Allowing for Heterogeneous Preferences on Unobserved Quality in Random Coefficient Demand Models

Mons Chan
University of Minnesota

Amit Gandhi
University of Wisconsin

Kyoo il Kim
Michigan State University

Amil Petrin
University of Minnesota

February 10, 2018

Abstract

This paper generalizes the random coefficient demand model by allowing for heterogeneous preferences over unobserved product quality. In empirical settings much of demand loads on unobserved product quality. Allowing for heterogeneous preferences over this unobserved quality has an obvious economic interpretation which is useful for the purposes of estimating product quality and demand. We establish conditions under which the generalized model is identified. We show that the standard computational method following Berry (1994) and BLP (1995) doesn't apply to this model, and develop an alternative estimator which we show to be globally convergent. We conclude with monte carlo evidence of the performance of our estimator, and demonstrate that applying the standard BLP estimator to data in which agents have heterogeneous preferences on unobserved quality generates biased results.

Keywords: Demand Estimation, Unobserved Product Quality, Discrete Choice Models, Horizontal Differentiation, Invertibility.

1 Introduction

In this paper we study a class of random utility models that allows for horizontal product differentiation to enter an otherwise purely vertically differentiated market. The model generalizes the standard random coefficient model by allowing for consumer heterogeneity to interact with a product j's unobserved attribute ξ_j. We seek to establish three basic results concerning this model that is relevant for empirical work. First, the discrete choice demand literature to date allows for heterogeneous preferences on observed characteristics, but not on unobserved characteristics. This is potentially problematic, as in empirical settings much of demand loads on unobserved product quality. Allowing for heterogeneous preferences over this unobserved quality (which might represent reliability, marketing, style, or other unmeasured product characteristics) thus has an obvious economic interpretation which is useful for the purposes of measuring product quality and demand. Second, a key issue in generalizing the model in this way is that the standard computation method, following Berry (1994) and Berry, Levinsohn and Pakes (1995) (henceforth BLP), does not apply to this model, as the conditions under which the proposed mapping is a contraction are no longer generally satisfied. Third, we build on work by Berry, Gandhi and Haile (2013) (henceforth BGH) and Berry and Haile (2014) (henceforth BH) and establish conditions under which the generalized model is both identified and has a globally convergent solution method.

To keep this draft short, we have omitted most of the formal proofs, some of the intermediate derivations, and do not formally describe that notation which is otherwise standard in random coefficient models of demand. This draft represents an early work in progress, so our current empirical/computational results are limited to preliminary Monte Carlo exercises. We are currently working on expanding these Monte Carlo results as well as applying the generalized framework/estimator to several micro data sets in which we believe marketing and other unobservable product characteristics play an important role in demand.

2 The Model

For simplicity, we consider a linear in random coefficients utility model[1]

$$u_{ij} = \mathbf{x}_j \beta_i - \alpha p_j + \lambda_i \xi_j + \varepsilon_{ij} \tag{1}$$

where $(\mathbf{x_j}, \xi_j) \in \mathbb{R}^K \times \mathbb{R}$ is product j's vector of characteristics and (β_i, λ_i) is consumer i's vector of random coefficients which characterize consumer i's tastes over a product's characteristics. The preference shock, ε_{ij}, is a standardized type-I extreme value shock which is independent of the random coefficients and across the set of products $j \in \mathcal{J} = \{0, 1, \ldots, J\}$. A market is then defined by $(\mathcal{J}, \mathcal{X})$, with the product and market characteristics represented by $\mathcal{X} = (x, p, \xi) \in \mathbb{R}^{KJ} \times \mathbb{R}^J \times \mathbb{R}^J$.

3 Identification and Computation

There are two key problems we need to solve in order to estimate the model described by (1). The first is to establish conditions under which the model (and in particular, ξ) is identified. We generally follow the identification and invertibility conditions outlined in BGH and BH. The second problem is computational, as we can no longer use the Berry/BLP contraction to recover ξ. To solve this, we establish conditions under which there exist globally convergent methods for recovering the unique vector of ξ_j terms in each market.

3.A The Computational Problem

Let $\hat{s}(\delta) = (\hat{s}_0(\delta), \ldots, \hat{s}_J(\delta))$ denote the vector of predicted shares in a market for any vector of mean utilities $\delta \in \mathbb{R}^J$ of the inside goods. Given the observed market shares $s \in R^{J+1}$ where $s_j > 0$ for all j and $\sum_j s_j = 1$, BLP propose iterating the mapping

$$f(\delta) = \delta + \ln(s) - \ln(\hat{s}(\delta)) \tag{2}$$

to solve $\hat{s}(\delta) = s$. Their argument is based on showing that f is a contraction mapping with modulus strictly less than 1 over a certain domain. Unfortunately, in the case of (1), the

[1] With the notable exception of Das, Olley and Pakes (1994), all existing empirical work which utilizes the mixed logit form of (1) restricts $\lambda_i = 1$ (i.e., no heterogeneity in the random coefficient on the unobservable).

mapping (2) is no longer a contraction and consequently does not converge to the solution. This can be readily confirmed in practice. The theoretical source of the problem is the fact that a product's "own demand" can now become more responsive to "own quality" when a random coefficient λ_i is allowed to interact with ξ_j. To see the issue more formally, recall that a key condition for f to be a contraction is that $\partial f_j(\delta)/\partial\delta_j > 0$ for all products $j > 0$. Of course,

$$\frac{\partial f_j(\delta)}{\partial\delta_j} = 1 - \frac{1}{\hat{s}(\delta)}\frac{\partial\hat{s}(\delta)}{\partial\delta_j}$$

and since $\hat{s}_j < 1$, the condition requires that $\partial\hat{s}/\partial\delta_j > 0$ be sufficiently small, i.e., that own demand is not "too responsive" to own quality. However, this requirement need not be satisfied in the model with a random coefficient on quality.

3.B A Solution

We now propose an alternative computational solution that can be used to estimate the model (1). First let us establish some notation. Let $\mathcal{S}$ be the $J+1$ dimensional simplex, and let $\mathcal{S}^*$ denote the interior of $\mathcal{S}$. Partition x_j as $(x_j^{(1)}, x_j^{(2)})$, with $x_j^{(1)} \in \mathbb{R}$. Let $x^{(1)} = (x_1^{(1)}, \ldots, x_J^{(1)})$ and $x^{(2)} = (x_1^{(2)}, \ldots, x_J^{(2)})$. Define $\delta_j = x_j^{(1)} + \xi_j$ and let $\delta = (\delta_1, \ldots, \delta_J)$. The model (1) gives rise to a demand system $\sigma : \mathbb{R}^J \to \mathcal{S}$ that maps any vector of qualities $\xi = (\xi_1, ..., \xi_J)$ over the J inside goods to a vector of market shares. We observe a particular vector of market shares $s \in \mathcal{S}^*$, and seek to solve the system $s = \sigma(\xi)$ in the underlying qualities ξ.[2] It is straightforward to show that demand derived from the model (1) satisfies the following properties:

Assumption 1. $\sigma : \mathbb{R}^J \to \mathcal{S}^*$ and is continuous.

Assumption 2. For any $j > 0, \sigma_j(\xi_j, \xi_{-j})$ is strictly increasing in ξ_j and approaches 1 as $\xi_j \to \infty$ and approaches 0 as $\xi_j \to -\infty$.

Assumption 3. For any $k \neq j$, $\sigma_k(\xi_j, \xi_{-j})$ is non-increasing in ξ_j.

Assumption 4. For any $\xi \in \mathbb{R}^J$ and any nonempty $\mathcal{K} \subseteq \{1, \ldots, J\}$, there exist $k \in \mathcal{K}$ and $\ell \notin \mathcal{K}$ such that $\sigma_\ell(\xi)$ is strictly decreasing in ξ_k.

Assumption 5. $F_u(\cdot|\mathcal{X}) = F_u\left(\cdot|\delta, x^{(2)}, p\right)$, where $F_u(u_{i1}, \ldots, u_{iJ}|\mathcal{X})$ is the conditional joint distribution of consumer utility, which is i.i.d across consumers and markets.

[2] Note that $\sigma(\xi)$ is more properly written $\sigma(\xi; \theta, x, p, s)$, but we suppress the (θ, x, p, s) to simplify notation.

Note that assumptions 2-4 are equivalent to the condition that goods $(0, 1, \ldots, J)$ are *connected substitutes in* ξ as defined in BGH and BH. Assumption 5 is a restatement of the index restriction from BH. Assumption 1 follows from the fact that the assumed extreme value distribution on ε_{ij} for $j \geq 0$ is continuous and has full support over $\mathbb{R}$. Assumption 2 follows from the additively separable nature of $\lambda_i \xi_j$ term in preferences. Assumption 3 follows from the discrete choice nature of demand, which rules out the products being complements. Assumption 4 was introduced by BGH as a general property of demand models and the inclusion of the ε_{ij} ensures that it holds (see their paper for various interpretations of the condition). To satisfy assumption 5, we will impose that $\lambda_i = \beta_{ik} \, \forall \, i$, for some observable characteristic $k \in \{1, \ldots, K_x\}$.

The ultimate goal is to solve the system $s = \sigma(\xi)$. Define $\phi : \mathbb{R}^J \to \mathbb{R}^J$ as $\phi(\xi) \equiv \sigma(\xi) - s$. We now establish an important key property concerning ϕ.

Definition 3.1. A mapping $F : D \subset \mathbb{R}^j \to \mathbb{R}^J$ is an M-function if it is inverse isotone and off-diagonally antitone.

Lemma 3.1. *The mapping ϕ is a continuous, surjective, M-function.*

Proof. Demand function σ (and thus ϕ) is continuous by construction and surjective under assumptions 1-3 following the existence argument in the Appendix of Berry (1994). Under assumptions 2-4, theorem 1 in BGH shows that ϕ is inverse isotone. Mapping ϕ is also off-diagonally antitone by assumption 3. □

The next theorem is a restatement of Theorem 3.3 in Rheinboldt (1970) and proves that the Gauss-Seidel and Gauss-Jacobi processes applied to our demand mapping ϕ computes the unique solution (ξ^*) in a globally convergent way.

Theorem 3.2. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous, surjective M-function, and $\epsilon \in (0, 1]$ a given number. Then for any $z \in \mathbb{R}^n$, any starting point $x^0 \in \mathbb{R}^n$, and any sequence $\{\omega_k\} \in [\epsilon, 1]$ of relaxation factors, the Gauss-Seidel process as well as the Jacobi process converge to the unique solution x^* of $Fx = z$.*

4 Computation

We estimate the model using a nested fixed point approach similar to that employed by BLP. Define $\theta^0 \in \mathbb{R}^{K_\theta}$ to be the vector of true mean and distributional parameters, common across individuals and markets. Then given the above assumptions, for a particular parameter guess θ there exists a unique vector ξ^* which solves the J-dimensional equation $\phi(\xi) = 0$. By theorem 3.2, we know that one globally convergent method of solving for ξ^* is the Gauss-Seidel method, defined as iterating over:

Solve: $\phi_j(\xi_1^{m+1}, ..., \xi_{j-1}^{m+1}, \xi_j, \xi_{j+1}^m, ..., \xi_J^m) = 0$ for ξ_j.

Update: $\xi_j^{m+1} = (1 - \omega_m)\xi_j^m + \omega_m \xi_j$

where $\{\omega_m\} \subset (0, 1], m = 0, 1, ...$ is any sequence of relaxation factors. This process can be very slow, so we also employ a variant of Newton's Method for finding the root of a vector-valued function, defined as iterating over:

$$\xi^{m+1} = \xi^m - \gamma^m \phi_\xi(\xi^m)^{-1}\phi(\xi^m) = \xi^m - \gamma^m \Delta^m$$

where $\phi_\xi(\xi^m)$ is the $J \times J$ Jacobian matrix for ϕ and $\gamma^m = (1 + 10 * \max(|\Delta^m|))^{-1}$ is a "smart" softening factor. In practice, our algorithm first employs Newton's Method, and switches to the guaranteed but slower Gauss-Seidel if it detects a problem.

To deal with objective functions which may have many local minima, we develop a global search algorithm which we use when estimating $\theta \in \mathbb{R}^{K_\theta}$. The algorithm is as follows. Select initial trust region $\Theta^* \subset \mathbb{R}^{K_\theta}$ where Θ^* is a K_θ-dimensional hypercube. Evaluate L initial guesses, where each parameter guess $\tilde{\theta}_\ell$ is drawn from a Sobol sequence defined on Θ^*. Choose $\bar{N}$ best guesses, ordered $\{\hat{\theta}_1, .., \hat{\theta}_{\bar{N}}\}$, and start the search algorithm (Nelder Mead etc) at $\hat{\theta}_1$, giving initial best guess θ_1^*. Iteratively update θ_n^* by restarting the search at $r_n\theta_{n-1}^* + (1 - r_n)\hat{\theta}_n$, where $r_n \in (r_{n-1}, 1]$, until convergence.

5 Monte Carlo Exercise and Results

We examine the performance of our estimator using Monte Carlo studies in a simple oligopoly demand/supply equilibrium setting. Since cost/demand parameters are constant across the M otherwise independent markets, we suppress the market subscript t to simplify notation. Every market contains a set N of multi-product firms which each produce J_f goods. Let J denote the set of goods offered in a particular market. Firm f

sets prices to maximize profits, defined as

$$\pi_f = \max_{p_f} \sum_{j}^{J_f} (p_{fj} - mc_{fj}) q_{fj}(p)$$

where p_f is firm f's vector of prices, $p = \{p_f\}_{f \in N}$ is the set of equilibrium prices in that market, and marginal costs are

$$mc_{fj} = \gamma w_{fj} + \zeta_{fj}$$

Here, γ is a vector of cost function coefficients common across firms, goods, and markets, $w_{fj} \sim U_{|0,1|}$ is a vector of observed product-level cost shifters and $\zeta_{fj} \sim U_{|-1/2,1/2|}$ is an unobserved cost shock. Demand is derived from a simple version of the indirect utility function considered above with only one observed demand shifter, where for customer i,

$$u_{ij} = \beta_0 + \beta_{i1}(x_{j1} + \xi_j) - \alpha p_j + \varepsilon_{ij}, \quad \beta_{i1} \sim N(\beta_1, \sigma_1^2) \tag{3}$$

We set $x_{j1} \sim U_{|1,3|}$ and $\xi_j \sim U_{|-1/2,1/2|}$. The true demand parameter values are $(\alpha, \beta_0, \beta_1, \sigma_1) = (5, 0, 3, 3)$. We generate the data with $M = 100$, $M = 500$ and $M = 5000$ in order to examine both small sample performance and the consistency of our estimator. Every market has two firms, each with a random number $J_f \in \{2, 3\}$ of products.

For now we perform two main exercises. First we estimate the model using our estimator. Second, we naively estimate the model using the BLP estimator, ignoring the random coefficient on ξ_j.[3] This gives us both an idea of our estimator's performance, and the bias that might result in frameworks which erroneously restrict $\lambda_i = 1$. In both cases we use two-step GMM with BLP instruments and report the standard errors based on 200 replications of the estimates (using 200 different data sets).

The Monte Carlo results are displayed in Table 1. Our estimator (which we label "CGKP") is consistent and clearly converges to the true parameter values as the number of markets M increases. The BLP estimator where λ_j is restricted to 1 is clearly biased. The random coefficient β_{i1} in particular is biased significantly downwards. This suggests that allowing for a random coefficient on ξ_j may be important for calculating own and cross-price elasticities. As mentioned above, our next step is to test *how* important by taking our estimator to the BLP automobile data, as well as several other micro data sets in which we believe unobservable product characteristics play an important role in demand.

[3]Note that we cannot test the bias/performance of the BLP estimator on the true model for the reasons detailed above.

Table 1: Preliminary Monte Carlo Experiment Results

Parameters	**True**		**CGKP**			**BLP**	
Alpha (α)	5.000	4.979	4.996	5.000	4.910	4.940	4.941
		(0.535)	(0.239)	(0.076)	(0.566)	(0.253)	(0.079)
Beta 0 (β_0)	0.000	0.062	0.016	0.007	0.003	0.061	0.085
		(1.032)	(0.462)	(0.146)	(1.112)	(0.497)	(0.157)
Beta 1 (β_1)	3.000	3.010	3.000	3.000	2.882	2.879	2.820
		(0.094)	(0.041)	(0.013)	(0.124)	(0.054)	(0.017)
Sigma 1 (σ_1)	3.000	2.987	2.999	2.999	2.867	2.862	2.853
		(0.309)	(0.137)	(0.043)	(0.349)	(0.156)	(0.048)
Markets		**100**	**500**	**5000**	**100**	**500**	**5000**
Mean Obj. Value		3.074	2.978	3.291	3.002	3.105	5.283
Replications		200	200	200	200	200	200
Performance							
AVG Time (mins)		0.5	0.9	10.1	3.0	5.4	53.2

Appendix

A Derivation of Newton's Method

Herein we describe and derive the "Fast Method" of solving for ξ_t for each market t. The Fast Method is actually an implementation of Newton's Method for finding the root of a vector-valued function. First we will review the model, then derive the theoretical and empirical versions of the root-finding algorithm. Throughout, we suppress the t notation for the market, since we solve for ξ_t in each market separately. Let J be the number of goods in the arbitrary market under consideration here.

Derivation of Market Shares

The probability of purchasing good j for individual i is expressed as

$$s_{ij} = \frac{e^{\delta_j + \mu_{ij}}}{1 + \sum_l e^{\delta_l + \mu_{il}}} \tag{4}$$

where

$$\begin{aligned} \delta_j &= x_j'\beta + \beta_c \xi_j \\ \mu_{ij} &= \sigma_c \nu_{ic} \xi_j + \sum_k \sigma_k \nu_{ik} x_{jk} - \alpha p_j e^{-m-\sigma_y \nu_{iy}} \end{aligned}$$

For notational ease, let $\theta = \{\beta, \sigma, \alpha\}$ be the vector of common taste parameters, and let ν_i be the vector of individual demographic taste parameters. Note that β_c and σ_c are the common taste parameters for some product characteristic $c \in \{1..K\}$. Note that for a given θ and ν_i, we can express these as functions of ξ:

$$s_{ij}(\xi; \theta, \nu_i) = \frac{e^{\delta_j(\xi_j;\theta) + \mu_{ij}(\xi_j;\theta,\nu_i)}}{1 + \sum_l e^{\delta_l(\xi_l;\theta) + \mu_{il}(\xi_l;\theta,\nu_i)}} \tag{5}$$

Then, for a given θ, the theoretical market share for good j is

$$s_j(\xi; \theta) = \int_\nu s_{ij}(\xi; \theta, \nu) dF(\nu) \tag{6}$$

and the empirical expression, where we simulate a market with R individuals, is

$$\hat{s}_j(\xi;\theta) = \frac{1}{R}\sum_{i=1}^{R} s_{ij}(\xi;\theta,\nu_i) \tag{7}$$

The partial derivative of the share for good j with respect to ξ_j is then

$$\begin{aligned}\frac{\partial}{\partial \xi_j} s_j(\xi;\theta) &= \frac{\partial}{\partial \xi_j}\int_\nu s_{ij}(\xi;\theta,\nu)dF(\nu) \\ &= \int_\nu \lambda(\nu) s_{ij}(\xi;\theta,\nu)(1 - s_{ij}(\xi;\theta,\nu))dF(\nu) \end{aligned} \tag{8}$$

and the derivative of the share for good j with respect to ξ_l for $l \neq j$ is

$$\frac{\partial}{\partial \xi_l} s_j(\xi;\theta) = \int_\nu -\lambda(\nu) s_{ij}(\xi;\theta,\nu) s_{il}(\xi;\theta,\nu)dF(\nu) \tag{9}$$

where $\lambda(\nu) = \beta_c + \sigma_c\nu_c$, the coefficients on ξ_j. Empirically, we can write these as

$$\frac{\partial}{\partial \xi_j} \hat{s}_j(\xi;\theta) = \frac{1}{R}\sum_{i=1}^{R} \lambda_i s_{ij}(\xi;\theta,\nu_i)(1 - s_{ij}(\xi;\theta,\nu_i)) \tag{10}$$

$$\frac{\partial}{\partial \xi_l} \hat{s}_j(\xi;\theta) = \frac{1}{R}\sum_{i=1}^{R} -\lambda_i s_{ij}(\xi;\theta,\nu_i) s_{il}(\xi;\theta,\nu_i) \tag{11}$$

where $\lambda_i = \beta_c + \sigma_c\nu_{ic}$.

Root Finding Algorithm

In order to solve for the market-specific vector ξ which contains the unobserved characteristics of all goods j in this particular market, we employ a version of Newton's Root Finding Method. Specifically, let

$$f(\xi;\theta) = s(\xi;\theta) - \bar{s} \tag{12}$$

where $s(\xi;\theta)$ is the $J \times 1$ vector of theoretical market shares, dependent on ξ and θ, and $\bar{s}$ is the $J \times 1$ vector of observed shares in the data. Note that from here on, I suppress the θ in order to simplify the notation. However, all functions of ξ are likewise conditional on θ. Thus, we want to find the $\tilde{\xi}$ which sets equation 12 to (a vector of) zero(s).

Newton's Method for finding the root of (12) is to iterate on the following equation:

$$\xi_{m+1} = \xi_m - \gamma_m [Jf(\xi_m)]^{-1} f(\xi_m) = \xi_m - \gamma_m \Delta_m \tag{13}$$

until convergence[4]. Here, γ_m is a "softening" factor[5], and $Jf(\xi_m)$ denotes the $J \times J$ Jacobian matrix evaluated at ξ_m. We construct the Jacobian as follows. First, note that

$$\frac{\partial f_j(\xi_m)}{\partial \xi_{jm}} = \frac{\partial (s(\xi_m) - \bar{s})}{\partial \xi_{jm}} = \frac{\partial s_j(\xi_m)}{\partial \xi_{jm}} \tag{14}$$

where $f_j(\xi_m)$ is the j'th element of $f(\xi_m)$. Thus, we can write the Jacobian as

$$Jf(\xi_m) = \begin{bmatrix} \frac{\partial f_1(\xi_m)}{\partial \xi_{1m}} & \cdots & \frac{\partial f_1(\xi_m)}{\partial \xi_{Jm}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_J(\xi_m)}{\partial \xi_{1m}} & \cdots & \frac{\partial f_J(\xi_m)}{\partial \xi_{Jm}} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_1(\xi_m)}{\partial \xi_{1m}} & \cdots & \frac{\partial s_1(\xi_m)}{\partial \xi_{Jm}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_J(\xi_m)}{\partial \xi_{1m}} & \cdots & \frac{\partial s_J(\xi_m)}{\partial \xi_{Jm}} \end{bmatrix} \tag{15}$$

or empirically,

$$\hat{J}f(\xi_m) = \frac{1}{R} \begin{bmatrix} \sum_{i=1}^{R} \lambda_i s_{i1}(\xi_m;\nu_i)(1 - s_{i1}(\xi;\theta,\nu_i)) & \cdots & \sum_{i=1}^{R} -\lambda_i s_{i1}(\xi_m;\nu_i) s_{iJ}(\xi;\nu_i) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{R} -\lambda_i s_{iJ}(\xi_m;\nu_i) s_{i1}(\xi;\nu_i) & \cdots & \sum_{i=1}^{R} \lambda_i s_{iJ}(\xi_m;\nu_i)(1 - s_{iJ}(\xi;\nu_i)) \end{bmatrix} \tag{16}$$

Constructing the Newton Steps

Our code constructs the Jacobian matrix in the following way. Let $\circ$ denote the Hadamard or entrywise product of two equal-dimension matrices, so that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \circ \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae & bf \\ cg & dh \end{bmatrix} \tag{17}$$

and

$$\begin{bmatrix} a & b \end{bmatrix} \circ \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae & bf \\ ag & bh \end{bmatrix} \tag{18}$$

[4]We consider the process to have converged when $\max(|\Delta_m - \Delta_{m-1}|) < T$ for some tolerance level T

[5]We use a "smart" softening factor, where $\gamma_m = 1/(1 + 10 * \max(|\Delta_m|))$

We make the following definitions:

$$P = \underset{J \times R}{\begin{bmatrix} s_{11}(\xi;\nu_1) & \cdots & s_{R1}(\xi;\nu_R) \\ \vdots & \ddots & \vdots \\ s_{1J}(\xi;\nu_1) & \cdots & s_{RJ}(\xi;\nu_R) \end{bmatrix}} \tag{19}$$

$$L = \underset{1 \times R}{\begin{bmatrix} \lambda_1 & \cdots & \lambda_R \end{bmatrix}} \tag{20}$$

$$\hat{A} = L \circ P \tag{21}$$

$$B = \text{diag}(\text{rowsum}(\hat{A})) \tag{22}$$

The empirical Jacobian is then constructed as

$$\hat{J}f(\xi_m) = \frac{1}{R}(B - \hat{A}P^T) \tag{23}$$

References

Berry, Steven. 1994. "Estimating Discrete Choice Models of Product Differentiation." *The RAND Journal of Economics*, 242–262.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected Substitutes and Invertibility of Demand." *Econometrica*, 81(5): 2087–2111.

Berry, Steven, and Philip Haile. 2014. "Identification in Differentiated Product Markets Using Market Level Data." *Econometrica*, 82(5): 1749–1797.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Das, Sanghamitra, Steven Olley, and Ariel Pakes. 1994. "Evolution of Brand Qualities of Consumer Electronics in the US."

Rheinboldt, Werner C. 1970. "On M-Functions and their Application to Nonlinear Gauss-Seidel iterations and to network flows." *Journal of Mathematical Analysis and Applications*, 32: 274–307.