced for estimator of the CCPs $\{P_{i\ell}\}$ using a McFadden's Conditional Logit model.
(b) Using the reduced form estimates in (a), obtain a two-step estimator of the vector of structural parameters θ.
(c) Interpret the results.