No. 3

OCTUBRE de 2019

quantil | Monografías

Edición electrónica

Structural Demand Models in Empirical Industrial Organization: Theory, Estimation, and Programming

Natalia Serna

Juan D. Martin

Álvaro J. Riascos

quantil | matemáticas aplicadas

Serie de Monografías Quantil, 2019-20
Edición electrónica.

OCTUBRE de 2019

Carrera 7 # 77 - 07. Oficina 901, Bogotá, D. C., Colombia
Teléfonos: 3718132 – (310)6791459 – (320)8461236
E-mail: info@quantil.com.co
http://www.quantil.com.co

Impreso en Colombia – Printed in Colombia

Structural Demand Models in Empirical Industrial Organization: Theory, Estimation, and Programming

Natalia Serna* Juan David Martin† Álvaro J. Riascos‡

"... it is the combination of an economic assumption and statistical assumptions that distinguishes a structural model from a descriptive one." (Heckman and Leamer, 2007)

*PhD Student at University of Wisconsin-Madison. Former member of Quantil. E-mail: nserna@wisc.edu
†Former member of Quantil. E-mail: juandavid.martin@quantil.com.co
‡Quantil - Universidad de los Andes. E-mail: alvaro.riascos@quantil.com.co

Contents

1 Introduction

This document presents a detailed description of the programming and estimation of structural demand models in economics. Applications have grown exponentially since they were first introduced in the 90s. In particular, they have grown popular in marketing because they allow estimating cross-price elasticities, in policy evaluation because they allow conducting counterfactual scenarios, and in finance because they allow measuring the effect of competition on risk hedging. Structural demand models are called "structural" mainly for two reasons: one is that they rely on the theoretical fundamentals of microeconomics, namely the existence of utility and production functions; and the second is that the parameters these models allow to estimate for describing preferences and technology are not expected to vary over the short-term (Berry et al., 1995). The estimation techniques usually employed for these models are flexible enough to allow for different functional forms in both the demand and the supply side. However, a good specification for the model requires a proper understanding regarding consumer choice over products (whether such choices are static or dynamic), imperfect competition in the market (whether firms compete in prices or quantities) and how products interact with each other according to the consumer preferences (whether they are homogeneous, vertically differentiated, or horizontally differentiated).

In this document we focus on the description of static choices in markets where firms compete by setting the prices that maximize their profits in markets with vertically and horizontally differentiated products. Although extensions to the other cases can be easily done.

The field of empirical industrial organization emerged from various needs. Modeling imperfect competition among firms, both from the theoretical and the empirical perspectives, started to be crucial as researchers began to study the structure of industries. Theoretical approaches were developed earlier than empirical ones thanks to advances in game theory and data limitations. Sources of imperfect competition such as barriers to entry, asymmetric information, product differentiation, and pricing strategies were introduced in the mid 1980s by the works of Eaton and Lipsey (1980), Stigler (1983), Tirole (1988), and Stiglitz (1977). However, empirical approaches linking theoretical models to data were not introduced until the late 1980s with works such as the ones developed by Strickland (1976), Martin (1979), Geroski (1982), and Caves et al. (1980), which mainly relied in -allegedly- exogenous industry cross-variation to identify casual effects of imperfect competition on outcomes measures such as profits, productivity, prices or costs. However, arguments about the exogeneity of outcome measures to cross-industry differences were not compelling. Schmalensee (1989), for example, argued:

> "[...] in the long-run equilibria with which cross-section studies must be primarily concerned, essentially all variables that have been employed in such studies are logically endogenous. This means that there are in general no theoretically exogenous variables that can be used as instruments to identify and estimate any structural equation".

It is important to notice that economists have long been motivated by the challenge of identifying causal effects and computing counterfactual scenarios. This motivation eventually lead, in the late 1980s, to what Bresnahan (1989) coined as the field of "New Empirical Industrial Organization" (NEIO). Such field started to be a reference for studies dedicated to the modeling of within-industry differences for explaining and testing hypotheses of consumer and firm behavior. Studies in the NEIO field were also developed thanks availability of new data sets. Researchers, however, were still concerned with identifying the structural parameters of theoretical models, which they argued remained unchanged when computing counterfactual scenarios.

During the 1990s, studies proposing estimation techniques and identification strategies of structural models proliferated Berry et al. (1995, 1999) as well as the number of applications in different markets from the ready-to-eat cereal market (Nevo, 2001) to the telecommunications market (Kim, 2006). Almost every industry with an imperfect competition setting was subject of study by NEIO researchers. Nevertheless, detractors grew large in number at the same time. The estimation and identification of most structural empirical models were only possible under strong assumptions which constrained the set of attainable equilibria that ware intended to be linked to the data afterwards (Caplin and Nalebuff, 1991; Chintagunta et al., 2006). Critics also argued heavily on the concern that numerical approaches for estimation could often lead to different results in the counterfactual analyses, which in turn, would depend on the convergence thresholds set by the researcher or on the algorithm used to compute the equilibrium. This raised many questions regarding how structural the estimations really were and how could results be extrapolated (Brown and Walker, 1989).

Such criticisms have forced NEIO researchers to begin relaxing several assumptions on which their models rely and studying the consequences of using different numerical algorithms for estimation (Knittel and Metaxoglou, 2008; Dubé et al., 2009a). The most difficult criticisms to address are those regarding dynamic structural models, since there are infinite number of equilibria from which it is crucial to select one for estimation to be feasible. This is where the knowledge frontier in NEIO is currently standing (Nevo and Rosen, 2012; Ho and Lee, 2013; Gowrisankaran et al., 2014). For further description on what has lead NEIO to its current state of art, the reader can refer to Einav and Levin (2010).

In the folloeing sections of this document we focus on the description of structural models, their estimation techniques and identification strategies as proposed first by Berry et al. (1995). We also describe how different specifications lead to different conclusions and then extend to how such models are programmed in several object-oriented statistical software.

2 Economic assumptions

To understand how statistical approaches link theoretical models to the data in the NEIO, first we have to go back to the basics of economic models of imperfect competition.

2.1 Imperfect competition in the supply side: simultaneous production and pricing decisions

Production decisions

Assume there are F firms competing setting quantities in an homogeneous product market. The problem of firm f is to maximize its profits function which is defined as the difference between its own revenue and costs:

$$\max_{q_f} \quad \pi_f = P(Q)q_f - C(q_f), \tag{1}$$

where $P(Q)$ is the inverse demand function and $Q \equiv \sum_{f=1}^{F} q_f$. Under the assumption that the profit function is concave in quantity ($\partial^2 \pi_f / \partial^2 q_f < 0$), the solution to the firm's problem is equivalent to computing the level of production that equals marginal income to marginal cost. Notice, on the one hand, that producing above such level would make each additional unit to cost more it than what the firm can obtain from selling it in the market, hence profits can increase by reducing production. On the other hand, producing below such level would make any additional unit to be more profitable in the margin so that the firm could achieve higher profits by selling more units as a whole. Therefore, the production level that maximizes profits is achieved when:

$$P(Q) + \frac{\partial P}{\partial Q}\frac{\partial Q}{\partial q_f} q_f = \frac{\partial C}{\partial q_f}. \tag{2}$$

More generally, marginal revenue, the left-hand side of equation (2), can be rewritten as:

$$MR_f = P(Q) + \theta \frac{\partial P}{\partial Q} q_f \quad \text{for} \quad \theta = \frac{\partial Q}{\partial q_f}, \tag{3}$$

where θ is a parameter that defines the firms' behavior which, in turn, describes the competitive structure of the market. For instance, $\theta = 0$ denotes the case perfect competition. We can also express θ equivalently as:

$$MR_f = P(Q) + \left(\theta_1 + \theta_2 \frac{1}{s_f}\right) \frac{\partial P}{\partial Q} q_f, \tag{4}$$

where $s_f = q_f/Q$. When $\theta_1 = \theta_2 = 0$, price equals marginal cost in the optimal production level and firms exhibit no market power; when $\theta_1 = 1$ and $\theta_2 = 0$ the first order condition denotes the standard Cournot model solution with firms competing simultaneously via quantities; and when $\theta_1 = 0$ and $\theta_2 = 1$, then $q_j = Q$ and there is market power equivalent to a monopoly structure.

Overall, notice that markups, which we denote by b_f, in the first order condition of the firm's problem is a function that depends only on the demand parameters: marginal income which is the derivative of the inverse demand function and units sold or the respective market shares.

$$b_f = P(Q) - \frac{\partial C}{\partial q_f} = -\theta \frac{\partial P}{\partial Q} q_f. \tag{5}$$

Pricing decisions

Now consider the case in which firms are choosing prices instead of quantities so that their profit maximization problem is as follows:

$$\max_{P_f} \pi_f = P_f q(P_f, P_{-f}) - C(q_f(P_f, P_{-f})), \tag{6}$$

where P_{-f} denotes the vector of prices set by all the firms other than f. The first order condition with respect to P_f, after rearranging terms, yields to the pricing equations of the Nash-Bertrand equilibrium:

$$P_f - \frac{\partial C}{\partial q_f} = -\frac{1}{\partial q_f / \partial P_f} q_f(P_f) = -\frac{\partial P_f}{\partial q_f} q_f(P_f). \tag{7}$$

As in the competition by quantities game, we can rewrite the expression above introducing term θ:

$$P_f - \frac{\partial C}{\partial q_f} = -\theta \frac{\partial P_f}{\partial q_f} q_f(P_f). \tag{8}$$

In this case, when $\theta = 0$, there is no product differentiation and we fall in the Bertrand's paradox with price being equal to marginal cost. But for $\theta > 0$ firms can exert market power through product differentiation charging positive markups. Markups in this case are a function of the same demand parameters as in the case of competition via quantities. So lets turn to the demand side for understanding of such parameters.

2.2 Consumer theory and demand derivation

2.2.1 Traditional approaches

Assume consumer preferences are rational and, thus, can be represented by an utility function. Consumer theory states that consumer i's demand for two goods, x_1 and x_2, with prices P_1 and P_2, respectively, is determined by her utility maximization problem subject to a budget constraint:

$$\max_{x_1,x_2} \; U_i(x_1, x_1) \;\; s.t \;\; Y_i = P_1 x_1 + P_2 x_2, \tag{9}$$

where Y_i is i's income. The solution to this problem yields the Marshallian demands which depend on income and prices, namely $x_1^*(P_1, P_2, Y_i)$ and $x_2^*(P_1, P_2, Y_i)$. Equivalently, Marshallian demands can be obtained from the derivatives of the indirect utility function $V_i^*(P_1, P_2, Y_i)$ with respect to prices.[1] This approach to demand derivation is often more convenient for empirical studies as the indirect utility function also allows us to obtain the share of each good on consumer i's total expenditure (or income) using the logarithmic version of Roy's identity:

$$s_j(P_1, P_2, Y_i) = -\frac{\partial \log(V_i^*(P_1, P_2, Y_i))/\partial \log(p_j)}{\partial \log(V_i^*(P_1, P_2, Y_i))/\partial \log(Y_i)} \;\; \text{for } j = 1, 2.$$

For example, assume i's utility is for consuming goods 1 and 2 can be represented by Cobb-Douglas function. In this case, the log of the indirect utility function will have a translog functional form as below:

$$\log(V_i(P_1, P_2, Y_i)) = \alpha_0 + \sum_{j=1,2} \alpha_j \log\left(\frac{P_j}{Y_i}\right) + \frac{1}{2} \sum_{j=1,2} \sum_{k=1,2} \beta_{jk} \log\left(\frac{P_j}{Y_i}\right) \log\left(\frac{P_k}{Y_i}\right),$$

where j and k denote goods' subindices. Given this indirect utility function, the share of good j on total consumer's i expenditure can be computed as:

$$w_j(P_1, P_2, Y_i) = \frac{\alpha_j + \sum_{j=1,2} \beta_{jk} \log\left(\frac{P_k}{Y_i}\right)}{\sum_{j=1,2} \alpha_j + \sum_{j=1,2} \sum_{k=1,2} \beta_{jk} \log\left(\frac{P_k}{Y_i}\right)}.$$

[1] In microeconomic theory this is known as Roy's identity.

Using the estimation methods already developed by the NEIO literature and assuming the availability of data on prices, income, and sales, we can obtain parameter estimates for the α's and β's. In particular, such estimates can be computed using only the shares function of one of the goods since shares add up to 1. Generalizing to the case of J goods, this means we would have to estimate $J-1$ equations and $J(J+1)$ parameters. Therefore, the estimation exercise becomes problematic as the number of goods increases. For example, having 100 products would require estimation of 100 cross price elasticities and one income elasticity for each of 100 products. In markets with highly differentiated products this is computationally infeasible. Hence, the need to use another identification strategy for deriving demand from consumer theory.

Aggregation of goods that permits finding reasonable structural parameters is one way of addressing the curse of dimensionality. Hedonic analysis and discrete choice models are some methodologies usually employed to derive reasonable substitution patterns and shares functions. Below we focus on the discrete choice case.

2.2.2 Discrete choice approach

Consider a mass of N consumers, each of them choosing among $j = 1, \ldots, J$ products to buy in a given market. Consumer i's indirect utility function satisfies all rationality conditions and is given by the following Cobb-Douglas specification:

$$U_{ij} = G(Y_i, X_j, P_j)e^{\epsilon_{ij}}, \tag{10}$$

where X_j is a vector describing product j's characteristics, P_j is the price of product j, Y_i is consumer i's income, ϵ_{ij} is a random shock to the utility function, and G is a function linear in logs. Taking logs on both sides of this equation, we can rewrite the indirect utility function as:

$$u_{ij} = \gamma Y_i + \beta X_j + \alpha P_j + \epsilon_{ij}, \tag{11}$$

where $\ln(U_{ij}) \equiv u_{ij}$ and $\{\gamma, \beta, \alpha\}$ are the parameters describing preferences. Consumers choose product j over all other alternatives if $u_{ij} \geq u_{ik}$ for all$j \neq k$, that is, when:

$$u_{ij} - u_{ik} = \beta X_j + \alpha P_j + \epsilon_{ij} - (\beta X_k + \alpha P_k + \epsilon_{ik}) \geq 0. \tag{12}$$

Therefore, the probability of consumer i choosing product j can be expressed as:

$$Prob[u_{ij} - u_{ik} \geq 0] = Prob[\epsilon_{ik} - \epsilon_{ij} \leq \delta_j - \delta_k], \tag{13}$$

where δ_j is denoted as the mean utility associated with product j.

If we have an assumption on the probability distribution of ϵ we can derive choice probabilities for each consumer i as:

$$s_{ij} = \int_{\epsilon} dP_{\epsilon}. \tag{14}$$

Notice that the discrete choice setting only requires estimation of α, β, and γ, which represents a significant reduction on the dimentionality of the problem. However, it is not clear whether the discrete choice approach and the choice probabilities implied by the model are consistent with the theory of utility maximization from which demands are obtained. Theorem I in McFadden (1978) states:

> Suppose $G(y_1, ..., y_J)$ is non-negative, homogenous-of-degree-1 function of $(y_1, ..., y_J) > 0$. Suposse $\lim_{y_i \to \infty} G(y_1, ..., y_J) = +\infty$ for $i = 1, ..., J$. Suppose for any distinct $(i_1, ..., i_k)$ from $\{1, ..., J\}$, $\partial^k G / \partial y_{i_1} ... \partial y_{i_k}$ is non-negative if k is odd and non-positive if k is even. Then $P_i = e^{V_i} G_i(e^{V_1}, ..., e^{V_J}) / G(e^{V_1}, ..., e^{V_J})$ defines a probabilistic choice model from alternatives $i = 1, ..., J$ which is consistent with utility maximization.

To prove this theorem, following McFadden (1978), first of all we have to show function $F(\varepsilon_1, ..., \varepsilon_J) = e^{-G(e^{-\varepsilon_1}, ..., e^{-\varepsilon_J})}$ is a multivariate extreme value distribution, or the distribution of the maximum values. Then, showing the choice probabilities generated by such distribution correspond to $P_i = e^{V_i} G_i(e^{V_1}, ..., e^{V_J}) / G(e^{V_1}, ..., e^{V_J})$, where V_i is the utility, would prove P_i are the choice probabilities that maximize the utility function for each alternative when ε

follows an extreme value type I distribution.

If $\varepsilon_i \to -\infty$, then $e^{\varepsilon_i} \to +\infty$, $G(\cdot) \to +\infty$ and $F \to 0$. If $(\varepsilon_1, ..., \varepsilon_J) \to +\infty$, then $(e^{\varepsilon_1}, ..., e^{\varepsilon_J}) \to 0$, $G(\cdot) \to 0$ and $F \to 1$. Differentiating F we have:

$$\frac{\partial F}{\partial \varepsilon_1} = e^{-G(e^{-\varepsilon_1},...,e^{-\varepsilon_J})}\frac{\partial G}{\partial \varepsilon_1}e^{-\varepsilon_1} = Fe^{-\varepsilon_1}\frac{\partial G}{\partial \varepsilon_1}.$$

A cross derivative of F, after rearranging terms, would be given by:

$$\begin{aligned}\frac{\partial^2 F}{\partial \varepsilon_1 \partial \varepsilon_2} &= e^{-\varepsilon_1}\left[F\frac{\partial^2 G}{\partial \varepsilon_1 \partial \varepsilon_2} + \frac{\partial F}{\partial \varepsilon_2}\frac{\partial G}{\partial \varepsilon_1}\right] \\ &= e^{-\varepsilon_1}e^{-\varepsilon_2}e^{-G(e^{-\varepsilon_1},...,e^{-\varepsilon_J})}\left[\frac{\partial G}{\partial \varepsilon_1}\frac{\partial G}{\partial \varepsilon_2} - \frac{\partial^2 G}{\partial \varepsilon_1 \partial \varepsilon_2}\right]\end{aligned} \tag{15}$$

Now, let $Q_1 = G_1 = \frac{\partial G}{\partial \varepsilon_1}$ and $Q_k = Q_{k-1}G_k - \partial Q_{k-1}/\partial Y_k$, then $\partial F/\partial \varepsilon_1 = FQ_1e^{-\varepsilon_1}$ and

$$\frac{\partial^{k-1} F}{\partial \varepsilon_1 ... \partial \varepsilon_{k-1}} = e^{-\varepsilon_1}e^{-\varepsilon_2}...e^{-\varepsilon_{k-1}}Q_{k-1}F.$$

Notice if Q_{k-1} is non-negative, given G_k is non-negative by assumption, then $Q_{k-1}G_k$ is non-negative. Each term in $\partial Q_{k-1}/\partial Y_k$ is non-positive because one term in the derivative has increased in order, therefore:

$$\frac{\partial^{k-1} F}{\partial \varepsilon_1 ... \partial \varepsilon_{k-1}} > 0,$$

and F is a cumulative distribution. Now consider the case $i = 1, 2$ and $\varepsilon_2 = +\infty$, then given the homogeneity of degree 1 of G:

$$F = e^{-G(e^{-\varepsilon_1}, e^{-\varepsilon_2})} = e^{-G(e^{-\varepsilon_1}, 0)} = e^{-G(e^{-\varepsilon_1})} = e^{-a_1 e^{-\varepsilon_1}},$$

where $a_1 = G(1, 0)$. More generally, for $j \neq i$ if $\varepsilon_j = +\infty$, then:

$$F = e^{-a_i e^{-\varepsilon_i}},$$

where $a_i = G(0, ..., 0, 1, 0..., 0)$ with 1 in the i-th place and this is the univariate extreme value distribution. Hence, F is a multivariate extreme value distribution. Having proved F denotes the cumulative distribution of maximum values, we turn to the probabilities generated by this distribution.

Assuming the utility function is $u_i = V_i + \varepsilon_i$ where $(\varepsilon_1, ..., \varepsilon_J)$ is distributed F. The probability of choosing the first alternative is:

$$P_1 = \int_{\varepsilon=-\infty}^{+\infty} F_1(\varepsilon, V_1 - V_2 + \varepsilon, ..., V_1 - V_J + \varepsilon) d\varepsilon$$
$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} Q_1 F d\varepsilon$$
$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} G_1(e^{-\varepsilon - V_1 + V_2}, ..., e^{-\varepsilon - V_1 + V_J}) e^{-G(e^{-\varepsilon}, e^{-\varepsilon - V_1 + V_2}, ..., e^{-\varepsilon - V_1 + V_J})} d\varepsilon$$

Recall $G(y_1, ..., y_J)$ is an homogeneous-of-degree-1 function of $(y_1, ..., y_J)$ by assumption and $therefore G_k(y_1, ..., y_J)$ is an homogeneous-of-degree-0 function of $(y_1, ..., y_J)$, then

$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} \left(e^{-\varepsilon}\right) G_1(e^{V_1}, e^{V_2}, ..., e^{V_J}) e^{-\left(e^{-\varepsilon} e^{-V_1}\right) G(e^{V_2}, ..., e^{V_J})} d\varepsilon$$
$$= \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} G_1(e^{V_1}, e^{V_2}, ..., e^{V_J}) e^{-\left(e^{-\varepsilon} e^{-V_1}\right) G(e^{V_1}, e^{V_2}, ..., e^{V_J})} d\varepsilon$$
$$= e^{V_1} G_1(e^{V_1}, e^{V_2}, ..., e^{V_J}) / G(e^{V_1}, e^{V_2}, ..., e^{V_J}). \tag{16}$$

Since the expression for the choice probability above can be done for every alternative and equals the probability in the theorem statement, then it is proved that choice probabilities generated by the extreme value distribution are consistent with the theory of utility maximization.

Knowing utility is maximum under the assumption ε follows a extreme value type I distribution, we can aggregate over the distribution of consumer characteristics (in this case, income) to find product's j market share.

$$s_j = \int_y s_{ij} dP_y. \tag{17}$$

These market shares and their derivatives with respect to prices enter the first order conditions of the firms' profit maximization problem, closing the simultaneity of supply and demand.

3 Structural demand models

Structural demand models in NEIO are designed to characterize imperfectly competitive markets. Although, there are many models we do not address in this document (e.g., auctions models), we focus on the ones that rely on the economic assumptions outlined in the previous section to attain identification strategies and estimate the first order conditions of the firms' profit maximization problem and the consumer's indirect utility function. Estimation strategies depend on the type of data that is available for the researcher. Since micro data or consumer-level data is usually unavailable, most estimation techniques have been developed in the context of aggregated or product-level data. To our knowledge, the most main article in this subject is the one by Steven Berry, James Levinsohn, and Ariel Pakes (BLP) published in 1995 in Econometrica. This study sets the base for estimation of a random coefficients discrete choice model that maps consumer's decision choices and firms' pricing strategies to product-level data. This section is dedicated to the description and analysis of the model proposed by BLP.

Consumer i's utility after choosing the product j available in market t be defined as:

$$u_{ijt} = \alpha_i f(y_{it} - p_{jt}) + \mathbf{x_{jt}}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{18}$$

where y_{it} is the consumer's income, p_j is the price of the product and $\mathbf{x_{jt}}$ and ξ_{jt} denote, respectively, observable and unobservable (by the researcher) characteristics of the product. The term ε_{ijt} represents the random shock to her utility which is assumed to follow a Type I Extreme Value distribution.

The assumed specification for $f(\cdot)$, which determines how available income[2] is introduced in the indirect utility, can take any functional form that maintains the regularity assumptions of consumer preferences. This specification choice, however, has important consequences on how income determines consumer choices. Consider the case in which available income enters the indirect utility function linearly as $f(y_{it} - p_{jt}) = y_{it} - p_{jt}$. As we show later, income in this case will have no linear effect that can be distinguished from the model's constant and

[2]Defined as the difference between income and price

hence vanishes from the utility specification. The economic intuition, on the other hand, suggests that the vanishing linear effect of income means product market shares are not linearly increasing in income. In other words, income does not shift the indifference curve to the right. However, if product observable characteristics are allowed to interact with income, then income does affect the slope of the indifference curve and moves the optimal product choice along the same level of indirect utility. Now consider the case available income enters the indirect utility function in logarithms as $f(y_{it} - p_{jt}) = log(y_{it} - p_{jt})$. Income effects in this case do not vanish from the utility specification and will both shift the indifference curve and change the optimal product choice.

Deciding on the functional form of the available income depends on the market being analyzed. If the researcher believes income effects are important when choosing a product from the market, then a functional form where income does not vanishes from the indirect utility level should be considered. This might be the case of markets such as the automobiles market, computers market, digital cameras market, etc. On the other hand, if income effects are really not as important as perhaps other consumer characteristics then it would be suitable to specify a linear functional form. This might be case in markets such as the ready-to-eat cereal market or the carbonated soft drinks market.

The distribution of coefficients over consumers in the indirect utility function is captured with interactions between the consumers' observable and unobservable characteristics with product prices. These interactions make up the random coefficients discrete choice model with which we began this section. The decision on which product observable traits to interact with consumer characteristics is also up to the researcher. In our specification all marginal utilities are allowed to vary over consumers (α_i, β_i), but it is also common to estimate only the mean price marginal utility $(\overline{\alpha})$ and leave the rest to vary over consumers. Specifically, random coefficients are defined as follows:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \overline{\alpha} \\ \overline{\beta} \end{pmatrix} + (\Sigma v_i + \Pi D_{it}), \tag{19}$$

where D_{it} is a $d \times N$ matrix of demographic variables, v_i is a mean zero normal variable of dimensions $(K+1) \times N$ with identity covariance matrix, Σ is a $(K+1) \times (K+1)$ matrix of coefficients, and Π is a $(K+1) \times d$ vector of coefficients, where N is the number of simulated consumers, K the number of observable characteristics in x_{jt} and d the number of demographic characteristics. Individual heterogeneity generates a distribution for each parameter characterized by a mean effect (first term of the right-hand side of equation (19)) and a deviation from the mean which can be due to differences in demographic characteristics or to the variance of the standard normal effect (second term of the right-hand side of equation (19)).

We can rewrite the indirect utility function as the sum of characteristics that vary over terminals and their interactions with consumer traits, as:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \tag{20}$$

where mean utilities are:

$$\begin{cases} \delta_{jt} = \mathbf{x}_{jt}\overline{\beta} - \overline{\alpha} p_{jt} + \xi_{jt} & \text{if price is linearly separable from income} \\ \delta_{jt} = \mathbf{x}_{jt}\overline{\beta} + \xi_{jt} & \text{otherwise} \end{cases} \tag{21}$$

And deviations from the mean utilities are:

$$\begin{cases} \mu_{ijt} = & p_{jt}(\sigma_p v_{ip} + \pi_p^1 D_{i1t} + \cdots + \pi_p^d D_{idt}) \\ & + \sum_{k=1}^{K} x_{jkt}(\sigma_k v_{ik} + \pi_k^1 D_{i1t} + \cdots + \pi_k^d D_{idt}) \quad \text{if price is linearly separable from income} \\ \mu_{ijt} = & f(y_{it} - p_{jt})(\overline{\alpha} + \sigma_p v_{ip} + \pi_k^1 D_{i1t} + \cdots + \pi_k^d D_{idt}) \\ & + \sum_{k=1}^{K} x_{jkt}(\sigma_k v_{ik} + \pi_k^1 D_{i1t} + \cdots + \pi_k^D D_{iDt}) \quad \text{otherwise} \end{cases} \tag{22}$$

Besides the set of interior products ($j = 1, ..., J$), consumers can choose an *outside option* ($j = 0$) which represents the possibility of not buying any of the interior products. Its indirect utility level is given by:

$$u_{i0t} = \alpha_i f(y_{it}) + \xi_{0t} + \sigma_0 v_{i0} + \varepsilon_{i0t}. \tag{23}$$

The existence of term $\sigma_0 v_{i0}$ means we are allowing for the possibility of there being more unobserved variance in the outside other than the inside alternatives. Also, since $\alpha_i f(y_{it})$ is common to all models and can not be identified from the model's constant, this is the same as normalizing the outside option's mean utility to zero.

Consumer i chooses product j over all other available alternatives if $u_{ijt} \geq u_{ilt} \quad \forall j \neq l$. Assuming, first, that there are no ties in utility; second, that consumers choose only one option from the available set; and, third, that ε follows an extreme value type I distribution. Then we can condition on the distribution of v_{ik} and D_{it} to integrate out the extreme value distribution and obtain choice probabilities in a closed form fashion as in equation (24) below:

$$s_{ijt}(x_{jt}, p_{jt}, \xi_{jt}, P_v, P_D; \theta) = \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{l=1}^{J} \exp(\delta_{lt} + \mu_{ilt})}, \tag{24}$$

where P_v denotes the distribution of v_{ik} and P_D the distribution of demographic characteristics. Now let A_j be the set of consumers that choose product j. Market shares are the aggregation of choice probabilities over A_j as equation (25) shows.

$$s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_v, P_D; \theta) = \int_{A_j} s_{ijt} dP_v dP_D. \tag{25}$$

Even if income and price are not linearly separable, one can assume income draws are lognormal and express their distribution as $\exp(m_t + \sigma_t^y v_{it}^y)$ where m_t is the average income in market t, σ_t^y is the standard deviation, and $v_{it}^y \sim N(0, 1)$. The two parameters of the distribution of income draws can be estimated, for example, from household surveys. Therefore, the only variable that generates variation across consumers is the standard normal effect v_{it}^y which we include in P_v assuming $v = (v_{i1}, v_{i2}, ..., v_{ik}, v_i^y)$ are independent. This means equation (25) serves for the cases where income and price are linearly separable and where they are not.

Recall from equation (7), the second parameter of demand that enters the pricing equations in the first order condition of the firm's problem are the derivatives of demand with respect to prices. The actual demand for product j in market t, q_{jt}, can be expressed as the product of its market share s_{jt} and the size of the potential market M_t:

$$q_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_v, P_D; \theta) = M_t s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_v, P_D; \theta). \tag{26}$$

Thus, derivatives of demand with respect to prices are the same as derivatives of market shares with respect to prices. Even if the joint estimation of supply and demand is not carried out, estimating price derivatives is fundamental to obtain the matrix of cross price elasticities. The price derivatives of market shares are shown in equation (27). The first line of the equation shows the own-price derivatives and the second line the cross-price derivatives. Own-price derivatives (in absolute value) are a measure of the amount of consumers that substitute away from product j when its price increases. Multiplying these expressions by p_{jt}/s_{jt} yields price elasticities:

$$\begin{aligned} \frac{\partial s_{jt}}{\partial p_{jt}} &= -\int_{A_j} s_{ijt}(1 - s_{ijt})(d\mu_{ijt}/dp_{jt})dP_v dP_D \\ \frac{\partial s_{jt}}{\partial p_{lt}} &= \int_{A_j} s_{ijt}s_{ilt}(d\mu_{ijt}/dp_{lt})dP_v dP_D \end{aligned} \tag{27}$$

Up to this point, estimation of the structural model is equivalent to estimating $\overline{\alpha}, \overline{\beta}, \Sigma, \Pi, \xi_{jt}$. If we introduce the supply side, the number of parameters to estimate increases but so does the observed variability used for identification.

After estimating price derivatives and market shares, we can recover marginal costs for each firm in the market as explained below. Assume there are F firms in the market, each producing a subset F_f of the set of interior products. The firms' problem is:

$$\max_{p_{jt}} \sum_{j \in F_f} (p_{jt} - mc_{jt}) M_t s_{jt}, \tag{28}$$

which yields the next first order condition:

$$s_{jt} + \sum_{r \in F_f} (p_{rt} - mc_{rt}) \frac{\partial s_{rt}}{\partial p_{jt}} = 0. \tag{29}$$

Let,

$$\Omega = \left\{ \begin{array}{ll} -\frac{\partial s_{rt}}{\partial p_{jt}} & \text{if } f \text{ produces both } r \text{ and } j \\ 0 & \text{o.w} \end{array} \right\}. \tag{30}$$

Rewriting the first order conditions in matrix form and rearranging terms, we obtain an expression for price markups, which depends only on demand parameters already derived from the model (price derivatives and market shares):

$$b = p - mc = \Omega^{-1} s_{jt}. \tag{31}$$

Therefore, marginal costs can be recovered from the model as the difference between observed prices and estimated markups, and then we can estimate a cost function for each producer as in equation (32) below:

$$mc_{jt} = p_{jt} - b_{jt} = \mathbf{w_{jt}}\gamma + \omega_{jt} \tag{32}$$

Any functional form for the marginal cost is also attainable if certain conditions hold like cost convexity. One could be interested, for instance, in estimating a logarithmic regression. Other specifications of the cost function like allowing for non-constant returns to scale or capturing risk aversion are yet to be studied.

3.1 A note on price elasticities and substitution to the outside option

To obtain own-price elasticities we should multiply the first line of equation (27) by p_{jt}/s_{jt} and to obtain cross-price elasticities we multiply the second line of the same equation by p_{kt}/s_{jt}. The resulting matrix with elements (i, j) where i denotes the row and j the column will indicate the percentage variation in the demand for good i when the price of good j increases by 1%. However a 1% increase in the price of good j can be higher or lower in magnitude than a 1% increase in the price of another good, making price elasticities difficult to compare among products. One way of accounting for differences in magnitude is reporting price semi-elasticities, which indicate the percentage variation in the demand for a good when its own price or the price of other goods increase by x dollars. Obtaining price semi-elasticities consists of multiplying each line of equation (27) by x/s_{jt}. So, for example, if prices are measured in thousand dollars and $x = 2$, then each element of the resulting price semi-elasticities matrix should be interpreted as the percentage variation in demand when prices increase by 2,000 dollars.

The substitution to the outside option is also an important measure of demand sensitivity. Deriving reasonable substitution patterns to the outside is non trivial given we do not observe its price nor its characteristics. In fact since the observable characteristics of the outside option are assumed to be zero, it is not reasonable to obtain substitution patterns to the outside in the same fashion as with price elasticities between interior products. Berry et al. (1995) report an alternative measure of substitution to the option defined as:

$$\frac{100 \times (ds_{0t}/dp_{jt})}{|ds_{jt}/dp_{jt}|}. \tag{33}$$

This expression measures the number of consumers that substitute away to the outside option as a percentage of total consumers who substitute away from product j when its price increases. The numerator includes changes in the outside option's market share when price j changes and the denominator is the absolute value of the own-price derivative of good j. Given consumers are assumed to buy one unit of the product from which they derive the highest utility, equation (33) can be interpreted as the percentage of consumers that go to the outside option.

4 Estimation

The integral of choice probabilities over A_j has no analytic solution and therefore has to be solved numerically. Numeric aggregation consists of making ns random draws from the distribution of v and then averaging to obtain market shares such that $\lim_{ns\to\infty} P_{ns}(v) \to P_v(v)$, where P_{ns} is the numerical approximation. The usual approximation employed is the following:

$$s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}, P_D; \theta) \approx \frac{1}{ns} \sum_{i=1}^{ns} s_{ijt}. \tag{34}$$

It is worth mentioning also that the same numeric approximation has to be taken for the computation of price derivatives:

$$\begin{aligned} \frac{\partial s_{jt}}{\partial p_{jt}} &\approx -\frac{1}{ns}\sum_{i=1}^{ns} s_{ijt}(1-s_{ijt})(d\mu_{ijt}/dp_{jt}) \\ \frac{\partial s_{jt}}{\partial p_{lt}} &\approx \frac{1}{ns}\sum_{i=1}^{ns} s_{ijt}s_{ilt}(d\mu_{ijt}/dp_{lt}). \end{aligned} \tag{35}$$

Now, define $\theta_1 = (\overline{\alpha}, \overline{\beta})$ and $\theta_2 = (\Sigma, \Pi)$. Conditional on θ_2, estimation of the linear parameters in θ_1 consists of minimizing the distance between observed market shares and predicted market shares:

$$\min_{\theta_1} \quad ||s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}, P_D; \theta_2) - s_{jt}||. \tag{36}$$

Inversion of this minimization strategy conditional on θ_2, yields the contraction mapping suggested by BLP. This contraction mapping converges to a fixed point solution in the vector of average utilities that equal observed to predicted market shares:

$$\delta' = \log(s_{jt}) - \log(s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}, P_D; \theta)) + \delta. \tag{37}$$

Setting the seed for δ in the first iteration, subsequent iterations update this vector based on the difference between the log of observed market shares s_{jt} and the log of predicted market shares $s_{jt}(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}; \theta)$ until convergence. After obtaining the vector of mean utilities, the regression model in equation (21) allows us to obtain estimates for the mean marginal utilities in θ_1 associated to each product's observable characteristics as well as the demand-side unobservable ξ_{jt}. Even though product differentiation is captured by the observable characteristics $[p_{jt}, x_{jt}]$ and the unobservable ξ_{jt}, the only element that sets this model apart from the homogenous products case, conditional on $[p_{jt}, x_{jt}]$, is ξ_{jt}.

Using ordinary least squares for the regression model in (21) would result in biased estimates of θ_1 because price is an endogenous variable. For example, if residuals in ξ_{jt} capture product quality, which is unobserved by the researcher, then higher prices are correlated with higher values of ξ_{jt}. Or if marketing efforts, which increase prices, also increase consumer average utility, then we can expect $corr(\xi_{jt}, \delta_{jt}) > 0$. This suggests instruments for price (z_{jt}) must be used in order to obtain unbiased estimates for θ_1 and ξ_{jt} given the independence assumption below:

$$E[\xi_{jt}|z_{jt}] = 0. \tag{38}$$

BLP mention optimal price instruments are: z_{jt}, $\sum_{r\neq j, r\in\mathcal{F}_f} z_{rt}$, and $\sum_{r\neq j, r\notin\mathcal{F}_f} z_{rt}$. For instance, if horse power is one of the observable characteristics of vehicles in the automobile market, then optimal price instruments for vehicle j manufactured by firm f are: its horse power, the sum of the horse power of the rest of automobiles manufactured by f, and the sum of the horse power of all other automobiles not manufactured by f. Estimation techniques should also depart from the OLS case to the instrumental variable regression case or two-stage least squares.

After estimation of θ_1, estimates of θ_2 are obtained by minimizing a function of the model's errors. Notice these errors equal ξ_{jt} if only the demand-side is estimated, but equal $[\xi_{jt}, \omega_{jt}]$ if joint estimation of supply and demand is performed. For the latter case, ω_{jt} is obtained from the residuals of a regression of marginal costs on product observable characteristics and instruments (as in equation (32)) given the following independence assumption:

$$E[\omega_{jt}|z_{jt}] = 0 \tag{39}$$

Marginal costs, on their hand, are obtained as the difference between observed prices and markups computed from estimated market shares and price derivatives.

The description above makes it obvious that the distribution of ξ_{jt} and ω_{jt} in the population has the mean independent properties of (38) and (39). Moreover, assuming $E[(\xi_{jt}, \omega_{jt})'(\xi_{jt}, \omega_{jt})|z_{jt}]$ is finite, then θ_2 can be chosen as the value that minimizes the equivalent sample moment conditions up to a constant:

$$\theta_2 = arg\min \begin{pmatrix} \xi(\theta_2, P_{ns}, s_{jt}) \\ \omega(\theta_2, P_{ns}, s_{jt}) \end{pmatrix}' ZWZ' \begin{pmatrix} \xi(\theta_2, P_{ns}, s_{jt}) \\ \omega(\theta_2, P_{ns}, s_{jt}) \end{pmatrix}, \tag{40}$$

where Z is the matrix of instruments and W is a weighting identity matrix.

Joint estimation of supply and demand can be summarized in the following steps:

1. For a vector of θ_2:
   (a) Set the seed for δ and build the matrix for μ_{ijt}.
   (b) Compute the fixed point algorithm in (37) to obtain the vector of δ that equals observed to predicted market shares. In this step it is crucial to hold fixed the simulation draws from $P_{ns}(v)$ as the parameters change, or otherwise changes in the objective function would be due to changes in simulation.
   (c) Compute ξ_{jt} from a regression of δ on product observable characteristics and instruments.
   (d) Compute market shares as in equation (34) and price derivatives as in equation (35).
   (e) From predicted market shares and price derivatives, compute markups as in equation (31) and recover marginal costs.
   (f) Compute ω_{jt} from a regression of marginal costs on product observable characteristics and instruments.
2. Compute the value of the Generalized Method of Moments (GMM) function in equation (40).
3. Repeat (1) and (2) for different values of θ_2 until the GMM function is minimized.

If only the demand-side is to be estimated, then the process is the same as above except for step (f), and the GMM function would be $(Z'\xi(\theta_2, P_{ns}, s_{jt}))' \, W \, (Z'\xi(\theta_2, P_{ns}, s_{jt}))$.

4.1 The variance of the GMM estimator

The asymptotic variance-covariance matrix for the GMM estimator is (see Berry et al. (1995)):

$$\hat{V}(\hat{\beta}_{GMM}) = (D'ZZ'D)^{-1}(D'ZSZ'D)(D'ZZ'D)^{-1}, \tag{41}$$

where Z is the matrix of instruments and D is the Jacobian of the moment conditions with respect to θ_1 and θ_2. For the linear parameters in θ_1, the derivatives equal x_{jt} but for the nonlinear parameters in θ_2 derivatives must be calculated as:

$$\frac{\partial \delta_{jt}}{\partial \theta_2} = \left(\frac{\partial s_{jt}}{\partial \delta_{kt}}\right)^{-1} \left(\frac{\partial s_{jt}}{\partial \theta_2}\right), \tag{42}$$

where:

$$\begin{aligned}
\frac{\partial s_{jt}}{\partial \delta_{jt}} &= \frac{1}{ns} \sum_{i=1}^{ns} s_{ijt}(1 - s_{ijt}) \\
\frac{\partial s_{jt}}{\partial \delta_{lt}} &= \frac{1}{ns} \sum_{i=1}^{ns} s_{ijt} s_{ilt} \\
\frac{\partial s_{jt}}{\partial \sigma_k} &= \frac{1}{ns} \sum_{i=1}^{ns} v_{ik} s_{ijt} \left(x_{jkt} - \sum_{k=1}^{J} x_{jkt} s_{ijt} \right) \\
\frac{\partial s_{jt}}{\partial \pi_d} &= \frac{1}{ns} \sum_{i=1}^{ns} D_{idt} s_{ijt} \left(x_{jkt} - \sum_{k=1}^{J} x_{jkt} s_{ijt} \right)
\end{aligned} \quad . \tag{43}$$

the term S, on the other hand, is the sum of the model's errors: the *estimation error*, the *simulation/sampling error*. The first arises due the variance in data across products and the second due to the sampling of consumers and the simulation from the distribution of unobserved consumer traits needed to compute the fixed point in the contraction mapping for δ. Failing to account for the second would yield in underestimated standard errors for the demand-side parameters. Since simulation errors enter the inversion for δ a non-linear way, they get worse as market shares get small. The problem is even more worrying when the number of products is large, in which case a small simulation error in the computation of market shares will cause large changes in ξ. To see why this happens, notice that in the random coefficients logit model each product is a substitute of all the rest of products in the market. Thus, when ξ_j decreases, consumers who substitute away from product j will distribute among the rest of products and, as the number of products increase, the number of consumers that go to each alternative also tends to zero, formally this implies $\partial s/\partial \xi \to 0$. Now, given the sampling error is in the shares, a measure of the impact of the sampling error on the GMM function is $\partial \xi/\partial s$. So when the number of products increases and the elements of $\partial s/\partial \xi$ tend to 0, the elements in $\partial \xi/\partial s$ will tend to grow large. One way to control for the impact of the sampling error is to increase (quadratically) the number of consumer samples as the number of products increase. Berry et al. (2004) provide a detailed description of the asymptotic distribution of the GMM estimators. In particular, they show estimators in the context of the random logit coefficients are consistent if $J \log(J/N)$ tend to zero as J increases, but are asymptotically normal if N grows as the square of the growth in J.

Rewriting the structural unobservable element in terms of the three sources of error in the GMM function we have:

$$\begin{aligned}
\xi(s_n, P_{ns}; \theta) = &\underbrace{\xi(s_o, P_o; \theta)}_{\text{Data-driven error}} + \underbrace{\{\xi(s_n, P_{ns}; \theta) - \xi(s_o, P_{ns}; \theta)\}}_{\text{Sampling error}} + \underbrace{\{\xi(s_o, P_{ns}; \theta) - \xi(s_o, P_o; \theta)\}}_{\text{Simulation error}} \\
&= \xi(s_o, P_o; \theta) + \underbrace{\{s_n - s_o\}}_{\epsilon_n} + \underbrace{\{s(\xi, P_{ns}, s_o; \theta) - s(\xi, P_o, s_o; \theta)\}}_{\epsilon_{ns}},
\end{aligned} \tag{44}$$

where s_o are the true market shares, s_n are the observed market shares, P_o is the true distribution of v, P_{ns} is the simulated distribution of P_o making ns random draws, $s(\xi, P_{ns}, s_o; \theta)$ are the predicted market shares over P_{ns} and $s(\xi, P_o, s_o; \theta)$ are the predicted market shares over P_o.

Recall the contraction mapping for the computation of ξ is $s_n = s(\xi, P_{ns}, s_o; \theta)$, then from equation (44) we can express it as:

$$s_o + \epsilon_n - \epsilon_{ns} = s(\xi, P_{ns}, s_o; \theta), \tag{45}$$

and then obtain expressions for ϵ_n and ϵ_{ns} in terms on ξ using the following matrix:

$$H^{-1}(\xi, P; \theta) = \left\{ \frac{\partial s(\xi, P; \theta)}{\partial \xi'} \right\}^{-1}. \tag{46}$$

Thus,

$$\xi(s_n, P_{ns}; \theta) \approx \xi(s_o, P_o; \theta) + H_o^{-1}\{\epsilon_n - \epsilon_{ns}\}, \tag{47}$$

where

$$H_o = H(s_o, P_o; \theta_o). \tag{48}$$

Since there is no easy expression for H_o we approximate it with H as defined by the inverse matrix of equation (46) and then define the variance S of the GMM estimator as the sum of the following three terms representing the variance of the data-driven error, the sampling error and the simulation error, respectively[3]:

$$S_1 = \frac{1}{J}\sum_{j=1}^{J} z_j z_j' \xi_j^2 \ , \quad S_2 = \frac{1}{nJ} z' H^{-1} V_2 H^{-1'} z \ , \text{ and } \ S_3 = \frac{1}{ns \ J} z' H^{-1} V_3 H^{-1'} z, \tag{49}$$

where $V_2 = S_n - s_n s_n'$ with $S_n = diag(s_n)$, $V_3 = \frac{1}{ns}\sum_{r=1}^{ns} \epsilon_r \epsilon_r'$, and the derivatives in H are:

$$\frac{\partial s_j(\xi, P; \theta)}{\partial \xi_k} = \begin{cases} \int s_j(1 - s_j) dP_{ns}(v) & j = k \\ -\int s_j s_k dP_{ns}(v) & k \neq j \end{cases} \tag{50}$$

Usually s_n is build from total sales data where n is so large that the sampling variance S_2 would have little impact on the GMM estimator. The standard errors of parameters in θ_1 and θ_2 are the square root of the diagonal of matrix $\hat{V}$.

For a final note on the variance of the GMM estimator, Berry et al. (2004) also mention that the traditional bootstrapping estimator for the standard errors is not well defined in the context of a random logit coefficients for mainly two reasons: the first is that bootstrap must be done over a sample of exogenous characteristics; but only after making some (strong) assumptions on the equilibrium we can move on to a different sample that can be used in estimation. The second is that such equilibrium assumption would have to yield a *unique* equilibrium vector of prices and market shares and uniqueness is not a property of the assumed Nash-Bertrand price setting.

4.2 The Logit demand

Let us now consider the case of a model with no interactions between product observable characteristics and consumer traits. The indirect utility function for inside alternatives reduces to:

$$u_{ijt} = \mathbf{x_{jt}}\overline{\beta} - \overline{\alpha} p_{jt} + \xi_{jt} + \varepsilon_{ijt} = \delta_{jt} + \varepsilon_{ijt}, \tag{51}$$

[3] The reader can refer to Berry et al. (2004) for understanding of how S_1, S_2 and S_3 are obtained as function of the model primitives.

and for the outside option it is:

$$u_{i0t} = \xi_{0t} + \varepsilon_{i0t}. \tag{52}$$

Since there is no individual heterogeneity, the probability of choosing product j equals its market share. So under the assumption ε follows an extreme value type I distribution, market shares are given by:

$$s_{jt}(p_{jt}, x_{jt}, \xi_{jt}; \theta_1) = \frac{exp(\delta_{jt})}{1 + \sum_{l=1}^{J} exp(\delta_{lt})}, \tag{53}$$

and their derivatives with respect to prices amount to the following expressions:

$$\begin{aligned} \frac{\partial s_{jt}}{\partial p_{jt}} &= - \overline{\alpha} s_{jt}(1 - s_{jt}) \\ \frac{\partial s_{jt}}{\partial p_{lt}} &= \overline{\alpha} s_{jt} s_{lt}. \end{aligned} \tag{54}$$

In this case, the minimization of distance between observed and predicted market shares has an analytical solution for δ. Since the quotient between predicted market shares and the outside option's market share is equivalent to $exp(\delta_{jt})$ as shown below:

$$exp(\delta_{jt}) \equiv \frac{exp(\delta_{jt})}{1 + \sum_{l=1}^{J} exp(\delta_{lt})} / \frac{1}{1 + \sum_{l=1}^{J} exp(\delta_{lt})}. \tag{55}$$

Then, taking logs on both sides of the equation yields the following solution for δ_{jt}:

$$\delta_{jt} \equiv \log(s_{jt}) - \log_{s_{0t}}. \tag{56}$$

Estimation of $\log(s_{jt}) - \log_{s_{0t}}$ on x_{jt}, p_{jt} and price instruments allows us to recover the product unobservable ξ_{jt}, and then we can include estimates of price derivatives and market shares within the pricing equations in the supply side to obtain estimates for γ and ω_{jt}, completing the estimation of the model without interactions.

Although computation of a model without interactions is simpler than the full model of the previous section, this simplicity comes at a cost: substitution patterns (or price elasticities) generated by the logit demand are somewhat misleading because depend only on market shares but not on product observable characteristics. Price elasticities of demand in the case of a model without interactions, would suggest products with similar market shares would have the same cross price elasticity with respect to a third product even if they differ significantly in their observable characteristics. Consider the market of automobiles. Suppose the Nissan Sentra and the Chevrolet Captiva have the same market share. One could expect cross price elasticities of the Nissan Sentra with respect to a Honda CRV to be lower than cross price elasticities between the Captiva and the CRV just because Captiva and CRV are trucks while Sentra is sedan. However, the logit demand would show cross price elasticities with respect to the CRV are the same for Captivas and Sentras, which is not reasonable. Substitution patterns in the model without interactions would also suggest products with the same market share would have the same own-price elasticity, this implies such products have the same price-cost margins, which is not necessarily true. Furthermore, if the market share of the outside option is large relative to the rest of products, then the logit demand will yield downward biased estimates of the price cross elasticities between interior products.

In terms of the substitution patterns to the outside option denoted by equation (33), the logit demand will yield:

$$\frac{100 \times (ds_{0t}/dp_{jt})}{|ds_{jt}/dp_{jt}|} = 100 \times s_{0t}/(1 - s_{jt}),$$

because

$$ds_{0t}/dp_{jt} = \alpha s_{0t} s_{jt} \quad \text{and} \quad |ds_{jt}/dp_{jt}| = \alpha s_{jt}(1 - s_{jt}).$$

Experiences so far also show the number of inelastic individual demands after estimating a model without interactions is larger than with the full model (Berry, 1994; Berry and Pakes, 1993). Demand elasticity is crucial when conducting counterfactual scenarios. The researcher must consider the trade-off between computational time and interpretability of results before deciding on a logit demand.

4.3 A model with interactions

For the full model with interactions, we will provide further description of the indirect utility specification and estimation presented in Nevo (2001) and Berry et al. (1995).

4.3.1 Nevo's specification

Nevo's application to the ready-to-eat cereal market focuses on measuring market power through market shares and price-cost margins under three scenarios: full price collusion, multi-product firm pricing, and single-product firm pricing. The author uses the following indirect utility specification:

$$u_{ijt} = \mathbf{x_{jt}}\beta_i - \alpha_i p_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}.$$

Notice this function is the same as the one outlined in section (3) already taking into account that the linear effect of income on the indirect utility level vanishes as it is common to all options available to the consumer. The extra term in the indirect utility, $\Delta\xi_{jt}$, arises because the author uses brand-specific dummy variables in the regression model for δ_{jt}. These dummies capture the fixed effect of ξ_j and, thus, the error term in the regression are the deviations from the brand-specific average utility in every city-quarter (markets). The definition of α_i and β_i are the same as in section (3). Although the linear effect of income vanishes from the indirect utility, the author captures its non linear effect by including income and squared income in the matrix of demographic characteristics, D, besides age and the number of children.

The price derivatives that this model generates are given by:

$$\frac{\partial s_{jt}}{\partial p_{jt}} = -\int_{A_j} \alpha_i s_{ijt}(1 - s_{ijt}) dP_v dP_D$$
$$\frac{\partial s_{jt}}{\partial p_{lt}} = \int_{A_j} \alpha_i s_{ijt} s_{ilt} dP_v dP_D.$$

Since $d\mu_{ijt}/dp_{jt} = \sigma_p v_{ipt} + \pi_p^1 D_{i1t} + \cdots + \pi_p^4 D_{i4t}$ and $d\delta_{jt}/dp_{jt} = -\overline{\alpha}$.

From the supply side, the author estimates a linear regression for the marginal cost as shown in the previous section. After jointly estimating the structural parameters for the demand-side and the supply-side using the observed structure for Ω in the firms' first order condition, the author conducts two counterfactual scenarios: one in which he assumes all firms collide, thus

Ω is a squared unit matrix of size J:

$$\Omega = \mathbf{J}_J.$$

And another in which he assumes each product is manufactured by a single firm, thus Ω is an identity matrix of size J:

$$\Omega = \mathbf{I}_J.$$

4.3.2 BLP's specification

The application of Berry et al. (1995) to the automobiles market focuses on finding reasonable structural parameters for price-cost margins, elasticities, and variable profits, comparing different specifications for the pricing equations in the supply-side. The authors use the following indirect utility specification:

$$u_{ijt} = \alpha \log(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt},$$

where

$$\beta_i = \overline{\beta} + \sigma v_{it}.$$

In this case, individual heterogeneity is modeled only through the standard normal effect v_{it}. They do not include consumer demographic characteristics other than income, which also interacts in a special way with price. Since income and price are not linearly separable, as in Nevo's application, the average utility and the deviation from that average are:

$$\begin{aligned}\delta_{jt} =& x_{jt}\overline{\beta} + \xi_{jt} \\ \mu_{ijt} =& \alpha \log(y_i - p_{jt}) + \sum_k \sigma_k x_{jkt} v_{ik}.\end{aligned}$$

Therefore, price only directly impacts the utility distribution through its variance but is not included in the regression model for δ_{jt}.

The indirect utility of the outside option is:

$$u_{i0t} = \alpha \log(y_i) + \xi_{0t} + \sigma_0 v_{i0} + \varepsilon_{i0t}.$$

Notice that since market shares depend on differences in utility $Prob[u_{ijt} - u_{i0t} > 0]$, the outside option's indirect utility ends up subtracting from the indirect utility of inside goods. In the estimation this amounts to normalizing u_{i0t} to zero and capturing $\sigma_0 v_{i0}$ with an interaction between the model's constant and v_i. In other words, $\alpha \log(y_i)$ ends up being common to all products but income is allowed to have a differential effect on the purchase probabilities of inside goods.

Even though $\alpha \log(y_i)$ is not common to all options because income and price are not linearly separable, which is an important conceptual difference with respect to the model outlined in section (3), normalization has the same implications over the logit purchase probabilities as before. In other words, income is allowed to have a differential effect over the purchase probabilities of interior goods.

One of the main questions that arise from the analysis of this utility specification is what happens when price is greater than income, indetermining the first term of the indirect utility function. A negative available income only suggests a consumer will derive no marginal utility from income after buying such product. This effect is the same as making the logarithm in the first term of the equation as negative as possible, some suggest using a taylor series expansion for the natural logarithm for it to defined in the $(-\infty, \infty)$ domain. To correct for such cases afterwards in the estimation, the authors use an importance sampling estimator to derive market shares. The importance sampling estimator consists of making draws from the distribution of v and accepting them with probability $\overline{f}_t(v,\theta) = \sum_{j=1}^{J} s_{ijt}$. Then the vector of predicted market shares would be:

$$s_j(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}; \theta) = \sum_{i=1}^{ns} \frac{\overline{s}_t}{\overline{f}_t(v,\theta)} s_{ijt}(p_{jt}, x_{jt}, \xi_{jt}, P_{ns}; \theta), \tag{57}$$

where $\overline{s}_t = 1 - s_{0t}$ and the sum is over accepted draws. This procedure oversamples from the region of P_{ns} from which consumers are more likely to buy a car and then weights choice probabilities by the inverse sampling weights to obtain market shares.

The derivatives of demand with respect to prices for this model are:

$$\begin{aligned}\frac{\partial s_{jt}}{\partial p_{jt}} =& -\alpha \int_{A_j} \frac{1}{y_{it}-p_{jt}} s_{ijt}(1-s_{ijt}) dP_v \\ \frac{\partial s_{jt}}{\partial p_{lt}} =& \alpha \int_{A_j} \frac{1}{y_{it}-p_{jt}} s_{ijt} s_{ilt} dP_v.\end{aligned}$$

Notice how prices can enter the indirect utility level in any functional form (for example, the reader can imagine including $\log(p_{jt})$ instead of p_{jt} in the regression for δ_{jt}) but derivatives must also reflect so. From the supply side, the authors estimate a log-log regression for the marginal cost as in the equation below:

$$\log(MC_{jt}) = \log(x_{jt})\gamma + \omega_{jt}.$$

In some cases, including $\log(q_{jt})$ as a regressor also.

4.3.3 Interpreting results

Estimation of the full model with interactions yields parameter estimates for $\overline{\alpha}$, $\overline{\beta}$, Σ, Π, ξ_{jt}, γ, ω_{jt}, and $\partial s_{jt}/\partial p_{jt}$. All the parameters that are estimated from the regression model for δ_{jt} are the mean marginal utilities of each of the products' observable characteristics. If income and price are linearly separable, then $\overline{\alpha}$ is expected to have negative sign suggesting marginal increases in prices generate marginal reductions in the utility level. $\overline{\beta}$ would be the vector of marginal utilities for each characteristic in $\mathbf{x_{jt}}$ but the expected sign depends on the intuition on which characteristics are associated to greater choice probabilities.

More interesting are the interpretations of the deviations from the average utility. If coefficients in vector Σ are significant, then variations in the product observable characteristics that interact with consumer traits affect the variance of the utility distribution. In terms of the substitution patterns, this implies that if a product with certain trait x increases its price, then consumers will tend to substitute disproportionately to other products with similar traits. Recall the discussion about the substitution patterns implied by the logit demand

and consider again the case of the automobiles market. Significant coefficients in Σ means cross price elasticity between the Chevrolet Captiva and the Honda CRV will be greater than the one between the Nissan Sentra and the Honda CRV because the type of vehicle (trucks or sedans) would have a significant effect on the variance of the utility distribution. If coefficients in Σ were not significant, this would be the same as having a model without interactions, i.e, consumers would substitute away from the Chevrolet Captiva to the Nissan Sentra in the same magnitude they substitute towards the Honda CRV after a price increase of the Captiva. Since Σ captures deviations from the average utility, its sign must be positive. Minimization of the GMM function can be constrained to the region of semi-positive values for Σ in order to obtain estimates of the expected sign.

The coefficients in Π capture the part of the variation of the marginal utilities due to differences in observable demographic characteristics. Consider the case of the ready-to-eat cereal market. Suppose the grams of sugar per serving size is one of the product observable characteristics and age one of the consumer traits included in D_{it}. A negative and significant coefficient associated to the interaction between these two variables would suggest the marginal utility for sugar grams in a serving size is decreasing with age. If the coefficient is insignificant from zero, then age would have no effect on the distribution of the marginal utility. In other words, an adult and a child would perceive the same level of utility from an additional gram of sugar per serving size conditional on v_i. The expected sign of the coefficients in Π also depend on the intuition or anecdotal evidence from the market.

In the case of substitution patterns or price elasticities, an element (i, j) of matrix $(\partial s_{jt}/\partial p_{lt}) \times (p_{lt}/s_{jt})$, where i denotes the row and j the column, indicates the percentage variation in the quantity demanded of product i, when the price of product j increases 1%. In some cases, price semi-elasticities are preferred for making comparisons because they represent percentage variations in demand over price increases of the same magnitude across products. For example, if prices are measured in 1,000 dollars, then the expression $(\partial s_{jt}/\partial p_{jt}) \times (1/s_{jt})$ would be the demand sensitivity over price increases of a thousand dollars.

Sensitivity of the outside option's market share to the price of interior goods is also an important measure. Recall the absolute value of own-price elasticities represents the amount of consumers that substitute away from a product when its price increases. Then, the quotient between $\partial s_{0t}/\partial p_{jt}$ and $|\partial s_{jt}/\partial p_{jt}|$ would be the amount of consumers that go to the outside option as a proportion of all consumers that substitute away from product j when its price increases.

5 Counterfactual analysis

One of the main properties of structural models in economics is that they allow the researcher to conduct counterfactual scenarios, i.e, to compute the market equilibrium under market conditions that differ from the ones that are observed in the data. In particular, the structural demand model we have described permits finding equilibrium prices and shares that would have resulted in the following scenarios to name a few:

- A new product is introduced in the market.
- An existing product is eliminated from the market.
- Firms merge.
- Merged firms separate.
- Product characteristics' change.
- Government imposes a per-unit tax.

Overall, counterfactuals are used to evaluate any kind of policy whether it affects the demand or the supply side. In this section of the document we will focus on how counterfactual scenarios are computed. First lets consider the case of counterfactuals that affect the supply side. For instance, the ones estimated in Nevo (2001). The competitive structure of a market is captured by matrix Ω. Estimation of the effects of mergers or merger dissolutions must rely on changing the structure of this matrix. Let the upper script C denote the counterfactual scenario and the upper script O the observed scenario. The pricing equation under the counterfactual is:

$$b^C_{jt} = p^C_{jt} - mc^O_{jt} = (\Omega^C)^{-1} s^C_{jt}.$$

The equation above is built upon the assumption marginal costs do not change in the counterfactual. This assumption has strong implications on how the market updates with the variation in the competitive structure. It implies firms in the counterfactual do not experience cost reductions nor cost increases after the merger or after the dissolution. Another underlying assumption is that the structural parameters estimated with the observed market conditions ($\overline{\alpha}$, $\overline{\beta}$, Σ, Π, ξ_{jt}, ω_{jt}) remain fixed in the counterfactual, which also suggests firms do not incur in marketing efforts that could change consumer preferences over certain product characteristics. This last assumption, however, is crucial for the computation of the market equilibrium.

Since market shares in the counterfactual depend on prices and, at the same time, prices charged in the counterfactual are obtained from price derivatives of demand and market shares, computation of equilibrium conditions will consist on finding the vector of prices that equal demand and supply using the contraction mapping of equation (58).

$$\mathbf{p}' = mc^O + \Omega^C s_{jt}(\mathbf{p}, x_{jt}, \xi_{jt}, P_{ns}, P_D; \theta). \tag{58}$$

We begin by setting a seed for the vector of prices and predicting u_{ijt}. This implies predicting δ_{jt}, adding ξ_{jt} to it, and then predicting μ_{ijt}. Afterwards, we compute choice probabilities, aggregate to obtain market shares, and estimate price derivatives to obtain the measure of price-cost margin. We add the margins to the marginal cost and recover a new vector of prices with which we repeat the procedure until convergence of prices.

Note the market equilibrium after an exogenous change in product characteristics results in the same estimation procedure as above. If there is a new vector of characteristics x^C_{jt}, this changes the vector of δ_{jt}, choice probabilities, and price derivatives, which result in new price-cost margins and new prices. Then, prices must be updated in each iteration until convergence:

$$\mathbf{p}' = mc^O + \Omega s_{jt}(\mathbf{p}, x^C_{jt}, \xi_{jt}, P_{ns}, P_D; \theta). \tag{59}$$

Also note the introduction of a new product or the elimination of an existing one results in a change of matrix Ω, therefore amounting to the contraction mapping of equation (58). In the former case, a new product must be modeled by assigning a vector of observable x_{jt}, p_{jt}

and unobservable ξ_{jt} characteristics to it. Although assigning values for x_{jt} only requires knowledge or certain anecdotal evidence on where will the product lie on the distribution of each x and assigning p_{jt} requires knowledge of the pricing equation of the manufacturer, what value of ξ_{jt} to assign is less straightforward. One of the most common approaches is to assign the value of ξ_{jt} of the existing product that most closely resembles to its observable characteristics. If there is no such product, then one could average the values of ξ_{jt} of the products that the firm of the new good manufactures and assign the resulting measure to the new product. Computing equilibrium prices and shares if the manufacturer is also an entrant must rely in stronger assumptions, since, first, both x_{jt} and Ω change in the counterfactual, and, second, the researcher must decide not only on which value of ξ_{jt} to assign to the new products but also on which value of ω_{jt} to assign to the manufacturer's marginal cost.

Now consider the case of a per-unit tax, τ, charged to the manufacturer. Equation (60) shows how taxes affect the first order condition of the firm's maximization problem. Although it seems as if there was a pass-through of taxes to prices, the competitive structure of the market might help alleviate proportional increases in prices due to taxation. Computation of the counterfactual scenario consists of the same procedure outlined before.

$$\mathbf{p}' = (1+\tau) \times mc^O + \Omega s_{jt}(\mathbf{p}, x^C_{jt}, \xi_{jt}, P_{ns}, P_D; \theta). \tag{60}$$

6 Welfare analysis

Policy evaluation is concerned with how consumer welfare and firm benefits vary between the observed scenario and the counterfactual. There are several measures of consumer welfare that include the equivalent variation and the compensating variation. Let $e_{ijt}(q_{jt}, p_{jt}, y_{it})$ be the expenditure function of consumer i in the observed scenario. The equivalent variation represents the additional expenditure in which the consumer has to incur in order to reach the level of welfare in the observed scenario under the equilibrium prices of the counterfactual.

$$EV = e(q_{jt}, p^C_{jt}, y^C_{it}) - e(q_{jt}, p_{jt}, y_{it}). \tag{61}$$

On the other hand, the compensating variation is a measure of the additional expenditure the consumer has to incur in order to reach the level of welfare in the counterfactual under the observed equilibrium prices:

$$CV = e(q_{jt}^C, p_{jt}^C, y_{it}^C) - e(q_{jt}^C, p_{jt}, y_{it}). \tag{62}$$

Given the budget constraint is binding in the utility maximization problem, equations (61) and (62) can be interpreted as the variation in consumer income necessary to keep the level of welfare constant under different equilibrium conditions. Nonetheless, one of the assumptions when computing counterfactual scenarios after estimation of the structural demand model, is that marginal utilities remain fixed in the counterfactual. If the marginal utility of income is fixed for every individual, i.e, there are no income effects, then the marshallian and the hicksian demand curves overlap, so the compensated version of consumer surplus when choice probabilities follow an extreme value type I distribution can be expressed as:

$$CV_i = \frac{log(\sum_{j=1}^J e^{V_{ijt}^O}) - log(\sum_{j=1}^J e^{V_{ijt}^C})}{\alpha_i}, \tag{63}$$

where $log(\sum_{j=1}^J e^{V_{ijt}^O})$ is the expected maximum utility in the observed scenario, $log(\sum_{j=1}^J e^{V_{ijt}^C})$ is the expected maximum utility in the counterfactual, and $V_{ijt} = \delta_{jt} + \mu_{ijt}$.

To prove $log(\sum_{j=1}^J e^{V_{ijt}})$ corresponds to the expected maximum, suppose the utility of a consumer of choosing product i is $u_i = V_i + \varepsilon_i$ and $\varepsilon \sim$ extreme value type I distribution. The expected maximum utility following the notation of section (2.2.2) is:

$$\overline{u} = \int_{\widetilde{\varepsilon}=-\infty}^{+\infty} \max_i (V_i + \varepsilon_i) f(\widetilde{\varepsilon}) d\widetilde{\varepsilon},$$

which can be rewritten as:

$$\overline{u} = \sum_i \int_{\varepsilon_i=-\infty}^{+\infty} (V_i + \varepsilon_i) F_i(V_i + \varepsilon_i - V_j) d\varepsilon_i$$

Let $a = G(e^{V_1}, ..., e^{V_J})$. Given $F_i(V_i + \varepsilon_i - V_j) = e^{-G(e^{V_i+\varepsilon_i-V_j})} G_i(e^{V_i+\varepsilon_i-V_j}) e^{-\varepsilon_i}$ and G is homogeneous of degree 1 and G_i homogenous of degree 0, then:

$$F_i(V_i + \varepsilon_i - V_j) = e^{-ae^{-V_i-\varepsilon_i}} G_i(e^{V_j}) e^{-\varepsilon_i}.$$

Now let $V_i + \varepsilon_i = w$, the expected maximum can be expressed as:

$$\begin{aligned}
\overline{u} &= \sum_i \int_{w=-\infty}^{+\infty} we^{-ae^{-w}} G_i(e^{V_j}) e^{-w} e^{V_i} dw \\
&\text{using Euler's law } \textstyle\sum_i e^{V_i} G_i(e^{V_j}) = G(e^{V_j}) \text{ we have that} \\
&= \int_{w=-\infty}^{+\infty} we^{-ae^{-w}} G(e^{V_j}) e^{-w} dw \\
&= \int_{w=-\infty}^{+\infty} we^{-ae^{-w}} ae^{-w} dw \\
&\text{which is the mean of the extreme value distribution } e^{-ae^{-\varepsilon}}\text{, hence:} \\
&= \log a + \gamma \\
&\text{where } \gamma \text{ is Euler's constant}
\end{aligned} \tag{64}$$

Having proved $log(\sum_{j=1}^{J} e^{V_{ijt}})$ corresponds to the expected maximum utility, aggregation over the distribution of consumer characteristics yields the total compensating variation between the observed scenario and the counterfactual in monetary terms:

$$\Delta C = M \int CV_i dP_o(v) \tag{65}$$

From the point of view of producers, changes in producer welfare can be computed as changes in benefits between the observed and the counterfactual situations as shown in equation (66):

$$\Delta E = \sum_f \pi_f^O - \pi_f^C = \sum_f \left(\sum_{j \in \mathcal{F}_f} (p_{jt}^O - cmg_{jt}^O) M s_{jt}^O - \sum_{j \in \mathcal{F}_f} (p_{jt}^C - cmg_{jt}^C) M s_{jt}^C \right) \tag{66}$$

Thus, changes in welfare from the societal perspective are:

$$\Delta S = \Delta C + \Delta E \tag{67}$$

7 Programming

This section is devoted to the matrix-like notation that should be considered for programming and estimation of the structural demand model of section (3), first, under linearly separable income and prices as in Nevo (2001), and then under non separable income and prices as in Berry et al. (1995).

7.1 Linearly separable income and prices: the case of Nevo

Suppose you have a matrix X of dimensions $J \times (K+1)$ where K are the number of product observable characteristics and the additional column vector stands for the model's constant. Allow p_{jt} to be included in X so that it can be interacted with consumer traits as shown further. v_i is a matrix of dimensions $(K+1) \times N$ where N are the number of simulated consumers. Each row vector in v_i is sampled from a standard normal distribution. D_{it} is a matrix of $d \times N$ where d are the number of demographic variables over which consumers are assumed to vary also. Now let Σ be a diagonal matrix of coefficients of size $(K+1) \times (K+1)$ and Π a matrix of coefficients of size $(K+1) \times d$. If income and prices are linearly separable in the function for available income, then computation of the contraction mapping for finding δ_{jt} holding Σ and Π fixed at any value, consists of:

7.1.1 The demand-side

For every market:

1. Set the seed for δ_j for instance as a column vector of 1's of dimensions $J \times 1$.

2. Build the matrix for μ_{ij} as:

$$\mu_{ij} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,K} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{J,1} & x_{J,2} & \cdots & x_{J,K} \end{pmatrix} \times \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{K+1} \end{pmatrix} \times \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,N} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{K+1,1} & v_{K+1,2} & \cdots & v_{K+1,N} \end{pmatrix}$$
$$+ \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,K} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{J,1} & x_{J,2} & \cdots & x_{J,K} \end{pmatrix} \times \begin{pmatrix} \pi_{1,1} & \pi_{1,2} & \cdots & \pi_{1,d} \\ \pi_{2,1} & \pi_{2,2} & \cdots & \pi_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{K+1,1} & \pi_{K+1,2} & \cdots & \pi_{K+1,d} \end{pmatrix} \times \begin{pmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,N} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ D_{d,1} & D_{d,2} & \cdots & D_{d,N} \end{pmatrix} \tag{68}$$

$$\begin{aligned} \mu_{ij_{J\times N}} = & \left(\; x_{j,k} \; \right)_{J\times(K+1)} \times \left(\; \sigma_k \; \right)_{(K+1)\times(K+1)} \times \left(\; v_{k,i} \; \right)_{(K+1)\times N} \\ & + \left(\; x_{j,k} \; \right)_{J\times(K+1)} \times \left(\; \pi_{k,d} \; \right)_{(K+1)\times d} \times \left(\; D_{d,i} \; \right)_{d\times N} \end{aligned} \tag{69}$$

3. Add δ_j to μ_{ij} to obtain u_{ij}:

$$u_{ij} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{J\times 1} \times (1 \; 1 \cdots \; 1)_{1\times N} + \mu_{ij_{J\times N}} \tag{70}$$

4. Attach a row vector of zeros to u_{ij} and exponentiate:

$$e^{u_{j+1,i}} = \begin{pmatrix} e^{u_{1,1}} & e^{u_{1,2}} & \cdots & e^{u_{1,N}} \\ e^{u_{2,1}} & e^{u_{2,2}} & \cdots & e^{u_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{u_{J,1}} & e^{u_{J,2}} & \cdots & e^{u_{J,N}} \\ e^0 & e^0 & \cdots & e^0 \end{pmatrix}_{(J+1)\times N} \tag{71}$$

5. Find the running column sums:

$$\left(\sum_{j=0}^{J} e^{u_{j,1}} \quad \sum_{j=0}^{J} e^{u_{j,2}} \quad \cdots \quad \sum_{j=0}^{J} e^{u_{j,N}} \right)_{1\times N} \tag{72}$$

6. Divide each row vector of matrix $e^{u_{j+1,i}}$ into the expression of step (4) to obtain choice

probabilities $s_{j,i}$:

$$s_{j,i} = \begin{pmatrix} \frac{e^{u_{1,1}}}{\sum_{j=0}^{J} e^{u_{j,1}}} & \frac{e^{u_{1,2}}}{\sum_{j=0}^{J} e^{u_{j,2}}} & \cdots & \frac{e^{u_{1,N}}}{\sum_{j=0}^{J} e^{u_{j,N}}} \\ \frac{e^{u_{2,1}}}{\sum_{j=0}^{J} e^{u_{j,1}}} & \frac{e^{u_{2,2}}}{\sum_{j=0}^{J} e^{u_{j,2}}} & \cdots & \frac{e^{u_{2,N}}}{\sum_{j=0}^{J} e^{u_{j,N}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{e^{u_{J,1}}}{\sum_{j=0}^{J} e^{u_{j,1}}} & \frac{e^{u_{J,2}}}{\sum_{j=0}^{J} e^{u_{j,2}}} & \cdots & \frac{e^{u_{J,N}}}{\sum_{j=0}^{J} e^{u_{j,N}}} \\ \frac{e^{0}}{\sum_{j=0}^{J} e^{u_{j,1}}} & \frac{e^{0}}{\sum_{j=0}^{J} e^{u_{j,2}}} & \cdots & \frac{e^{0}}{\sum_{j=0}^{J} e^{u_{j,N}}} \end{pmatrix}_{(J+1)\times N} \tag{73}$$

7. Find the running row mean of matrix $s_{j,i}$; this will yield the vector of predicted market shares s_j:

$$s_j = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} s_{1,i} \\ \frac{1}{N}\sum_{i=1}^{N} s_{2,i} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} s_{J,i} \\ \frac{1}{N}\sum_{i=1}^{N} s_{0,i} \end{pmatrix}_{(J+1)\times 1} \tag{74}$$

8. With the vector of observed market shares s_j^o, compute the contraction mapping:

$$\delta_j' = \log(s_j^o) - \log(s_j) + \delta_j \tag{75}$$

   and update the vector of δ_j.

9. Repeat steps 2 to 8 until $||\delta_j' - \delta_j|| < tolerance$.

The vector of average utilities that equal observed to predicted market shares is obtained using a fixed point algorithm. After convergence, we estimate a regression of δ_{jt} on product observable characteristics, price, and price instruments. Note the contraction mapping above is computed for each market separately; however, we use the full vector of δ_{jt} including all markets to estimate the regression model and obtain structural parameter estimates for θ_1 and ξ_{jt}. Then, we can estimate price markups *per market* as follows.

7.1.2 The supply-side

After computing $s_{j,i}$ and s_j with the true vector of δ_j:

1. Compute own-price derivatives as the running row mean of the following matrix:

$$\frac{\partial s_{j,i}}{\partial p_j} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,N} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{J,1} & \alpha_{J,2} & \cdots & \alpha_{J,N} \end{pmatrix} \odot \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \odot \begin{pmatrix} 1-s_{1,1} & 1-s_{1,2} & \cdots & 1-s_{1,N} \\ 1-s_{2,1} & 1-s_{2,2} & \cdots & 1-s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 1-s_{J,1} & 1-s_{J,2} & \cdots & 1-s_{J,N} \end{pmatrix} \tag{76}$$

or

$$\frac{\partial s_{j,i}}{\partial p_j} = \left(\alpha_{j,i} \right)_{J\times N} \odot \left(s_{j,i} \right)_{J\times N} \odot \left(1 - s_{j,i} \right)_{J\times N} \tag{77}$$

where $\odot$ is the element-wise multiplication operator and

$$\alpha_i = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_N \end{pmatrix}$$
$$= \begin{pmatrix} \overline{\alpha} + \sigma_p p_1 v_{p,1} + \pi_{1,p} p_1 D_{1,1} + \cdots + \pi_{d,p} p_1 D_{d,1} & \cdots & \overline{\alpha} + \sigma_p p_1 v_{p,N} + \pi_{1,p} p_1 D_{1,N} + \cdots + \pi_{d,p} p_1 D_{d,N} \\ \overline{\alpha} + \sigma_p p_2 v_{p,1} + \pi_{1,p} p_2 D_{1,1} + \cdots + \pi_{d,p} p_2 D_{d,1} & \cdots & \overline{\alpha} + \sigma_p p_2 v_{p,N} + \pi_{1,p} p_2 D_{1,N} + \cdots + \pi_{d,p} p_2 D_{d,N} \\ \vdots & \ddots & \vdots \\ \overline{\alpha} + \sigma_p p_J v_{p,1} + \pi_{1,p} p_J D_{1,1} + \cdots + \pi_{d,p} p_J D_{d,1} & \cdots & \overline{\alpha} + \sigma_p p_J v_{p,N} + \pi_{1,p} p_J D_{1,N} + \cdots + \pi_{d,p} p_J D_{d,N} \end{pmatrix} \tag{78}$$

Hence, own-price derivatives are:

$$\frac{\partial s_j}{\partial p_j} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial p_1 \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial p_2 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial p_J \end{pmatrix}_{J\times 1} \tag{79}$$

2. Compute cross price derivatives as:

$$\frac{\partial s_j}{\partial p_k} = \left(\begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,N} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{J,1} & \alpha_{J,2} & \cdots & \alpha_{J,N} \end{pmatrix} \odot \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \right) \times \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,J} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \cdots & s_{N,J} \end{pmatrix} \tag{80}$$

or,

$$\frac{\partial s_j}{\partial p_k} = \Big(\big(\ \alpha_{j,i} \ \big)_{J\times N} \odot \big(\ s_{j,i} \ \big)_{J\times N} \Big) \times \big(\ s_{i,j} \ \big)_{N\times J} \tag{81}$$

3. Replace the diagonal vector of $\partial s_j/\partial p_k$ by $\partial s_j/\partial p_j$.
4. Build matrix Ω. This the element-wise product of a diagonal matrix where the elements in the diagonal are squared matrices of ones of size $\mathcal{F}_f$ and the matrix of price derivatives.

$$\Omega = \begin{pmatrix} \mathbf{J}_{\mathcal{F}_1} & 0 & \cdots & 0 \\ 0 & \mathbf{J}_{\mathcal{F}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{J}_{\mathcal{F}_F} \end{pmatrix} \odot \frac{\partial s_j}{\partial p_k} \tag{82}$$

For example, if there are 30 products in the market and 3 firms producing 15, 10, and 5 of those products respectively, then Ω is:

$$\begin{pmatrix} \begin{matrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{matrix}_{15\times 15} & 0 & \cdots & 0 \\ 0 & \begin{matrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{matrix}_{10\times 10} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{matrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{matrix}_{5\times 5} \end{pmatrix} \odot \frac{\partial s_j}{\partial p_k}$$

5. Estimate markups as:

$$b_j = (-\Omega)^{-1} \times s_j \tag{83}$$

With estimated markups we can obtain marginal costs and the supply-side unobservable. Again notice the vector of markups is computed separately for each market but then bound together in a single column vector to obtain marginal costs as $mc_{jt} = p_{jt} - b_{jt}$. After computing marginal costs we can estimate a regression of $f(mc_{jt})$ on product observable characteristics and cost instruments to obtain structural parameter estimates for γ and ω_{jt}. Estimation of the demand and supply linear parameters is done while holding non-linear parameters in Σ and Π fixed. Therefore, to obtain optimal estimates of non-linear parameters we have to compute the GMM function and minimize it over Σ and Π as follows:

7.1.3 The GMM function

1. With the estimates of ξ_{jt}, ω_{jt}, and the matrix of instruments, build the sample analog of the moment conditions of equations (38) and (39).

$$g(\theta_2) = \epsilon' \times Z \times \mathbf{I}_{K \times K} \times Z' \times \epsilon \tag{84}$$

where

$$Z = \left(\begin{Bmatrix} 1 & x_{1,1} & \cdots & x_{1,K-1} & z_{1,1} & \cdots & z_{1,Q} \\ 1 & x_{2,1} & \cdots & x_{2,K-1} & z_{2,1} & \cdots & z_{2,Q} \\ \vdots & \vdots & \ddots & \vdots & & & \\ 1 & x_{J,1} & \cdots & x_{J,K-1} & z_{2,1} & \cdots & z_{2,Q} \\ 1 & x_{1,1} & \cdots & x_{1,K-1} & z_{1,1} & \cdots & z_{1,Q} \\ 1 & x_{2,1} & \cdots & x_{2,K-1} & z_{2,1} & \cdots & z_{2,Q} \\ \vdots & \vdots & \ddots & \vdots & & & \\ 1 & x_{J,1} & \cdots & x_{J,K-1} & z_{2,1} & \cdots & z_{2,Q} \end{Bmatrix} \right) \tag{85}$$

$$\epsilon = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_J \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_J \end{pmatrix} \tag{86}$$

and $z_{j,q}$ are the additional instrumental variables. Z includes all exogenous variables used in the estimation, therefore, in addition to instruments in $z_{j,q}$, it includes all product observable characteristics in X *except* for price (this is the reason for the $K-1$ column dimension for X). Moreover, the estructure of Z in equation (85) implies price and cost instruments used in the regression of δ_{jt} and mc_{jt} are the same. If this is not the case, then the appropriate matrix of instruments must be build, for example assigning a vector of zeros in the lower matrix of Z when an instrumental variable for the demand-side is not used in the supply-side and viceversa.

$g(\theta_2)$ yields the value of the GMM function we wish to minimize over θ_2. To do so, we nest the demand side and the supply side into a minimization algorithm such as the Nelder-Mead simplex routine. In every iteration of the minimization algorithm, which is the outer loop of the model, a new vector of θ_2 is computed based on the function gradient of g(θ_2), and then the inner loop computes ξ_{jt} and ω_{jt} as described before conditional on θ_2.

7.2 Non separable income and prices: the case of BLP

To derive the programming structure of the model in Berry et al. (1995), suppose as before you have a matrix X of product observable characteristics with dimensions $J \times (K+1)$. In this case, let p_{jt} be separate vector of prices *not included* in X. Again, let v_i be a matrix of dimensions $(K+1) \times N$ with each row vector sampled from a standard normal distribution. Let v_{iy} be a row vector also sampled from a standard normal distribution and obtain estimates of the average log income m_t and the standard deviation of log income σ_t^y from household surveys. As in Nevo's application, Σ is a diagonal matrix of coefficients of size $(K+1) \times (K+1)$ but, unlike Nevo, there are no demographic variables that interact with consumer traits other than income, which is only allowed to interact with price but not with the rest of observable characteristics. Also unlike Nevo, non-linear parameters in BLP include the coefficient for the available income α, therefore μ_{ijt} includes $\alpha \log(y_{it} - p_{jt})$. If income and prices are not linearly separable in the function of the available income as in the utility specification in Berry et al. (1995), then computation of the contraction mapping for finding δ_{jt} holding Σ and α fixed, consists of:

7.2.1 The demand side

For each market:

1. Set the seed for δ_j for instance as a column vector of 1's of dimensions $J \times 1$.
2. Build the matrix for μ_{ij} as:

$$\mu_{ij} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,K} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{J,1} & x_{J,2} & \cdots & x_{J,K} \end{pmatrix} \times \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{K+1} \end{pmatrix} \times \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,N} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{K+1,1} & v_{K+1,2} & \cdots & v_{K+1,N} \end{pmatrix}$$
$$+ \alpha \odot \begin{pmatrix} \log(y_1 - p_1) & \log(y_2 - p_1) & \cdots & \log(y_N - p_1) \\ \log(y_1 - p_2) & \log(y_2 - p_2) & \cdots & \log(y_N - p_2) \\ \vdots & \vdots & \ddots & \vdots \\ \log(y_1 - p_J) & \log(y_2 - p_J) & \cdots & \log(y_N - p_J) \end{pmatrix} \tag{87}$$

$$\mu_{ij_{J\times N}} = \left(\; x_{j,k} \; \right)_{J\times(K+1)} \times \left(\; \sigma_k \; \right)_{(K+1)\times(K+1)} \times \left(\; v_{k,i} \; \right)_{(K+1)\times N} + \alpha \left(\; \log(y_i - p_j) \; \right)_{J\times N}$$

where

$$y_i = m + \sigma_y \times \begin{pmatrix} v_{1,y} & v_{2,y} & \cdots & v_{N,y} \end{pmatrix}$$

3. Compute δ_j and market shares as in steps (3) to (9) of the demand side of section (7.1.1). But, for the computation of market shares in step (7) perform the importance

sampling estimator as follows:

(a) Find the running column sums of the matrix of choice probabilities over interior products:

$$\overline{f} = \left(\sum_{j=1}^{J} \frac{e^{u_{1,i}}}{\sum_{k=0}^{J} e^{u_{k,i}}} \quad \sum_{j=1}^{J} \frac{e^{u_{2,i}}}{\sum_{k=0}^{J} e^{u_{k,i}}} \quad \cdots \quad \sum_{j=1}^{J} \frac{e^{u_{J,i}}}{\sum_{k=0}^{J} e^{u_{k,i}}} \right) \tag{88}$$

(b) Compute the sum of observed market shares of interior products $\overline{s} = \sum_{j=1}^{J} s_j$

(c) Weight each element of the matrix of choice probabilities by $\overline{s}/\overline{f}$.

(d) Estimate market shares as the running row means of the resulting matrix.

In BLP's specification, after obtaining the vector of δ_{jt} that equal observed to predicted market shares, the regression model to obtain estimates for the linear structural parameters in the demand side includes as regressors all of the product observable characteristics *except* for price. The coefficient for price (α) is non-linear and therefore its estimation must be done in the outer loop rather than in the inner loop. Since price can not be separated linearly from the function of available income, price derivatives will also have a special notation in this model. The next subsection shows the matrix-like notation.

7.2.2 The supply side

After computing $s_{j,i}$ and s_j, for each market:

1. Compute own-price derivatives as the running row mean of the following matrix:

$$\begin{aligned} \frac{\partial s_{j,i}}{\partial p_j} = -\alpha \odot & \begin{pmatrix} 1/(y_1 - p_1) & 1/(y_2 - p_1) & \cdots & 1/(y_N - p_1) \\ 1/(y_1 - p_2) & 1/(y_2 - p_2) & \cdots & 1/(y_N - p_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1/(y_1 - p_J) & 1/(y_2 - p_J) & \cdots & 1/(y_N - p_J) \end{pmatrix} \\ \odot & \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \odot \begin{pmatrix} 1 - s_{1,1} & 1 - s_{1,2} & \cdots & 1 - s_{1,N} \\ 1 - s_{2,1} & 1 - s_{2,2} & \cdots & 1 - s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 - s_{J,1} & 1 - s_{J,2} & \cdots & 1 - s_{J,N} \end{pmatrix} \end{aligned} \tag{89}$$

or

$$\frac{\partial s_{j,i}}{\partial p_j} = -\alpha \odot \left(\ 1/(y_i - p_j)\ \right)_{J\times N} \odot \left(\ s_{j,i}\ \right)_{J\times N} \odot \left(\ 1 - s_{j,i}\ \right)_{J\times N} \tag{90}$$

Hence, own-price derivatives are:

$$\frac{\partial s_j}{\partial p_j} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial p_1 \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial p_2 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial p_J \end{pmatrix}_{J\times 1} \tag{91}$$

2. Compute cross price derivatives as:

$$\begin{aligned} \frac{\partial s_j}{\partial p_k} = & \left(\alpha \begin{pmatrix} 1/(y_1 - p_1) & 1/(y_2 - p_1) & \cdots & 1/(y_N - p_1) \\ 1/(y_1 - p_2) & 1/(y_2 - p_2) & \cdots & 1/(y_N - p_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1/(y_1 - p_J) & 1/(y_2 - p_J) & \cdots & 1/(y_N - p_J) \end{pmatrix} \odot \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \right) \\ & \times \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,J} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \cdots & s_{N,J} \end{pmatrix} \end{aligned} \tag{92}$$

or,

$$\frac{\partial s_j}{\partial p_k} = \Big(\left(\ \alpha(1/(y_i - p_j))\ \right)_{J\times N} \odot \left(\ s_{j,i}\ \right)_{J\times N} \Big) \times \left(\ s_{i,j}\ \right)_{N\times J} \tag{93}$$

3. Replace the diagonal vector of $\partial s_j/\partial p_k$ by $\partial s_j/\partial p_j$.
4. To compute markups follow steps (4) to (5) of section (7.1.2).

7.2.3 The GMM function

The sample analog of the GMM estimator in the case of BLP has the same matrix-like notation as in Nevo, but minimization is performed over a different vector of non-linear parameters. Recall in Nevo, non-linear parameters are Σ and Π. The first are the coefficients of the interactions between product observable characteristics and the consumer's unobservable characteristic v_i, and the second are the coefficients of the interactions between product observable characteristics and consumer observable demographics. In the case of BLP, the non-linear parameters of the model are Σ and α. Thus,

$$g(\theta_2) = \epsilon' \times Z \times \mathbf{I}_{K\times K} \times Z' \times \epsilon$$

with Z and ϵ as in Nevo's case.

7.3 The variance-covariance matrix

Recall from section (43), to estimate the variance-covariance matrix for the GMM estimator we have to estimate the Jacobian of the moment conditions with respect to the linear and the non-linear parameters. In the first case this amounts to x_{jt}, but in the second case, if product characteristics interact with the unobservable v_i and observable demographic traits, then the derivatives of δ_{jt} with respect to σ_k and π_d are given by equation (43). The matrix-like notation for programming of the variance-covariance matrix is outlined in the steps below.

For each market:

1. Compute the partial derivative of s_j with respect to δ_j as the running row mean of the following matrix:

$$\frac{\partial s_{j,i}}{\partial \delta_j} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \odot \begin{pmatrix} 1-s_{1,1} & 1-s_{1,2} & \cdots & 1-s_{1,N} \\ 1-s_{2,1} & 1-s_{2,2} & \cdots & 1-s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 1-s_{J,1} & 1-s_{J,2} & \cdots & 1-s_{J,N} \end{pmatrix} \tag{94}$$

   or

$$\frac{\partial s_{j,i}}{\partial \delta_j} = \left(\ s_{j,i}\ \right)_{J\times N} \odot \left(\ 1 - s_{j,i}\ \right)_{J\times N}$$

Hence, the derivatives are:

$$\frac{\partial s_j}{\partial \delta_j} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \delta_1 \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \delta_2 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \delta_J \end{pmatrix}_{J\times 1} \tag{95}$$

2. Compute the cross derivatives of s_j with respect to δ_k as:

$$\frac{\partial s_j}{\partial \delta_k} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \times \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,J} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \cdots & s_{N,J} \end{pmatrix} \tag{96}$$

or,

$$\frac{\partial s_j}{\partial \delta_k} = \begin{pmatrix} s_{j,i} \end{pmatrix}_{J\times N} \times \begin{pmatrix} s_{i,j} \end{pmatrix}_{N\times J} \tag{97}$$

3. Replace the diagonal vector of $\partial s_j/\partial \delta_k$ with $\partial s_j/\partial \delta_j$.
4. For every $x_k \in X$, compute the derivative of s_j with respect to σ_k. This is consists of:
   (a) Calculate the element-wise product:

$$x_{j,k}s_{j,i} = \begin{pmatrix} x_{1,k} & x_{1,k} & \cdots & x_{1,k} \\ x_{2,k} & x_{2,k} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{J,k} & x_{J,k} & \cdots & x_{J,k} \end{pmatrix} \odot \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \tag{98}$$

   (b) Find the running column sums of the matrix above,

$$\sum_{j=1}^{J} x_{j,k}s_{j,i} = \left(\sum_{j=1}^{J} x_{j,k}s_{j,1} \quad \sum_{j=1}^{J} x_{j,k}s_{j,2} \quad \cdots \quad \sum_{j=1}^{J} x_{j,k}s_{j,N} \right) \tag{99}$$

(c) Compute the derivative of choice probabilities with respect to σ_k as:

$$\frac{\partial s_{j,i}}{\partial \sigma_k} = \left(\begin{pmatrix} v_{k,1} & v_{k,2} & \cdots & v_{k,N} \\ v_{k,1} & v_{k,2} & \cdots & v_{k,N} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k,1} & v_{k,2} & \cdots & v_{k,N} \end{pmatrix} \odot \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \right) \odot \left(\begin{pmatrix} x_{1,k} & x_{1,k} & \cdots & x_{1,k} \\ x_{2,k} & x_{2,k} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{J,k} & x_{J,k} & \cdots & x_{J,k} \end{pmatrix} - \begin{pmatrix} \sum_{j=1}^{J} x_{j,k}s_{j,1} & \sum_{j=1}^{J} x_{j,k}s_{j,2} & \cdots & \sum_{j=1}^{J} x_{j,k}s_{j,N} \\ \sum_{j=1}^{J} x_{j,k}s_{j,1} & \sum_{j=1}^{J} x_{j,k}s_{j,2} & \cdots & \sum_{j=1}^{J} x_{j,k}s_{j,N} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{J} x_{j,k}s_{j,1} & \sum_{j=1}^{J} x_{j,k}s_{j,2} & \cdots & \sum_{j=1}^{J} x_{j,k}s_{j,N} \end{pmatrix} \right) \tag{100}$$

or

$$\frac{\partial s_{j,i}}{\partial \sigma_k} = \left(\left(v_{k,i} \right)_{J\times N} \odot \left(s_{j,i} \right)_{J\times N} \right) \odot \left(\left(x_{j,k} \right)_{J\times N} - \left(\sum_{j=1}^{J} x_{j,k}s_{j,i} \right)_{J\times N} \right)$$

(d) Compute the running row mean of $\frac{\partial s_{j,i}}{\partial \sigma_k}$ to obtain $\frac{\partial s_j}{\partial \sigma_k}$:

$$\frac{\partial s_j}{\partial \sigma_k} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \sigma_k \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \sigma_k \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \sigma_k \end{pmatrix}_{J\times 1} \tag{101}$$

5. Bind each column vector resulting from the procedure above to obtain:

$$\frac{\partial s_j}{\partial \Sigma} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \sigma_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \sigma_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \sigma_{K+1} \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \sigma_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \sigma_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \sigma_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \sigma_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \sigma_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \sigma_{K+1} \end{pmatrix}_{J\times (K+1)} \tag{102}$$

6. Find the derivative of δ_j with respect to Σ as

$$\frac{\partial \delta_j}{\partial \Sigma} = \left(\frac{\partial s_j}{\partial \delta_k} \right)^{-1} \frac{\partial s_j}{\partial \Sigma} \tag{103}$$

7. Repeat steps (4) to (6) for the derivatives with respect to π_d. Notice the only difference in computation lies in replacing matrix $\left(\, v_{k,i} \, \right)_{J\times N}$ by $\left(\, D_{d,i} \, \right)_{J\times N}$. The result must be a matrix as the one below:

$$\frac{\partial s_j}{\partial \Pi} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \pi_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \pi_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \pi_d \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \pi_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \pi_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \pi_d \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \pi_1 & \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \pi_2 & \cdots & \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \pi_d \end{pmatrix}_{J\times d} \tag{104}$$

and then the derivative of δ_j with respect to Π would be:

$$\frac{\partial \delta_j}{\partial \Pi} = \left(\frac{\partial s_j}{\partial \delta_k}\right)^{-1} \frac{\partial s_j}{\partial \Pi} \tag{105}$$

8. Build the Jacobian matrix by binding matrices $x_{j,k}$, $\frac{\partial \delta_j}{\partial \Sigma}$, $\frac{\partial \delta_j}{\partial \Pi}$, and $w_{j,q}$:

$$D = \begin{pmatrix} x_{j,k} & \frac{\partial \delta_j}{\partial \Sigma} & \frac{\partial \delta_j}{\partial \Pi} & w_{j,q} \end{pmatrix}_{J\times(2(K+1)+d+Q)} \tag{106}$$

where $w_{j,q}$ are the variables used in the regression for the marginal cost.

9. Compute the variance of the estimators following equation (49) as:

   (a) Calculate $\partial s_j/\partial \xi_j$ as the running row mean of the following matrix:

$$\frac{\partial s_{j,i}}{\partial \xi_j} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \odot \begin{pmatrix} 1-s_{1,1} & 1-s_{1,2} & \cdots & 1-s_{1,N} \\ 1-s_{2,1} & 1-s_{2,2} & \cdots & 1-s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ 1-s_{J,1} & 1-s_{J,2} & \cdots & 1-s_{J,N} \end{pmatrix} \tag{107}$$

   or,

$$\frac{\partial s_{j,i}}{\partial \xi_j} = \left(\, s_{j,i} \, \right)_{J\times N} \odot \left(\, 1 - s_{i,j} \, \right)_{J\times N} \tag{108}$$

then

$$\frac{\partial s_j}{\partial \xi_j} = \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} \partial s_{1,i}/\partial \xi_1 \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{2,i}/\partial \xi_2 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} \partial s_{J,i}/\partial \xi_J \end{pmatrix}_{J\times 1} \tag{109}$$

(b) Calculate $\partial s_j/\partial \xi_k$ as:

$$H = \frac{\partial s_j}{\partial \xi_k} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{J,1} & s_{J,2} & \cdots & s_{J,N} \end{pmatrix} \times \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,J} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N,1} & s_{N,2} & \cdots & s_{N,J} \end{pmatrix} \tag{110}$$

(c) Replace the diagonal of matrix $\frac{\partial s_j}{\partial \xi_k}$ by $\frac{\partial s_j}{\partial \xi_j}$

(d) Compute S_1 of equation (49) for a particular market, namely S_1^t, as:

$$S_1^t = (Z'\xi\xi'Z)$$

(e) Compute S_2 of equation (49) for market t as:

$$S_2^t = Z'H^{-1}V_2H^{-1'}Z$$

where

$$V_2 = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_J \end{pmatrix} - \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_J \end{pmatrix} \times \begin{pmatrix} s_1 & s_2 & \cdots & s_J \end{pmatrix} \tag{111}$$

or,

$$V_2 = diag(s)_{J\times J} - ss'_{J\times J}$$

(f) Compute S_3 of equation (49) for market t as:

$$S_3^t = \frac{1}{ns} Z' H^{-1} V_3 H^{-1'} Z$$

where

$$V_3 = \begin{pmatrix} s_1 - s_1^o \\ s_2 - s_2^o \\ \vdots \\ s_J - s_J^o \end{pmatrix} \times \begin{pmatrix} s_1 - s_1^o & s_2 - s_2^o & \cdots & s_J - s_J^o \end{pmatrix} \tag{112}$$

and s_j^o are observed market shares. If several simulations are carried out in the estimation, then V_3 should be averaged between simulations, which explains the $1/ns$ in the equation above.

For every market, the procedure above will yield a $K \times K$ matrix for S_1^t, S_2^t, and S_3^t. To aggregate markets and obtain one single variance covariance matrix do the following:

1. Average S_1^t, S_2^t, and S_3^t through markets as:

$$\begin{aligned} S_{1_{K\times K}} =& \frac{\xi'\xi}{JT-K} \odot \sum_{t=1}^{T} S_1^t \\ S_{2_{K\times K}} =& \frac{1}{nJT} \sum_{t=1}^{T} S_2^t \\ S_{3_{K\times K}} =& \frac{1}{JT} \sum_{t=1}^{T} S_2^t \end{aligned} \tag{113}$$

2. Add S_1, S_2 and S_3 to obtain the aggregate measure of variance due to the three sources of error:

$$S_{1_{K\times K}} + S_{2_{K\times K}} + S_{3_{K\times K}} = S_{K\times K} \tag{114}$$

3. Compute the variance-covariance matrix:

$$\hat{V}(\hat{\beta}_{GMM}) = (D'ZZ'D)^{-1}(D'ZSZ'D)(D'ZZ'D)^{-1} \tag{115}$$

The squared root of the diagonal of this matrix yields the standard error of each parameter.

7.4 Computation of counterfactuals

Estimation of counterfactual scenarios rely on the contraction mapping of prices for finding the market equilibrium under new market conditions. The contraction mapping can be solved through fixed point algorithms as in the case of average utilities that equal observed to predicted market shares. However, the functional form to approach the fixed point has strong implications over the vector of equilibrium prices. In other words, counterfactual scenarios can have multiple equilibriums. This does not occur for the contraction mapping of average utilities in which there is a unique fixed point. The existence of multiple equilibriums or multiple equilibrium prices that solve the pricing equation of each firm, renders difficulties for hypothesis testing in the counterfactuals. One way to approach this issue is to try to compute every equilibrium, test the hypothesis in each of them, and then decide upon the conclusion the majority of them arrive to. In this subsection we focus on the matrix-like notation and steps to compute equilibrium prices and market shares in the counterfactual analysis and leave the discussion on the different algorithms to approach the fixed point to the next section where we also revise the literature. We would focus on the counterfactuals that change the competitive structure of the market such as mergers in the context of separable income and prices as in Nevo. Analogs to the non-separable case only require changing matrices μ_{ijt} and $\partial s_{jt}/\partial p_{kt}$ accordingly.

Suppose after estimation of structural parameters with an observed market structure of various multi-product firms, the market turns into a monopoly with all firms colluding. The vector of new equilibrium prices and market shares under the monopolistic structure is computed in following steps:

For every market:

1. Set the seed for the vector of prices, p^*, for instance, a column vector of 1's.

2. Predict the average utilities as:

$$\delta^* = X\hat{\theta}_1 + \hat{\xi} \tag{116}$$

   where X includes p^*.

3. Compute matrix μ_{ijt} as in equation (68) and market shares, $s(p^*)$ as in equation (74).
4. Calculate price derivatives as in equation (80).
5. Compute the vector of markups under the new market structure:

$$b^* = \Omega(p^*)^{-1}s(p^*) \tag{117}$$

   where

$$\Omega(p^*) = -\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \frac{\partial s_j}{\partial p_k^*} \tag{118}$$

6. Add b^* to $\widehat{mc}$ to obtain a new vector of prices p'.
7. Make $p^* = p'$.
8. Repeat steps (2) to (7) until $||p^* - p'|| < tolerance$

Estimation of counterfactual scenarios in light of Nevo's application has strong assumptions on the cost function and the producer/retailer relations. As the author mentions, the fact that marginal costs are held constant between the observed scenario and the counterfactual means, on the one hand, that he is ruling out any potential cost savings due to the merger and, on the other hand, that producer/retailer relations which also determine marginal costs do not change after the merger, in other words, the merger has no effect on the price negotiation between producers and retailers. Another key strategy for the computation of counterfactuals is holding fixed the structural parameters estimated from the observed scenario. This implies the counterfactual situation does not affect consumer taste for any product observable characteristic, which might not be the case if firms in the counterfactual, for example, carry out marketing efforts that change consumer preferences. The lesson here is that the researcher must consider all implications of the assumptions on which counterfactuals are build, and analyze how relaxing those assumptions could affect results.

8 Applications

Structural demand models are usually used in the analysis of industries with market power, policy evaluation and understanding of consumer preferences. Applications abound in all these matters and they have become important tools for the design of competition policy in different sectors. Take, for instance, the health care market. Ho and Lee (2013) study the impact of insurer competition on negotiated hospital prices by estimating the demand for hospitals using the random coefficients logit outlined in this document but modelling the supply side in two steps: first insurers set premiums charged to consumers and then they engage in a Nash bargaining game with health service providers from which service prices are determined. Using the estimates of the demand of patients for hospitals, the authors can compute counterfactual scenarios assuming a hospital is drop from the provider network of each insurer and derive a measure of the patient's willingness to pay for each hospital. Willingness to pay helps in the identification of the bargaining power parameters in the supply side. Grennan (2013) is another example of how structural demand models help in the identification of bargaining parameters if prices are not determined by direct competition between producers but by bilateral negotiations between producers and retailers. The author studies the impact of bargaining power on the price discrimination of medical devices, in particular, coronary stents. He models the demand of every patient/doctor combination for coronary stents in every market defined as a hospital in a month. Since preferences for coronary stents might change due to recruitment of new doctors or changes in the allocation of patients to physicians, then the author models the unobserved stent quality ξ as an autoregressive process and further captures aggregated measures of stent quality by using a nested logit model with nests defined by the type of stent (of which there are only two). Gowrisankaran et al. (2014) also estimate the demand of patients for hospitals to examine the effects of hospital mergers in negotiated health service prices but add the demand of enrollees for health insurance companies for identification of the bargaining power. The timing of the game between insurers and providers is: 1) insurers negotiate prices with hospitals, 2) insurer set premiums, 3) patients choose to which insurer enroll, 4) patients get sick and go to the hospital. The solution to this game is found using backwards induction, solving the demand of patients for hospitals first, then the profit maximization problem of health insurers choosing premiums, and finally estimating the Nash bargaining game between insurers and provider.

Most applications have been developed in markets of products of mass consumption where product differentiation is crucial. Berry et al. (1995) introduce the empirical framework in the U.S. automobiles market to derive reasonable substitution patterns between products and reasonable markups. Nevo (2001) studies the ready-to-eat cereal market and uses the static utility and oligopolistic structure of the model outlined above to measure the impact of the competitive structure on price-cost margins. Petrin (2002) uses the empirical framework to measure changes in consumer welfare due to the introduction of minivans in the automobiles market. Dubé (2005) also measures changes in consumer welfare due to mergers in the carbonated soft drink industry. Applications in trade, merger and environmental policy include Berry and Pakes (1993) and Berry et al. (1999). The empirical framework is flexible enough to accommodate nonlinear pricing strategies, sequential consumer choices and demand for product segments in a nested logit context. For example, Slade (2004) studies the brewing industry in the U.K using a nested logit approach.

In the telecommunications market, applications of structural demand models have been used mainly to measure the extent of switching costs due to handset bundling and the effects of number portability on consumer welfare. Kim (2006) models the static utility a consumer gets from using a certain amount of minutes under a voice plan. Consumers choose network operators and then their minutes consumption. However, network operator choice is dynamic. Every period, the probability that the consumer switches to a new operator or remains with the same operator is modeled as a first-order Markov process. Grzybowski (2008) relies only on the static utility specification to identify switching costs in the mobiles market in the UK. The author models consumer utility from choosing a network operator and includes in the utility specification a matrix with a set of dummy variables accounting for switching from one operator to the other. In the case of Dubé et al. (2009b) the dummy variables in the utility function denote brand loyalty and the consumer only derives additional satisfaction if she remains buying the same brand in the next period. The probability that a consumer loyal to brand j will become loyal to another brand, as before, is modeled as a first-order Markov chain.

Applications of structural demand models to the telecommunications markets make explicit the fact that consumer choices for certain products are dynamic rather than static. Besides the choice of network operator, choices of durable, storable, or experience goods are also determined by a dynamic indirect utility level. We do not discuss the dynamic context of structural demand models but we have to mention that treatment of dynamics have important implications on the parameter estimates, in particular, on the substitution patterns. Gowrisankaran and Rysman (2012) state modeling dynamics turns the consumer purchase problem into a capital investment problem that depends on the difference in the present-value price of the product between this period and the next. For example, in the automobile industry, if prices are falling, static models will predict a steady increase in sales, while dynamic models would predict sales increase the most as prices stop falling. In other words, in a dynamic context, consumers who value high quality products might prefer waiting several periods before buying until they see prices stop falling and then level-off. Thus, in a static model price sensitivity for low quality products that are bought when prices are low, will be high, but price sensitivity for high quality products which are bought as prices level-off, will be low. This suggests price sensitivity in a static model when goods are durable will be biased downwards. Estimation of dynamic demand models rely mainly on the specification of a fixed point algorithm on the Bellman equation indicating the present-value utility in the current period is a function of the present-value utility of the previous period. Applications of dynamic demand models were firstly motivated by Rust (1987) who studied optimal bus engine replacement decisions and ever since have abounded in different markets such as: the market for digital cameras (Carranza, 2007), the camcorder industry (Gowrisankaran and Rysman, 2012), the US computer printer market (Melnikov, 2000), the laundry detergents industry (Hendel and Nevo, 2006), the market for ketchup sauce (Erdem et al., 2003), the automobiles market (Esteban and Shum, 2007; Schiraldi, 2011), etc.

Since estimation of structural demand models imply certain computational burden in particular when calculating numeric integrals, many authors have studied the impact of using different numeric algorithms to approach both the minimization of the GMM function, the fixed points, and the integrals for market shares and price derivatives. Knittel and Metaxoglou (2008) study criterion functions that are not globally concave or convex and measure the impact of choosing different local extrema in own and cross price elasticities. Since the minimization procedure may stop in regions where the GMM function reaches a saddle point or a local extrema, choosing which solution to report in non trivial, specially when different solutions may yield differences by factors of over 100 in own and cross price elasticities and differences in the direction of changes in welfare. The authors compare parameters obtained with the Nelder-Mead simplex, the Generalized Pattern Search, the Mesh Adaptive Direct Search and the SolvOpt algorithms, among others, and show that the SolvOpt algorithm arrives to the global minimum across many different sets of starting values. Their overall conclusion is that "for an exhaustive nonlinear search, researchers will need to use multiple starting values, at least 50, and multiple algorithms". Dubé et al. (2009a) focus on the impact of nested fixed point algorithms on parameter estimates and on how the inner loop error can propagate to the outer loop. The authors find fixing loose inner loop tolerances and standard outer loop tolerances will yield convergence problems. They also show having loose tolerances in both loops will derive in biased parameter estimates that do not minimize the objective function.

9 Conclusions

This document is concerned with the theory, estimation and programming of structural demand models in economics. First it gives a brief overview of how economic models of markets with imperfect competition have evolved from early theoretical studies of asymmetric information, pricing strategies, and product differentiation to advanced empirical approaches for the estimation of structural parameters such as price elasticities and marginal costs. Second, it outlines the basic economic intuition behind the random coefficients discrete choice models that were introduced in the mid 1990's to undergo hypothesis testing in industries with market power and highly differentiated products. Third, it gives a detailed description of the random coefficients logit model for the demand side of the market and then introduces the supply side under the argument it helps identify the endogenous effect of price on equilibrium market shares. Both the demand and the supply side rely on important assumptions for identification and equilibrium. For instance, it assumes a particular distribution for the random shock to the utility function of a consumer and a Nash-Bertrand equilibrium in firm competition. Fourth, it outlines the estimation procedure proposed by Berry et al. (1995) and then it gives the matrix-like notation that should be taken to computer programming for estimation of these models. Finally a brief literature review on applications of such models is provided in the last section.

References

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25:242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S., Levinsohn, J., and Pakes, A. (1999). Voluntary Export Restraints on Automobiles: Evaluation a Strategic Trade Policy. *The American Economic Review*, 89(3):400–430.

Berry, S., Linton, O., and Pakes, A. (2004). Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems. *The Review of Economic Studies*, 71(3):613–654.

Berry, S. and Pakes, A. (1993). Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis. *The American Economic Review, Papers and Proceedings*, 83:247–252.

Bresnahan, T. (1989). *Empirical Studies of Industries with Market Power*, volume 2. North Holland.

Brown, B. and Walker, M. (1989). The Random Utility Hypothesis and Inference in Demand Systems. *Econometrica*, 57(4):815–829.

Caplin, A. and Nalebuff, B. (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, pages 25–59.

Carranza, J. E. (2007). Estimation of demand for differentiated durable goods. *Manuscript, University of Wisconsin*.

Caves, R., Porter, M., and Spence, A. (1980). *Competition in the open economy: A model applied to Canada*. Harvard University Press.

Chintagunta, P., Erdem, T., Rossi, P., and Wedel, M. (2006). Structural Modeling in Marketing: Review and Assessment. *Marketing Science*, 25(6):604–616.

Dubé, J. (2005). Product differentiation and merger in the carbonated soft drink industry. *Journal of Economics and Management Strategy*, 14(4):879–904.

Dubé, J., Fox, J., and Su, C. (2009a). Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand. *National Bureau of Economic Research*, (14991).

Dubé, J., Hitsch, G., and Rossi, P. (2009b). Do Switching Costs Make Markets Less Competitive. *Journal of Marketing Research*, 46(4):435–445.

Eaton, C. and Lipsey, R. (1980). Exit barriers and entry barriers: The durability of capital as a barrier to entry. *The Bell Journal of Economics*, 11(2):721–729.

Einav, L. and Levin, J. (2010). Empirical Industrial Organization: A Progress Report. *Journal of Economic Perspectives*, 24(2):145–162.

Erdem, T., Imai, S., and Keane, M. (2003). Brand and Quantity Choice Dynamics Under Price Unceritainty. *Quantitative Marketing and Economics*, 1(1):5–64.

Esteban, S. and Shum, M. (2007). Durable-goods oligopoly with secondary markets: the case of automobiles. *The RAND Journal of Economics*, 38(2):332–354.

Geroski, P. (1982). Simultaneous equations models of the structure-performance paradigm. *European Economic Review*, 19(1):145–158.

Gowrisankaran, G., Nevo, A., and Town, R. (2014). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *The American Economic Review*, 205(1):172–203.

Gowrisankaran, G. and Rysman, M. (2012). Dynamics of Consumer Demand for New Durable Goods. *Journal of Political Economy*, 120(6):1173–1219.

Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *The American Economic Review*, 103(1):145–177.

Grzybowski, L. (2008). Estimating Switching Costs in Mobile Telephiny in the UK. *Journal of Industry, Competition and Trade*, 8(2):113–132.

Heckman, J. and Leamer, E. (2007). *Handbook of Econometrics*, volume 6. North Holland.

Hendel, I. and Nevo, A. (2006). Measuring the Implications of Sales and Consumer Inventory Behavior. *Econometrica*, 74(6):1637–1673.

Ho, K. and Lee, R. (2013). Insurer competition in health care markets. *National Bureau of Economic Research*, (19401).

Kim, J. (2006). Conusmer' Dynamic Switching Desicions in the Cellular Service Industry. *NET Institute*, (0624).

Knittel, C. and Metaxoglou, K. (2008). Estimation of random coefficient demand models: Challenges, difficulties and warnings. *National Bureau of Economic Research*, (14080).

Martin, S. (1979). Advertising, concentration and profitability: the simultaneity problem. *The Bell Journal of Economics*, 10:639–647.

McFadden, D. (1978). Modelling the Choice of Residential Location. *Cowles Foundation Discussion Papers*, (477).

Melnikov, O. (2000). Demand for Differentiated Durable Products: The Case of the U.S Computer Printer Market. *Manuscript, Yale University*.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industrial. *Econometrica*, 69(2):307–342.

Nevo, A. and Rosen, A. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, 94(3):659–671.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *The*

Journal of Political Economy, 110(4):705–729.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Schiraldi, P. (2011). Automobile replacement: a dynamic structural approach. *The RAND Journal of Economics*, 42(2):266–291.

Schmalensee, R. (1989). *Inter-industry studies of structure and performance*, volume 2. Handbook of Industrial Organization.

Slade, M. (2004). Market power and joint dominance in UK brewing. *The Journal of Industrial Economics*, 52(1):133–163.

Stigler, G. (1983). *The organization of industry.* University of Chicago Press Economic Books.

Stiglitz, J. (1977). Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies*, 44:407–430.

Strickland, A. an Weiss, L. (1976). Advertising, concentration and price-cost margins. *The Journal of Political Economy*, 84(5):1109–1121.

Tirole, J. (1988). *The Theory of Industrial Organization.* MIT Press.