

A Computationally Efficient Fixed Point Approach to Dynamic Structural Demand Estimation^{*†}

Yutec Sun
KU Leuven

Masakazu Ishihara
New York University

June 13, 2017

Abstract

This paper develops a computationally efficient approach to the estimation of random coefficients logit model of dynamic consumer demand using product panel data. The conventional GMM estimation relies on two computationally intensive fixed point algorithms, each developed by Rust (1987) and Berry, Levinsohn, and Pakes (1995), nested within an optimization routine. We transform the GMM estimation into a quasi-Bayesian (Laplace type) framework and develop a Markov Chain Monte Carlo (MCMC) method that solves the fixed point problems incrementally with Markov chain simulation. The proposed approach has two main advantages. First, it reduces the computational burden of the nested fixed point (NFP) algorithm employed by GMM without sacrificing the model flexibility. Our Monte Carlo experiments demonstrate that the new method outperforms both NFP and MPEC in computational speed by substantial margin, particularly in the most computationally intensive estimations. Second, the proposed method requires only moment restrictions as GMM, thereby avoiding the risk of misspecification bias in equilibrium frameworks.

JEL Classification: C11, C13, C51

Keywords: Nested fixed point, BLP, dynamic, MCMC, Laplace, random coefficients logit

^{*}We thank Victor Aguirregabiria, Andrew Ching, J.P. Dubé, Avi Goldfarb, Susumu Imai, Matt Shum, and Frank Verboven for their valuable comments. We thank Telecom-ParisTech for the provision of computing facilities.

[†]A sample program code used in the Monte Carlo analysis can be obtained at <https://github.com/yutec/LTE.git>.

1 Introduction

In the empirical industrial organization literature, there is a growing need for a more computationally efficient strategy for estimating dynamic structural models of aggregate demand. Since the seminal work of Melnikov (2013), the model structure has become increasingly complex as the subsequent literature has developed to account for unobserved consumer heterogeneity and price endogeneity.

As a strategy to estimate these models with increased complexity, Gowrisankaran and Rysman (2012) propose a nested fixed point (NFP) algorithm. The NFP algorithm embeds Rust's (1987) dynamic programming algorithm as a subroutine under Berry, Levinsohn and Pakes's (1995) market share inversion procedure, which is commonly called as BLP contraction mapping. By incorporating these two fixed point algorithms within a generalized method of moments (GMM) framework, the NFP method estimates random coefficients logit models without specifying either the distribution of unobservables or the reduced form function of endogenous variables. However, it faces a substantial challenge due to the need to repeatedly solve the two fixed point problems, each computationally intensive, at every step of parameter search. Hence, the high computational cost has limited the use of the random coefficients specification in empirical studies.

To overcome the limitation of the NFP algorithm, we develop an alternative estimation method inspired by the Bayesian dynamic programming algorithm of Imai, Jain and Ching (2009) (henceforth IJC algorithm). They solve the Bellman equation incrementally with MCMC iterations instead of finding exact solutions at every iteration. This incremental approach motivates our fixed point computation for both the BLP and the Bellman procedures. We call it *pseudo fixed point* (PFP) algorithm since the algorithm uses partially solved fixed points to simulate the Markov chain. The partial solution approach employs a standard MCMC procedure that requires only a few fixed point operations per iteration in order for the partial solutions to converge to the exact fixed points. Hence, by replacing the computationally intensive fixed point computation with the partial solution procedure, the PFP algorithm can significantly reduce the computational burden of the NFP algorithm.

However, there exist important tradeoffs in employing the MCMC method for structural demand estimation. One of the main criticisms of Bayesian structural estimation is aimed at the potential bias that may arise from the misspecification of likelihood models on the supply side.¹ Equilibrium frameworks may yield ambiguous endogenous outcomes, providing incomplete information for fully characterizing the likelihood functions. Furthermore, the Bayesian approach needs to numerically compute a Jacobian matrix for likelihood evaluation in most cases, which can be time-consuming. Hence, these tradeoffs may offset the advantages of the proposed MCMC method if implemented within the Bayesian framework.

¹See the comments of Bajari (2003), Berry (2003), and Dubé and Chintagunta (2003) on the Bayesian demand estimation of Yang, Chen and Allenby (2003).

Instead, we incorporate the PFP algorithm in a classical estimation setting to circumvent the disadvantages of the Bayesian method. We adopt a quasi-Bayesian approach that transforms the GMM objective function into a quasi-posterior distribution. The estimation approach is based on the Laplace type (LT) estimation framework (Chernozhukov and Hong, 2003), which needs only a statistical criterion function instead of a likelihood to form a posterior-like distribution that resembles the Bayesian posterior. Hence, the LT estimation (LTE) framework allows us to compute the GMM estimator using the MCMC method with only minimal modification; this implies that our PFP approach can achieve computational efficiency without relying on *ad hoc* supply-side models. This contrasts with alternative approaches such as simulated maximum likelihood, control function, and Bayesian methods that need to specify either the joint or the conditional distribution of unobservables in the supply and demand.

This paper contributes to the growing literature on dynamic demand estimation. Dubé, Fox and Su (2012) propose to recast the classical extremum estimation into a constrained optimization problem, which can be solved by the mathematical program with equilibrium constraints (MPEC) algorithm. They use Monte Carlo analysis to demonstrate that the MPEC algorithm can offer a numerically efficient and reliable alternative to NFP in the static and dynamic BLP demand estimations. However, one has to exercise caution in generalizing their results beyond their formulation of dynamic demand since it shows some important departures from the standard framework: for example, the random coefficient distribution is degenerated to have two point masses, the demand system contains only two products, and a likelihood-based approach is employed for estimation. Thus, it is unclear whether MPEC would maintain its advantage in the settings for which the NFP algorithm was originally developed.

We conduct Monte Carlo experiments to analyze the computational performance of NFP, MPEC, and PFP in the dynamic BLP demand estimation. While we do not employ field data, we adopt a data generating process similar to the Monte Carlo example of Dubé et al. (2012) with necessary changes made to bring the model closer to the dynamic BLP framework commonly estimated by the NFP method. We focus on the GMM estimation of a dynamic programming logit demand that contains fully continuous distribution of random coefficients. Within the range of parameter values tested, our experiments find that MPEC is outperformed by NFP in computational speed unless the BLP modulus is exceedingly high to the extent that NFP would fail to converge. Even in such cases, MPEC does not deliver superior performance relative to PFP despite spending far more time than the two fixed point algorithms. The Monte Carlo experiments highlight the fact that the intertemporal dependence of the dynamic BLP demand renders the constraint Jacobian matrix almost fully dense, such that it creates a substantial challenge to MPEC, particularly when the demand system is large. It is the lack of sparsity in the large scale estimation that creates the main computational bottleneck, a problem that MPEC did not face in the estimation of static or a restricted form of dynamic demand considered by Dubé et al. (2012).

On the other hand, we find that the PFP algorithm exhibits a computational advantage

over NFP particularly when the fixed point convergence is costly. This feature makes the PFP algorithm potentially useful for cases where a large state space or a high-dimensional unobserved consumer heterogeneity may cause the estimation to become prohibitively costly, a unique problem that exist exclusively in the dynamic BLP demand estimation. We also find that PFP can still perform more efficiently than MPEC with a comparable rate of convergence even when the BLP contraction mapping condition fails to hold in large part of the parameter space.

The paper also adds to the static BLP demand estimation literature. Recently, Lee and Seo (2015) have developed the approximate BLP (ABLP) method, a new GMM estimation algorithm that outperforms both NFP and MPEC. The improvement in computational performance is made possible by replacing the fixed point procedure with a derivative-based solver for the BLP fixed point. While ABLP can deliver faster convergence than NFP in the static model where the derivatives can be analytically evaluated, it is not obvious whether it would maintain the computational advantage in dynamic BLP that incurs dramatically increased cost of derivatives evaluation.

In the Bayesian literature, Jiang, Manchanda and Rossi (2009) develop a Bayesian approach using a limited information likelihood function. Romeo (2007) and Duan and Mela (2009) construct a Bayesian GMM estimator by assuming a parametric distribution for the GMM moment. However, their approaches are subject to the same criticisms of the likelihood-based inference; furthermore, they tend to have a heavier computational burden than the GMM estimation of Berry et al. (1995) since they employ the same NFP procedure inside the MCMC algorithm.

Beyond the structural demand literature, the LTE has been used to deal with other types of computational problems. It has previously been applied to the estimation of static and dynamic oligopoly games that relies on simulated moments (Bajari, Hong and Ryan, 2010) or moment inequalities (Ryan, 2012; Wan, 2013), where the moments may not possess differentiability, convexity, or even continuity. Chernozhukov and Hong (2003) show that LTE can provide a computationally attractive method for a wide range of classical estimation problems. However, in structural demand estimation with well-behaved moments, the application of LTE has been rather limited due to the relatively high computational burden associated with MCMC simulations.

The rest of the paper proceeds as follows. The next section provides a brief background on the LT estimation in a canonical example of dynamic demand model. Section 3 describes the proposed algorithm, and Section 4 presents theoretical results on the convergence of the pseudo fixed points. In Section 5, we compare the empirical performance of the PFP, NFP and MPEC algorithms in Monte Carlo experiments. We discuss the strengths and weaknesses of each algorithm and conclude in the final section.

2 Estimation of dynamic aggregate demand

2.1 A dynamic model of differentiated product demand

We illustrate the new estimation approach for a dynamic model of consumer demand for differentiated durable products that has been widely used in the empirical analysis of the adoption of new technology products (Melnikov, 2013; Carranza, 2010; Dubé, Hitsch and Chintagunta, 2010; Lee, 2013). Following Melnikov (2013), the model is characterized by forward-looking consumers making an optimal decision on the timing of adoption as well as on the choice of product for adoption based on their expectation of the future product evolution. The consumer decision is formulated as a solution to a discrete choice dynamic programming problem, which is a special case of the Markov decision process (Rust, 1994). The aggregate demand system for differentiated products is modeled by a random coefficients logit framework developed by Berry (1994) and Berry et al. (1995).

Suppose that the researcher has a panel data that tracks the sales and characteristics of J products over T discrete time periods. In each period, the consumers, indexed by $i = 1, \dots, N$, choose either to purchase a product that gives the highest utility or to delay the purchase decision to the next period. The consumers exit the market once a purchase is made. Upon purchasing product j at time t , consumer i receives a terminal utility net of adoption cost

$$U_{ijt} = x'_{jt}\theta_i + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, \dots, J, \quad (1)$$

where the vector x_{jt} consists of all observed product characteristics including prices, ξ_{jt} is a scalar of time-varying random shock to product quality unobserved to the econometrician, and ϵ_{ijt} is a random preference shock independently drawn from an extreme value distribution with its scale normalized for all i, j , and t . The consumer receives a one-period flow utility $U_{i0t} = \epsilon_{i0t}$ from choosing the outside option.

Let $\omega_{it} \in \Omega$ represent the state of market for consumer i at time t , where Ω is a complete separable metric space. The state ω_{it} may be a scalar (Melnikov, 2013; Gowrisankaran and Rysman, 2012) or a multi-dimensional vector (Lee, 2013). For illustration, we take the first approach by defining a scalar state variable ω_{it} as

$$\omega_{it} = \log \left[\sum_{j=1}^J \exp (x'_{jt}\theta_i + \xi_{jt}) \right]. \quad (2)$$

The literature usually imposes a stationary first-order Markov process $\mathcal{F}_i(\omega_{it+1}|\omega_{it})$ to describe the consumer's expectation about the future state of the market (Melnikov, 2013; Hendel and Nevo, 2006; Gowrisankaran and Rysman, 2012; Aguirregabiria and Nevo, 2013).² It is also

²Additionally, \mathcal{F}_i is specified such that it has a regular conditional density as a Radon-Nikodym derivative with respect to the Lebesgue measure on Ω .

commonly assumed that the consumers have rational expectations, i.e., the distribution \mathcal{F}_i is consistent with the realized observations of $\{\omega_{it}\}_t$.

Let $\tilde{V}_i(\omega_{it}, \epsilon_{it})$ denote the consumer's intertemporal utility generated by optimal stationary decision rules conditional on state ω_{it} and random taste shocks $\epsilon_{it} = [\epsilon_{i0t}, \dots, \epsilon_{iJt}]$ over infinite time horizon. Under regularity conditions (Rust, 1988, Theorem 3.1), \tilde{V}_i is a unique solution to the Bellman equation

$$\tilde{V}_i(\omega_{it}, \epsilon_{it}) = \max \left\{ \epsilon_{i0t} + \beta \mathbb{E}_{\omega_{it+1}, \epsilon_{it+1}} \left[\tilde{V}_i(\omega_{it+1}, \epsilon_{it+1}) | \omega_{it} \right], \max_{j=1, \dots, J} \{U_{ijt}\} \right\},$$

where $\beta \in (0, 1)$ is a discount factor and \mathbb{E} is the expectation operator for $(\omega_{it+1}, \epsilon_{it+1})$ conditional on the consumer's information ω_{it} . Then the above equation can be integrated with respect to ϵ_{it} on both sides to obtain the expected value function

$$V_i(\omega_{it}) \equiv \mathbb{E}_{\epsilon_{it}} \tilde{V}_i(\omega_{it}, \epsilon_{it}) = \log \left[\exp \left(\beta \mathbb{E}_{\omega_{it+1}} [V_i(\omega_{it+1}) | \omega_{it}] \right) + \exp(\omega_{it}) \right]. \quad (3)$$

Suppose the taste parameter θ_i is independently drawn from a distribution $F(\cdot | \theta)$ characterized by a parameter vector $\theta = (\bar{\theta}, \Sigma)$ that consists of a mean $\bar{\theta}$ and a variance Σ . Then for those who have not adopted any of the products in the past, the probability of choosing product j at time t endowed with taste θ_i is

$$\pi_{ijt}(\delta, V_i, \theta_i) = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\exp \left(\beta \mathbb{E}_{\omega_{it+1}} [V_i(\omega_{it+1}) | \omega_{it}] \right) + \sum_{k=1}^J \exp(\delta_{kt} + \mu_{ikt})},$$

where $\delta_{jt} = x'_{jt} \bar{\theta} + \xi_{jt}$, and $\mu_{ijt} = x'_{jt}(\theta_i - \bar{\theta})$, with (δ, V_i) denoting the collection of δ_{jt} and $V_i(\omega_{it})$: $\delta = [\delta_{11}, \dots, \delta_{JT}]$ and $V_i = [V_i(\omega_{i1}), \dots, V_i(\omega_{iT})]$. For consumer i , we define h_{it} as the probability of the consumer's participation as a potential buyer in the market at time t , i.e., the probability of not purchasing any of the products prior to time t . We assume that the market starts with zero prior adoption in the initial period: for each i , $h_{i1} = 1$, with the subsequent participation rate evolving over time as

$$h_{it}(\delta, V_i, \theta_i) = h_{it-1}(\delta, V_i, \theta_i) \left[1 - \sum_{j=1}^J \pi_{ijt-1}(\delta, V_i, \theta_i) \right] \quad \text{for } t \geq 2.$$

Then consumer i 's unconditional probability of adopting product j at time t is $\pi_{ijt} h_{it}$, which is then aggregated across consumers to yield the market share of product j :

$$S_{jt}(\delta, V, \theta) = \int \pi_{ijt}(\delta, V_i, \theta_i) h_{it}(\delta, V_i, \theta_i) F(d\theta_i | \theta). \quad (4)$$

The market share function takes as inputs $\delta = [\delta_{11}, \dots, \delta_{JT}]$ and $V = [V_1, \dots, V_N]$. The GMM estimator solves for δ and V that match the above market share function with the observed

shares at each step of parameter search, which constitutes the most computationally intensive part of the NFP algorithm. Given this dynamic demand model, we now turn to the description of the GMM and the LT estimation approaches.

2.2 GMM and Laplace type estimators

Without loss of generality, we focus on estimating the true structural parameters $\theta_0 = [\bar{\theta}_0, \Sigma_0]$.³ In the literature, the standard estimation strategy assumes that the unobserved demand shock ξ_{jt} is mean independent of exogenous instruments denoted by Z_{jt} for each j and t (Berry et al., 1995). Formally, it assumes

$$E[\xi_{jt}|Z_{jt}] = 0. \quad (5)$$

The mean independence condition implies that the moment function $g(\theta) = E[Z_{jt}\xi_{jt}(\theta)]$ has zero value at $\theta = \theta_0$, where $\xi_{jt}(\theta) = \delta_{jt}(\Sigma) - x'_{jt}\bar{\theta}$. Then the GMM objective function can be formed as

$$L_n(\theta) = -\frac{n}{2}g_n(\theta)' \Phi_n^{-1}(\theta)g_n(\theta), \quad (6)$$

where n denotes the sample size, g_n is the sample analog of g , and Φ_n is a consistent estimate of $\Phi(\theta_0) = E[g(\theta_0)g(\theta_0)']$. Instead of finding θ that maximizes $L_n(\theta)$ as in GMM, the LT estimator computes the mean of a *quasi-posterior* $p_n(\theta)$, i.e., $\hat{\theta} = E[\theta|p_n(\theta)]$,⁴ where the quasi-posterior $p_n(\theta)$ is defined as

$$p_n(\theta) = \frac{e^{L_n(\theta)}\pi(\theta)}{\int e^{L_n(\theta')}\pi(\theta')d\theta'},$$

where $\pi(\theta)$ is a prior distribution (Chernozhukov and Hong, 2003). We adopt a diffuse prior ($\pi(\theta) \propto 1$) throughout the paper, which is canceled out of the quasi-posterior. The quasi-posterior can be simulated by the MCMC method in the same way as the Bayesian posterior.

Since formulating the quasi-posterior requires only the conditional moment restrictions as in the GMM estimation, the LT approach circumvents the potential risk of imposing inappropriate structure on the supply-side model. In contrast, the simulated maximum likelihood, Bayesian, and control function methods rely on a fully specified likelihood, whether structural or reduced-form, for the supply side model (Romeo, 2007; Jiang et al., 2009; Train, 2009; Petrin and Train, 2010). Even with the most flexible model form, however, the alternative approaches cannot avoid the misspecification bias that is likely to arise in the vast majority of game theoretic

³The parameters of $\{\mathcal{F}_i(\omega_{it+1}|\omega_{it})\}_i$ and β are omitted from θ for a simplified illustration since the former set of parameters is determined by θ and the discount factor is usually fixed *a priori* by the econometrician.

⁴ $\hat{\theta}$ minimizes the quasi-posterior risk under the squared loss function.

models admitting multiple equilibria.

Another important property of the LT estimator is that it is asymptotically equivalent to the GMM estimator under general regularity conditions (Chernozhukov and Hong, 2003):

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left[\nabla_{\theta} g(\theta_0)' \Phi^{-1}(\theta_0) \nabla_{\theta} g(\theta_0)\right]^{-1}\right).$$

This asymptotic result follows immediately from Lemmas 1 and 2 of Chernozhukov and Hong (2003) when the GMM objective function is twice continuously differentiable. The smoothness condition is implicitly assumed in the gradient-based optimization routines such as the KNITRO package (Byrd, Nocedal and Waltz, 2006; Dubé et al., 2012) and the quasi-Newton method. However, the smoothness is not a necessary condition for the asymptotics of LTE. Proposition 1 of Chernozhukov and Hong (2003) shows that it is sufficient to satisfy *stochastic equicontinuity*, a weaker condition than the almost sure continuity.⁵ Later, we formally show the first-order smoothness of the GMM model in the dynamic programming framework of Rust (1988) (see Section 4 and Appendix C). Having established the smoothness, it is straightforward to verify the regularity conditions of LTE (Appendix D).

Lastly, it should be emphasized that the LTEs belong to the classical estimation framework, and thus it is possible to form the confidence interval by differentiating the GMM moment. However, Chernozhukov and Hong (2003) highlight that the asymptotic confidence interval can also be obtained directly from the MCMC output as in the standard Bayesian MCMC inference. The MCMC output provides asymptotically valid confidence intervals under the optimal GMM weight matrix, which is well known to obtain the generalized information equality (see Theorems 3-4 of Chernozhukov and Hong (2003)).

3 Fixed point algorithms

So far, we have seen that the LT estimation can use a standard MCMC algorithm to simulate the quasi-posterior. While it produces an estimate asymptotically equivalent to GMM without likelihood assumptions, this approach is unlikely to alleviate the computational cost if it relies on the same fixed point solution procedure. Hence, we develop the PFP algorithm to overcome the bottleneck by modifying the recursive fixed point operations. Before describing the proposed algorithm, we briefly review the NFP algorithm first.

3.1 Nested fixed point

The NFP approach consists of three loops. The same middle and inner loops can be employed for both the GMM and the LT estimators. The two estimators differ only in the outer loop

⁵Analogous assumptions for the general asymptotics of GMM can be found in Theorem 7.2 of Newey and McFadden (1994).

where GMM searches for the parameter that maximizes $L_n(\theta)$ while LTE computes the mean of the quasi-posterior $p_n(\theta)$. For illustration, we describe both the NFP and the PFP algorithms under a LT framework. To simplify the notation, we suppress the time index t whenever possible.

In the middle loop, the NFP approach solves for $\delta(\theta) = [\delta_{11}(\theta), \dots, \delta_{JT}(\theta)]$ by recursively iterating a BLP fixed point operator $T_V: \Delta \times \Theta \rightarrow \Delta$, where Δ is a subset of R^{JT} with L^∞ -norm. For the model in Section 2.1, we consider a fixed point mapping T_V defined as

$$T_V(\delta, \theta) = \delta(\theta) + \phi(\log s - \log S(\delta, V, \theta)), \quad \theta \in \Theta, \quad (7)$$

for some fixed $\phi \in (0, 1]$, where s is a vector of observed market shares $[s_{11}, \dots, s_{JT}]$, V is a set of value functions $[V_1, \dots, V_N]$ defined in Equation 3, and $S(\delta, V, \theta)$ is a vector of the market share functions in Equation 4.⁶ The parameter ϕ regulates the step size in the output sequence of $T_V(\cdot, \theta)$ to satisfy the contraction property.

In each cycle of the BLP fixed point iteration, solving for $[V_1, \dots, V_N]$ constitutes an inner-loop procedure. Let $\mathcal{C}(\Omega)$ denote a Banach space of continuous and bounded functions $V_i: \Omega \rightarrow R$ under L^∞ -norm. For a given $(\delta, \theta) \in \Delta \times \Theta$, the NFP algorithm solves for V_i in the inner loop by recursively iterating the Bellman operator $\Gamma: \mathcal{C}(\Omega) \times \Delta \times \Theta \rightarrow \mathcal{C}(\Omega)$ defined as

$$\Gamma(V_i, \delta, \theta)(\omega) = \log \left[\exp \left(\beta E_{\omega'} [V_i(\omega') | \omega] \right) + \exp(\omega) \right], \quad \omega \in \Omega.$$

It is worth emphasizing that the Bellman operator Γ depends on δ as well as θ . This is because δ , together with θ , determines not only the state variable ω but also the state transition distribution $\mathcal{F}(\omega' | \omega)$.

BLP procedure (Middle loop) Suppose the MCMC algorithm is at the r -th iteration. Given a candidate parameter θ^{*r} drawn from a proposal distribution, the NFP algorithm iterates the BLP fixed point operator to obtain the fixed point at θ^{*r} satisfying

$$\delta(\theta^{*r}) = T_V(\delta, \theta^{*r}).$$

Applying the operator $T_V(\cdot, \theta^{*r})$ recursively until convergence yields the mean utility δ that satisfies $S_{jt}(\delta, V, \theta^{*r}) = s_{jt}$ for $\forall j, t$. Since the BLP operator involves the value functions $\{V_i\}_i$, they need to be solved at each iteration of the middle loop using the following recursive procedure.

⁶In other applications, T_V may need to be constructed differently depending on the estimation problem. In the appendix, we show that the fixed point operator defined above generates convergent sequences under some regularity conditions, which will be discussed in the next section.

Bellman procedure (Inner loop) For a given input of (δ, θ^{*r}) in the BLP iteration, the NFP algorithm computes the value functions $\{V_i\}_i$ by iterating the Bellman operator Γ until it satisfies

$$V_i(\omega; \delta, \theta^{*r}) = \Gamma(V_i, \delta, \theta^{*r})(\omega), \quad \forall \omega \in \Omega \quad \text{for } i = 1, \dots, N. \quad (8)$$

This inner loop is invoked at every cycle of the middle loop procedure since the consumer expectation has to be consistent with the movement of δ under the rational expectations assumption. Since convergence must be achieved in the inner loop for all simulated taste draws $i = 1, \dots, N$, the two procedures collectively constitute the most computationally intensive part of the NFP algorithm.

Outer loop We can use the standard Metropolis-Hastings (M-H) algorithm to draw MCMC samples from the quasi-posterior. For example, we can draw θ^{*r} from a random-walk proposal distribution $q(\theta^{*r}|\theta^{r-1})$ and accept it by letting $\theta^r = \theta^{*r}$ with acceptance probability

$$\lambda(\theta^{r-1}, \theta^{*r}) = \min \left\{ \frac{\pi(\theta^{*r}) e^{L_n(\theta^{*r})} q(\theta^{r-1}|\theta^{*r})}{\pi(\theta^{r-1}) e^{L_n(\theta^{r-1})} q(\theta^{*r}|\theta^{r-1})}, 1 \right\}. \quad (9)$$

On the other hand, the conventional GMM estimation involves an optimization routine in the outer loop instead of the M-H procedure. Usually each element of the parameters is perturbed to find a direction of the algorithm path at each iteration. Compared to the optimization approach, the MCMC simulation usually has a higher computational burden because it involves more frequent evaluation of the computationally intensive nested loops.⁷

3.2 Pseudo fixed point

The PFP algorithm replaces the middle and the inner loops each with a partial solution procedure. In the BLP procedure, we partially solve for the fixed point by iterating the BLP operator only once during each MCMC cycle. Even though we obtain only partial solutions to the BLP fixed point problem, they converge to the exact solution as the MCMC algorithm proceeds. This approach is analogous to the IJC algorithm, which incrementally solves for the value function via a one-step iteration of the Bellman operator.

In the Bellman step, we modify the IJC algorithm by obtaining a partial solution within a loose convergence threshold. We increasingly tighten this threshold such that the pseudo value function will converge to the exact fixed point. This can be understood as a hybrid approach between Rust's full solution and IJC's partial solution methods. We impose this restriction to ensure the convergence of the BLP operator.

⁷Jiang et al. (2009) report that their Bayesian MCMC method takes about three times longer time than the standard GMM approach in a static demand setting.

Let δ^r and V^r denote the pseudo fixed points of the BLP and the Bellman steps corresponding to the r -th MCMC iteration, respectively. We retain the history H^r that contains $N(r)$ most recent MCMC samples and their corresponding pseudo fixed points: $H^r = \{(\theta^{*s}, \delta^s, V^s) : s = r - N(r), \dots, r - 1\}$. Given θ^{*r} and H^r , we obtain δ^r and V^r from the following steps.

Modified BLP step First, we define the pseudo BLP operator $T_{V^r} : \Delta \times \Theta \rightarrow \Delta$ as

$$T_{V^r}(\delta, \theta) \equiv \delta(\theta) + \phi(\log s - \log S(\delta, V^r, \theta)),$$

where V^r is the pseudo value function obtained in the modified Bellman step. Given the operator T_{V^r} , we compute the pseudo BLP fixed point δ^r via a one-step iteration as

$$\delta^r(\theta^{*r}) = T_{V^r}(\hat{\delta}^r, \theta^{*r}), \quad (10)$$

where the input $\hat{\delta}^r$ is a nonparametric approximation of δ at θ^{*r} . We can obtain $\hat{\delta}^r$ by taking a weighted average over the past pseudo fixed points stored in the history H^r . Specifically, its value at θ is given by

$$\hat{\delta}^r(\theta) = \sum_{s=r-N(r)}^{r-1} \delta^s(\theta^{*s}) W_s(\theta), \quad (11)$$

where $W_s(\theta)$ is a weight function that assigns more weight to δ^s s closer to θ . The choice of the weights depends on different approximation methods.⁸ By growing the size of past history $N(r)$ as r increases, we obtain a better approximation of δ , i.e., the convergence of $\hat{\delta}^r$ to δ . After the modified BLP step, the output δ^r is stored in H^{r+1} for the subsequent MCMC iteration.

Modified Bellman step The modified BLP step takes as input the pseudo value function V^r . Given $\hat{\delta}^r$ and θ^{*r} , the pseudo value function V^r is obtained by iterating the corresponding Bellman operator multiple times until it is within some error bound. Specifically,

$$V_i^r(\omega) = \Gamma^K(\hat{V}_i^r, \hat{\delta}^r, \theta^{*r})(\omega), \quad \omega \in \Omega \quad \text{for } i = 1, \dots, N,$$

for some $K \in \mathbb{N}$ such that $\|V - V^r\| < \epsilon(r)$, where the input value function \hat{V}_i^r is an approximation of the value function at θ^{*r} , which is defined as

$$\hat{V}_i^r(\omega) = \sum_{s=r-N(r)}^{r-1} V_i^s(\omega) W_s(\theta). \quad (12)$$

⁸Imai et al. (2009) used a Gaussian kernel instead of the nearest neighbor method.

The convergence threshold $\epsilon(r)$ is a sequence decreasing toward zero as $r \rightarrow \infty$. This ensures the convergence of the pseudo value function V^r to V and thus the convergence of the modified BLP operator T_{V^r} to T_V as well. On the other hand, setting a loose error bound accelerates the modified Bellman step during the burn-in period of the Markov chain simulation. Even though the error bound is eventually tightened, the modified Bellman step still needs less iteration than the NFP procedure because the input of the Bellman operator, \hat{V}^r , converges to V .

This completes the PFP algorithm for the LT estimator. In summary, the PFP algorithm achieves computational gain over the NFP procedure by obtaining partial solutions to the fixed point problems. The sequence of pseudo fixed points generated by the PFP algorithm converges to the exact solution as a growing number of historical pseudo fixed points increase the precision of the subsequent partial solutions over the course of Markov chain simulation. The complete details on the procedure are provided in Appendix A.

4 Theoretical results

This section establishes the almost sure convergence of the pseudo BLP fixed point under regularity conditions commonly assumed in the literature. The convergence is stochastic since the pseudo fixed point is approximated at parameter values randomly drawn from the MCMC sampler. In the remainder of the paper, we state the results without explicitly conditioning on the observed sample of data whenever possible.

For modeling the dynamic decision of individual consumers, we postulate the dynamic programming framework of Rust (1988) and the associated regularity conditions therein. Together with the standard assumptions made by the NFP algorithm, Rust's framework allows us to establish the continuity of the BLP fixed point curve $\delta(\theta)$, which serves to establish the main convergence result.

We begin by making the following assumptions.

Assumption 1.

- (i) *The parameter space Θ is a compact subset of the Euclidean space, which contains the true parameter θ_0 in the interior.*
- (ii) *The state space Ω is a Borel subset of a complete separable metric space.*
- (iii) *The vector of value functions $V(\omega; \delta, \theta) = [V_1(\omega; \delta, \theta), \dots, V_N(\omega; \delta, \theta)]$ exists and is a unique solution to Equation 8 for all $(\delta, \theta) \in \Delta \times \Theta$.*
- (iv) *$g(\theta) = 0$ if and only if $\theta = \theta_0$.*

Assumptions 1(i) and 1(ii) are common in the literature. Assumption 1(iii) directly follows from Theorem 3.1. of Rust (1988). We will assume that the regularity conditions implied in

Assumption 1(*iii*) hold throughout the paper. Assumption 1(*iv*) is a standard identification condition.

Assumption 2. *For each $\theta \in \Theta$ and observed market shares s , there exists a unique $\delta_0(\theta) \in \Delta$ that satisfies $S(\delta_0, V, \theta) = s$, where $S(\delta, V, \theta)$ is the vector of market share functions in Equation 4.*

Together with Assumption 1(*iv*), Assumption 2 provides identification for the estimation approaches that invert the demand system. While we cannot always rule out the possibility of multiple fixed points, Berry, Gandhi and Haile (2013) show that the condition holds in a wide range of demand systems.

Specifically, Berry et al. (2013) validate Assumption 2 in demand systems characterized by a structure called “connected substitutes.” The connected substitutes conditions impose weak restrictions that inside goods are weak substitutes and each inside good is a strict substitute to an outside alternative. In addition to inverting the demand system, the connected substitutes conditions are also useful for constructing a convergent fixed point mapping.

Assumption 3.

- (i) $V(\omega; \delta, \theta)$ is continuously differentiable in $(\delta, \theta) \in \Delta \times \Theta$ for all $\omega \in \Omega$.
- (ii) The fixed point $\delta_0(\theta)$ is continuous with respect to the Lebesgue measure in Θ .

Assumption 3 provides a basis for numerically efficient optimization algorithms. Both the KNITRO package of MPEC (Dubé et al., 2012; Byrd et al., 2006) and the quasi-Newton method for NFP (Press, Teukolsky, Vetterling and Flannery, 1992) rely on twice-differentiable GMM moments, a stronger condition than Assumption 3(*ii*). Similarly, Rust (1988) derives Assumption 3(*i*) to develop a more efficient NFP algorithm (Rust, 1988, Theorem 4.1).

Assumption 3 is also fundamental to important theoretical results in the literature. The second-order differentiability of the moment has been used by Dubé et al. (2012) and Kim and Park (2010) to characterize the bound of the numerical error in the NFP algorithm. Assumption 3(*i*) is used by Rust (1988) to establish the large sample properties of the value function.

In the static BLP specification, Nevo (2001) takes advantage of the derivatives of $\delta_0(\theta)$ for the quasi-Newton algorithm. Observing that $\delta_0(\theta)$ is implicitly defined by a system of differentiable equations, Nevo (2000) invokes the implicit function theorem to obtain the analytic derivatives. Likewise, we use a similar argument to verify Assumption 3 in Lemma 1 (Appendix C).

Assumption 4. *For each $\theta \in \Theta$ and $\delta_0(\theta) \in \Delta$, there exists $\phi > 0$ such that a recursive operation of $T_V(\cdot, \theta)$ starting from a subset of Δ generates a sequence that converges to $\delta_0(\theta)$: i.e., for each $\theta \in \Theta$, there exists $D(\theta) \subset \Delta$ and $\phi > 0$ such that $T_V^n(\delta, \theta) \rightarrow \delta_0(\theta) \in D(\theta)$ as $n \rightarrow \infty$ for all $\delta \in D(\theta)$.*

For each $\theta \in \Theta$ and $\phi > 0$, $\delta_0(\theta)$ is the fixed point of the mapping T_V in Equation 7 by construction. Assumption 4 requires the existence of a domain of attraction D where a successive iteration of the mapping generates a sequence converging to the fixed point. In dynamic models, it is usually difficult to verify this assumption a priori in the global space of $\Delta \times \Theta$. Instead, we restrict our attention to the class of demand systems characterized by the invertibility conditions of Berry et al. (2013), under which Assumption 4 is satisfied (Appendix G).

If the fixed point algorithm starts from outside D , it may yield a sequence of outputs that does not converge to the true fixed point. This will likely cause the Markov chain to exhibit a nonstationary behavior even after long iterations. However, Assumption 4 does not necessarily mean that the convergence failure will always occur outside D ; the algorithm may still reach convergence even from the outside of the domain of attraction after a finite number of iterations.⁹

Assumption 5. *Let $B = \Omega \times \Delta \times \Theta$. Then for any Borel set $C \subset \Theta$, there exists $\underline{\epsilon} > 0$ such that*

$$P(\theta^{*r} \in C \mid \{(\theta^s, \theta^{*s}, \delta^s, V^s) : s = 1, \dots, r-1\}) \geq \underline{\epsilon} \nu(C)$$

*for a given sample path $\{(\theta^s, \theta^{*s}, \delta^s, V^s) \in \Theta \times \Theta \times \Delta \times \mathcal{C}(B) : s = 1, \dots, r-1\}$ at the r -th MCMC iteration, where ν is the Lebesgue measure.*

Assumption 5 plays a central role in the convergence of our pseudo fixed point algorithm, as in the IJC algorithm (Imai et al., 2009; Norets, 2009). It implies that the Markov chain has a strictly positive probability of visiting the neighborhood of any point in the parameter space. This condition holds for most standard MCMC methods by the compactness condition in Assumption 1. Hence we can achieve a desired level of precision for the pseudo fixed point δ^r at θ^{*r} by simulating the Markov chain long enough to obtain many samples in the neighborhood of θ^{*r} .

Now we specify the weight function $W_s(\theta)$ that appears in Equations 11 and 12. We use the nearest neighbor method of Norets (2009) to formulate the weight function, which greatly simplifies the proof. Among the past $N(r)$ MCMC samples, we choose $\tilde{N}(r)$ nearest neighbors of θ^{*r} from the history H^r defined in the previous section. Let k_i denote the index for the i -th nearest neighbor, i.e.,

$$k_1 = \arg \min_{j \in \{r-N(r), \dots, r-1\}} \|\theta - \theta^{*j}\|, \quad \text{and}$$

$$k_i = \arg \min_{j \in \{r-N(r), \dots, r-1\} \setminus \{k_1, \dots, k_{i-1}\}} \|\theta - \theta^{*j}\|, \quad i = 2, \dots, \tilde{N}(r).$$

The nearest neighbor method assigns nonzero weights only to the $\tilde{N}(r)$ nearest neighbors

⁹This is analogous to the equilibrium mapping of the nested pseudo likelihood algorithm for estimating dynamic games (Kasahara and Shimotsu, 2012).

selected above. Then $\hat{\delta}^r$ in Equation 11 is obtained by

$$\hat{\delta}^r(\theta) = \sum_{i=1}^{\tilde{N}(r)} \delta^{k_i}(\theta^{*k_i}) W_{k_i,r}(\theta), \quad (13)$$

where $W_{k_i,r}(\theta)$ satisfies that $W_{k_i,r}(\theta) \in [0, 1]$ and $\sum_{i=1}^{\tilde{N}(r)} W_{k_i,r}(\theta) = 1$. In the same way, for given $(\hat{\delta}^r, \theta)$, we obtain \hat{V}^r in Equation 12 as

$$\hat{V}^r(\omega; \hat{\delta}^r, \theta) = \sum_{i=1}^{\tilde{N}(r)} V^{k_i}(\omega; \hat{\delta}^{k_i}, \theta^{*k_i}) W_{k_i,r}(\theta).$$

Assumption 6. $\epsilon(r) \rightarrow 0$ as $r \rightarrow \infty$.

The error threshold $\epsilon(r)$ for the pseudo value function needs to be tightened as the MCMC iteration proceeds in order to achieve the convergence of the pseudo fixed point δ^r . We later show that the approximation error in δ^r decreases at the same rate as $\epsilon(r)$ (Proposition 1 in Appendix F).¹⁰

Finally, we make the following assumption for the size of history $N(r)$ and the number of neighbors $\tilde{N}(r)$.

Assumption 7. Let $0 \leq \tau_2 < \tau_1 < \tau_0 < 1$, $N(r) = \lfloor r^{\tau_1} \rfloor$, and $\tilde{N}(r) = \lfloor r^{\tau_2} \rfloor$, where $\lfloor x \rfloor$ is a floor function of x .

For the theoretical convergence results, we need to let $N(r)$ grow with r . However, in the actual implementation, we need to choose an upper bound for $N(r)$. The choice of τ_1 and the upper bound for $N(r)$ will depend on specific applications, but they should be chosen to ensure a good approximation for the true fixed point. Ching, Imai, Ishihara and Jain (2012) propose an iterative procedure where researchers check the distance between the pseudo-fixed point and the true fixed points during the estimation and adjust the upper bound for $N(r)$ adaptively.¹¹

Theorem 1. Under Assumptions 1–7, the pseudo fixed point δ^r in Equation 10 converges almost

¹⁰In the static BLP literature, there is a concern about the propagation of the numerical error from the inner loop into the optimization procedure of the outer loop in the NFP algorithm (Dubé et al., 2012). Dubé et al. (2012) show that the approximation error may increase with the modulus of the BLP contraction mapping. Hence, tightening the threshold in proportion to the inverse of β is recommended when the discount factor is high. On the other hand, Lee and Seo (2016) derive a sharper bound for the numerical error than Dubé et al. (2012) to argue for the sufficiency of moderately tight thresholds. Nonetheless, Dubé, Fox and Su's (2012) argument for tightening the inner-loop threshold remains valid when the Lipschitz constant is expected to be high. Hence, we conjecture that a reasonably conservative error threshold would be sufficient in many applications.

¹¹To do this, researchers can store a larger history and use only a subset of it for approximation so that it is easy to adjust the upper bound adaptively. In terms of the trade-off between convergence speed and approximation accuracy, they also suggest that researchers set a smaller $N(r)$ at the beginning to take advantage of convergence speed and subsequently increase $N(r)$ in the later iterations to ensure the approximation accuracy.

surely to the exact fixed point δ_0 uniformly in $\theta \in \Theta$. In other words,

$$\sup_{\theta \in \Theta} \|\delta_0(\theta) - \delta^r(\theta)\| \rightarrow 0 \quad \text{almost surely as } r \rightarrow \infty.$$

Proof. The proof is provided in Appendix E. □

Theorem 1 completes the proof for the convergence of the PFP algorithm. Based on Theorem 1, the convergence of the Markov chain can be shown using a similar argument as in the proof for Theorem 2 of Imai et al. (2009).

5 Monte Carlo analysis

So far, we have established the theoretical convergence of the PFP algorithm. However, it is difficult to characterize the convergence rate of the Markov chain in application. This difficulty motivates the Monte Carlo experiments to evaluate the sampling and computational performance of the PFP algorithm.

5.1 Convergence of the pseudo fixed points

In the first study, we aim to understand how closely the PFP method can approximate NFP in the LT estimation by examining the MCMC samples generated by the two algorithms. Using the same estimation framework provides a benchmark MCMC sample by which the PFP algorithm can be evaluated.

We simulate consumer demand for 100 products for 100 time periods in a single market using the specification in Equation 1, where products are differentiated only by price p_{jt} , i.e., $x_{jt} = [1, -p_{jt}]$. We assume that the price coefficient is heterogeneous, i.e., $\theta_i = [\beta_0, \alpha_i]$, where $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$. The model specification follows Dubé et al. (2012), except the distribution of the random coefficient. For the state transition law $\mathcal{F}_i(\omega_{it+1}|\omega_{it})$, we assume a first order autoregressive process: $\omega_{it} = \lambda_{0i} + \lambda_i \omega_{it-1} + \psi_{it}$.¹²

The prices are generated from the following equation:

$$p_{jt} = \gamma_0 + \gamma_z z_{jt} + \gamma_w w_{jt} + \gamma_\xi \xi_{jt} + u_{jt}, \quad (14)$$

where z_{jt} is observed price/cost shifter evolving over time as $z_{jt} = \rho_0 + \rho_z z_{jt-1} + \eta_{jt}$, where η_{jt} is a measurement error. w_{jt} is an observable cost shifter independently distributed over time, and u_{jt} is an unobserved random shock. We use z_{jt} and w_{jt} as the instruments for price and formulate the GMM moment as given in Equation 5.

For the value function, we obtain exact solutions to the Bellman equation (Equation 8) at 50 evenly spaced grid points in Ω for the data generating process and the NFP estimation.

¹²The R^2 value of the autoregression at the true parameter was 0.9946 on average.

Throughout the data simulation and the estimation, the market share function (Equation 4) is evaluated with the Monte Carlo integration using the same 50 Halton draws, and the discount factor is held fixed at 0.99 and assumed to be known to the researcher. We provide full details on the data generating process in Appendix B.

We draw MCMC samples using a random-walk M-H algorithm started with instrumental variables (IV) logit regression estimates for the parameters that enter the linear part of the utility function. To ensure the convergence of PFP, we simulate a longer Markov chain (50,000 draws) than NFP (11,000 draws). Then we reserve the last 10,000 draws for inference in both algorithms. We compute the one-step estimator by using the GMM weighting matrix $W = (Z'Z)^{-1}$. We limit the parameter search to σ_α by recovering the linear parameters through an IV regression with the known weight following Nevo (2000). For the prior distribution, a diffuse prior is used, i.e., $\pi(\theta) \propto 1$, such that it is canceled out of the quasi-posterior. We omit the prior in the rest of the paper as the same prior is used throughout the analysis. For implementing the PFP algorithm, we use the nearest neighbor method of Norets (2009) by setting $\tau_2 = 0$ and allowing the size of history $N(r)$ at the rate of $\tau_1 = 0.75$. To minimize the simulation error from MCMC sampling, we fix the seed for the pseudo random number generator and start the simulation from $\sigma_\alpha = 0$ in both MCMC algorithms. For the convergence threshold for the NFP algorithm, we use 10^{-6} for the middle loop and 10^{-8} for the inner loop. The error bound $\epsilon(r)$ in Assumption 6 is set as $\epsilon(r) = \min\{5/(r - 5000), 10^{-3}\}$.

Parameter	True	LTE-NFP			LTE-PFP		
		Mean	RMSE	MAD	Mean	RMSE	MAD
σ_α	0.50	0.4893	0.0606	0.0502	0.4887	0.0601	0.0497
β_0	5.00	5.0813	0.3710	0.2883	5.0811	0.3696	0.2874
α	2.00	2.0443	0.1190	0.0866	2.0432	0.1176	0.0857
Total MCMC draws		11,000			50,000		
Burn-in samples		1,000			40,000		
CPU time (sec)		20,963			2,350		
Bellman iteration		1,924 million			35 million		

Laplace type estimation results from 100 Monte Carlo experiments.

Data generated with $T=100$ periods, $J=100$ products, $N=50$ simulation draws.

Discount factor: $\beta=0.99$. Average BLP modulus: 0.908.

Table 1: Laplace type estimation of dynamic BLP demand

Table 1 summarizes the LT estimation results from 100 Monte Carlo experiments for the NFP and the PFP algorithms. For all parameters, both algorithms yield very similar estimates and biases. From this result, 50,000 draws appear to be sufficient for PFP to closely approximate the MCMC generated by NFP. Despite the extended Markov chain simulation, we find the PFP algorithm to spend about 88% less time or to run 8.68 times faster than NFP on average.¹³

¹³The experiment was conducted by a single-threaded C++ code on a 64-bit Linux system (Ubuntu 15.10) equipped with Intel CPU Xeon E5-2643. The code was compiled by the GNU compiler collection (GCC) software 5.2.1 with -O3

Since PFP iterates the BLP operator only once for each MCMC cycle, it executes considerably fewer inner loop iterations than NFP. Yet, the total computation time is not exactly proportional to the inner loop iterations since the NFP program code has been optimized to minimize the overhead involved with the repeated evaluation of the BLP operator.

5.2 Computational performance of NFP, PFP, and MPEC

In the next Monte Carlo analysis, we examine the computational performance of PFP in comparison to the alternative GMM estimation algorithms: the standard NFP and MPEC.

We simulate 20 datasets using the same model in Section 2.1, where x_{jt} in Equation 1 includes three exogenous product characteristics χ , price p , and an intercept: $x_{jt} = [1, \chi_{jt}, -p_{jt}]$, where $\chi_{jt} = [x_{1jt}, x_{2jt}, x_{3jt}]$. The corresponding coefficients are denoted by $\theta_i = [\theta^{x0}, \theta_i^{x1}, \theta_i^{x2}, \theta_i^{x3}, \theta_i^p]$, where $[\theta_i^{x1}, \theta_i^{x2}, \theta_i^p]$ are normally distributed, and $[\theta^{x0}, \theta^{x3}]$ are constants. The prices are generated from

$$p_{jt} = \gamma_0 + \gamma'_x \chi_{jt} + \gamma_z z_{jt} + \gamma_w w_{jt} + \gamma_\xi \xi_{jt} - \gamma'_p \sum_{k \neq j} \chi_{kt} + u_{jt},$$

where z_{jt} and w_{jt} are observed cost shifters: specifically, $z_{jt} = \rho_0 + \rho_z z_{jt-1} + \eta_{jt}$ for a Gaussian measurement error η_{jt} , and $w_{jt} \sim \text{i.i.d. } N(0, \sigma_w^2)$. u_{jt} is an unobserved random shock distributed as $u_{jt} \sim \text{i.i.d. } N(0, \sigma_u^2)$. We use polynomial expansion of the baseline instruments $[\chi_j, z_j, w_j, \sum_{k \neq j} \chi_k]$ to form 66 instrumental variables. The GMM moment restrictions are given in Equation 5.

As in the previous Monte Carlo exercise, the discount factor is assumed to be known and fixed at 0.99, and 50 pseudo random numbers are used for the market share integration throughout the experiments. The exact solution to the Bellman equation is obtained at 50 evenly spaced grid points in the data simulation and the estimation with NFP. The remaining details on the data generating process are available in Appendix B.

For MPEC, we follow Dubé et al. (2012) to formulate the estimation problem as

$$\begin{aligned} \min_{\{\theta, \xi, V, g\}} g' W g \quad & \text{subject to} \quad S(\xi, V, \theta) = s, \\ V(\omega) &= \log(e^{\beta E[V(\omega')|\omega]} + e^\omega), \\ g &= Z' \xi, \end{aligned} \tag{15}$$

where the weighting matrix is given by $W = (Z'Z)^{-1}$. Under this formulation, the constraint

-march=native optimization flags.

Jacobian takes the form of

$$J_{\theta, \xi, V, g} = \begin{bmatrix} \frac{\partial S}{\partial \theta} & \frac{\partial S}{\partial \xi} & \frac{\partial S}{\partial V} & 0 \\ \frac{\partial \phi_V}{\partial \theta} & \frac{\partial \phi_V}{\partial \xi} & \frac{\partial \phi_V}{\partial V} & 0 \\ 0 & -Z' & 0 & I_g \end{bmatrix}, \quad (16)$$

where ϕ_V is a mapping defined as $\phi_V(\omega) = V(\omega) - \log(e^{\beta E[V(\omega')|\omega]} + e^\omega)$ for $\omega \in \Omega$.

To compute the GMM-MPEC estimator, we use the C++ callable library of the KNITRO solver developed by Artelys. We use the interior-point conjugate gradient algorithm and follow the literature (Dubé et al., 2012; Lee and Seo, 2015, 2016) in the tuning parameters setting.¹⁴ To achieve maximum performance, we use the C/C++ automatic differentiation package ADOL-C to supply the exact Jacobian and Hessian for KNITRO (Griewank, Juedes and Utke, 1996; Walther and Griewank, 2012). ADOL-C was chosen to overcome the limitation of TOMLAB’s MAD package previously used by Dubé et al. (2012), which occupied the memory exceeding 32GBs causing the system to fail. To further accelerate MPEC, we use the ColPack library to help ADOL-C and KNITRO exploit the sparsity in the constraint Jacobian and the Lagrangian Hessian (Gebremedhin, Nguyen, Patwary and Pothen, 2013).¹⁵

For GMM-NFP, we use the same interior point algorithm available in KNITRO. We compute the Jacobian and Hessian by the numerical differentiation method provided by KNITRO.¹⁶ For convergence criteria, we use relatively conservative thresholds, 10^{-12} and 10^{-14} in the L^∞ norm for the middle and the inner loops, respectively.

For the LTE-PFP estimator, the MCMC samples are obtained using the M-H algorithm by which the parameters are jointly drawn from a multivariate normal distribution. The PFP history is allowed to grow at the rate of $\tau_1 = 0.65$ and $\tau_2 = 0$. The error bound is set to $1/r$, starting from an upper bound of 10^{-4} . The MCMC convergence was determined by the Heidelberger-Welch diagnostic, which sequentially tests for the stationarity and controls the length of the run for a single MCMC simulation (Heidelberger and Welch, 1983).¹⁷ To alleviate high autocorrelations in the MCMC samples, we retain one out of every 10 iterations for analysis.¹⁸

In each estimation run, all three algorithms started from the same five points randomly drawn in a bounded parameter space $(0, 5) \times (0, 5) \times (0, 2.5)$. For the initial fixed point at the start of the algorithm, MPEC was provided with the exact solution predetermined at the

¹⁴blasoption=0, presolve=0, linsolver=4, bar_switchrule=1, bar_directinterval=100000.

¹⁵By default, ADOL-C starts to write on the disk space when it generates temporary files in excess of the preset size of the buffer memory. We ensured that ADOL-C ran at the full speed by forcing it to use the system memory exclusively rather than the much slower disk storage.

¹⁶Central difference method was used for the Jacobian, and the L-BFGS option for the Hessian computation. We found forward difference to be inferior to central difference since it yielded considerably more convergence failures than the latter.

¹⁷The diagnostic was implemented using the CODA library with $\epsilon = 0.05$ and $p = 0.1$ (Plummer, Best, Cowles and Vines, 2006). Following the recommendation of Heidelberger and Welch (1983), we set $j_1 = 0.1j_{\max}$ and $I = 1.5$.

¹⁸No sample was discarded in the small scale experiment (top panel in Table 2) since the thinning did not improve the sampling efficiency.

starting point while NFP and PFP were given a log ratio of observed market shares. In order to accelerate NFP after the start, each outer loop cycle was supplied with the exact BLP fixed point solved at the previous iteration as a new starting point.¹⁹

Table 2 summarizes the computation time for the three algorithms under varying problem sizes and intercept values. To save the space, the root mean squared errors are reported in Appendix B Table 4. For each given dimension of the estimation problem, the dataset is generated with two levels of the intercept parameter θ^{x0} to produce different BLP modulus following the literature (Dubé et al., 2012; Lee and Seo, 2015, 2016). The modulus of the BLP contraction mapping was estimated by measuring the ratio of distances between successive fixed point operations.²⁰ The CPU time is measured in seconds for a single-threaded C++ program code executed on the same system used in the previous experiment. The convergence rate for NFP and MPEC denotes the total number of successful convergences among the total 100 replicated experiments that satisfied the convergence criteria of KNITRO.²¹

Period T	Product J	R. draws N	Grid G	Intercept θ^{x0}	BLP modulus	Estimation method	Run time (seconds)	Conv. (%)
25	25	50	50	6	0.820	NFP	839	100
						MPEC	5,346	100
						LTE	409	100
				12	0.957	NFP	1,517	96
						MPEC	5,321	95
						LTE	462	100
50	50	50	50	6	0.934	NFP	1,591	100
						MPEC	29,722	99
						LTE	814	100
				12	0.988	NFP	5,612	100
						MPEC	30,444	98
						LTE	1,996	96
100	50	50	50	6	0.978	NFP	7,588	100
						MPEC	82,338	93
						LTE	2,844	98
				10	1.004	NFP	31,283	10
						MPEC	99,212	44
						LTE	3,456	41

20 datasets were generated for each level of T, J, N, G , and θ^{x0} . For each dataset, the estimation algorithm was started from five randomly drawn parameter values. The discount factor β was set to 0.99 in the data generation and estimation.

Table 2: The computation time of NFP, MPEC, and PFP

¹⁹We make available the source codes used in the exercise at <https://github.com/yutec/LTE.git>.

²⁰Dubé et al. (2012) report the Lipschitz constant of Berry et al. (1995) evaluated at the true parameter. However, the modulus regulating the speed of the NFP contraction is a function in $\Delta \times \Theta$ since it depends on the fixed point operator $T_V(\delta, \theta)$. Furthermore, the Lipschitz constant does not necessarily provide a sharp upper bound. Alternatively, we measure the “effective” modulus of the contraction mapping.

²¹The successful convergence is indicated by an output message `exit==0` in KNITRO.

For each given size of the estimation problem, the run time for LTE-PFP increases more slowly than GMM-NFP as the BLP modulus becomes higher. In the top panel of Table 2 ($T = 25$, $J = 25$, $N = 50$, $G = 50$), the NFP run time doubles from 839 seconds to 1,517 seconds while LTE sees a moderate increase from 409 seconds to 462 seconds. Similarly, the relative performance gap is widened with the higher intercept in other experiments as well. This result is intuitive since the MCMC inference typically requires at least a certain number of samples to achieve a sufficient level of precision even though the sampler may have already converged. Consequently, the acceleration in the fixed point procedure can be outweighed by the overhead of MCMC sampling from the stationary distribution when the exact fixed point is not costly to obtain.

On the other hand, MPEC can no longer replicate the computational advantage over NFP as achieved in the prior literature. The primary difference with Dubé et al. (2012) is found in the large number of nonzero elements of $J_{\theta, \xi, V, g}$. In Equation 16, it is straightforward to see that the submatrices of $J_{\theta, \xi, V, g}$ are either fully dense or block diagonal.²² Hence, the total number of derivatives increases sharply with the scale of demand system and grid points of the market share and value functions. To put the difference into perspective, the smallest estimation in our experiment has 3,706,941 nonzero elements in $J_{\theta, \xi, V, g}$ while the number would be 93,660 in the equivalent GMM estimation of the dynamic BLP demand of Dubé et al. (2012).²³ Nonetheless, for a given size of the optimization problem, MPEC demonstrates highly robust performance with respect to the change in the BLP modulus. These results are consistent with the conclusion of Dubé et al. (2012) that MPEC's performance depends primarily on the size and sparsity of the optimization problem rather than on the BLP modulus.

Except for one experiment, all three algorithms demonstrate very high convergence rates throughout the experiments. Despite using the numerical differentiation, NFP does not suffer from convergence failures, a problem formerly experienced by Dubé et al. (2012). On the other hand, we find the exact derivatives for the Hessian, along with the Jacobian, to be indispensable for MPEC; a numerically computed Hessian led to almost 100% convergence failure for MPEC even with the exact Jacobian input throughout the tested specifications.

In the last rows of Table 2 where the BLP modulus exceeds 1 on average, it is not surprising that NFP fails to converge successfully from most starting points. Compared to NFP, MPEC shows significantly superior convergence rate while its speed improves relative to the previous exercises. Yet, its performance still lags behind the two algorithms and also experiences steep decline in the convergence rate due to excessively slow improvement in the optimization: many runs of MPEC reached the maximum time limit before reaching convergence.²⁴ On the

²²The matrices $\partial S/\partial\theta$, $\partial S/\partial\xi$, $\partial S/\partial V$, $\partial\phi_V/\partial\theta$, and $\partial\phi_V/\xi$ have no zero elements while $\partial\phi_V/\partial V$ is block diagonal. This contrasts with the static BLP where $\partial S/\partial\xi$ is sparse due to the lack of intertemporal dependence of consumer demand.

²³The number is only an approximation to what is considered by Dubé et al. (2012) because of the different formulation of estimator and instruments.

²⁴Exclusively in the last Monte Carlo experiment, KNITRO was terminated after 75 major iterations, which took about 2-3 days. MPEC could have achieved significantly higher number of successful convergence without the

other hand, PFP often manages to escape the region where the contraction mapping is violated, resulting in better convergence than NFP.

Overall, the BLP modulus is a strong predictor for the performance of the fixed point algorithms exclusively. While MPEC does not rely on the fixed point operations, its performance suffers from the high dimension and density of dynamic BLP estimation since the demand system creates a large number of non-zero elements in the constraint Jacobian and the Lagrangian Hessian. Even when NFP is no longer functional, MPEC shows no considerable advantage over PFP in computational speed. Hence, we conclude that LTE-PFP can complement the two algorithms in estimating the dynamic BLP demand that requires a costly fixed point solution.

Period T	Product J	R. draws N	Grid G	BLP modulus	Estimation method	Run time (seconds)	Conv. (%)	Time ratio to MPEC
25	25	25	25	0.941	NFP	266	100	0.433
					MPEC	615	100	1.000
					LTE	122	100	0.198
25	25	50	50	0.957	NFP	1,559	96	0.293
					MPEC	5,321	95	1.000
					LTE	462	100	0.087
25	25	75	75	0.959	NFP	4,016	100	0.164
					MPEC	24,530	100	1.000
					LTE	1,001	100	0.041

20 datasets were generated for three combinations of (T, J, N, G) with $\theta^{x^0} = 12$. For each dataset, the estimation algorithm was started from five randomly drawn parameter values. The discount factor β was set to 0.99 in the data generation and estimation.

Table 3: The computation time and the dimension of MPEC problem

To identify the impact of the scale of the MPEC constraints separately from the BLP modulus, we repeat the same Monte Carlo exercise where we vary only the number of the random draws and grid points for the Monte Carlo integration while holding the remaining model primitives fixed. This allows us to test the performance under various problem sizes without affecting the BLP modulus in principle. We focus on the small scale estimation listed at the top panel of Table 2 with $\theta^{x^0} = 12$ and generate two additional experimental configurations, each with 20 replications starting from five randomly drawn parameter values.

Table 3 confirms that the performance of the fixed point algorithms is not as much subject to the dimension of the estimation problem as MPEC. The time ratio, defined as a proportion of run times with respect to MPEC's, is declining for both NFP and PFP with the increasing scale in the two dimensions while the three configurations show negligible differences in the BLP modulus estimates. This result indicates that the scale has a disproportionately large impact on MPEC's performance. Therefore, we conclude that the dimension and density can jointly generate a computational bottleneck for MPEC in the dynamic BLP demand estimation.

interruption, but it would have also resulted in considerably increased computation time overall.

Our Monte Carlo analysis is bounded by the limited scope of the experiments. In particular, there exist various alternatives to the numerical method used in the analysis. For example, sparse grids integration can be employed to reduce the number of consumer draws and grid points (Heiss and Winschel, 2008; Skrainka and Judd, 2011; Reynaert and Verboven, 2014), which may accelerate MPEC in particular. However, the method has a drawback of unrestricted weights that may produce negative market shares (Heiss and Winschel, 2008). Furthermore, even advanced algorithms may still require a nontrivial number of grid points in real-world applications, often in larger quantities than used in our experimental setting.

While we have used an exact solution algorithm for the dynamic programming problem, there are alternatives available to help alleviating the computational burden depending on the application. Sieve approximation has proved to be effective in dealing with the curse of dimensionality in large state spaces (Arcidiacono, Bayer, Bugni and James, 2013). In a smaller state space, one may find Chebyshev approximation useful especially for MPEC to the extent that it reduces the number of constraints (Dubé et al., 2012). However, these algorithms typically use a relaxed metric (L^2 norm) to determine convergence, which may have adverse effects on the convergence of the BLP fixed point if used without caution.²⁵

Lastly, the MCMC performance may often suffer from poor implementations. For example, one may have to experiment with different jump sizes to obtain an efficient M-H algorithm. While we have used fixed jump sizes in the Monte Carlo analysis, we expect that a superior sampling efficiency can be achieved with adaptive methods following Roberts and Rosenthal (2007) or Winschel and Krätzig (2010).

6 Conclusion

As a summary, we develop a computationally efficient fixed point algorithm for estimating dynamic models of aggregate demand using the random coefficients logit framework. In order to reduce the computational burden of NFP, we develop the PFP algorithm to solve for the BLP and Bellman fixed points incrementally with MCMC simulation. Then we embed the PFP algorithm within the Laplace type framework to avoid the usual criticisms faced by the conventional Bayesian estimation. We show that the pseudo fixed points converge to the true fixed points under mild conditions satisfied in a wide class of demand systems. We conduct Monte Carlo experiments to evaluate the empirical performance of PFP, NFP, and MPEC in dynamic demand estimation. We find that our approach yields consistent estimates while reducing the computational cost of NFP substantially, particularly when the BLP modulus is high. On the other hand, while MPEC exhibits robust computational performance under high BLP modulus, it faces a bottleneck imposed by the vastly increased scale and density of the dynamic demand estimation. Our Monte Carlo results suggest that the PFP algorithm

²⁵Alternatively, L^∞ norm can be used to achieve high accuracy. However, doing so will likely increase the computational cost dramatically.

can complement the existing approaches in the most computationally intensive dynamic BLP demand estimations.

References

- Aguirregabiria, Victor and Aviv Nevo**, “Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games,” in “Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress,” Vol. 3, Cambridge University Press, May 2013, pp. 53–122.
- Arcidiacono, Peter, Patrick Bayer, Federico A Bugni, and Jonathan James**, “Approximating High-Dimensional Dynamic Models: Sieve Value Function Iteration,” in “Structural Econometric Models,” Emerald Group Publishing Limited, 2013, pp. 45–95.
- Bajari, Patrick**, “Comment,” *Quantitative Marketing and Economics*, 2003, 1 (3), 277–283.
- , **Han Hong, and Stephen P. Ryan**, “Identification and Estimation of a Discrete Game of Complete Information,” *Econometrica*, 2010, 78 (5), 1529–1568.
- Berry, Steven**, “Comment,” *Quantitative Marketing and Economics*, 2003, 1 (3), 285–291.
- , **Amit Gandhi, and Philip Haile**, “Connected Substitutes and Invertibility of Demand,” *Econometrica*, September 2013, 81 (5), 2087–2111.
- , **James Levinsohn, and Ariel Pakes**, “Automobile Prices in Market Equilibrium,” *Econometrica*, July 1995, 63 (4), 841–90.
- , **Oliver B. Linton, and Ariel Pakes**, “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems,” *Review of Economic Studies*, 2004, 71 (4), 613–654.
- Berry, Steven T.**, “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, Summer 1994, 25 (2), 242–262.
- Byrd, Richard H., Jorge Nocedal, and Richard A. Waltz**, “KNITRO: An Integrated Package for Nonlinear Optimization,” in “in,” Vol. 83, Springer US, 2006, pp. 35–59.
- Carranza, Juan Esteban**, “Product Innovation and Adoption in Market Equilibrium: The Case of Digital Cameras,” *International Journal of Industrial Organization*, 2010, 28 (6), 604–618.
- Chernozhukov, Victor and Han Hong**, “An MCMC Approach to Classical Estimation,” *Journal of Econometrics*, 2003, 115 (2), 293–346.
- Ching, Andrew, Susumu Imai, Masakazu Ishihara, and Neelam Jain**, “A Practitioner’s Guide to Bayesian Estimation of Discrete Choice Dynamic Programming Models,” *Quantitative Marketing and Economics*, 2012, 10 (2), 151–196.
- Duan, Jason A. and Carl F. Mela**, “The Role of Spatial Demand on Outlet Location and Pricing,” *Journal of Marketing Research*, 2009, 46 (2), 260–278.
- Dubé, Jean-Pierre and Pradeep K. Chintagunta**, “Comment,” *Quantitative Marketing and Economics*, 2003, 1 (3), 293–298.
- Dubé, Jean-Pierre H., Günter J. Hitsch, and Pradeep K. Chintagunta**, “Tipping and Concentration in Markets with Indirect Network Effects,” *Marketing Science*, 2010, 29 (2), 216–249.
- , **Jeremy T. Fox, and Che-Lin Su**, “Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation,” *Econometrica*, 2012, 80 (5), 2231–2267.

- Gebremedhin, Assefaw H., Duc Nguyen, Md. Mostofa Ali Patwary, and Alex Pothén**, “Col-Pack: Software for Graph Coloring and Related Problems in Scientific Computing,” *ACM Transactions on Mathematical Software (TOMS)*, 2013, 40 (1), 1–31.
- Gowrisankaran, Gautam and Marc Rysman**, “Dynamics of Consumer Demand for New Durable Goods,” *Journal of Political Economy*, December 2012, 120 (6), 1173–1219.
- Griewank, Andreas, David Juedes, and Jean Utke**, “Algorithm 755: ADOL-C: A Package for the Automatic Differentiation of Algorithms Written in C/C++,” *ACM Transactions on Mathematical Software (TOMS)*, 1996, 22 (2), 131–167.
- Hansen, Lars Peter**, “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 1982, 50 (4), 1029–1054.
- Heidelberger, Philip and Peter D. Welch**, “Simulation Run Length Control in the Presence of an Initial Transient,” *Operations Research*, 1983, 31 (6), 1109–1144.
- Heiss, Florian and Viktor Winschel**, “Likelihood Approximation by Numerical Integration on Sparse Grids,” *Journal of Econometrics*, 2008, 144 (1), 62–80.
- Hendel, Igal and Aviv Nevo**, “Measuring The Implications of Sales and Consumer Inventory Behavior,” *Econometrica*, 2006, 74 (6), 1637–1673.
- Imai, Susumu, Neelam Jain, and Andrew Ching**, “Bayesian Estimation of Dynamic Discrete Choice Models,” *Econometrica*, 2009, 77 (6), 1865–1899.
- Jiang, Renna, Puneet Manchanda, and Peter E. Rossi**, “Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data,” *Journal of Econometrics*, 2009, 149 (2), 136–148.
- Kasahara, Hiroyuki and Katsumi Shimotsu**, “Sequential Estimation of Structural Models with a Fixed Point Constraint,” *Econometrica*, 2012, 80 (5), 2303–2319.
- Kim, Kyooil and Minjung Park**, “Approximation Error in the Nested Fixed Point Algorithm for BLP Model Estimation,” *Working Paper*, 2010.
- Lee, Jinhyuk and Kyoungwon Seo**, “A Computationally Fast Estimator for Random Coefficients Logit Demand Models Using Aggregate Data,” *RAND Journal of Economics*, 2015, 46 (1), 86–102.
- and —, “Revisiting the Nested Fixed-Point Algorithm in BLP Random Coefficients Demand Estimation,” *Economics Letters*, 2016, 149, 67–70.
- Lee, Robin S.**, “Vertical Integration and Exclusivity in Platform and Two-Sided Markets,” *American Economic Review*, 2013, (3), 2960–3000.
- Melnikov, Oleg**, “Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market,” *Economic Inquiry*, April 2013, 51 (2), 1277–1298.
- Nevo, Aviv**, “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand,” *Journal of Economics & Management Strategy*, December 2000, 9 (4), 513–548.
- , “Measuring Market Power in the Ready-To-Eat Cereal Industry,” *Econometrica*, March 2001, 69 (2), 307–342.
- Newey, Whitney K. and Daniel McFadden**, “Large Sample Estimation and Hypothesis Testing,” in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, Elsevier Science B.V., 1994, pp. 2111–2245.

- Norets, Andriy**, “Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables,” *Econometrica*, September 2009, 77 (5), 1665–1682.
- , “Continuity and Differentiability of Expected Value Functions in Dynamic Discrete Choice Models,” *Quantitative Economics*, 2010, 1 (2), 305–322.
- Petrin, Amil and Kenneth Train**, “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research*, 2010, 57 (1), 3–13.
- Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines**, “CODA: Convergence Diagnosis and Output Analysis for MCMC,” *R News*, 2006, 6 (1), 7–11.
- Press, William H., Saul A. Teukolsky, Willaim T. Vetterling, and Brian P. Flannery**, *Numerical Recipes in C: The Art of Scientific Computing*, second ed., Cambridge: Cambridge University Press, 1992.
- Reynaert, Mathias and Frank Verboven**, “Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments,” *Journal of Econometrics*, 2014, 179 (1), 83–98.
- Roberts, Gareth O. and Jeffrey S. Rosenthal**, “Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms,” *Journal of Applied Probability*, 2007, 44 (2), 458–475.
- Romeo, Charles J.**, “A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Markets Using Aggregate Data,” *Computational Economics*, 2007, 29 (1), 33–68.
- Rudin, Walter**, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, 1976.
- Rust, John**, “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 1987, 55 (5), 999–1033.
- , “Maximum Likelihood Estimation of Discrete Control Process,” *SIAM Journal on Control and Optimization*, 1988, 26 (5), 1006–1024.
- , “Structural Estimation of Markov Decision Processes,” in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, Elsevier Science B.V., 1994, pp. 3081–3143.
- Ryan, Stephen P.**, “The Cost of Environmental Regulation in a Concentrated Industry,” *Econometrica*, 2012, 80 (3), 1019–1061.
- Skrainka, Benjamin S. and Kenneth L. Judd**, “High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation,” *Working Paper*, February 2011.
- Train, Kenneth E.**, *Discrete Choice Methods with Simulation*, Cambridge University Press, 2009.
- Walther, Andrea and Andreas Griewank**, “Getting Started with ADOL-C,” in Uwe Naumann and Olaf Schenk, eds., *Combinatorial Scientific Computing*, Chapman & Hall/CRC Computational Science, CRC Press, 2012, chapter 7, pp. 181–202.
- Wan, Yuanyuan**, “An Integration-Based Approach to Moment Inequality Models,” *Working paper*, 2013.
- Winschel, Viktor and Markus Krätzig**, “Solving, Estimating, and Selecting Nonlinear Dynamic Models Without the Curse of Dimensionality,” *Econometrica*, 2010, 78 (2), 803–821.
- Yang, Sha, Yuxin Chen, and Greg M. Allenby**, “Bayesian Analysis of Simultaneous Demand and Supply,” *Quantitative Marketing and Economics*, 2003, 1, 251–275.

A The pseudo fixed point algorithm

At the r -th MCMC iteration, the proposed M-H algorithm draws a candidate parameter θ^{*r} at which the pseudo BLP fixed point operator is iterated once to obtain the new pseudo fixed points (δ^r, V^r) . The pseudo fixed point δ^r is used for evaluating the quasi-posterior at θ^{*r} , and the proposed draw is accepted as an MCMC sample, i.e., $\theta^r = \theta^{*r}$ with probability $\lambda(\theta^{r-1}, \theta^{*r})$ in Equation 9. Finally, the new solutions are collected in the set H^{r+1} for subsequent MCMC iterations.

At the r -th MCMC iteration:

1. Draw a candidate parameter θ^{*r} .
2. Find θ^{k_1} nearest to θ^{*r} among the $N(r)$ past draws in H^r and set $\hat{\delta}^r(\theta^{*r}) = \delta^{k_1}(\theta^{k_1})$.
3. Estimate the state transition distribution $F_\omega(\omega_{it+1}|\omega_{it})$ given θ^{*r} .
4. Iterate the Bellman operator K times until $\|V_i^{k_1} - V_i^r\| < \epsilon(r)$ for all i to obtain V^r :

$$V_i^r(\omega; \hat{\delta}^r, \theta^{*r}) = \Gamma^K(V_i^{k_1})(\omega; \hat{\delta}^r, \theta^{*r}), \quad \omega \in \Omega \quad \text{for } i = 1, \dots, N.$$

5. Iterate the pseudo BLP fixed point operator once:

$$\delta^r(\theta^{*r}) = \hat{\delta}^r(\theta^{*r}) + \log s - \log S(\theta^{*r}, \hat{\delta}^r, V^r).$$

6. Compute the GMM objective function $L_n(\theta)$ in Equation 6 at $\delta^r(\theta^{*r})$ and $\delta^{r-1}(\theta^{r-1})$.
7. Accept θ^{*r} with probability $\lambda(\theta^{r-1}, \theta^{*r})$ in Equation 9: $\theta^r = \theta^{*r}$
8. Store the new pseudo fixed point for the next MCMC iteration:

$$H^{r+1} = \{(\theta^{*s}, \delta^s, \{V_i^s\}_i) : s = r - N(r+1) + 1, \dots, r\}.$$

Optionally, the PFP algorithm can repeat the steps 3-5 for θ^{r-1} to obtain an updated pseudo fixed point $\delta^r(\theta^{r-1})$, which is used for evaluating $L_n(\theta^{r-1})$ in step 6. For more information on the practical implication of this optional procedure, the reader is advised to refer to Ching et al. (2012).

B The Monte Carlo experiments

B.1 Table 1

We simulate the consumer demand from Equation 1, where $\bar{\theta} = [\beta_0, \alpha]$ with $\beta_0 = 5$, $\alpha = 2$, and $\sigma_\alpha = 0.5$. The prices are drawn from Equation 14 with $\gamma_0 = 1$, $\gamma_z = 1$, $\gamma_w = 0.2$, and $\gamma_\xi = 0.8$. z_{jt} is simulated from the AR(1) process with $\rho_0 = 0.1$ and $\rho_z = 0.95$. We set $z_{j0} = 8$ for all j and obtain the i.i.d. random draws of ξ, w, u, η from a normal distribution with the mean zero and the standard deviation given as: $\sigma_\xi = 1.5$, $\sigma_w = 1$, $\sigma_u = 0.01$, $\sigma_\eta = 0.1$. For constructing the price instruments (z_{jt}, w_{jt}) , we add small perturbations that are normally distributed as $N(0, 0.1^2)$ and $N(0, 0.01^2)$, respectively. The contraction mapping parameter ϕ is set to 1 for both algorithms throughout the analysis.

B.2 Tables 2 and 3

Throughout the Monte Carlo experiments conducted in Table 2, we use the same specification for generating datasets except the intercept θ^{x0} . The fixed coefficient θ^{x3} is set to 0.5, and the random coefficients are distributed as

$$\begin{bmatrix} \theta_i^{x1} \\ \theta_i^{x2} \\ \theta_i^p \end{bmatrix} \sim \text{i.i.d. } N \left(\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.5^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 0.25^2 \end{bmatrix} \right).$$

The variables are generated as

$$\chi_{jt} = \begin{bmatrix} x_{1jt} \\ x_{2jt} \\ x_{3jt} \end{bmatrix} \sim \text{i.i.d. } N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5^2 & 0 & 0 \\ 0 & 0.5^2 & 0 \\ 0 & 0 & 0.5^2 \end{bmatrix} \right), \quad \xi_{jt} \sim \text{i.i.d. } N(0, 1),$$

and

$$p_{jt} = \gamma_0 + \gamma'_x \chi_{jt} + \gamma_z z_{jt} + \gamma_w w_{jt} + \gamma_\xi \xi_{jt} - \gamma'_p \sum_{k \neq j} \chi_{kt} + u_{jt},$$

where $[\gamma_0, \gamma_{x1}, \gamma_{x2}, \gamma_{x3}, \gamma_z, \gamma_w, \gamma_\xi] = [1, 0.2, 0.2, 0.1, 1, 0.2, 0.7]$, and $\gamma_p = [0.1, 0.1, 0.1]$. The cost shifters are generated as

$$z_{jt} = \rho_0 + \rho_z z_{jt-1} + \eta_{jt}, \quad w_{jt} \sim \text{i.i.d. } N(0, 1), \quad u_{jt} \sim \text{i.i.d. } N(0, 0.01^2),$$

with $z_{j0} = 8$, $[\rho_0, \rho_z] = [0.1, 0.95]$, and $\eta_{jt} \sim \text{i.i.d. } N(0, 0.1^2)$.

For the MCMC simulation, the jump size is set to be $[0.05, 0.05, 0.025]$ for the largest dataset (bottom panel) in Table 2. For the medium and small scale experiments, it is $[0.1, 0.1, 0.05]$ and

$[0.2, 0.2, 0.1]$, respectively. The same jump size for the small experiment is used for all MCMC simulations in Table 3. After the MCMC sampling, we retain one out of every 10 samples for the convergence test and inference in the large and medium scale experiments to alleviate high autocorrelations while all the samples are used in the small experiment. Like in the previous experiment, ϕ is set to 1 for both fixed point algorithms throughout the analysis.

B.3 MPEC estimation

To apply the MPEC approach, we follow Dubé et al. (2012) to formulate the constraint optimization problem as follows:

$$\begin{aligned} \min_{\{\theta, \xi, V, g\}} g'Wg \quad \text{subject to} \quad & S(\xi, V, \theta) = s, \\ & V(\omega) = \log(e^{\beta E[V(\omega')|\omega]} + e^\omega), \\ & g = Z'\xi. \end{aligned} \tag{17}$$

Then the Lagrangian takes the form as

$$L(\theta, \xi, V, g, \lambda_S, \lambda_V, \lambda_g) = g'Wg + \lambda'_S(S(\xi, V, \theta) - s) + \lambda'_V(V(\omega) - \Gamma(V)(\omega, \xi, \theta)) + \lambda'_g(g - Z'\xi),$$

and the gradient of the objective function is

$$\nabla_{(\theta, \xi, V, g)} g'Wg = \begin{bmatrix} 0 & 0 & 0 & 2Wg \end{bmatrix}'.$$

The constraint Jacobian is

$$\begin{bmatrix} \frac{\partial S}{\partial \theta} & \frac{\partial S}{\partial \xi} & \frac{\partial S}{\partial V} & 0 \\ \frac{\partial \phi_V(\omega)}{\partial \theta} & \frac{\partial \phi_V(\omega)}{\partial \xi} & \frac{\partial \phi_V}{\partial V} & 0 \\ 0 & -Z' & 0 & I_g \end{bmatrix},$$

where $\phi_V(\omega) = V(\omega) - \log(e^{\beta E[V(\omega')|\omega]} + e^\omega)$, and $\frac{\partial \phi_V}{\partial V}$ is a block diagonal matrix of size $(NN^S) \times (NN^S)$.

The Lagrangian Hessian is

$$\nabla^2 L(\theta, \xi, V, g, \lambda_S, \lambda_V, \lambda_g) = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta \partial \theta'} & \frac{\partial^2 L}{\partial \theta \partial \xi'} & \frac{\partial^2 L}{\partial \theta \partial V'} & 0 \\ \frac{\partial^2 L}{\partial \xi \partial \theta'} & \frac{\partial^2 L}{\partial \xi \partial \xi'} & \frac{\partial^2 L}{\partial \xi \partial V'} & 0 \\ \frac{\partial^2 L}{\partial V \partial \theta'} & \frac{\partial^2 L}{\partial V \partial \xi'} & \frac{\partial^2 L}{\partial V \partial V'} & 0 \\ 0 & 0 & 0 & \frac{\partial^2 L}{\partial g \partial g'} \end{bmatrix}.$$

For the relative feasibility and optimality tolerances in MPEC, we use 10^{-6} following Dubé et al. (2012).

B.4 Estimation results of Table 2

RMSE	Intercept θ^{x0}					
	Low			High		
	NFP	PFP	MPEC	NFP	PFP	MPEC
Small scale: $T = 25, J = 25, N = 50, G = 50$						
σ_{x1}	0.284	0.158	0.307	0.199	0.107	0.213
σ_{x2}	0.217	0.137	0.222	0.232	0.144	0.244
σ_{α}	0.211	0.118	0.212	0.200	0.115	0.202
θ^{x0}	1.577	1.207	1.575	1.704	1.278	1.686
θ^{x1}	0.108	0.084	0.109	0.096	0.098	0.098
θ^{x2}	0.087	0.081	0.087	0.072	0.070	0.074
θ^{x3}	0.064	0.064	0.064	0.045	0.045	0.045
α	0.264	0.200	0.264	0.220	0.164	0.218
Medium scale: $T = 50, J = 50, N = 50, G = 50$						
σ_{x1}	0.089	0.093	0.166	0.088	0.092	0.124
σ_{x2}	0.125	0.129	0.136	0.126	0.134	0.144
σ_{α}	0.085	0.074	0.102	0.067	0.066	0.078
θ^{x0}	0.734	0.693	0.859	0.787	0.737	0.811
θ^{x1}	0.042	0.044	0.046	0.042	0.045	0.045
θ^{x2}	0.048	0.047	0.050	0.048	0.048	0.048
θ^{x3}	0.039	0.039	0.039	0.039	0.039	0.039
α	0.101	0.096	0.119	0.085	0.082	0.087
Large scale: $T = 100, J = 50, N = 50, G = 50$						
σ_{x1}	0.098	0.111	0.151	0.087	0.101	0.101
σ_{x2}	0.055	0.086	0.092	0.036	0.079	0.052
σ_{α}	0.029	0.029	0.047	0.025	0.025	0.023
θ^{x0}	0.348	0.343	0.470	0.398	0.465	0.335
θ^{x1}	0.030	0.030	0.046	0.026	0.033	0.032
θ^{x2}	0.033	0.037	0.040	0.044	0.036	0.030
θ^{x3}	0.035	0.035	0.034	0.039	0.031	0.036
α	0.051	0.051	0.066	0.046	0.047	0.040

The RMSEs are obtained from the successfully converged estimation runs.

Table 4: The root mean squared errors (RMSEs) of NFP, PFP, and MPEC for the Monte Carlo Analysis in Table 2

C Continuity of the BLP fixed point

The continuity of the fixed point $\delta_0(\theta)$ in Assumption 3 is one of the regularity conditions that guarantee the convergence of the pseudo fixed points. In the following Lemma 1, we present a slightly stronger result instead, which will help simplifying the proof of the regularity conditions for the LT estimation.

Lemma 1. $\delta_0(\theta)$ is continuously differentiable in $\theta \in \Theta$.

Proof. Theorem 4.1 of Rust (1988) proves that $V(\omega; \delta, \theta)$ is continuously differentiable in $\Delta \times \Theta$ for all $\omega \in \Omega$ under Assumptions A10-A13 of Rust (1988) (since $\delta = [\delta_{11}, \dots, \delta_{JT}]$ are contained in parameter θ_3 in Rust (1988)). Then it is straightforward to show that

$E[V(\omega'; \delta, \theta)|\omega]$ is continuously differentiable in $(\delta, \theta) \in \Delta \times \Theta$, where $\omega = \omega(\delta, \theta)$ for some function $\omega \in C^1(\Delta \times \Theta)$ with slight abuse of notation.²⁶

Since the market share function $S(\delta, V, \theta)$ in Equation 4 is completely characterized by (δ, θ) , we can define a mapping $\Psi : \Delta \times \Theta \rightarrow R^{JT}$ such that

$$\Psi(\delta, \theta) = s - S(\delta, V, \theta).$$

By construction, Ψ is a C^1 -mapping such that $\Psi(\delta_0, \theta) = 0$ at some fixed point $\delta_0 \in \Delta$ for each $\theta \in \Theta$. On the other hand, the differentiability of $\Psi(\delta, \theta)$ implies that the Jacobian $\nabla_{\delta}\Psi(\delta_0, \theta)$ is invertible under the connected substitutes conditions by Theorem 2 of Berry et al. (2013).

Then, it follows from the implicit function theorem of Rudin (1976) (Theorem 9.28) that the fixed point $\delta_0(\theta)$ is unique and continuously differentiable in $\theta \in \Theta$. \square

D Regularity conditions of the LT estimator

We verify the regularity conditions for the asymptotic results of LTE presented in Theorems 1-4 of Chernozhukov and Hong (2003). We postulate the assumptions of Berry, Linton and Pakes (2004) on the structure of unobserved demand shocks and the construction of GMM instruments and weighting matrix.

Assumption 1 of Chernozhukov and Hong (2003) follows from Assumption 1 of Section 4 in the paper. Their Assumption 2 holds for the squared loss function with a diffuse prior: $\pi(\theta) \propto 1$. Their identifiability condition (Assumption 3) is implied by Assumption 1 (iv) in Section 4. Hence, it remains to be seen that Assumption 4 of Chernozhukov and Hong (2003) is satisfied. Instead of showing it directly, we simply verify the sufficient condition provided by Proposition 1 of Chernozhukov and Hong (2003).

Assumptions (4.1)-(4.6) of Chernozhukov and Hong (2003) are standard in the GMM estimation of the BLP demand that employs an optimal GMM weighting matrix (Berry et al., 1995, 2004), under which the generalized information equality is satisfied (Newey and McFadden, 1994). Condition (ii) of Proposition 1 holds by Lemma 1 in the appendix. Condition (iii) is implied by the standard GMM theory (Hansen, 1982; Newey and McFadden, 1994). Lastly, the stochastic equicontinuity condition in (iv) trivially holds for the continuous sample moment. Therefore, Assumption 4 follows from Proposition 1, thus satisfying the conditions of Theorems 1-4 of Chernozhukov and Hong (2003).

E Convergence of pseudo fixed point algorithm

First we establish the following lemma to prove Theorem 1.

²⁶It is because the partial derivative of $E[V(\omega'; \delta, \theta)|\omega]$ w.r.t. (δ, θ) and ω is continuous, respectively. The latter result has been proved in Proposition 3 of Norets (2010).

Lemma 2. For any given $\theta \in \Theta$ and $\delta \in \Delta$ satisfying Assumption 2, $\|\delta(\theta) - \delta^r(\theta)\| \rightarrow 0$ almost surely as $r \rightarrow \infty$.

Proof. First we decompose the approximation error into the following three parts:

$$\begin{aligned}
\|\delta(\theta) - \delta^r(\theta)\| &= \|(T_V \delta)(\theta) - (T_{V^r} \hat{\delta}^r)(\theta)\| \\
&\leq \|(T_V \delta)(\theta) - (T_V \hat{\delta}^r)(\theta)\| + \|(T_V \hat{\delta}^r)(\theta) - (T_{V^r} \hat{\delta}^r)(\theta)\| \\
&\leq \beta_\delta \|\delta(\theta) - \hat{\delta}^r(\theta)\| + A_3^r(\theta) \quad (\text{by Corollary 1}) \\
&= \beta_\delta \left\| \delta(\theta) - \sum_{i=1}^{\tilde{N}(r)} \delta^{k_i}(\theta^{*k_i}) W_{k_i, r}(\theta) \right\| + A_3^r(\theta) \\
&\leq \beta_\delta \left[\left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta) - \delta(\theta^{*k_i})] W_{k_i, r}(\theta) \right\| + \left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta^{*k_i}) - \delta^{k_i}(\theta^{*k_i})] W_{k_i, r}(\theta) \right\| \right] + A_3^r(\theta) \\
&= \beta_\delta [A_1^r(\theta) + A_2^r(\theta)] + A_3^r(\theta),
\end{aligned}$$

where $A_3^r(\theta) = \|(T_V \hat{\delta}^r)(\theta) - (T_{V^r} \hat{\delta}^r)(\theta)\|$, and

$$A_1^r(\theta) = \left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta) - \delta(\theta^{*k_i})] W_{k_i, r}(\theta) \right\|, \quad A_2^r(\theta) = \left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta^{*k_i}) - \delta^{k_i}(\theta^{*k_i})] W_{k_i, r}(\theta) \right\|.$$

In the following, we show that each of the three parts converges to zero almost surely as $r \rightarrow \infty$ for any given θ . The convergence of $A_3^r(\theta)$ follows from Assumption 6 and the continuity of $S(\delta, V, \theta)$ in V . Lemmas 3 and 4 establish the convergence of $A_1^r(\theta)$ and $A_2^r(\theta)$, respectively. Hence, the proof of Lemma 2 is complete. \square

Lemma 3. For any $\epsilon > 0$, there exist $\delta_1 > 0$ and $T \in \mathbb{N}$ such that for all $r \geq T$,

$$P\left(\left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta) - \delta(\theta^{*k_i})] W_{k_i, r}(\theta) \right\| > \epsilon\right) \leq e^{-\delta_1(N(r) - \tilde{N}(r))} \leq e^{-\delta_1 r^{\tau_1/2}}.$$

Proof. We show that as we draw more MCMC samples, the nearest neighbors $\{\theta^{*k_i}\}$ become closer to θ , such that the approximation error converges toward zero. First, we have

$$\begin{aligned}
P\left(\left\| \sum_{i=1}^{\tilde{N}(r)} [\delta(\theta) - \delta(\theta^{*k_i})] W_{k_i, r}(\theta) \right\| > \epsilon\right) &\leq P\left(\sum_{i=1}^{\tilde{N}(r)} \|\delta(\theta) - \delta(\theta^{*k_i})\| W_{k_i, r}(\theta) > \epsilon\right) \\
&\leq P\left(\bigcup_{i=1}^{\tilde{N}(r)} \left\{ \|\delta(\theta) - \delta(\theta^{*k_i})\| > \epsilon \right\}\right),
\end{aligned}$$

since $\{W_{k_i, r}(\theta), i = 1, \dots, \tilde{N}(r)\}$ are the positive weights that sum to 1, i.e., $W_{k_i, r}(\theta) \in [0, 1]$ and $\sum_{i=1}^{\tilde{N}(r)} W_{k_i, r}(\theta) = 1$. By the continuity of $\delta(\theta)$ in Lemma 1, there exists $\delta_\epsilon > 0$ such that if

$\|\theta - \theta'\| < \delta_\epsilon$, then $\|\delta(\theta) - \delta(\theta')\| < \epsilon$. Therefore,

$$\mathbb{P}\left(\bigcup_{i=1}^{\tilde{N}(r)} \left\{\|\delta(\theta) - \delta(\theta^{*k_i})\| > \epsilon\right\}\right) \leq \mathbb{P}\left(\bigcup_{i=1}^{\tilde{N}(r)} \left\{\|\theta - \theta^{*k_i}\| > \delta_\epsilon\right\}\right).$$

Since

$$\begin{aligned} \left\{\exists k_i : \|\theta - \theta^{*k_i}\| > \delta_\epsilon\right\} &\subset \left\{\forall j \in \{r - N(r), \dots, r - 1\} \setminus \{k_1, \dots, k_{\tilde{N}(r)}\} : \|\theta - \theta^{*j}\| > \delta_\epsilon\right\} \\ &\subset \bigcup_{(j_1, \dots, j_{N(r) - \tilde{N}(r)}) \in J_{N(r) - \tilde{N}(r)}} \bigcap_{i=1}^{N(r) - \tilde{N}(r)} \left\{\|\theta - \theta^{*j_i}\| > \delta_\epsilon\right\}, \end{aligned}$$

for $J_{N(r) - \tilde{N}(r)} = \{(j_1, \dots, j_{N(r) - \tilde{N}(r)}) : j_i \in \{r - N(r), \dots, r - 1\}, j_i < j_k \text{ if } i < k, \forall i, k\}$, we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\tilde{N}(r)} \left\{\|\theta - \theta^{*k_i}\| > \delta_\epsilon\right\}\right) \leq \sum_{(j_1, \dots, j_{N(r) - \tilde{N}(r)}) \in J_{N(r) - \tilde{N}(r)}} \mathbb{P}\left(\bigcap_{i=1}^{N(r) - \tilde{N}(r)} \left\{\theta^{*j_i} \in B_{\delta_\epsilon}^c(\theta)\right\}\right), \quad (18)$$

where $B_{\delta_\epsilon}(\theta) = \{\theta' \in \Theta : \|\theta - \theta'\| < \delta_\epsilon\}$. Let $\delta_1 = -\frac{1}{2} \log \left[1 - \underline{\epsilon} \nu(B_{\delta_\epsilon}(\theta))\right]$. Then for $N \geq 1$,

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^N \left\{\theta^{*j_i} \in B_{\delta_\epsilon}^c(\theta)\right\}\right) &= \int_{B^c \times \dots \times B^c} \int \mathbb{P}(\theta^{*j_N} \in B^c | H^{j_N-1}) \mathbb{P}(dH^{j_N-1} | \theta^{*j_{N-1}}, \dots, \theta^{*j_1}) \\ &\quad \times d\mathbb{P}(\theta^{*j_{N-1}}, \dots, \theta^{*j_1}) \\ &\leq \left[1 - \underline{\epsilon} \nu(B_{\delta_\epsilon}(\theta))\right] \int_{B^c \times \dots \times B^c} d\mathbb{P}(\theta^{*j_{N-1}}, \dots, \theta^{*j_1}) \quad \text{by Assumption 5} \\ &\leq \dots \leq \left[1 - \underline{\epsilon} \nu(B_{\delta_\epsilon}(\theta))\right]^{N-1} \int \mathbb{P}(\theta^{*j_1} \in B_{\delta_\epsilon}^c | H^{j_1-1}) \mathbb{P}(dH^{j_1-1}) \\ &\leq \left[1 - \underline{\epsilon} \nu(B_{\delta_\epsilon}(\theta))\right]^N = e^{-2\delta_1 N}. \end{aligned}$$

Hence it follows from the above inequality that

$$\begin{aligned} \text{RHS of (18)} &= \frac{N(r)!}{\tilde{N}(r)!(N(r) - \tilde{N}(r))!} \exp \left[-2\delta_1 (N(r) - \tilde{N}(r)) \right] \\ &\leq \exp \left[-\delta_1 (N(r) - \tilde{N}(r)) \right], \quad \forall r \geq T_1, \end{aligned}$$

for sufficiently large $T_1 \in \mathbb{N}$. The last inequality follows from Proposition 7 of Norets (2009). Since $N(r) - \tilde{N}(r) = \lfloor r^{\tau_1} \rfloor - \lfloor r^{\tau_2} \rfloor \geq r^{\tau_1} - 1 - r^{\tau_2} \geq r^{\tau_1}/2$, $\forall r \geq T_2$ for sufficiently large $T_2 \in \mathbb{N}$ by Assumption 7, the proof of Lemma 3 is complete. \square

Lemma 4. For $\epsilon > 0$, there exists $\delta_2 > 0$ and $T \in \mathbb{N}$ such that for all $r \geq T$ and all $\theta \in \Theta$,

$$P\left(\left\|\sum_{i=1}^{\tilde{N}(r)} [\delta(\theta^{*k_i}) - \delta^{k_i}(\theta^{*k_i})] W_{k_i,r}(\theta)\right\| > \epsilon\right) \leq e^{-\delta_2 r^{\tau_0 \tau_1}}.$$

Proof. First, we find an upper bound of the above approximation error.

$$\begin{aligned} A_2^r(\theta) &\leq \sum_{i=1}^{\tilde{N}(r)} \left\| \delta(\theta^{*k_i}) - \delta^{k_i}(\theta^{*k_i}) \right\| W_{k_i,r}(\theta) \\ &\leq \sum_{i=1}^{\tilde{N}(r)} \left[\beta_\delta (A_1^{k_i}(\theta^{*k_i}) + A_2^{k_i}(\theta^{*k_i})) + A_3^{k_i}(\theta^{*k_i}) \right] W_{k_i,r}(\theta) \\ &\leq \beta_\delta \max_{i=r-N(r), r-1} A_1^i(\theta^{*i}) + \beta_\delta \max_{i=r-N(r), r-1} A_2^i(\theta^{*i}) + \max_{i=r-N(r), r-1} A_3^i(\theta^{*i}), \end{aligned} \quad (19)$$

where the second inequality uses the relation $\|\delta(\theta) - \delta^r(\theta)\| \leq \beta_\delta [A_1^r(\theta) + A_2^r(\theta)] + A_3^r(\theta)$ in the proof of Lemma 2.

Following Imai et al. (2009) and Norets (2009), our strategy is to apply the above inequality recursively to show the convergence of the right hand side. Define $M(r, 0) = r$ and $M(r, k) = M(r, k-1) - N(M(r, k-1))$. Then for any given $m \in \mathbb{N}$,

$$\begin{aligned} \text{RHS of (19)} &\leq \sum_{k=1}^m \beta_\delta^k \max_{i=M(r,k), r-k} A_1^i(\theta^{*i}) + \beta_\delta^m \max_{i=M(r,m), r-m} A_2^i(\theta^{*i}) + \max_{i=r-N(r), r-1} A_3^i(\theta^{*i}) \\ &\leq \frac{\beta_\delta}{1-\beta_\delta} \max_{i=r-mN(r), r-1} A_1^i(\theta^{*i}) + \beta_\delta^m \max_{i=r-mN(r), r-1} A_2^i(\theta^{*i}) + \max_{i=r-N(r), r-1} A_3^i(\theta^{*i}). \end{aligned}$$

We use the fact that $A_2^i(\cdot)$ is bounded above by some $\bar{A}_2 < \infty$, which follows from its continuity in the compact space Θ . Let $m(r) = \lfloor (r - r^{\tau_0})/N(r) \rfloor$. Then since $\max_{i=r-N(r), r-1} A_3^i(\theta^{*i})$ converges to zero as $r \rightarrow \infty$ by Assumption 6, for any given $\epsilon > 0$, we have $\epsilon_1, \epsilon_2 > 0$ and $T_1 \in \mathbb{N}$ such that

$$\begin{aligned} P(A_2^r(\theta) > \epsilon) &\leq P\left(\frac{\beta_\delta}{1-\beta_\delta} \max_{i=r-m(r)N(r), r-1} A_1^i(\theta^{*i}) + \beta_\delta^{m(r)} \max_{i=r-m(r)N(r), r-1} A_2^i(\theta^{*i}) > \epsilon_1\right), \quad \forall r \geq T_1, \\ &\leq P\left(\frac{\beta_\delta}{1-\beta_\delta} \max_{i=r-m(r)N(r), r-1} A_1^i(\theta^{*i}) > \epsilon_1/2\right) + P\left(\beta_\delta^{m(r)} \bar{A}_2 > \epsilon_1/2\right) \\ &\leq \sum_{i=r-m(r)N(r)}^{r-1} P(A_1^i(\theta^{*i}) > \epsilon_2), \quad \forall r \geq \max(T_1, T_2), \quad \text{for sufficiently large } T_2 \in \mathbb{N}, \\ &\leq \sum_{i=r-m(r)N(r)}^{r-1} e^{-\delta_2(N(i)-\tilde{N}(i))} \quad (\text{by Lemma 3}) \\ &\leq e^{-\delta_2 r^{\tau_0 \tau_1}}, \quad \forall r \geq \max(T_1, T_2, T_3), \quad \text{for sufficiently large } T_3 \in \mathbb{N}. \end{aligned}$$

The third inequality holds since $m(r) \rightarrow \infty$ as $r \rightarrow \infty$ by construction. The last inequality follows from Proposition 8 of Norets (2009). \square

Now we show that the approximation error converges uniformly in Θ in Theorem 1. It assumes the measurability of the least upper bound of the approximation error, which can be easily verified by the same argument of Norets (2009).

Theorem 1. *Under Assumptions 1–7, the pseudo fixed point δ^r converges almost surely to the exact fixed point δ uniformly in $\theta \in \Theta$. In other words,*

$$\sup_{\theta \in \Theta} \|\delta(\theta) - \delta^r(\theta)\| \rightarrow 0 \quad \text{a.s. as } r \rightarrow \infty.$$

Proof. First, we decompose the approximation error into three parts as in the proof of Lemma 2:

$$\|\delta(\theta) - \delta^r(\theta)\| \leq \beta_\delta [A_1^r(\theta) + A_2^r(\theta)] + A_3^r(\theta).$$

By Assumption 6, $\sup_\theta A_3^r(\theta) \rightarrow 0$ a.s. Furthermore, $\sup_\theta A_2^r(\theta) \rightarrow 0$ a.s. since the RHS of Equation 19 does not depend on θ . Hence, it suffices to show that $\sup_\theta A_1^r(\theta) \rightarrow 0$ a.s. By using the same steps in the proof of Lemma 3, we obtain

$$\begin{aligned} \left\{ \sup_\theta A_1^r(\theta) > \epsilon \right\} &\subset \bigcup_{\theta \in \Theta} \bigcup_{(j_1, \dots, j_{N(r)-\tilde{N}(r)}) \in J_{N(r)-\tilde{N}(r)}} \bigcap_{i=1}^{N(r)-\tilde{N}(r)} \left\{ \theta^{*j_i} \in B_{\delta_\epsilon}^c(\theta) \right\} \\ &\subset \bigcup_{m=1}^M \bigcup_{\theta \in B_{\delta_\epsilon/2}(\theta_m)} \bigcup_{(j_1, \dots, j_{N(r)-\tilde{N}(r)}) \in J_{N(r)-\tilde{N}(r)}} \bigcap_{i=1}^{N(r)-\tilde{N}(r)} \left\{ \theta^{*j_i} \in B_{\delta_\epsilon}^c(\theta) \right\} \text{ for some } \{\theta_m \in \Theta\}_{m=1}^M \\ &\subset \bigcup_{m=1}^M \bigcup_{(j_1, \dots, j_{N(r)-\tilde{N}(r)}) \in J_{N(r)-\tilde{N}(r)}} \bigcap_{i=1}^{N(r)-\tilde{N}(r)} \left\{ \theta^{*j_i} \in B_{\delta_\epsilon/2}^c(\theta_m) \right\}, \end{aligned}$$

where the second relation holds since Θ can be covered by a finite number of open balls, $\{B_{\delta_\epsilon/2}(\theta_m)\}_{m=1}^M$. Then as in Lemma 3, there exists some $C > 0$ such that for sufficiently large r ,

$$\begin{aligned} \mathbb{P}\left(\sup_\theta A_1^r(\theta) > \epsilon\right) &= \frac{MN(r)!}{\tilde{N}(r)!(N(r)-\tilde{N}(r))!} e^{-2C(N(r)-\tilde{N}(r))} \\ &\leq M e^{-C(N(r)-\tilde{N}(r))}, \\ &\leq M e^{-Cr^{\tau_1}/2}, \end{aligned}$$

where the second inequality follows from the Proposition 7 of Norets (2009), and the third

inequality from the last part of the proof in Lemma 3. Therefore we obtain

$$\sum_{r=1}^{\infty} P \left(\sup_{\theta \in \Theta} \|\delta(\theta) - \delta^r(\theta)\| > \epsilon \right) < \infty,$$

and Theorem 1 follows from the Borel-Cantelli lemma. \square

F The approximation error of the pseudo fixed point algorithm

Proposition 1. *For given $\hat{\delta}^r$ and θ , denote $\epsilon = \epsilon(r)$ in Assumption 6 of Section 4. Then For any set of value functions $V = [V_1, \dots, V_N]$ satisfying Assumption 1 (iii) under the model described in Section 2.1,*

$$\|T_V(\hat{\delta}^r)(\theta) - T_{V^r}(\hat{\delta}^r)(\theta)\| = O(\epsilon).$$

Proof. Suppose $V' = V + \epsilon$, i.e., $V'_i(\omega) = V_i(\omega) + \epsilon$ for all $\omega \in \Omega$ and $i = 1, \dots, N$. Then for a given $(\delta, \theta) \in \Delta \times \Theta$, $j = 1, \dots, J$, and $t = 1, \dots, T$, the market share function $S_{jt}(\delta, V', \theta)$, following Equation 4, can be written with an abuse of notation as

$$S_{jt}(\delta, V', \theta) = \int \pi_{ijt}(\epsilon) h_{it}(\epsilon) F_{\nu}(d\nu_i),$$

where $h_{it}(\epsilon)$ is equivalent to the market participation rate, h_{it} , in Section 2.1 with V replaced by V' , and

$$\begin{aligned} \pi_{ijt}(\epsilon) &= \frac{\exp(\delta_{jt} + \mu_{ijt})}{\exp\left(\beta \mathbb{E}[V'_i(\omega') | \omega_{it}]\right) + \exp(\omega_{it})} \\ &= \frac{\exp(\delta_{jt} + \mu_{ijt})}{\exp\left(\beta \mathbb{E}[V_i(\omega') | \omega_{it}] + \beta\epsilon\right) + \exp(\omega_{it})}. \end{aligned}$$

For given (j, t) , define $F(\epsilon) = \log S_{jt}(\delta, V + \epsilon, \theta)$, and $\pi_{i0t} = 1 - \sum_{j=1}^J \pi_{ijt}$. Since $F(\epsilon)$ is an analytic function of ϵ everywhere, we can expand it as

$$\begin{aligned} F(\epsilon) &= F(0) + F'(0)\epsilon + O(\epsilon) \\ &= F(0) + \frac{\epsilon}{S_{jt}(\delta, V, \theta)} \int \left[\pi'_{ijt}(0) h_{it}(0) + \pi_{ijt}(0) h'_{it}(0) \right] F_{\nu}(d\nu_i) + O(\epsilon). \end{aligned}$$

Since $h_{it}(\epsilon) = \prod_{s=1}^{t-1} \pi_{i0s}(\epsilon)$ by definition,

$$h'_{it}(0) = \sum_{\tau=1}^{t-1} \beta \pi_{i0\tau} (1 - \pi_{i0\tau}) \prod_{\substack{s=1 \\ s \neq \tau}}^{t-1} \pi_{i0s},$$

where we use an abbreviated notation $\pi_{i0s} = \pi_{i0s}(0)$. By the triangular inequality,

$$\begin{aligned} |F(\epsilon) - F(0)| &\leq \frac{\beta\epsilon}{S_{jt}(\delta, V, \theta)} \left[\int \pi_{ijt} \pi_{i0t} h_{it} F_{\nu}(d\nu_i) + \int \pi_{ijt} \frac{t-1}{2^2} F_{\nu}(d\nu_i) \right] + O(\epsilon) \\ &\leq \epsilon \left[1 + \frac{t-1}{4S_{jt}(\delta, V, \theta)} \right] + O(\epsilon) = O(\epsilon), \end{aligned}$$

since the first inequality follows from $\pi'_{ijt}(0) = -\beta \pi_{ijt}(0) \pi_{i0t}(0)$. We can assume without loss of generality that $|\log S(\delta, V, \theta) - \log S(\delta, V', \theta)| \leq |\log S(\delta, V, \theta) - \log S(\delta, V', \theta)|$ for sufficiently large r since $S(\delta, V, \theta)$ is continuous in V . Hence Proposition 1 follows from the definition of T_V and T_{V^r} . \square

G Convergence of fixed point iterations

In this section, we show that the connected substitutes conditions of Berry et al. (2013) imply the local convergence of T_V defined in Section 3.1, which was assumed in Assumption 4.

Lemma 5. *Suppose that Assumptions 1–3 of Berry et al. (2013) hold in the model specified in Section 2.1. Then there exists a fixed point mapping T_V that has a Jacobian $\nabla_{\delta} T_V$ satisfying $\|\nabla_{\delta} T_V(\delta_0)\|_{\infty} < 1$.*

Proof. Given the differentiability of the demand system, the connected substitutes conditions imply that $\partial S_{kt}(\delta_0)/\partial \delta_{jt} \leq 0$ for all $j \neq k$, and $\partial S_{0t}(\delta_0)/\partial \delta_{jt} < 0$. Then we obtain $\partial S_{jt}(\delta_0)/\partial \delta_{jt} > 0$ since $S_{0T} + \sum_{t=1}^T \sum_{j=1}^J S_{jt} = 1$.

The Jacobian $\nabla_{\delta} T_V(\delta)$ has elements $\{\partial T_{jt}(\delta)/\partial \delta_{ks} : 1 \leq j \leq J, 1 \leq k \leq J, 1 \leq t \leq T, 1 \leq s \leq T\}$ given as

$$\frac{\partial T_{jt}(\delta)}{\partial \delta_{ks}} = \begin{cases} 1 - \frac{\phi}{S_{jt}(\delta)} \frac{\partial S_{jt}(\delta)}{\partial \delta_{jt}} & \text{if } k = j \text{ \& } s = t, \\ -\frac{\phi}{S_{jt}(\delta)} \frac{\partial S_{jt}(\delta)}{\partial \delta_{ks}} & \text{otherwise.} \end{cases}$$

By Equation 4 and Theorem 4.1 of Rust (1988), the partial derivatives $\{\partial S_{jt}(\delta_0)/\partial \delta_{ks} : \forall j, k, s, t\}$ are continuous on the bounded space Δ . Hence, the derivatives are bounded at δ_0 . Since the own partial derivatives $\{\partial S_{jt}(\delta_0)/\partial \delta_{jt} : \forall j, t\}$ are strictly positive, we can always find a sufficiently small $\phi > 0$ such that $\|\nabla_{\delta} T_V(\delta_0)\|_{\infty} < 1$. \square

Corollary 1. *Suppose that T_V satisfies the assumptions of Lemma 5 for the model of Section 2.1. Then there exists a neighborhood D containing the fixed point δ_0 such that $T_V^n(\delta) \rightarrow \delta_0$ as $n \rightarrow \infty$ for all $\delta \in D$.*

Proof. By Lemma 5, there exists $\epsilon > 0$ such that $\|\nabla_\delta T_V(\delta_0)\|_\infty + \epsilon < 1$. Then for a given $\delta \in D$, we can expand $T_V(\delta)$ around δ_0 as

$$T_V(\delta) = \delta_0 + \nabla T_V(\delta_0)(\delta - \delta_0) + O(\|\delta - \delta_0\|^2).$$

By rearranging the terms and taking the vector norm $\|\cdot\|_\infty$ defined below, we obtain

$$\begin{aligned} \|T_V(\delta) - \delta_0\| &\leq \|\nabla T_V(\delta_0)(\delta - \delta_0)\| + O(\|\delta - \delta_0\|^2) \\ &\leq \|\delta - \delta_0\| \|\nabla T_V(\delta_0)\|_\infty + O(\|\delta - \delta_0\|^2) \\ &\leq \|\delta - \delta_0\| (\|\nabla T_V(\delta_0)\|_\infty + \epsilon) \\ &< \|\delta - \delta_0\|, \end{aligned}$$

where the second inequality follows from the relation between vector and matrix norms: $\|A\delta\| = \|A[\delta \ 0 \ \cdots \ 0]\|_\infty \leq \|A\|_\infty \|\delta\|$ for any matrix A . The third inequality is obtained by finding a sufficiently small neighborhood D . The last inequality follows by construction of ϵ . By recursively substituting δ with $T_V(\delta)$, we achieve the convergence result. \square