A New Computational Algorithm for Random Coefficients Model with Aggregate-level Data*

Jinyoung Lee†
UCLA

February 9, 2011

Abstract

The norm for estimating differentiated product demand with aggregate-level data is the BLP (1995) model. However, the estimation process BLP suggest requires a nested contraction mapping, which can make the process complicated and time-consuming. In this paper, we propose an alternative estimation routine that avoids the nested contraction mapping. This routine relies on a linear approximation to the market share function. Thus, we call it Approximate BLP (ABLP). If ABLP is performed once, it provides an approximation to the BLP estimate. However, if the ABLP procedure is iterated to convergence, it provides estimates identical to the BLP estimate. Therefore, converged ABLP can be interpreted as a new computational algorithm for the BLP estimation. We show in Monte Carlo experiments that converged ABLP is faster than other computational algorithms, especially in datasets with a large number of products or markets: ABLP is typically faster than BLP because ABLP avoids a contraction mapping. Moreover, ABLP is typically faster than Mathematical Programming with Equilibrium Constraints (MPEC) because the dimension of the space of unknown variables is smaller for ABLP. Therefore, ABLP is potentially useful to empirical researchers who study problems that involve large datasets.

1 Introduction

In the estimation of differentiated product demand with aggregate-level data, Berry, Levinsohn, and Pakes (1995), hereafter BLP, have been widely used and become a gold standard. The model deals with the endogeneity of prices in a random coefficients logit model, which has flexible substitution patterns and can produce realistic demand elasticities. On the other hand, there is no

*I am grateful to my advisor Daniel Ackerberg for his insightful guidance and continuous encouragement. I also thank Jinyong Hahn, Connan Snider, Rosa Matzkin, Moshe Buchinsky, and seminar participants at UCLA for their helpful discussions and comments. All remaining errors are my own.
†Contact: Department of Economics, University of California at Los Angeles, leejy@ucla.edu.

analytic solution to invert unobserved product characteristics, which are needed to form the BLP moment condition. To solve this problem, BLP suggest a nested fixed point (NFP) algorithm that numerically inverts unobserved product characteristics for each conjectured parameter value using a proposed contraction mapping. This nested algorithm can be time-consuming, as the recent literature shows (e.g., Dube, Fox, and Su 2009; and Kalouptsidi 2010).

Dube, Fox, and Su (2009), hereafter DFS, propose a new computational algorithm, which is called Mathematical Programs with Equilibrium Constraints (MPEC), that eliminates the nested algorithm entirely. Instead, they add a system of market share equations as a constraint in their optimization problem. This algorithm is potentially faster than the BLP method because the rate of local convergence of MPEC is higher than that of the BLP method. Moreover, the MPEC approach can be applied to several settings (e.g., single-agent dynamic discrete-choice models or dynamic games). The MPEC approach, however, is a large-scale problem, especially in a random coefficients logit model: the number of unknown variables is the number of both parameters and unobserved product characteristics, which grows with the number of products and markets. Large-scale problems can be time-consuming, but DFS show that MPEC is generally faster than the BLP method.

Petrin and Train (2010) propose a control function technique. They assume on the supply side that price is a linear function of own observed and unobserved product characteristics. The control variable is the residual of the price equation. The assumption on the pricing equation implies that price is the sum of the marginal cost and a *fixed* markup (Park and Gupta 2009); in other words, the price of one product is not affected by prices and characteristics of other products in the same market. This implication seems unlikely to hold for many industries. However, Kim and Petrin (2010) relax the assumption on the pricing equation and deal with non-separability between observed and unobserved factors in demand, cases for which the Berry (1994) correction may not work. Nonetheless, their identifying assumption may be too strong: the residual of the price equation is correlated with unobserved product characteristics, but both the residual and unobserved product characteristic are independent of instrumental variables.

Kalouptsidi (2010) assumes consumer tastes for products are drawn from a discrete distribution and finds a duality between consumer types and product market shares. This duality enables us to transform the system of market share equations to an equivalent system of consumer type equations we can solve for some new consumer unobservables. In a market with a large number of products and a small number of consumer types, computations in the dual domain can be faster. On the other hand, in markets with more consumer types than products, the dual method may lose its computational advantages.

In this paper, we propose a new computational approach for estimating the BLP model. We call this method Approximate BLP, henceforth ABLP. ABLP (1) avoids the nested contraction mapping: ABLP uses a linear approximation to the market share function, which implies that, unlike the BLP estimation approach, unobserved product characteristics can be inverted analytically;[1]

[1] Our MC results suggest ABLP is much quicker, though ABLP does require inversion of a matrix of dimension

(2) is a small-scale problem: unlike DFS, the number of unknown variables is fixed regardless of the number of observations; (3) has no additional assumption on the supply side unlike Petrin and Train (2010); and (4) keeps the type of consumer tastes continuous unlike Kalouptsidi (2010).

ABLP proceeds as follows. We conduct the ABLP inversion of unobserved product characteristic using a linear approximation to the market share function. Then we search over parameter values to minimize a GMM objective function identical to BLP's. We can improve the estimate of ABLP by iterating the procedure. If we do not iterate (or iterate a small number of times), ABLP can be interpreted as a simple[2] and quick approximation to BLP estimation. Alternatively, if we iterate until convergence, ABLP is a new computational method to the BLP estimation that is potentially faster than the method defined above.

Aguirregabiria and Mira (2002) propose an iterative method in the single-agent dynamic model. In general, the iterative approach in ABLP is similar to that in Aguirregabiria and Mira (2002) with two main differences. First, they propose a nested pseudo-likelihood algorithm (NPL) for the estimation of a class of discrete Markov decision models, but ABLP is a nested pseudo-GMM algorithm for the estimation of random coefficients logit models. Second, NPL sequential estimators are consistent if the conditional choice probability can be consistently estimated prior to estimation. However, ABLP estimators, in the iterations before convergence, are not consistent because we cannot find any consistent nonparametric estimator for unobserved product characteristics. However, on convergence, the ABLP estimator is consistent because, as we show, in this case, the ABLP estimator is numerically equivalent to the BLP estimator under some regularity conditions.

Su and Judd (2008) and DFS argue that iterative methods generally have lower rates of local convergence than MPEC. In the random coefficients logit model, the ABLP algorithm turns out to have a superlinear rate of local convergence. However, a faster theoretical rate of local convergence does not necessarily imply lower computational time in a given setting. For example, the theoretical result does not take into account the large-scale nature of the MPEC approach. In fact, we show in MC experiments using the same setup as DFS that ABLP iterated until convergence is 3 to 10 times faster than MPEC.

This paper contributes to the literature on numerical methods for applied econometric models. The BLP model is widely used for modeling aggregated discrete-choice data in empirical IO research, but the estimation procedure in BLP can be computationally burdensome, especially when an empirical researcher needs to estimate multiple specifications of a model (e.g., to check robustness, or for model-selection purposes). Thus, finding computationally efficient ways of estimating these models seems useful (e.g., Dube, Fox and Su 2009; Su and Judd 2008; Petrin and Train 2010; Park and Gupta 2009; and Kalouptsidi 2010). Many procedures are optimal from particular theoretical perspectives (e.g., rate of convergence) in the literature (e.g., Su and Judd 2008; and Dube, Fox, and Su 2009), but these theoretical arguments do not imply that a numerical method with a fast rate of

equal to the number of products in a market.

[2] Programming wise, ABLP is essentially the same as the BLP estimation, replacing the contraction mapping with a matrix inverse.

convergence will work faster than a method with a slow rate of convergence in any specific situation. In some cases, an estimator has a special structure, such as linearity (or approximate linearity), that can help one tailor computationally fast estimation procedures, which is exactly what the ABLP estimation procedure, based on a linear approximation around the well-known logit model, does. In at least the MC specifications we tried (based on DFS's MC setup), the Converged ABLP runs 3 to 10 times faster than MPEC. The K-step ABLP runs even faster, though the estimator has approximation error. As in any Monte-Carlo study, our results are subject to the caveat that many other possible specifications exist, and predicting how ABLP would do in those is difficult. However, ABLP presumably will do worse for models that are "further" from the logit model.

We organize the remainder of the paper as follows. In section 2, we discuss the BLP model and moment conditions. Section 3 presents both the BLP and MPEC algorithm. Section 4 provides an explanation of the ABLP algorithm. Section 5 shows Monte-Carlo experiments for the relative performances of the ABLP and MPEC algorithms. We conclude in section 6.

2 BLP Model[3]

Following BLP (1995) and Nevo (2000), in each market[4] $t = 1, ..., T$, where $T \geq 1$, the utility of purchasing product $j = 1, ..., J$ of consumer $i = 1, ..., I$ is

$$\begin{aligned} U_{ijt} &= X_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \\ U_{i0t} &= \varepsilon_{i0t} \quad \text{for outside good,} \end{aligned} \tag{1}$$

where X_{jt} is a K-dimensional vector of product- and market-varying attributes (including a constant and price that is typically interpreted as endogenous), β_i is the preference of individual i for observed product attributes, ξ_{jt} is the unobserved product characteristic or demand shocks, and ε_{ijt} is an extreme value deviate that is i.i.d. across agents, choices, and time periods.

Under the assumption that ε_{ijt} is drawn from the type I extreme value distribution, the probability of consumer i in market t purchasing product j is given by

$$s_j\left(X_t, \xi_t; \beta_i\right) = \frac{\exp\left(X_{jt}\beta_i + \xi_{jt}\right)}{1 + \sum_{j'=1}^{J} \exp\left(X_{j't}\beta_i + \xi_{j't}\right)}, \tag{2}$$

where $X_t \equiv \left(X'_{1t}, ..., X'_{Jt}\right)'$, $\xi_t \equiv \left(\xi_{1t}, ..., \xi_{Jt}\right)'$, and the random coefficients, β_i, are drawn from a cumulative density function, $\beta_i \sim F\left(\beta_i; \theta\right)$ where θ is the parameter (means and standard deviations) that determines the distribution of random coefficients, $F\left(\beta_i; \theta\right)$. The predicted market

[3] In this paper, we consider a panel setting just for generality. ABLP can be used in the original BLP setting, which is purely cross sectional ($T = 1$).

[4] Market means different regions or times.

share of product j is then

$$s_j\left(X_t, \xi_t; \theta\right) = \int \frac{\exp\left(X_{jt}\beta_i + \xi_{jt}\right)}{1 + \sum_{j'=1}^{J} \exp\left(X_{j't}\beta_i + \xi_{j't}\right)} dF\left(\beta_i; \theta\right).$$

Often this integral needs to be simulated, for example,

$$s_j\left(X_t, \xi_t; \theta, I\right) = \frac{1}{I} \sum_{i=1}^{I} \frac{\exp\left(X_{jt}\beta_i + \xi_{jt}\right)}{1 + \sum_{j'=1}^{J} \exp\left(X_{j't}\beta_i + \xi_{j't}\right)}, \tag{3}$$

where I is the number of simulated individuals in each market.
Define $s\left(X_t, \xi_t; \theta\right) = \left(s_1\left(X_t, \xi_t; \theta\right), ..., s_J\left(X_t, \xi_t; \theta\right)\right)'$. For simplicity of notation, we rewrite $s\left(X_t, \xi_t; \theta\right)$ as $s\left(\xi_t; \theta\right)$ from now on.

2.1 BLP Moment Condition

The BLP estimator utilizes the following population moment condition:

$$E\left[g_{jt}\left(\theta\right)\right] = 0 \text{ where } g_{jt}\left(\theta\right) = \xi_{jt}\left(\theta\right) z_{jt}; \tag{4}$$

z_{jt} are the D-dimensional instrumental variables for prices, and $\xi_{jt}\left(\theta\right)$ is the unobserved product characteristic for a parameter θ. $\xi_{jt}\left(\theta\right)$ is obtained by the inverse mapping, $\xi_t\left(\theta\right) \equiv s^{-1}\left(S_t, \theta\right)$, from the market share equations, $S_t = s\left(\xi_t; \theta\right)$ where S_t is the $(J \times 1)$ vector of *observed* market shares at market t and $\xi_t\left(\theta\right) = \left(\xi_{1t}\left(\theta\right), ..., \xi_{Jt}\left(\theta\right)\right)'$.

Define $\overline{g}_J\left(\theta\right) = \frac{1}{J} \sum_{j=1}^{J} E\left[g_{jt}\left(\theta\right)\right]$. Then the population criterion function is

$$Q\left(\theta\right) = \overline{g}_J\left(\theta\right)' W \overline{g}_J\left(\theta\right),$$

where W is a weight matrix.

Note that the sample analog to the population moment is

$$\widehat{\overline{g}}_{JT}\left(\theta\right) = \frac{1}{JT} \sum_j \sum_t \xi_{jt}\left(\theta\right) z_{jt} = \frac{1}{JT} Z'\xi\left(\theta\right), \tag{5}$$

where $\xi\left(\theta\right) = \left(\xi_1\left(\theta\right)', ..., \xi_T\left(\theta\right)'\right)'$ is the $(JT \times 1)$ vector of unobserved product characteristics for a parameter, θ, and Z is the $(JT \times D)$ matrix of instrumental variables. We can write the sample analogue of the population criterion function as

$$Q_{JT}\left(\theta\right) = \widehat{\overline{g}}_{JT}\left(\theta\right)' \widehat{W}_{JT} \widehat{\overline{g}}_{JT}\left(\theta\right) = \left(\frac{1}{JT}\right)^2 \xi\left(\theta\right)' Z \widehat{W}_{JT} Z' \xi\left(\theta\right) \tag{6}$$

over $\theta \in \Theta$ and $\widehat{W}_{JT} \to_p W$. Here we assume that a unique true parameter, $\theta_0 \in int\left(\Theta\right)$ exists such that $\theta_0 = \arg\min Q\left(\theta\right)$.

3 Computational Algorithm

Now we can estimate the parameter, θ, once we obtain the inverse mapping, $\xi_t(\theta) \equiv s^{-1}(S_t, \theta)$. Generally the inverse mapping, $\xi_t(\theta)$, does not have an analytic solution, which has stimulated interest in how to compute ξ or how to estimate ξ and θ at the same time. Among all methods, we will explain three solutions to the problem in the following.

3.1 BLP Algorithm

The BLP algorithm is a nested fixed point algorithm in the random coefficients logit model. The BLP algorithm has two layers:

[Inner loop] BLP numerically invert out $\xi_t(\theta)$ using the following contraction mapping:

$$\xi_t^H = \xi_t^{H-1} + \ln S_t - \ln s\left(\xi_t^{H-1}; \theta\right) \tag{7}$$

for a given θ as $H \to \infty$, where H denotes the index of iterations: $\xi_t^{BLP}(\theta) \equiv \xi_t^{\infty}$. In practice, we use a tolerance level for the discrepancy between ξ_t^H and ξ_t^{H-1}.

[Outer loop] The BLP minimization problem is

$$\min_{\theta} \widehat{\overline{g}}_{JT}(\theta)' \widehat{W}_{JT} \widehat{\overline{g}}_{JT}(\theta), \tag{8}$$

where $\widehat{\overline{g}}_{JT}(\theta) = \frac{1}{JT} Z' \xi^{BLP}(\theta)$.

The BLP estimator is defined as the parameter, θ, with the lowest value of the BLP GMM objective function among the local roots of the first-order condition of the BLP minimization in (8).

In sum, the outer loop searches over the parameter space to minimize the GMM objective function in (8). In the inner loop, a contraction mapping exists for unobserved product characteristic, ξ, for each conjectured value of parameter, θ. This inner loop is time-consuming,[5] which motivated the development of other alternative estimation methods or computational algorithms. In the next subsection, we will look at the MPEC approach as a well-known alternative to the BLP algorithm.

3.2 MPEC Algorithm[6]

Mathematical Programs with Equilibrium Constraints (MPEC) is an optimization problem with the equilibrium or complementary constraints (Luo, Pang, and Ralph 1996). Su and Judd (2008) first apply MPEC in estimating structural models in economics. Dube, Fox, and Su (2009) use it in BLP random coefficients logit models.

[5] In some Monte Carlo experiments, BLP contraction mapping, given a set of parameters, does not seem to satisfy the tight stopping criterion $\left(10^{-14}\right)$, which DFS suggest.

[6] The explanation of MPEC in BLP estimation is mainly extracted from Dube, Fox, and Su (2009).

The DFS constrained minimization problem is

$$\begin{aligned} \min_{\theta,\ \xi} \quad & \widehat{g}_{JT}(\xi)'\widehat{W}_{JT}\widehat{g}_{JT}(\xi) \\ s.t. \quad & S = s(\xi;\theta), \end{aligned} \tag{9}$$

where $\widehat{g}_{JT}(\xi) = \frac{1}{JT}Z'\xi$; $S = (S_1', ..., S_T')'$ is a $(JT \times 1)$ vector of observed market shares, and $s(\xi;\theta) = \left(s(\xi_1;\theta)', ..., s(\xi_T;\theta)'\right)'$ is a $(JT \times 1)$ vector of market share functions. MPEC introduces the market share equations as nonlinear constraints to the optimization problem. They optimize over both the unobserved product characteristic, ξ, and the parameter, θ. DFS show the equivalence of both BLP and MPEC methods. MPEC approach has a quadratic local convergence rate (Luo, Pang, and Ralph 1996).

We can solve the constrained optimization problem defined above using a modern nonlinear optimization package researchers in numerical optimization developed. DFS use KNITRO in their Monte Carlo simulation. Although the MPEC algorithm has a quadratic rate of local convergence, it has one disadvantage that can slow down its computational speed: many unknown variables in its optimization problems. For example, with $J = 25$ products and $T = 200$ markets,[7] it has $5,000$ unobserved product characteristics, ξ_{jt}, as unknown variables. As the number of observations grows, MPEC becomes a larger-scale problem that might be hard to deal with and time-consuming. The section on Monte-Carlo experiments tests this conjecture.

4 ABLP Algorithm

4.1 ABLP Inversion

We take a different approach from DFS to avoid the time-consuming contraction mapping. Instead, we use a first-order approximation to the market share function. The advantage of approximation is that we can easily invert out ξ_t, the unobserved product characteristic in market t, using the analytic solution in (11) (which requires matrix inversions).

The first-order approximation of the log market share function, $\ln s(\xi_t;\theta)$, around a point of approximation, ξ_t^0, is[8]

$$\ln s(\xi_t;\theta) \approx \ln s\left(\xi_t^0;\theta\right) + \frac{\partial \ln s\left(\xi_t^0;\theta\right)}{\partial \xi_t'}\left(\xi_t - \xi_t^0\right),$$

where $\approx$ denotes the first-order Taylor series expansion, $\ln s = (\ln s_1, ..., \ln s_J)'$, $\frac{\partial \ln s}{\partial \xi_t'} = \left(\frac{\partial \ln s}{\partial \xi_{1t}}, ..., \frac{\partial \ln s}{\partial \xi_{Jt}}\right)$, $\xi_t = (\xi_{1t}, ..., \xi_{Jt})'$, and $\xi_t^0 = \left(\xi_{1t}^0, ..., \xi_{Jt}^0\right)'$. Let $\ln s^A(\xi_t;\theta)$ denote the first-order approximation of

[7] Here markets mean "regions" and "times". Marketing has many long panel datasets with weekly time series. They define regional market at the level of states, census tracts, or ZIP codes.

[8] Alternatively, we can take the first-order approximation to the market share function $s(\xi_t;\theta)$.

the log market share function:

$$\ln s^A(\xi_t;\theta) \equiv \ln s\left(\xi_t^0;\theta\right) + \frac{\partial \ln s\left(\xi_t^0;\theta\right)}{\partial \xi_t'}\left(\xi_t - \xi_t^0\right). \tag{10}$$

Instead of using the (exact) market share equations, $\ln S_t = \ln s(\xi_t;\theta)$, we use the approximate market share equations, $\ln S_t = \ln s^A(\xi_t;\theta)$, to establish the following relationship:

$$\xi_t = \Phi_t\left(\theta,\xi_t^0\right) \equiv \xi_t^0 + \left[\frac{\partial \ln s\left(\xi_t^0;\theta\right)}{\partial \xi_t'}\right]^{-1}\left[\ln S_t - \ln s\left(\xi_t^0;\theta\right)\right] \tag{11}$$

for $t = 1,...,T$. We call the mapping $\Phi_t\left(\theta,\xi_t^0\right)$ the ABLP inversion of unobserved product characteristic, ξ_t, at market t given θ with a point of approximation, ξ_t^0.

The ABLP inversion has a contrasting point with the BLP inversion. The ABLP inversion is analytic: given ξ_t^0, we can easily calculate $\Phi_t\left(\theta,\xi_t^0\right)$ for any parameter, θ. Thus the numerical error in the BLP contraction mapping does not exist in the ABLP inversion. On the other hand, the ABLP inversion has a first-order approximation error. However, this approximation error will disappear when we iterate the K-step ABLP estimator in subsection 4.3 until convergence.

We have two natural choices to find the initial point of approximation, $\xi^0 = \left(\xi_1^{0\prime},...,\xi_T^{0\prime}\right)'$: one is zero vector and the other is the estimated unobserved product characteristics from a logit model. If variation of the true ξ is very small, zero vector seems a good candidate for ξ^0. Otherwise, we can use the estimated $\widehat{\xi}$ from a logit model for ξ^0. However, we cannot tell the amount of variation of ξ in advance, that is, prior to estimation.

Our preferred solution to this issue of finding ξ^0 is to generate an arbitrary vector of unobserved product characteristic using the BLP contraction mapping just once. The procedure is as follows. First, posit a specific parameter, θ^0. Then, find the corresponding ξ^0 using the BLP contraction mapping, i.e. $\xi^0 = \xi^{BLP}\left(\theta^0\right)$. With this procedure, we can find any arbitrary initial point of approximation for ABLP. In other words, if we have a good conjecture of parameter, we can also obtain a good set of unobserved product characteristics.

4.2 ABLP Estimation

The ABLP minimization problem starts with an initial point of approximation, ξ^0, for the unobserved product characteristics. Then we can obtain a GMM estimate of θ as

$$\theta^1 = \arg\min_{\theta\in\Theta} \Phi\left(\theta,\xi^0\right)' Z\widehat{W}_{JT}Z'\Phi\left(\theta,\xi^0\right), \tag{12}$$

where

$$\Phi_t\left(\theta,\xi_t^0\right) = \xi_t^0 + \left[\frac{\partial \ln s\left(\xi_t^0;\theta\right)}{\partial \xi_t'}\right]^{-1}\left[\ln S_t - \ln s\left(\xi_t^0;\theta\right)\right] \tag{13}$$

for $t = 1, ..., T$ and $\Phi\left(\theta, \xi^0\right) = \left(\Phi_1\left(\theta, \xi^0\right)', ..., \Phi_T\left(\theta, \xi^0\right)'\right)'$.

We can obtain the estimate of parameters from the GMM estimation with the analytic ABLP inversion of unobserved product characteristic as described above. We name the solution of the ABLP minimization problem the (one-step) ABLP estimator. The accuracy of the ABLP estimate hinges on that of the initial unobserved product characteristic, ξ^0. The closer the point of approximation, ξ^0, is to the true one, the more accurate an estimate of parameter, θ^1, we can get. From this idea, we can naturally think of *an iterative approach* to obtain a better point of approximation and, as a result, a better estimate. In the next subsection, we introduce an iterative method with which we can obtain the BLP estimate from the ABLP algorithm.

4.3 K-step ABLP Estimation and ABLP mapping

The steps of iteration are as follows. First, we can obtain the estimate of parameter, θ^1, from the first estimation with the initial point of approximation, ξ^0. Second, we update the point of approximation for unobserved product characteristics with the mapping $\xi^1 = \Phi\left(\theta^1, \xi^0\right)$, using the estimate θ^1 from the first iteration. Then we can repeat the same procedure several times until the estimate or the unobserved product characteristics converge. The details are the following.

The ABLP algorithm starts with an initial point of approximation, ξ^0, for the unobserved product characteristics. At each iteration $K \geq 1$, we take the following two stages:

□ **Stage 1** [**ABLP Minimization** (θ)]: Obtain a new GMM estimate θ^K as

$$\theta^K = \Gamma\left(\xi^{K-1}\right) \equiv \arg\min_{\theta \in \Theta} \Phi\left(\theta, \xi^{K-1}\right)' Z \widehat{W}_{JT} Z' \Phi\left(\theta, \xi^{K-1}\right), \tag{14}$$

where

$$\Phi_t\left(\theta, \xi_t^{K-1}\right) = \xi_t^{K-1} + \left[\frac{\partial \ln s\left(\xi_t^{K-1}; \theta\right)}{\partial \xi_t'}\right]^{-1} \left[\ln S_t - \ln s\left(\xi_t^{K-1}; \theta\right)\right] \tag{15}$$

for $t = 1, ..., T$.

□ **Stage 2** [**ABLP Updating** (ξ)]: Update ξ using θ^K of stage 1, that is,

$$\xi^K = \Phi\left(\theta^K, \xi^{K-1}\right). \tag{16}$$

We assume that, given ξ^{K-1}, θ^K uniquely solve equation (14). By the theorem of the maximum,[9] $\Gamma : \Xi \to \Theta$ is also continuously differentiable in ξ. For each θ, $\Phi(\theta, \xi)$ maps the space of unobserved product characteristics into itself. We call the solution θ^K the K-step ABLP estimator.

[9] $s(\xi_t; \theta)$ is continuous in (ξ_t, θ), so $\Phi_t(\theta, \xi_t)$ is also continuous. Therefore, the GMM objective function in (14) is continuous. By theorem of maximum, if the objective function is continuous and if the domain Θ is compact, then the maximum-value function and the solution function are continuous.

For convenience sake, define *the ABLP mapping* as both (1) the solution function of the ABLP minimization, $\theta^K = \Gamma\left(\xi^{K-1}\right)$, and (2) the ABLP updating function, $\xi^K = \Phi\left(\theta^K, \xi^{K-1}\right)$. Now we are interested in what the limit of the K-step ABLP estimator is, as K goes to infinity. If the limit exists, its candidate is a fixed point of the ABLP mapping. Fixed points of the ABLP mapping provide key reference points for the characterization of the ABLP mapping. A fixed point of the ABLP mapping is defined as follows.

Definition 1 *A fixed point of the ABLP mapping is a pair* (θ^*, ξ^*) *such that*

$$\theta^* = \Gamma(\xi^*) \text{ and } \xi^* = \Phi(\theta^*, \xi^*).$$

Multiple fixed points could exist in the ABLP mapping. Denote the set of fixed points of the ABLP mapping by $T = \{(\theta^*, \xi^*) \,|\theta^* = \Gamma(\xi^*) \text{ and } \xi^* = \Phi(\theta^*, \xi^*)\}$. Among all fixed points of the ABLP mapping, we are interested in a fixed point with the lowest value of the GMM objected function, which is defined as follows.

Definition 2 *The Converged ABLP estimator is the parameter* $\widehat{\theta}$ *such that*

$$\left(\widehat{\theta}, \widehat{\xi}\right) = \arg \min_{(\theta^*, \xi^*) \in T} \Phi(\theta^*, \xi^*)' Z \widehat{W}_{JT} Z' \Phi(\theta^*, \xi^*).$$

The following lemma shows a local property of a fixed point of the ABLP mapping.

Lemma 1 *For any fixed point* (θ^*, ξ^*), $\frac{\partial \Phi(\theta^*, \xi^*)}{\partial \xi'} = 0.$

Proof. All proofs are in the appendix. ■

Lemma 1 implies that at a fixed point of the ABLP mapping, (1) updating the unobserved product characteristic ξ (i.e., $\xi = \Phi(\theta, \xi)$) and (2) minimizing the GMM objective function in (14) is impossible. These features are key properties of the ABLP mapping: it leads to the equivalence between the Converged ABLP estimator and the BLP estimator, and the local convergence property of the K-step ABLP estimator.

Proving the equivalence between the Converged ABLP estimator and the BLP estimator requires three steps. In the first step, we compare the first-order condition of the BLP minimization with the first-order condition of the ABLP minimization with no updating of ξ, and show they are equivalent. This result implies that the set of the local root of the first-order condition of the BLP minimization is identical to the set of fixed points of the ABLP minimization with no updating of ξ.[10] In the second step, we show that for any pair $(\xi^\circ, \theta^\circ)$ that is both a local root of the

[10] The set of fixed points of the ABLP minimization with no updating of ξ includes the set of fixed points of the ABLP mapping. Although the former set includes local maxima, minima, and saddle points, the latter set includes only local minima. In the same way, the set of local roots of the FOC of the BLP minimization includes the set of the solution of the BLP minimization. The former set includes local maxima, minima, and saddle points, whereas the latter set includes local minima.

BLP minimization and a fixed point of the ABLP mapping, the value of the GMM objective function in the BLP minimization is equal to the value of the GMM objective function in the ABLP minimization. Finally, in the third step, according to the definition of the BLP estimator and the Converged ABLP estimator, we find they are identical. The details are in the proof of the following proposition.

Proposition 1 *The Converged ABLP estimator and the BLP estimator are numerically equivalent.*

Proof. All proofs are in the appendix. ■

Proposition 1 shows the Converged ABLP estimator and the BLP estimator are the same statistical estimator. Therefore, the asymptotics of the Converged ABLP estimator follows that of the BLP estimator: the Converged ABLP estimator is consistent and asymptotically normal.

Now the remaining thing to be done is finding the Converged ABLP estimator, which we conjecture is possible to obtain, if we iterate the K-step ABLP estimator until convergence. In the next subsection, we examine the local convergence of the K-step ABLP estimator to the Converged ABLP estimator.

4.4 Local Convergence of K-step ABLP Estimator

Now we want to know whether the K-step ABLP estimator locally converges to the Converged ABLP estimator as $K \to \infty$ and whether the Converged ABLP estimator is asymptotically stable. By the Hartman-Grobman theorem, the characterization of the local behavior of the ABLP mapping in a neighborhood of its fixed point can be obtained by the local behavior of its linearization under some regularity conditions.

Consider a nonlinear system $\mathbf{x}_{k+1}= \mathbf{f}\left(\mathbf{x}_k\right)$ where $\mathbf{f}: \mathbf{R}^n \to \mathbf{R}^n$ is a continuously differentiable single-value function. Suppose a fixed point $\mathbf{x}^*$ exists such that $\mathbf{x}^*= \mathbf{f}\left(\mathbf{x}^*\right)$. The function $\mathbf{x}_{k+1}= \mathbf{f}\left(\mathbf{x}_k\right)$ can be approximated around the fixed point $\mathbf{x}^*$. Let $\mathbf{x}_k \equiv \mathbf{x}^*+\boldsymbol{\eta}_k$. A Taylor expansion around $\mathbf{x}^*$ yields

$$\mathbf{x}^*+\boldsymbol{\eta}_{k+1}= \mathbf{f}\left(\mathbf{x}^*+\boldsymbol{\eta}_k\right) = \mathbf{f}\left(\mathbf{x}^*\right) + \boldsymbol{\nabla}\mathbf{f}\left(\mathbf{x}^*\right)\boldsymbol{\eta}_k + \mathbf{O}\left(\left|\boldsymbol{\eta}_k\right|^2\right),$$

where $\mathbf{O}\left(\left|\boldsymbol{\eta}_k\right|^2\right)$ denotes terms of second and higher order in the deviation $\boldsymbol{\eta}_k$. For small enough $\left|\boldsymbol{\eta}_k\right|$, we have the linearized map $\boldsymbol{\eta}_{k+1}= \boldsymbol{\nabla}\mathbf{f}\left(\mathbf{x}^*\right)\boldsymbol{\eta}_k$. The maximum of the absolute of eigenvalues of the Jacobian matrix $\boldsymbol{\nabla}\mathbf{f}\left(\mathbf{x}^*\right)$ is called the spectral radius of $\nabla f\left(\mathbf{x}^*\right)$, denoted by $\sigma\left(\mathbf{x}^*\right)$. If $\sigma\left(\mathbf{x}^*\right)$ is less than 1, the fixed point $\mathbf{x}^*$ is asymptotically stable (Ch.4 in Galor 2007 and Ch.1 in Wiggins 1997). We apply the above stability theory to the ABLP mapping and prove its local stability and rate of local convergence.

Now let's look at the ABLP mapping, which consists of the solution function of the ABLP minimization, $\theta^K = \Gamma\left(\xi^{K-1}\right)$, and the updating function, $\xi^K = \Phi\left(\theta^K, \xi^{K-1}\right)$. Let $\left(\widehat{\theta}, \widehat{\xi}\right)$ be the Converged ABLP estimator.[11] Let $\|\cdot\|$ denote the Euclidian norm. The Taylor series expansion to

[11] In the exactly-identified model, (θ^*, ξ^*) can be any fixed point of the ABLP map for the proof of Proposition 2.

the ABLP mapping around $\left(\widehat{\theta},\widehat{\xi}\right)$ gives

$$\begin{aligned} \xi^K &= \Phi\left(\widehat{\theta},\widehat{\xi}\right)+\frac{\partial\Phi}{\partial\theta'}\left(\theta^K-\widehat{\theta}\right)+\frac{\partial\Phi}{\partial\xi'}\left(\xi^{K-1}-\widehat{\xi}\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right)+O\left(\left\|\theta^K-\widehat{\theta}\right\|^2\right) \\ \Longrightarrow\ & \xi^K-\xi^*=\frac{\partial\Phi}{\partial\theta'}\left(\theta^K-\widehat{\theta}\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right)+O\left(\left\|\theta^K-\widehat{\theta}\right\|^2\right) \end{aligned}$$

because $\Phi\left(\widehat{\theta},\widehat{\xi}\right)=\widehat{\xi}$ by the definition of a fixed point, and $\frac{\partial\Phi}{\partial\xi'}=0$ by Lemma 1. The remaining thing to do is check the relationship between $\theta^K-\widehat{\theta}$ and $\xi^{K-1}-\widehat{\xi}$.

Proposition 2 *Suppose $\Gamma:\Xi\rightarrow\Theta$ is a continuously differentiable single-value function in a neighborhood of $\widehat{\xi}$. For iteration $K=1,2,...$*
In the exactly-identified model,

$$\xi^K-\widehat{\xi}=O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right),$$

and in the over-identified model,

$$\xi^K-\widehat{\xi}=O_p\left((JT)^{-\frac{1}{2}}\left\|\xi^{K-1}-\widehat{\xi}\right\|\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).$$

Proof. All proofs are in the appendix. ■

Proposition 2 establishes the local convergence of the K-step ABLP estimator under some regularity conditions: as $K\rightarrow\infty$, the K-step ABLP estimator, θ^K, locally converges to the Converged ABLP estimator, $\widehat{\theta}$, at a superlinear rate of convergence, and the Converged ABLP estimator is asymptotically stable. The result in Proposition 2 is intuitive in a sense that the updating function of the ABLP mapping has a form of Newton update.

The Converged ABLP estimator is one of the fixed points of the ABLP mapping. That is, the K-step ABLP estimator could converge to some fixed points of the ABLP mapping, and the fixed point with the lowest value of the GMM objective function in the ABLP minimization becomes the Converged ABLP estimator. This procedure is exactly analogous to the local minima in the BLP estimation (as well as in the MPEC estimation). We can find several local minima in the BLP minimization, and the local minimum with the lowest value of the GMM objective function in the BLP minimization is the BLP estimator.

4.5 Features of ABLP Algorithm

The ABLP algorithm contrasts with the BLP algorithm. The BLP estimator is the nested fixed point (NFP) estimator in random coefficients logit models: they have to first find the unobserved product characteristics for each conjectured value of parameter and then search for the parameter that best minimizes the GMM objective function in (8). In contrast, the ABLP estimator is the nested pseudo-GMM (NPGMM) estimator in random coefficients logit models: we first search for the parameter that best satisfies the moment condition given the point of approximation of the unobserved product characteristic. Then we update the point of approximation and estimate the parameter again. A trade-off exists between the ABLP algorithm and BLP algorithm: the BLP algorithm spends time on its contraction mapping given a specific parameter, but the ABLP algorithm takes time for parameter search given specific unobserved product characteristics.

The ABLP algorithm could be faster than other methods for several reasons. First, the matrix inversion in the ABLP inversion and multiple parameter searches may be faster than the BLP contraction mapping and one-time parameter search. Second, ABLP could be faster than MPEC because ABLP remains a small-scale problem, whereas MPEC becomes a larger-scale problem as we add more markets or products. In the next section, we can see the relative performance of the ABLP and MPEC algorithm.

Moreover, the ABLP algorithm has an advantage in over-identified models. When doing GMM estimation, one typically wants to do two-steps of GMM estimation for efficiency (i.e., with optimal weight matrix). Since we can update the weight matrix in each step ABLP estimation, we can obtain the optimal matrix for free. In practice, we update the weight matrix when the ABLP algorithm checks the convergence. For example, in Figure 1, the ABLP algorithm checks the convergence from the fourth iteration. We can update the weight matrix from the fifth iteration and get the optimal matrix and the efficient estimator on convergence. In this sense, if we want to obtain the efficient estimator, the ABLP algorithm can save more computational time.

5 Monte Carlo Experiments

In this section, we use various synthetic datasets to compare the speed of the ABLP algorithm with that of the MPEC algorithm. (Hereafter, we use the term ABLP [MPEC or BLP] instead of the ABLP [MPEC or BLP] algorithm for simplicity, if there is no confusion.) We follow the same data-generating process as in DFS. Note that DFS compare MPEC to BLP and find MPEC is 1.3 to 5 times faster than BLP. Therefore, we compare ABLP only with MPEC. We test the speed of each algorithm by altering some features, such as the level of mean intercept $E\left[\beta_i^0\right]$, the number of markets T, the number of products J, and the variance of the random coefficients.

5.1 Data-generating Process[12]

The (base) data-generating process for Monte-Carlo experiments is as follows.

$$U_{ijt} = X_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \quad U_{i0t} = \varepsilon_{iot},$$

for $i = 1, ..., I,\ j = 1, ...J,\ t = 1, ..., T$, where there are 50 markets (T) and the same 25 products (J) in each market. Let $X_{jt} = \{1, x_{j1}, x_{j2}, x_{j3}, p_{jt}\}$. Observed product characteristics $\{x_{j1}, x_{j2}, x_{j3}\}$

[12] For more details, please, refer to subsections 6.2 and 7.2 in Dube, Fox, and Su (2009).

follow multivariate normal distribution with zero means and the covariance matrix:

$$\begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix}.$$

Unobserved product characteristic, ξ_{jt}, is distributed i.i.d. standard normal, $N(0,1)$. Finally, price is generated as follows:

$$p_{jt} = 3 + \xi_{jt} \cdot 1.5 + u_{jt} + \sum_{k=1}^{3} x_{kjt},$$

where u_{jt} follows a uniform distribution $U[0,5]$.

Denote consumer preference by $\beta_i = \left\{\beta_i^0, \beta_i^1, \beta_i^2, \beta_i^3, \beta_i^p\right\}$, each distributed independently normal: $E[\beta_i] = \{1, 1.5, 1.5, 0.5, -3\}$ and $Var[\beta_i] = \{0.5, 0.5, 0.5, 0.5, 0.2\}$. We use the same $I = 100$ individuals both in the data-generating process and in estimation. The same set of J products exists in each market, so the only variation is due to prices, which vary by market.

5.2 Speed Comparison of Converged ABLP and MPEC

In our Monte Carlo experiments, we use "ABLP with Concentration on Nonlinear Parameter" in the Appendix C, which has the same properties as ABLP and computes quickly because we can concentrate out the linear parameters. However, in MPEC, the concentration may be insignificant because δ is not a function of nonlinear parameter during parameter search.

For MPEC, we use DFS's original Matlab code, which we can download from Su's website.[13] As DFS do, we use Tomlab interface to run KNITRO in Matlab program. For ABLP, we use Matlab programming and its optimization function "fmincon".[14] DFS use an interior-point algorithm in KNITRO and we also use the same algorithm in fmincon. For the convergence criterion, ABLP uses 10^{-6} for both the update of unobserved product characteristics and the change in the value of the objective function in each iteration. MPEC uses 10^{-6} for the constraint of the market share equations and the optimality condition of its objective function, or the change in the value of the objective function in each run if it doesn't satisfy the optimality condition. For a fair comparison with MPEC, we run ABLP until convergence and use the same starting points for both ABLP and MPEC. For each experiment, we use 80 different datasets and five starting points in each dataset.

In our first experiment in Table 1, we manipulate the level of mean intercept, $E\left[\beta_i^0\right]$. DFS find the Lipschitz constant,[15] affects the speed of BLP and increases as the level of mean intercept gets

[13] Here is the link: http://faculty.chicagobooth.edu/jean-pierre.dube/vita/MPEC%20code.htm.

[14] We can use "fminunc" instead of "fmincon": "fminunc" seems faster than "fmincon", but it sometimes has an operational problem. The ABLP algorithm calculates the inverse of the matrix $\frac{\partial \ln S\left(\xi^{K-1};\theta\right)}{\partial \xi'}$ when it searches over parameters. The matrix is rarely not invertible with some parameters. (The BLP algorithm has a similar problem with calculating the gradient of its objective function.) In this case, "fminunc" stops during an optimization run, so we cannot handle it directly. However, "fmincon" can handle the problem by itself by searching nearby feasible sets of parameters. For this reason, we use "fmincon" for convenience and stability at the cost of some speed advantages.

[15] In BLP approach, the Lipschitz constant measures the rate of convergence of the contraction mapping. As it gets close to 1, the contraction mapping converges slowly (Dube, Fox, and Su 2009).

bigger. In general, ABLP is faster than MPEC by 3 to 5 times. ABLP slows by 50 percent from $E\left[\beta_i^0\right] = -1$ to 2. But MPEC gets a little faster and then seems to be stable as the Lipschitz constant grows. This result implies that ABLP seems to be monotonically affected by the Lipschitz constant.

In the second experiment, we change the number of markets, T, and check the relative performance of ABLP and MPEC. DFS argue the Lipschitz constant grows as we include more markets in a dataset. In that sense, this experiment is similar to the first one but different in that we can additionally test the problem of MPEC with scale. In MPEC, the unobserved product characteristics, ξ, are unknown variables to find. Thus, as we have more observations, such as more products and markets, MPEC has more unknown variables. For example, with $J = 25$ products and $T = 200$ markets, MPEC has at least $5,000$ unknown variables except parameters. Table 2 has four cases with $T = 25, 50, 100,$ and 200. With $T = 25$, ABLP is faster than MPEC by 3 times. This gap, however, widens as the number of markets, T, increases. With $T = 200$, ABLP is 10 times faster than MPEC. This result shows that MPEC, a large-scale problem, slows as the number of markets increases.[16]

In the third experiment, we change the number of products, J. The ABLP inversion requires a matrix inverse in each market at each conjectured value of θ (see equation [11]). The dimension of the matrix inverse is $J \times J$. Thus, as J increases, the cost of the inversion increases much more. To know the amount of this computing cost, we test different values of J in the experiment. In general, ABLP is faster than MPEC in this experiment. The time cost of ABLP increases by around 8 times from $J = 25$ to $J = 100$, whereas that of MPEC increases by about 29 times. This result implies the time cost of the ABLP inversion is less than that of the MPEC large-scale problem in this example.

Finally, we test the time cost of the ABLP inversion with respect to linear approximation of the market share function. In the logit model, the unobserved product characteristics are linear in the log of the market share function. Hence we expect the linear approximation to perform well with small variance of random coefficients. In the fourth experiment, we change the variance of random coefficients and check the computing time of ABLP and MPEC by multiplying $Var\left[\beta_i\right]$ by $0.5, 1,$ and 2. Unlike what we expect, the level of variances of random coefficients may not affect ABLP.

5.3 Performance of the K-step ABLP Estimator

Figure 1 shows the performance of the K-step ABLP estimator at each iteration in a Monte Carlo experiment. We measure the root mean-squared error (RMSE) between the true own price elasticity and the estimated elasticity. ABLP converges with six iterations in this experiment. After three iterations, no change seems to occur in RMSE at a significant level. In other words, RMSE improves

[16] With $J = 25$ and $T = 200$, we have $5,000$ observations. However, this is not an extreme case. Many researchers use more observations in their researches. For example, Nevo (2001) uses $27,862$ observations, and Davis (2006) uses $20,008$ observations. Moreover, many panel datasets in marketing have much more observations than these numbers because they include weekly sales and prices of many products in many small regional markets such as zipcode-level markets in the US.

quickly in the early iterations and then checks its convergence in the later iterations. Thus, this result implies that we could get a good estimate of parameter quickly even though we iterate ABLP only a few times.

Now let's look at the performance of the K-step ABLP estimator when $K = 2$. To measure how close the K-step ABLP estimator is to the Converged ABLP estimator, we use the mean absolute error (MAE) of the own price elasticities per each dataset: Let e_{jt} and $\widehat{e}_{jt}$ denote the Converged ABLP and the K-step ABLP own price elasticity of product j at market t, respectively. The (normalized) MAE of the own price elasticity (per dataset) is calculated by $MAE_{ds} = \frac{1}{JT}\sum_j \sum_t \left| \left(\widehat{e}_{jt} - e_{jt} \right) / e_{jt} \right|$, where ds is the index of dataset. The MAE in a table is the simple average of MAE's across 80 datasets: $MAE_{mean} = 1/80 \sum_{ds=1}^{80} MAE_{ds}$. We find in Table 5 to 8 that the 2-step ABLP estimator performs well: we can save the computational time by 30 to 50% using the 2-step ABLP estimator instead of using the Converged ABLP estimator, whereas the MAE is less than 1% in each case. This result implies we can save a large amount of time at the expense of a very small amount of accuracy using the K-step ABLP estimator. Thus, the K-step ABLP estimator, which is an even quicker approximation, may be particularly useful when running many specifications, e.g. for robustness checks.

6 Conclusion

In this paper, we propose a new computational algorithm, ABLP, for the random coefficients logit model. ABLP avoids the nested BLP contraction mapping and appears to save computational time. The ABLP inversion of unobserved product characteristic relies on an approximation to the market share function. However, we show that if we iterate the K-step ABLP estimator until convergence, the ABLP estimation exactly replicates the BLP estimation because the approximation error vanishes on convergence. In addition, we show the K-step ABLP estimator locally converges to the Converged ABLP estimator.

To measure the speed improvement of the ABLP algorithm, we conduct Monte Carlo experiments to investigate the relative performance of ABLP and MPEC. In DFS, MPEC was shown to be 1.3 to 5 times faster than the BLP algorithm. In general, ABLP appears to be faster than MPEC by 3 to10 times. The speed advantage of ABLP is highest with a large number of products or markets. ABLP is presumably faster than MPEC because ABLP is a small-scale problem, whereas MPEC is a large-scale problem. In addition, ABLP appears to be faster than the BLP algorithm because ABLP avoids a contraction mapping. Therefore, ABLP is potentially useful to empirical researchers who study problems that involve large datasets.

References

[1] Ackerberg, D., J. Geweke, and J. Hahn (2009): "Comments on "Convergence Properties of the Likelihood of Computed Dynamic Models," Econometrica, 77, 2009-2017.

[2] Aguirregabiria, V. and P. Mira (2002): "Swapping The Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," Econometrica, 70, 1519-1543.

[3] Aguirregabiria, V. and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," Econometrica, 75, 1–53.

[4] Bajari, P., J. T. Fox, K.-I. Kim and S. P. Ryan (2009): "The Random Coefficients Logit Model Is Identified," Working Paper, University of Chicago.

[5] Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," RAND Journal of Economics, 25(2), 242–262.

[6] Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," Econometrica, 63(4), 841–890.

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

[8] Davis, P. (2006): "Spatial Competition in Retail Markets: Movie Theaters," RAND Journal of Economics, 37(4), 964–982.

[9] Dub e, J. P., J. T. Fox, C. L. Su (2009): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coe cients Demand Estimation," Working paper, University of Chicago.

[10] Galor, O. (2007): Discrete Dynamical Systems. Springer, Berlin.

[11] Gandhi, A. (2008): "On the Nonparametric Foundations of Product Differentiated Demand Systems," Working Paper, University of Wisconsin-Madison.

[12] Kalouptsidi (2010): "From Market Shares to Consumer Types: Duality in Differentiated Product Demand Estimation," Journal of Applied Econometrics, forthcoming.

[13] Kasahara, H. and K. Shimotsu (2008): "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models", Journal of Econometrics 146(1), 92–106.

[14] Kasahara, H. and K. Shimotsu (2010): "Sequential Estimation of Structural Models with a Fixed Point Constraint," Working Paper, Queen's University.

[15] Kim, K. and M. Park (2010): "Approximation Error in the Nested Fixed Point Algorithm for BLP Model Estimation," Working Paper, University of Minnesota.

[16] Kim, K. and A. Petrin (2010): "Control Function Corrections for Unobserved Factors in Differentiated Product Models," Working Paper, University of Minnesota.

[17] Knittel, C. R. and K. Metaxoglou (2008): "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings," Working Paper, U.C. Davis.

[18] Luo, Z., J. Pang, and D. Ralph (1996): Mathematical Programs with Equilibrium Constraints. Cambridge University Press.

[19] Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," Journal of Economics and Management Strategy, 9(4), 513–548.

[20] Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," Econometrica, 69(2), 307–342.

[21] Nocedal, J. and S. J. Wright (2006): Numerical Optimization. Springer, New York, NY.

[22] Moon, R., M. Shum, and M. Weidner (2010): "Estimation of Random Coefficients Logit Demand Models with Interactive Fixed Effects," Working Paper, California Institute of Technology.

[23] Park, S. and S. Gupta (2009): "Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," Journal of Marketing Research, 46(4), 531–542.

[24] Pesendorfer, M. and P. Schmidt-Dengler (2010): "Sequential Estimation of Dynamic Discrete Games: A Comment," Econometrica, 78(2), 833-842.

[25] Petrin, A. and K. Train (2010): "Control Function Corrections for Omitted Attributes in Differentiated Products Markets,"Journal of Marketing Research, 47(1), 3–13.

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

[27] Strogatz, S. (2001): Nonlinear Dynamics And Chaos. Westview Press.

[28] Su, C.-L. and K. L. Judd (2008): "Constrained Optimization Approaches to Estimation of Structural Models," Working Paper, University of Chicago.

[29] Waltz, R. A. and T. D. Plantenga (2009): KNITRO 6.0 Users's Manual. Ziena Optimization,Inc.

[30] Wiggins, S. (1990): Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer, New York, NY.

[31] Zeidler, E. (1986): Nonlinear Functional Analysis and its Applications. Springer, New York, NY.

Table 1. Time Costs Varying the Mean Intercept $E\left[\beta_i^0\right]$

Mean Intercept $E\left[\beta_i^0\right]$	Implementation	Time 1* (median)	Time 2* (mean)	Iterations° (median)
-1	Converged ABLP	16.9	19.5	5
	MPEC	82.7	105.9	
0	Converged ABLP	17.8	19.7	5
	MPEC	74.6	99.3	
1	Converged ABLP	20.9	23.2	5
	MPEC	74.1	103.7	
2	Converged ABLP	25.0	27.5	6
	MPEC	72.9	102.0	

∗ Time 1 & 2 : the CPU time (second) per starting point.

∘ The median of the numbers of iterations on convergence per starting point.

Table 2. Time Costs Varying the Number of Markets T

# of Markets (T)	Implementation	Time 1* (median)	Time 2* (mean)	Iterations° (median)
25	Converged ABLP	10.5	11.6	5
	MPEC	30.3	45.9	
50	Converged ABLP	20.9	23.2	5
	MPEC	74.1	103.7	
100	Converged ABLP	51.2	54.7	6
	MPEC	220.5	278.9	
200	Converged ABLP	125.0	125.4	6
	MPEC	1,295.9	1,357.7	

∗ Time 1 & 2 : the CPU time (second) per starting point.

∘ The median of the numbers of iterations on convergence per starting point.

Table 3. Time Costs Varying the Number of Products J

# of Products (J)	Implementation	Time 1* (median)	Time 2* (mean)	Iterations° (median)
25	Converged ABLP	20.9	23.2	5
	MPEC	74.1	103.7	
50	Converged ABLP	65.4	70.4	6
	MPEC	219.0	300.4	
100	Converged ABLP	174.3	182.6	7
	MPEC	2,189.2	2,221.7	

$*$ Time 1 & 2 : the CPU time (second) per starting point.

$\circ$ The median of the numbers of iterations on convergence per starting point.

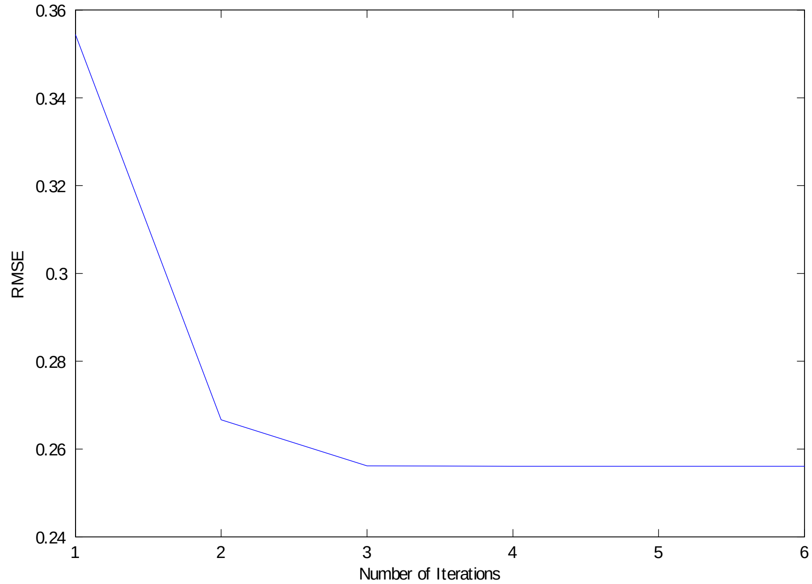
Table 4. Time Costs Varying the Variance of Random Coefficients $Var[\beta_i]$

Variance of Random Coefficients ($level * Var[\beta_i]$)	Implementation	Time 1* (median)	Time 2* (mean)	Iterations° (median)
0.5	Converged ABLP	20.6	22.4	5
	MPEC	69.2	93.1	
1	Converged ABLP	20.9	23.2	5
	MPEC	74.1	103.7	
2	Converged ABLP	23.0	26.3	7
	MPEC	81.4	109.9	

$*$ Time 1 & 2 : the CPU time (second) per starting point.

$\circ$ The median of the numbers of iterations on convergence per starting point.

Figure 1: Movement of RMSE with Number of Iterations*



∗ This result is from a dataset from the base data-generating process.

Table 5. Performance of the 2-step ABLP estimator Varying the Mean Intercept $E\left[\beta_i^0\right]$

Mean Intercept $E\left[\beta_i^0\right]$	2-step ABLP* (second)	Converged ABLP* (second)	MAE° (%)
-1	8.7	16.9	0.08
0	9.2	17.8	0.09
1	10.4	20.9	0.21
2	12.7	25.0	0.27

∗ The median of the CPU times per starting point

○ Mean absolute error (MAE) is the average difference between the own price elasticities of the 2-step ABLP and those of the Converged ABLP. It measures accuracy of the 2-step ABLP estimator.

Table 6. Performance of the 2-step ABLP estimator Varying the Number of Markets T

# of Markets (T)	2-step ABLP* (second)	Converged ABLP* (second)	MAE° (%)
25	5.6	10.5	0.14
50	10.4	20.9	0.21
100	32.1	51.2	0.33
200	77.2	125.0	0.11

∗ The median of the CPU times per starting point

○ Mean absolute error (MAE) is the average difference between the own price elasticities of the 2-step ABLP and those of the Converged ABLP. It measures accuracy of the 2-step ABLP estimator.

Table 7. Performance of the 2-step ABLP estimator Varying the Number of Products J

# of Products (J)	2-step ABLP* (second)	Converged ABLP* (second)	MAE° (%)
25	10.4	20.9	0.21
50	33.0	65.4	0.27
100	98.8	174.3	0.19

∗ The median of the CPU times per starting point

○ Mean absolute error (MAE) is the average difference between the own price elasticities of the 2-step ABLP and those of the Converged ABLP. It measures accuracy of the 2-step ABLP estimator.

Table 8. Performance of the 2-step ABLP estimator Varying the Variance of Random Coefficients

Variance of Random Coefficients ($level * Var[\beta_i]$)	2-step ABLP* (second)	Converged ABLP* (second)	MAE° (%)
0.5	12.5	20.6	0.50
1	10.4	20.9	0.21
2	10.3	23.0	0.89

∗ The median of the CPU times per starting point

○ Mean absolute error (MAE) is the average difference between the own price elasticities of the 2-step ABLP and those of the Converged ABLP. It measures accuracy of the 2-step ABLP estimator.

7 Appendix A : ABLP Setting

The ABLP moment function is

$$E\left[g_{jt}\left(\theta,\xi\right)\right]=0 \text{ where } g_{jt}\left(\theta,\xi\right)=\Phi_{jt}\left(\theta,\xi\right)z_{jt},$$

and $\Phi_{jt}\left(\theta,\xi\right)$ is the jt-th element of the mapping $\Phi\left(\theta,\xi\right)$ with an abuse of notation. Define $\overline{g}_J\left(\theta,\xi\right)=\frac{1}{J}\sum_{j=1}^{J}E\left[g_{jt}\left(\theta,\xi\right)\right]$. Then the population criterion function is

$$Q_0\left(\theta,\xi\right)=\overline{g}_J\left(\theta,\xi\right)' W\overline{g}_J\left(\theta,\xi\right),$$

where W is a weight matrix. We also write the sample analogue of the population criterion function as

$$Q_{JT}\left(\theta,\xi\right)=\widehat{\overline{g}}_{JT}\left(\theta,\xi\right)'\widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\theta,\xi\right),$$

where $\widehat{\overline{g}}_{JT}\left(\theta,\xi\right)=\frac{1}{JT}\sum_j\sum_t\Phi_{jt}\left(\theta,\xi\right)z_{jt}=\frac{1}{JT}Z'\Phi\left(\theta,\xi\right)$ and $\widehat{W}_{JT}\rightarrow_p W$.

With an abuse of notation, define $\Gamma_0\left(\xi\right)\equiv\arg\min_\theta Q_0\left(\theta,\xi\right)$, $\phi_0\left(\xi\right)\equiv\Phi\left(\Gamma_0\left(\xi\right),\xi\right)$. The ABLP population mapping consists of both $\theta=\Gamma_0\left(\xi\right)$ and $\xi=\phi_0\left(\xi\right)$. Let ∇_θ and ∇_ξ denote the first-order derivative with respect to θ and ξ, respectively. Let $\|\cdot\|$ denote the Euclidian norm. We start with imposing standard regularity conditions for the consistency and asymptotic normality of the nested pseudo-GMM estimator (i.e., Proposition 11 in Kasahara and Shimotsu 2009 and Condition 1 and 2 in Kim and Park 2010). We impose the regularity condition on $\Phi_t\left(\theta_0,\xi_0\right)$.

Condition 1 (a) $\widehat{W}_{JT}\rightarrow_p W$; (b) $\Phi\left(\theta,\xi\right)$ is three times continuously differentiable in the neighborhood of $\left(\theta_0,\xi_0\right)$; (c) Θ and B_ξ are compact: $\left(\theta,\xi\right)\in\Theta\times B_\xi$; (d) a unique $\theta_0\in int\left(\Theta\right)$ exists such that $\xi_0=\Phi\left(\theta_0,\xi_0\right)$; (e) $\Gamma_0\left(\xi\right)$ is a single-valued continuous function of ξ in a neighborhood of ξ_0; (f) the mapping $f\left(\xi\right)\equiv\phi_0\left(\xi\right)-\xi$ has a nonsingular Jacobian matrix at ξ_0; (g) $\overline{g}_J\left(\theta_0,\xi_0\right)=0$, $E\left[\|z_t\|\right]/\sqrt{J}<\infty$, $E\left[\|z_t\|^2\right]/\sqrt{J}<\infty$, $E\left[\|\Phi_t\left(\theta_0,\xi_0\right)\|\right]/\sqrt{J}<\infty$, $E\left[\|\Phi_t\left(\theta_0,\xi_0\right)\|^2\right]/\sqrt{J}<\infty$; (h) $E\left[\sup_{\theta\in\Theta}\|\nabla_\xi\Phi_t\left(\theta,\xi_0\right)\|^2\right]/\sqrt{J}<\infty$, $E\left[\sup_{\theta\in\Theta}\|\nabla_\theta\Phi_t\left(\theta,\xi_0\right)\|^2\right]/\sqrt{J}<\infty$; and
(i) $\nabla_\theta\overline{g}_J\left(\theta,\xi\right)' W\nabla_\theta\overline{g}_J\left(\theta,\xi\right)$ is nonsingular at $\left(\theta_0,\xi_0\right)$.

8 Appendix B : Proofs

Proof of Lemma 1. Assume without loss of generality that there is one market and J products exist in it. For $k = 1, 2, \cdots, J$,

$$\begin{aligned}
\Phi_k(\theta^*, \xi^*) &= \xi_k^* + \sum_{j=1}^{J} [\ln S_j - \ln s_j(\xi^*; \theta^*)] h_{kj} \\
\frac{\partial \Phi_k(\theta^*, \xi^*)}{\partial \xi'} &= 1 + \sum_{j=1}^{J} \left\{ [\ln S_j - \ln s_j(\xi^*; \theta^*)] \frac{\partial h_{kj}}{\partial \xi'} - h_{kj} \frac{\partial \ln s_j(\xi^*; \theta^*)}{\partial \xi'} \right\} \\
&= 1 - \sum_{j=1}^{J} h_{kj} \frac{\partial \ln s_j(\xi^*; \theta^*)}{\partial \xi'},
\end{aligned}$$

where $\ln S_j - \ln s_j(\xi^*; \theta^*) = 0$ for $j = 1, ..., J$, and h_{kj} is (k, j) element of the matrix $\left[\frac{\partial \ln s(\xi^*;\theta^*)}{\partial \xi'}\right]^{-1}$; that is,

$$\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1J} \\ h_{21} & h_{22} & \cdots & h_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ h_{J1} & h_{J2} & \cdots & h_{JJ} \end{pmatrix} \equiv \left[\frac{\partial \ln s(\xi^*; \theta^*)}{\partial \xi'}\right]^{-1}.$$

Thus we have

$$\begin{aligned}
\frac{\partial \Phi(\theta^*, \xi^*)}{\partial \xi'} &= I_J - \left[\frac{\partial \ln s(\xi^*; \theta^*)}{\partial \xi'}\right]^{-1} \frac{\partial \ln s(\xi^*; \theta^*)}{\partial \xi'} \\
&= I_J - I_J = 0_J.
\end{aligned}$$

■

Proof of Proposition 1.
Assume without loss of generality that there is one market and J products exist in it.
(BLP) The BLP minimization problem is as follows:

$$\min_{\theta \in \Theta} \ \xi(\theta)' Z \widehat{W}_{JT} Z' \xi(\theta),$$

where (1) $\xi(\theta)$ is obtained by the BLP contraction mapping: $\xi(\theta)$ satisfies $\ln S = \ln s(\xi(\theta); \theta)$. Suppose a local root $\widetilde{\theta}_{BLP}$ such that (2) the first-order condition (FOC) of the BLP minimization problem holds at $\widetilde{\theta}_{BLP}$:

$$2 \frac{d\xi(\theta)'}{d\theta} Z \widehat{W}_{JT} Z' \xi(\theta) |_{\theta = \widetilde{\theta}_{BLP}} = 0.$$

By the implicit function theorem and the market share equations, $\ln S = \ln s(\xi(\theta); \theta)$, for any θ,

$$\frac{d\xi(\theta)}{d\theta'} = -\left[\frac{\partial \ln s(\xi(\theta); \theta)}{\partial \xi'}\right]^{-1} \left[\frac{\partial \ln s(\xi(\theta); \theta)}{\partial \theta'}\right].$$

Then we can rewrite the FOC of the BLP minimization problem as

$$2\left\{-\left[\frac{\partial \ln s\left(\xi\left(\theta\right);\theta\right)}{\partial \xi'}\right]^{-1}\left[\frac{\partial \ln s\left(\xi\left(\theta\right);\theta\right)}{\partial \theta'}\right]\right\}' Z\widehat{W}_{JT}Z'\xi\left(\theta\right)|_{\theta=\widetilde{\theta}_{BLP}}=0. \tag{A.1}$$

(ABLP) The ABLP minimization problem with $\widetilde{\xi}_{ABLP}$ is as follows:

$$\min_{\theta\in\Theta}\ \Phi\left(\theta,\widetilde{\xi}_{ABLP}\right)Z\widehat{W}_{JT}Z'\Phi\left(\theta,\widetilde{\xi}_{ABLP}\right).$$

Suppose a fixed point $\left(\widetilde{\xi}_{ABLP},\widetilde{\theta}_{ABLP}\right)$ such that (1) (no updating) $\widetilde{\xi}_{ABLP}=\Phi\left(\widetilde{\theta}_{ABLP},\widetilde{\xi}_{ABLP}\right)$, and (2) the following first-order condition holds at $\widetilde{\theta}_{ABLP}$:

$$2\frac{\partial\Phi\left(\theta,\widetilde{\xi}_{ABLP}\right)'}{\partial\theta}Z\widehat{W}_{JT}Z'\Phi\left(\theta,\widetilde{\xi}_{ABLP}\right)|_{\theta=\widetilde{\theta}_{ABLP}}=0. \tag{A.2}$$

(The set of these fixed points includes the set of the fixed points of the ABLP mapping.)
Now we will show a fixed point $\left(\widetilde{\xi}_{ABLP},\widetilde{\theta}_{ABLP}\right)$ satisfies the market share equations, $\ln S=\ln s\left(\xi;\theta\right)$, and then that the FOC of the BLP minimization problem is equivalent to the FOC of the minimization problem in the ABLP mapping with no updating of ξ.
First, the equation, $\widetilde{\xi}_{ABLP}=\Phi\left(\widetilde{\theta}_{ABLP},\widetilde{\xi}_{ABLP}\right)$, implies no change or updating occurs in ξ at the fixed point as follows:

$$\begin{aligned}
\Phi\left(\widetilde{\theta}_{ABLP},\widetilde{\xi}_{ABLP}\right) &\equiv \widetilde{\xi}_{ABLP}+\left[\frac{\partial\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)}{\partial\xi'}\right]^{-1}\left[\ln S-\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)\right]\\
\Rightarrow\widetilde{\xi}_{ABLP} &= \widetilde{\xi}_{ABLP}+\left[\frac{\partial\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)}{\partial\xi'}\right]^{-1}\left[\ln S-\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)\right]\\
\Rightarrow 0 &= \left[\frac{\partial\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)}{\partial\xi'}\right]^{-1}\left[\ln S-\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)\right]\\
\Rightarrow 0 &= \ln S-\ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right).
\end{aligned} \tag{A.3}$$

Next, look at the first-order condition of the ABLP minimization problem in $(A.2)$:

$$2\frac{\partial\Phi\left(\theta,\widetilde{\xi}_{ABLP}\right)'}{\partial\theta}Z\widehat{W}_{JT}Z'\Phi\left(\theta,\widetilde{\xi}_{ABLP}\right)|_{\theta=\widetilde{\theta}_{ABLP}}=0,$$

where $\widetilde{\xi}_{ABLP} = \left(\widetilde{\xi}_{ABLP,1}, \cdots, \widetilde{\xi}_{ABLP,J}\right)'$, and for $i = 1, 2, \cdots, J$,

$$
\begin{aligned}
\Phi_i\left(\theta, \widetilde{\xi}_{ABLP}\right) &= \widetilde{\xi}_{ABLP} + \sum_{j=1}^{J}\left[\ln S_j - \ln s_j\left(\widetilde{\xi}_{ABLP};\theta\right)\right] f_{ij} \\
\frac{\partial \Phi_i\left(\theta, \widetilde{\xi}_{ABLP}\right)}{\partial \theta'} &= \sum_{j=1}^{J}\left\{\left[\ln S_j - \ln s_j\left(\widetilde{\xi}_{ABLP};\theta\right)\right]\frac{\partial f_{ij}}{\partial \theta'} - f_{ij}\frac{\partial \ln s_j\left(\widetilde{\xi}_{ABLP};\theta\right)}{\partial \theta'}\right\}
\end{aligned}
$$

where f_{ij} is (i,j) element of the matrix $\left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\theta\right)}{\partial \xi'}\right]^{-1}$; in other words,

$$
\begin{pmatrix}
f_{11} & f_{12} & \cdots & f_{1J} \\
f_{21} & f_{22} & \cdots & f_{2J} \\
\vdots & \vdots & \ddots & \vdots \\
f_{J1} & f_{J2} & \cdots & f_{JJ}
\end{pmatrix} \equiv \left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\theta\right)}{\partial \xi'}\right]^{-1}.
$$

If we evaluate both $\Phi\left(\theta, \widetilde{\xi}_{ABLP}\right)$ and $\frac{\partial \Phi\left(\theta, \widetilde{\xi}_{ABLP}\right)}{\partial \theta'}$ at $\widetilde{\theta}_{ABLP}$,

$$
\begin{aligned}
\Phi\left(\theta, \widetilde{\xi}_{ABLP}\right)|_{\theta=\widetilde{\theta}_{ABLP}} &= \widetilde{\xi}_{ABLP} && \text{(A.4)} \\
\frac{\partial \Phi\left(\theta, \widetilde{\xi}_{ABLP}\right)}{\partial \theta'}|_{\theta=\widetilde{\theta}_{ABLP}} &= -\left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\theta\right)}{\partial \xi'}\right]^{-1}\left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\theta\right)}{\partial \theta'}\right]|_{\theta=\widetilde{\theta}_{ABLP}}
\end{aligned}
$$

because $\ln S - \ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right) = 0$. Then the first-order condition of the ABLP minimization problem becomes

$$
2\left\{-\left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)}{\partial \xi'}\right]^{-1}\left[\frac{\partial \ln s\left(\widetilde{\xi}_{ABLP};\widetilde{\theta}_{ABLP}\right)}{\partial \theta'}\right]\right\}' ZAZ'\widetilde{\xi}_{ABLP} = 0. \tag{A.5}
$$

From the BLP contraction mapping and $(A.3)$, we can say both $\left(\xi\left(\widetilde{\theta}_{ABLP}\right), \widetilde{\theta}_{ABLP}\right)$ and $\left(\widetilde{\xi}_{ABLP}, \widetilde{\theta}_{ABLP}\right)$ satisfy the market share equations, $\ln S = \ln s(\xi;\theta)$. And from equations $(A.1)$ and $(A.5)$, we can say both $\left(\xi\left(\widetilde{\theta}_{ABLP}\right), \widetilde{\theta}_{ABLP}\right)$ and $\left(\widetilde{\xi}_{ABLP}, \widetilde{\theta}_{ABLP}\right)$ satisfy the same form of the first-order condition. This result implies that the set of the local roots of the BLP minimization, $\left\{\widetilde{\theta}_{BLP}, \xi\left(\widetilde{\theta}_{BLP}\right)\right\}$, is identical to the set of the fixed points of the ABLP minimization with no updating, $\left\{\widetilde{\theta}_{ABLP}, \widetilde{\xi}_{ABLP}\right\}$. In addition, for any pair $(\xi^\circ, \theta^\circ)$ (which is both a local root of the BLP minimization and a fixed point of the ABLP mapping), the value of the GMM objective function in the BLP minimization is equal to the value of the GMM objective function in the ABLP minimization: $\xi^\circ = \Phi(\theta^\circ, \xi^\circ)$ $\Rightarrow \xi^{\circ\prime} Z\widehat{W}_{JT}Z'\xi^\circ = \Phi(\theta^\circ, \xi^\circ)' Z\widehat{W}_{JT}Z'\Phi(\theta^\circ, \xi^\circ)$.

The BLP estimator, denoted by $\widehat{\theta}_{BLP}$, is the parameter θ with the lowest value of the BLP GMM objective function among the local roots of the FOC of the BLP minimization. The BLP estimator $\widehat{\theta}_{BLP}$ and its corresponding $\xi\left(\widehat{\theta}_{BLP}\right)$ become the fixed point with the lowest value of the ABLP GMM objective function among the fixed points of the FOC of the ABLP minimization with no updating. And the reverse is also true. Therefore, the BLP estimator is numerically equivalent to the Converged ABLP estimator. ■

Proof of Proposition 2.
Let ∇_θ and ∇_ξ denote the first-order derivative with respect to θ and ξ, respectively. Let $\|\cdot\|$ denote the Euclidian norm. Let $Q_{JT}\left(\theta^K,\xi^{K-1}\right) \equiv \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)' \widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right).$
By definition, θ^K satisfies the first-order condition (FOC):

$$\nabla_\theta Q_{JT}\left(\theta^K,\xi^{K-1}\right) = 0. \tag{A.6}$$

Expanding $\nabla_\theta Q_{JT}\left(\theta^K,\xi^{K-1}\right)$ in $(A.6)$ around $\left(\widehat{\theta},\widehat{\xi}\right)$ using the mean value theorem,

$$0 = \nabla_{\theta\theta} Q_{JT}\left(\overline{\theta},\overline{\xi}\right)\left(\theta^K - \widehat{\theta}\right) + \nabla_{\theta\xi} Q_{JT}\left(\overline{\theta},\overline{\xi}\right)\left(\xi^{K-1} - \widehat{\xi}\right), \tag{A.7}$$

where $\left(\overline{\theta},\overline{\xi}\right)$ lie between $\left(\theta^K,\xi^{K-1}\right)$ and $\left(\widehat{\theta},\widehat{\xi}\right)$, and $\nabla_\theta Q_{JT}\left(\widehat{\theta},\widehat{\xi}\right) = 0$. From $(A.7)$, we get the bound of $\theta^K - \widehat{\theta}$ as follows:

$$\begin{aligned}\theta^K - \widehat{\theta} &= \left[\nabla_{\theta\theta} Q_{JT}\left(\overline{\theta},\overline{\xi}\right)\right]^{-1} \nabla_{\theta\xi} Q_{JT}\left(\overline{\theta},\overline{\xi}\right)\left(\xi^{K-1} - \widehat{\xi}\right) \\ &= O\left(\left\|\xi^{K-1} - \widehat{\xi}\right\|\right).\end{aligned} \tag{A.8}$$

For the bound of $\xi^K - \widehat{\xi}$, we need to expand the right-hand side of $\xi^K = \Phi\left(\theta^K,\xi^{K-1}\right)$ around $\left(\widehat{\theta},\widehat{\xi}\right)$:

$$\begin{aligned}&\xi^K = \Phi\left(\widehat{\theta},\widehat{\xi}\right) + \nabla_\theta\Phi\left(\widehat{\theta},\widehat{\xi}\right)\left(\theta^K - \widehat{\theta}\right) + \nabla_\xi\Phi\left(\widehat{\theta},\widehat{\xi}\right)\left(\xi^K - \widehat{\xi}\right) + O\left(\left\|\xi^{K-1} - \widehat{\xi}\right\|^2\right) \\ \Rightarrow\ &\xi^K - \widehat{\xi} = \nabla_\theta\Phi\left(\widehat{\theta},\widehat{\xi}\right)\left(\theta^K - \widehat{\theta}\right) + O\left(\left\|\xi^{K-1} - \widehat{\xi}\right\|^2\right)\end{aligned} \tag{A.9}$$

because $\Phi\left(\widehat{\theta},\widehat{\xi}\right) = \widehat{\xi};\ \nabla_\xi\Phi\left(\widehat{\theta},\widehat{\xi}\right) = 0$ by Lemma 1 and $\theta^K - \widehat{\theta} = O\left(\left\|\xi^{K-1} - \widehat{\xi}\right\|\right)$ in $(A.8)$.

Expanding $\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)$ around $\left(\widehat{\theta},\widehat{\xi}\right)$ and using $\nabla_\theta\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)' \widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right) =$

$O_p\left((JT)^{-\frac{1}{2}}\left\|\theta^K-\widehat{\theta}\right\|\right)+O_p\left((JT)^{-\frac{1}{2}}\left\|\xi^{K-1}-\widehat{\xi}\right\|\right)$ in the over-identified model,

$$\begin{aligned}
& \nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right) \\
= \; & \nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)+\nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\nabla_\theta\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)\left(\theta^K-\widehat{\theta}\right) \\
& +\nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\nabla_\xi\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)\left(\xi^{K-1}-\widehat{\xi}\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).
\end{aligned}$$

$$\Rightarrow 0=\nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\nabla_\theta\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)\left(\theta^K-\widehat{\theta}\right)+O_p\left((JT)^{-\frac{1}{2}}\left\|\xi^{K-1}-\widehat{\xi}\right\|\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).$$

because $\nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)=0$ (FOC) and $\nabla_\xi\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)=0$ by Lemma 1. Thus we have

$$\theta^K-\widehat{\theta}=O_p\left((JT)^{-\frac{1}{2}}\left\|\xi^{K-1}-\widehat{\xi}\right\|\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right). \tag{A.10}$$

The result in $(A.10)$ substitutes for $\theta^K-\widehat{\theta}$ in $(A.9)$. Therefore, the bound of $\xi^K-\widehat{\xi}$ is

$$\xi^K-\widehat{\xi}=O_p\left((JT)^{-\frac{1}{2}}\left\|\xi^{K-1}-\widehat{\xi}\right\|\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).$$

By the way, in the exactly-identified model, $\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)=0$. Thus, Expanding $\widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)$ around $\left(\widehat{\theta},\widehat{\xi}\right)$ yields

$$\begin{aligned}
0 \; = \; & \nabla_\theta \widehat{\overline{g}}_{JT}\left(\theta^K,\xi^{K-1}\right)'\widehat{W}_{JT}\nabla_\theta\widehat{\overline{g}}_{JT}\left(\widehat{\theta},\widehat{\xi}\right)\left(\theta^K-\widehat{\theta}\right)+O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right). \\
\Rightarrow \; & \theta^K-\widehat{\theta}=O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).
\end{aligned}$$

In the same way as above, we have

$$\xi^K-\widehat{\xi}=O\left(\left\|\xi^{K-1}-\widehat{\xi}\right\|^2\right).$$

■

9 Appendix C : ABLP with Concentration on Nonlinear Parameters [17]

Similar to BLP (see Nevo 2000), ABLP restricts its parameter search over nonlinear parameters $\theta \in \Theta_{NL}$, and IV regression estimates its linear parameters β separately. The ABLP with a concentration on nonlinear parameters has the same properties as ABLP. Therefore, all proofs of ABLP will be applied to this version of ABLP. Similar to ABLP, ABLP with concentration starts with an initial guess δ^0 for the unobserved product characteristic. At each iteration $K \geq 1$, the following two stages occur:

Stage 1 [**Update of** θ]: Obtain a new GMM estimate of θ as

$$\theta^K = \Gamma\left(\delta^{K-1}\right) \equiv \underset{\theta \in \Theta_{NL}}{\arg\min}\ \Phi\left(\theta, \delta^{K-1}\right)' Z\widehat{W}_{JT}Z'\Phi\left(\theta, \delta^{K-1}\right),$$

where

$$\Phi\left(\theta, \delta^{K-1}\right) = \delta\left(\theta, \delta^{K-1}\right) - X\beta,$$
$$\delta\left(\theta, \delta^{K-1}\right) = \delta^{K-1} + \left[\frac{\partial \ln S\left(\theta, \delta^{K-1}\right)}{\partial \delta'}\right]^{-1}\left[\ln S - \ln S\left(\theta, \delta^{K-1}\right)\right],$$
$$\beta = (X'P_{ZW}X)^{-1}X'P_{ZW}\delta\left(\theta, \delta^{K-1}\right), \quad P_{ZW} = Z\widehat{W}_{JT}Z'.$$

Stage 2 [**Update of** δ **]:** Update δ using the θ^K of step 1; that is,

$$\delta^K = \Phi\left(\theta^K, \delta^{K-1}\right).$$

Iterate in K until convergence in δ and θ is reached.

[17] Recently, Moon, Shum, and Weidner (2010) (MSW) proposed a complementary estimation method to BLP by adding interactive fixed effects and using a nested fixed point algorithm. ABLP in Appendix C is applicable to their model. To our knowledge, however, the MPEC approach cannot be applied to the MSW model.

10 Appendix D : Partial Derivatives

- The following derivatives are calculated by market t. Thus, we omit subscript t for simple notation.
- $\frac{\partial \ln s(\theta,\xi)}{\partial \xi'} \left(= \frac{\partial \ln s(\theta,\delta)}{\partial \delta'}\right)$, where $\ln s = (\ln s_1, ..., \ln s_J)'$, and $\frac{\partial \ln s}{\partial \xi'} = \left(\frac{\partial \ln s}{\partial \xi_1}, ..., \frac{\partial \ln s}{\partial \xi_J}\right)$.

$$\begin{aligned}
\frac{\partial \ln s(\xi;\theta)}{\partial \xi'} &= \begin{bmatrix} \frac{\partial s_1}{s_1 \partial \xi_1} & \frac{\partial s_1}{s_1 \partial \xi_2} & \cdots & \frac{\partial s_1}{s_1 \partial \xi_J} \\ \frac{\partial s_2}{s_2 \partial \xi_1} & \frac{\partial s_2}{s_2 \partial \xi_2} & \cdots & \frac{\partial s_2}{s_2 \partial \xi_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial s_J}{s_J \partial \xi_1} & \frac{\partial s_J}{s_J \partial \xi_2} & \cdots & \frac{\partial s_J}{s_J \partial \xi_J} \end{bmatrix}_{\theta,\ \xi} \\
&= \begin{bmatrix} \frac{1}{s_1}\frac{1}{I}\sum_{i=1}^{I}\left(s_{i1} - s_{i1}^2\right) & \frac{1}{s_1}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{i1}s_{i2}\right) & \cdots & \frac{1}{s_1}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{i1}s_{iJ}\right) \\ \frac{1}{s_2}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{i2}s_{i1}\right) & \frac{1}{s_2}\frac{1}{I}\sum_{i=1}^{I}\left(s_{i2} - s_{i2}^2\right) & \cdots & \frac{1}{s_2}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{i2}s_{iJ}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{s_J}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{iJ}s_{i1}\right) & \frac{1}{s_J}\frac{1}{I}\sum_{i=1}^{I}\left(-s_{iJ}s_{i2}\right) & \cdots & \frac{1}{s_J}\frac{1}{I}\sum_{i=1}^{I}\left(s_{iJ} - s_{iJ}^2\right) \end{bmatrix}_{\theta,\ \xi},
\end{aligned}$$

where $s_{ij} = \frac{\exp\left(X_j\beta_i+\xi_j\right)}{1+\sum_{j'=1}^{J}\exp\left(X_{j'}\beta_i+\xi_{j'}\right)}$; $s_j = \frac{1}{I}\sum_{i=1}^{I} s_{ij}$; $\frac{\partial s_j}{\partial \xi_j} = \frac{1}{I}\sum_{i=1}^{I}\left(s_{ij} - s_{ij}^2\right)$ for $j = 1, ..., J$; and $\frac{\partial s_j}{\partial \xi_k} = \frac{1}{I}\sum_{i=1}^{I}\left(-s_{ij}s_{ik}\right)$ for $j \neq k$ and $k = 1, ..., J$.

- $\frac{\partial \ln s_j}{\partial \theta_k}$:

  $\frac{\partial s_j}{\partial \theta_k}$ is as follows:

$$\begin{aligned}
\frac{\partial s_j}{\partial \beta_k} &= \frac{1}{I}\sum_{i=1}^{I}\frac{\partial s_{ij}}{\partial \beta_k} = \frac{1}{I}\sum_{i=1}^{I} s_{ij}\left(x_j^k - \sum_{j'=1}^{J} x_{j'}^k s_{ij'}\right). \\
\frac{\partial s_j}{\partial \sigma_k} &= \frac{1}{I}\sum_{i=1}^{I}\frac{\partial s_{ij}}{\partial \sigma_k} = \frac{1}{I}\sum_{i=1}^{I} v_{ik} s_{ij}\left(x_j^k - \sum_{j'=1}^{J} x_{j'}^k s_{ij'}\right).
\end{aligned}$$

  We can get $\frac{\partial \ln s_j}{\partial \theta_k}$ using $\frac{\partial s_j}{s_j \partial \theta_k}$.

11 Appendix E.

11.1 Nonparametric Specification

The ABLP approach is not limited to a parametric distribution. We can approximate any well-behaved distribution using the histogram approach that Kamakura (2001) proposes and that Bajari, Fox, Kim, and Ryan (2009) apply for a discrete-choice model..

First, we can model the distribution of the random coefficients as a mixture of point masses. The market share function is approximated as follows:

$$s_j(\xi_t, \theta) = \sum_{r=1}^{R} \theta^r s_j(\xi_t, \beta^r), \tag{C.1}$$

where there are R basis points $\beta = (\beta^1, ..., \beta^R)$, β^r is the K-dimensional preference parameter and θ^r is the probability of basis r, respectively. The $\theta = (\theta^1, ..., \theta^R)$ must satisfy

$$\sum_{r=1}^{R} \theta^r = 1, \quad \theta^r \geq 0. \tag{C.2}$$

We could interpret that R types of agents with the proportion θ exist.

Second, we could instead model the density function of the random coefficients as a mixture of normal densities in order to have a smooth distribution of the random coefficients. In this model, basis r denotes a joint normal function with K independent normal densities as follows.

$$N(\beta^r | \mu^r, \sigma^r) = \prod_k N(\beta_k^r | \mu_k^r, \sigma_k^r),$$

where $\mu^r = (\mu_1^r, ..., \mu_K^r)$, and $\sigma^r = (\sigma_1^r, ..., \sigma_K^r)$. Let θ^r denote the probability weight to basis r. With basis r, we simulate I draws from $N(\beta^r | \mu^r, \sigma^r)$. Let a particular draw i be denoted by $\beta^{r,i}$. The market share function is

$$s_j(\xi_t, \theta) = \sum_{r=1}^{R} \theta^r \left(\frac{1}{I} \sum_{i=1}^{I} s_j\left(\xi_t, \beta^{r,i}\right) \right). \tag{C.3}$$

The $\theta = (\theta^1, ..., \theta^R)$ must satisfy $(C.2)$.

As in the case of the parametric market share function, we can apply the ABLP algorithm to the nonparametric market share function.

11.2 Estimation

The Convergent ABLP's estimation of parameters and standard errors is the same as BLP's. A straightforward approach to this model is to find parameters satisfying the following condition

$$g_{jt}(\theta) = \xi_{jt}(\theta) z_{jt} \text{ and } E[g_{jt}(\theta)] = 0,$$

where ξ_{jt} is the unobserved product characteristic and z_{jt} is instrumental variables for price p_{jt} for $j = 1, ..., J,\ t = 1, ..., T$. Note the sample analog to the population moment is

$$g_N(\theta) = \frac{1}{JT}\sum_j \sum_t \xi_{jt}(\theta) Z_{jt} = \frac{1}{N}\sum_n h_n(\theta),$$

where N denotes JT and $h_n(\theta) \equiv \xi_{jt}(\theta) Z_{jt}$.

If we have more instrumental variables than parameters, we can run GMM estimation by minimizing

$$g_N(\theta)' W g_N(\theta),$$

where W is a weight matrix and $V \equiv Var(g_N(\theta))$ in order to minimize variance of the GMM estimate. Note that the sample analog to V is given by

$$\begin{aligned}
\widehat{V} &\equiv Var\left(g_N\left(\widehat{\theta}\right)\right) = Var\left(\frac{1}{N}\sum_n h_n\left(\widehat{\theta}\right)\right) = \frac{1}{N^2}\sum_n Var\left(h_n\left(\widehat{\theta}\right)\right) \\
&= \frac{1}{N} Var\left(h_n\left(\widehat{\theta}\right)\right) = \frac{1}{N} E\left[h_n\left(\widehat{\theta}\right) h_n\left(\widehat{\theta}\right)'\right] = \frac{1}{N^2}\sum_n h_n\left(\widehat{\theta}\right) h_n\left(\widehat{\theta}\right)'
\end{aligned}$$

since $h_n(\theta)$ is i.i.d. and $E[h_n(\theta)] = 0$.

In estimation, we can start with some initial θ, compute W and then find the θ that minimizes the objective function. We can iteratively recompute the weight matrix W and re-estimate until the estimates don't change much between iterations. If we have the same number of moments as parameters, the choice of the weighting matrix W doesn't affect estimates of θ, in principle. However, if there is over-identification, this iteration procedure obtains efficient estimates conditional on instrumental variables Z.

11.3 Standard Error

Let $Q_N(\theta) = \frac{1}{2} g_N(\theta)' W g_N(\theta)$. GMM estimates $\widehat{\theta}$ satisfy

$$\sqrt{N}\left(\widehat{\theta} - \theta\right) = -\left(\frac{\partial^2 Q_N\left(\widetilde{\theta}\right)}{\partial\theta\partial\theta'}\right)^{-1} \sqrt{N}\frac{\partial Q_N(\theta)}{\partial\theta},$$

where $\widetilde{\theta}$ is between $\widehat{\theta}$ and θ.

$$\sqrt{N}\left(\widehat{\theta}-\theta\right) \xrightarrow{d} N\left(0,\left(\Gamma' W \Gamma\right)^{-1} \Gamma' W \Omega W \Gamma\left(\Gamma' W \Gamma\right)^{-1}\right),$$

where

$$\begin{aligned}
\Gamma &\equiv E\left[\frac{\partial g_N(\theta)}{\partial \theta'}\right]=E\left[\frac{\partial h_n(\theta)}{\partial \theta'}\right] \\
\Omega &\equiv Var\left(\sqrt{N} g_N(\theta)\right)=E\left[h_n(\theta) h_n(\theta)'\right] \\
\frac{\partial^2 Q_N\left(\widetilde{\theta}\right)}{\partial \theta \partial \theta'} &\xrightarrow{p} \Gamma' W \Gamma \\
\sqrt{N} \frac{\partial Q_N(\theta)}{\partial \theta} &\xrightarrow{d} N\left(0, \Gamma' W \Omega W \Gamma\right).
\end{aligned}$$

If we ignore the simulation error, the asymptotic variance matrix of a GMM estimate $\widehat{\theta}$ is given by

$$AVar\left(\widehat{\theta}\right)=\left(\Gamma' W \Gamma\right)^{-1} \Gamma' W \frac{\Omega}{N} W \Gamma\left(\Gamma' W \Gamma\right)^{-1}.$$

The asymptotic variance matrix can be approximated as

$$\widehat{AVar}\left(\widehat{\theta}\right)=\left(\widehat{\Gamma}' W \widehat{\Gamma}\right)^{-1} \widehat{\Gamma}' W \frac{\widehat{\Omega}}{N} W \widehat{\Gamma}\left(\widehat{\Gamma}' W \widehat{\Gamma}\right)^{-1}$$

with

$$\begin{aligned}
\widehat{\Gamma} &= \frac{1}{N} \sum_n \frac{\partial h_n\left(\widehat{\theta}\right)}{\partial \theta'} \\
\widehat{\Omega} &= \frac{1}{N} \sum_n h_n\left(\widehat{\theta}\right) h_n\left(\widehat{\theta}\right)'
\end{aligned}$$

where these derivatives are numerically obtained and the weight matrix W is fixed.

For a parametric specification of market share function, the standard errors of parameters are the diagonal elements of $AVar\left(\widehat{\theta}\right)^{\frac{1}{2}}$. For a nonparametric specification, however, θ contains the probability weights, which are not the parameters of interest. Let $B(\theta)$ be the parameters of interest. By the delta method,

$$\sqrt{N}\left(B\left(\widehat{\theta}\right)-B(\theta)\right) \xrightarrow{d} N\left(0, \nabla B(\theta)'\left(\Gamma' W \Gamma\right)^{-1} \Gamma' W \Omega W \Gamma\left(\Gamma' W \Gamma\right)^{-1} \nabla B(\theta)\right).$$

The asymptotic distribution of $B\left(\widehat{\theta}\right)$ provides the standard errors of interest.