Revisiting the Nested Fixed-Point Algorithm in BLP Random Coefficients Demand Estimation*

Jinhyuk Lee Kyoungwon Seo

September 19, 2015

Abstract

This paper examines the numerical properties of the nested fixed-point algorithm (NFP) in the estimation of Berry, Levinsohn, and Pakes's (1995) random coefficient logit demand model. We theoretically derive an upper bound on the numerical bias in the NFP estimates with contraction mappings (NFP/CTR) which, under our assumptions, is sharper than that derived by Dubé, Fox, and Su (2012). We also show that, compared with NFP/CTR, NFP using Newton's method has a smaller bound of the estimate error and converges more often. We illustrate actual numerical performance of NFP in our Monte Carlo experiments.

1 Introduction

The nested fixed-point approach (NFP) has been widely used for structural estimation in economics. In empirical industrial organization, for example, Rust (1987) uses NFP to estimate the single-agent dynamic discrete choice model. Berry, Levinsohn, and Pakes (1995; BLP hereafter) apply NFP to estimate the random coefficients logit demand model. To estimate structural parameters, NFP iterates two nested loops: The inner loop numerically solves a fixed-point problem given a parameter value, and the outer loop optimizes the objective function over a parameter space. Dubé, Fox, and Su (2012; DFS hereafter) argue that numerical errors from the inner loop propagate into the objective function, which leads to less convergence or to convergence to a point that is not a local minimum if a loose stopping criterion is used. As an alternative to NFP, DFS propose a

*Corresponding author: Kyoungwon Seo, Graduate School of Finance, Korea Advanced Institute of Science and Technology, Seoul, South Korea, kseo@kaist.ac.kr. Jinhyuk Lee is at Department of Economics, Korea University, Seoul, South Korea, jinhyuklee@korea.ac.kr. Seo gratefully acknowledges the financial support of National Research Foundation of Korea (NRF-2014S1A5A8018374) funded by the Korean Government. We thank Kyoo il Kim, Daniel Ackerberg and in particular Jeremy Fox for very helpful discussions.

mathematical programming with equilibrium constraints (MPEC) approach for the BLP estimation. There is no nested inner loop in MPEC, and thus it does not suffer from the error propagation problem.

This paper examines the numerical properties of NFP in the estimation of BLP's random coefficient logit demand model. In particular, we study two versions of NFP. The first is BLP's nested contraction mapping algorithm, which DFS also study.[1] We refer to this algorithm as NFP/CTR and find that the upper bound on the estimate error of NFP/CTR has the same order of the inner loop tolerance, and thus it is sharper under our assumptions than that derived by DFS. For MPEC, the numerical error bound is of the same order as the feasibility tolerance. This implies that NFP/CTR's numerical error in the parameter estimate is of the same order as MPEC's numerical error arising from the feasibility constraints when the inner loop and feasibility tolerances are of the same order.

Our error bound of NFP/CTR estimates should be interpreted carefully. Unlike in DFS, our error bound relies on the assumption that an estimate is obtained after an optimization routine converges. As DFS argue, this assumption often fails when the inner loop tolerance is too loose. With a loose inner loop tolerance, the outer loop objective function is evaluated inaccurately, and the outer loop may take a wrong direction toward a local optimum. Our result on a numerical error bound for an NFP/CTR estimate indicates that when a researcher is lucky enough to find convergence under a not-very-tight inner loop tolerance with a reasonable outer loop tolerance, the numerical error of the estimate is less problematic than suggested by the numerical error bound in DFS. The numerical error bound of DFS is larger than ours but may be applied even when an optimization routine does not report convergence like in their example with a loose inner loop tolerance. In addition, we have a different set of technical assumptions from DFS, which leads to a sharper error bound than that of DFS. We believe that all our assumptions as well as theirs are satisfied in most applications of the BLP model.

Our second version of NFP, which we refer to as NFP/NT, uses Newton's method to solve the inner fixed-point problem.[2] In NFP/NT, we combine contraction mapping iterations and Netwon's method, following Rust (1987) to ensure global convergence. The original version of Newton's method converges locally only. One way to guarantee global convergence is to start with contraction mapping iterations and then switch to Newton's method. A similar implementation to ours appears in Iskhakov et al. (2015), who estimate the single-agent dynamic discrete choice model. They show that NFP with Newton's method is as fast as MPEC.

We argue that NFP/NT has good numerical properties as well. First, the numerical error in the

[1] We note that their error bound (Theorem 3 in DFS) is an interesting result but is not their main criticism of loose inner loop tolerances.

[2] When we refer to both NFP/CTR and NFT/NT, we will simply write NFP.

parameter estimate of NFP/NT is smaller than that of NFP/CTR. We show that the error bound of NFP/NT has an order of the square of the inner loop tolerance. Second, due to the quadratic convergence rate of Newton's method, only a few more iterations are necessary to reduce the inner loop error, for example, from 10^{-6} to 10^{-12}. This property makes it easy to minimize the error propagation into the outer loop objective function, and the inner loop tolerance does not need to be loosened. With a tight tolerance, an optimization routine is more likely to converge to a local optimum.

We illustrate the actual numerical performance of NFP in our Monte Carlo experiments. We vary the level of inner loop tolerance but use a tight outer loop tolerance, 10^{-6}, to observe the proportion of convergence and the accuracy of the estimates. As in DFS, we find many cases of nonconvergence of NFP/CTR when using a loose inner loop tolerance. We also study the numerical error of the NFP/CTR estimate conditional on convergence. The experiment shows that, when NFP/CTR converges, the algorithm finds the estimates properly even under a loose inner loop tolerance. We show that NFP/NT converges more often and achieves a tighter inner loop tolerance with a few more iterations, as expected.

Our implementation of NFP/NT is not the first to use Newton's method for the inversion of the market share equations in the BLP estimation. Patel (2012) proposes Newton's method as a supplement to contraction mapping when contraction mapping does not perform efficiently. Reynaerts, Varadhan, and Nash (2012) propose Newton's method as an alternative to the contraction mapping iterations. Houde (2012) proposes combining a quasi-Newton method with the contraction mapping iterations. However, none of these studies discusses the effect of the change in tolerance on the numerical error in the estimate.

We organize the rest of this paper as follows. In Section 2, we discuss a random coefficients logit demand model using aggregate data and present the estimation procedures, NFP/CTR and NFP/NT. Section 3 provides a theoretical analysis of the numerical error in the estimate computed with NFP. In Section 4, we present the Monte Carlo experiments. Section 5 concludes. Finally, we provide proofs in Appendix A and discuss technical details of our implementation of NFP/NT in Appendix B.

2 Model and Estimation

In this section, we briefly explain the random coefficient logit demand model using aggregate data. Then, we discuss NFP/CTR and NFP/NT estimation procedures.

2.1 Random Coefficients Logit Demand Model

We assume a set of independent markets, $t = 1, ..., T$. For simplicity, we assume each market to have the same set of products, $j = 1, ..., J$. The utility of consumer i from consuming product j in market t is

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

where X_{jt} is a vector of observed product characteristics, β_i is a vector of consumer i's preference for observed product characteristics, ξ_{jt} is a product characteristic or a demand shock that is unobserved by the econometrician, and ε_{ijt} is an idiosyncratic shock. We also denote the option not to purchase as $j = 0$. The utility of consumer i from not purchasing in market t is $U_{i0t} = \varepsilon_{i0t}$. Following the standard random coefficient logit demand model, we further assume that the vector of random coefficients β_i is drawn independently from the distribution $F(\beta_i; \theta)$, and ε_{ijt} follows the Type I extreme value distribution.

Under the assumptions, the market share function of product j in market t is

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

where $X_t \equiv (X_{1t}', ..., X_{Jt}')'$ and $\xi_t \equiv (\xi_{1t}, ..., \xi_{Jt})'$. In practice, we approximate the integral using simulation; that is, we generate n_s draws of β_i and then evaluate (1) by calculating

$$\frac{1}{n_s} \sum_{i=1}^{n_s} \frac{\exp\left(X_{jt}\beta_i + \xi_{jt}\right)}{1 + \sum_{j'=1}^{J} \exp\left(X_{j't}\beta_i + \xi_{j't}\right)}. \tag{2}$$

For simplicity, we write $s_j(\xi_t; \theta)$ instead of $s_j(X_t, \xi_t; \theta)$. We define $s(\xi_t; \theta) \equiv (s_1(\xi_t; \theta), ..., s_J(\xi_t; \theta))'$ as the predicted market share functions in market t, and $S_t \equiv (S_{1t}, ..., S_{Jt})'$, where S_{jt} is the observed market share of product j in market t.

2.2 Estimation Procedures

Prices are likely to be correlated with the unobserved product characteristics ξ_{jt}. To address the endogeneity, BLP propose the generalized method of moments (GMM) estimation with the moment conditions

$$E\left[\xi_{jt} | z_{jt}\right] = 0,$$

where z_{jt} is a vector of instrumental variables. The moment conditions are often implemented as $E\left[\xi_{jt} h(z_{jt})\right] = 0$, where $h(\cdot)$ is a vector-valued function. To form the moment, BLP invert the

market share equations $S_t = s(\xi_t; \theta)$ for a given θ, and obtain the solution, denoted as $\xi_t(\theta) \equiv (\xi_{1t}(\theta), ..., \xi_{Jt}(\theta))'$. The sample moment is

$$g_T(\xi(\theta)) = \frac{1}{T}\sum_{t=1}^{T} g(\xi_t(\theta)) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J} \xi_{jt}(\theta) h(z_{jt}),$$

where $\xi(\theta) \equiv \left(\xi_1(\theta)', ..., \xi_T(\theta)'\right)'$ and should be close to 0 when T is large.

We can write the BLP GMM problem as follows:

$$\min_{\theta\in\Theta} Q(\xi(\theta)) = \min_{\theta\in\Theta} g_T(\xi(\theta))' W g_T(\xi(\theta)),$$

where W is a weight matrix. The BLP estimator θ^* is the minimizer of the BLP GMM problem in the finite sample.

BLP propose the following NFP approach: invert the market share equations for a given θ to obtain $\xi_t(\theta)$, $t = 1, ..., T$ in the inner loop, and search for θ that minimizes the GMM objective function $Q(\xi(\theta))$ in the outer loop. To obtain $\xi_t(\theta)$ in the inner loop, BLP suggest the contraction mapping iterations:

$$\xi^h_{CTR} = \xi^{h-1}_{CTR} + \ln S_t - \ln s\left(\xi^{h-1}_{CTR}; \theta\right), \quad h = 1, 2, \cdots$$

such that

$$\left\|\xi^{h+1}_{CTR} - \xi^h_{CTR}\right\| \leq L(\theta)\left\|\xi^h_{CTR} - \xi^{h-1}_{CTR}\right\|, \ \forall \xi^h_{CTR}, \quad h = 1, 2, \cdots \tag{3}$$

with a Lipschitz constant $L(\theta) \in [0, 1)$. They show that iterative applications of the contraction mapping converge to $\xi_t(\theta)$. This is what we refer to as NFP/CTR.

As an alternative, we propose NFP/NT, which starts with contraction mapping iterations and switches to Newton's method to ensure global convergence. Newton's method in NFP/NT begins with an initial guess ξ^0_{NT} obtained by the contraction mapping, and the subsequent iterate is computed using the iteration rule:

$$\xi^h_{NT} = \xi^{h-1}_{NT} + \left[\nabla_\xi s\left(\xi^{h-1}_{NT}; \theta\right)\right]^{-1}\left[S_t - s\left(\xi^{h-1}_{NT}; \theta\right)\right], \quad h = 1, 2, \cdots \tag{4}$$

where $\nabla_\xi s\left(\xi^{h-1}_{NT}; \theta\right)$ is the matrix of first partial derivatives of $s(\xi; \theta)$ with respect to ξ at the iterate ξ^{h-1}_{NT}. The Newton iterate converges if $s(\xi; \theta)$ is continuously differentiable with respect to ξ, $\nabla_\xi s(\xi; \theta)$ is invertible, and the initial guess is sufficiently close to a solution $\xi_t(\theta)$.

3 Bounds on Errors in the Estimate

In this section, we derive upper bounds on the errors in the BLP estimate when it is computed with NFP/CTR and NFP/NT. We follow the same notations as in DFS. For notational simplicity, we omit the market index t.

When we solve for $\xi(\theta)$ in the inner loop, the exact value of the solution is not available. Instead, we impose a stopping rule of the inner loop, which requires the change of two successive iterates to be less than a given inner loop tolerance, ϵ_{in}:

$$\left\|\xi^h - \xi^{h-1}\right\| \leq \epsilon_{in}. \tag{5}$$

This rule is commonly used to check convergence in the inner loop.

Following DFS, let $\xi(\theta, \epsilon_{in})$ be the first iterate ξ^h satisfying the stopping rule (5). When the contraction mapping iterations are applied, we write $\xi_{CTR}(\theta, \epsilon_{in})$. Similarly, we write $\xi_{NT}(\theta, \epsilon_{in})$ for Newton's method.

We are ready to study an effect of the inner loop and outer loop tolerances on the estimate errors for NFP. An optimization routine stops and reports θ if the norm of the gradient of the objective function is less than the outer loop tolerance, ϵ_{out}. Computing the gradient of the objective function, $\nabla_\theta Q(\xi(\theta))$, is not free of numerical error from the inner loop. To deal with this explicitly, define, for any ξ and θ (even when $\xi \neq \xi(\theta)$),

$$\Gamma(\xi, \theta) \equiv \left[-\left(\frac{\partial s(\xi;\theta)}{\partial \xi} \right)^{-1} \frac{\partial s(\xi;\theta)}{\partial \theta} \right]' \frac{\partial Q(\xi)}{\partial \xi}.$$

We can show that $\nabla_\theta Q(\xi(\theta)) = \Gamma(\xi(\theta), \theta)$ using the implicit function theorem. However, the error-free value of $\xi(\theta)$ $(= \xi(\theta, 0))$ is infeasible in most applications, and so is $\nabla_\theta Q(\xi(\theta))$. Thus, in practice, the approximated value, $\Gamma(\xi(\theta, \epsilon_{in}), \theta)$, is used instead of the true value $\nabla_\theta Q(\xi(\theta))$ and is understood as an analytical gradient of the objective function. Then, the optimizer reports θ if

$$\|\Gamma(\xi(\theta, \epsilon_{in}), \theta)\| \leq \epsilon_{out}. \tag{6}$$

Let $\widetilde{\theta}(\epsilon_{in}, \epsilon_{out})$ be a parameter θ satisfying (6). As noted previously, $\widetilde{\theta}_{CTR}(\epsilon_{in}, \epsilon_{out})$ and $\widetilde{\theta}_{NT}(\epsilon_{in}, \epsilon_{out})$ denote the computed estimates from the contraction mapping iterations and Newton's method, respectively. Recall that θ^* is the true value of the BLP estimate in the finite sample, and $\theta^* = \widetilde{\theta}(0, 0)$.

In what follows, we often suppress $(\epsilon_{in}, \epsilon_{out})$ in $\widetilde{\theta}(\epsilon_{in}, \epsilon_{out})$, and just write $\widetilde{\theta}$.

Theorem 1. *Suppose that (a) θ^* is an interior solution, (b) $\nabla_{\theta\theta'} Q(\xi(\theta))|_{\theta=\theta^*}$ is nonsingular, and (c) $\frac{\partial \Gamma(\xi,\theta)}{\partial \xi'}|_{(\xi,\theta)=(\xi(\widetilde{\theta}),\widetilde{\theta})}$ is bounded. Then,*

$$\left\|\widetilde{\theta} - \theta^*\right\| \leq O\left(\left\|\Gamma\left(\xi\left(\widetilde{\theta}, \epsilon_{in}\right), \widetilde{\theta}\right)\right\|\right) + O\left(\left\|\xi\left(\widetilde{\theta}, \epsilon_{in}\right) - \xi\left(\widetilde{\theta}\right)\right\|\right). \tag{7}$$

This theorem states a property of the estimate errors of both NFP/CTR and NFP/NT. The first term in the right-hand side (RHS) of (7) is due to the errors in the evaluation of the gradient and the second is due to the inner loop error. Because $\left\|\Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right)\right\| \leq \epsilon_{out}$, it is possible to replace the first term with $O\left(\epsilon_{out}\right)$. The second term depends on how the inner loop is solved. Subsequently, we investigate the effect of the inner loop errors on the estimates for NFP/CTR and NFP/NT.

3.1 Numerical Errors in NFP/CTR

DFS begin their analysis of the error propagation of NFP/CTR by observing the following bound on the error from (5):

$$\left\|\xi_{CTR}^{h}-\xi\left(\theta\right)\right\| \leq \frac{L\left(\theta\right)}{1-L\left(\theta\right)}\left\|\xi_{CTR}^{h}-\xi_{CTR}^{h-1}\right\| \leq \frac{L\left(\theta\right)}{1-L\left(\theta\right)}\epsilon_{in}. \tag{8}$$

That is, $\left\|\xi_{CTR}\left(\theta,\epsilon_{in}\right)-\xi\left(\theta\right)\right\| \leq \frac{L(\theta)}{1-L(\theta)}\epsilon_{in}$.

DFS set $\epsilon_{out}=0$ when defining $\widehat{\theta}\left(\epsilon_{in}\right) \equiv \widetilde{\theta}_{CTR}\left(\epsilon_{in},0\right)$. We use the same definition in the following corollary.

Corollary 2. *Suppose that the assumptions of Theorem 1 hold. Then,*

$$\left\|\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right\| \leq O\left(\frac{L\left(\widehat{\theta}\left(\epsilon_{in}\right)\right)}{1-L\left(\widehat{\theta}\left(\epsilon_{in}\right)\right)}\epsilon_{in}\right). \tag{9}$$

Theorem 3 in DFS states that, under mild conditions,

$$O\left(\left\|\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right\|^{2}\right) \leq \left|Q\left(\xi_{CTR}\left(\widehat{\theta}\left(\epsilon_{in}\right),\epsilon_{in}\right)\right)-Q\left(\xi\left(\theta^{*},0\right)\right)\right| + O\left(\frac{L\left(\widehat{\theta}\left(\epsilon_{in}\right)\right)}{1-L\left(\widehat{\theta}\left(\epsilon_{in}\right)\right)}\epsilon_{in}\right). \tag{10}$$

The first term in the RHS of (10) is the bias in the objective function values, and the second is the numerical bias in the demand shocks, ξ. In contrast, (9) has only the numerical bias in the demand shocks. Our error bound does not have the bias in the objective function values because we apply the Taylor expansion on $\nabla_{\theta}Q\left(\xi\left(\theta\right)\right)$ which vanishes at θ^{*} due to the first order condition. But DFS do so on $Q\left(\xi\left(\theta\right)\right)$ which does not necessarily vanish. It is worth noting that Corollary 2, unlike (10), relies on the assumption that the outer loop is solved exactly and (10) may be applied even when an optimization routine does not report convergence.

A more important difference lies in an order of magnitude of the error bounds. The error bound in (9) has the same order as the inner loop tolerance when the contraction mapping iterations are applied. This is sharper than that in (10), which has an order of the square-root of the inner

loop tolerance. Furthermore, the error in the estimate computed by MPEC has the same order as the optimality and feasibility tolerances (DFS, p. 2242 and references therein). Therefore, if the tolerances of NFP/CTR are set to the same levels as those of MPEC, the upper bounds on the estimate errors for both approaches are of the same order.

Why do orders of error bounds differ? DFS argue that $Q\left(\xi\left(\theta,\epsilon_{in}\right)\right)$ is not differentiable with respect to ϵ_{in}, and conjecture that $\xi\left(\theta,\epsilon_{in}\right)$ and $Q\left(\xi\left(\theta,\epsilon_{in}\right)\right)$ are not differentiable with respect to θ for $\epsilon_{in}>0$, even though the denoted differentiability can induce a sharper bound, as they discuss (see also Theorem 2 in Ackerberg, Geweke, and Hahn (2009)).[3] Taking this nondifferentiability into account, they approximate the objective function $Q\left(\xi\left(\theta\right)\right)$ with the Taylor expansion. In one part of their proof (Theorem 3 in DFS), they derive

$$\begin{aligned} & Q\left(\xi\left(\widehat{\theta}\left(\epsilon_{in}\right),0\right)\right)-Q\left(\xi\left(\theta^{*},0\right)\right) \\ = & \left[\nabla_{\theta}Q\left(\xi\left(\theta^{*}\right)\right)\right]'\left(\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right)+O\left(\left\|\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right\|^{2}\right). \end{aligned}$$

Then, because $\nabla_{\theta}Q\left(\xi\left(\theta^{*}\right)\right)=0$, the term involving $\left(\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right)$ vanishes, and the error bound is expressed in terms of $\left\|\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right\|^{2}$. They assume differentiability of $Q\left(\xi\left(\theta\right)\right)$ with respect to θ, which is a reasonable assumption.

We agree with DFS on the aforementioned nondifferentiability, but it is also reasonable to assume that the gradient of the objective function, $\nabla_{\theta}Q\left(\xi\left(\theta\right)\right)$, is differentiable with respect to θ.[4] Then, applying the Taylor approximation to the gradient, we obtain

$$\begin{aligned} & \nabla_{\theta}Q\left(\xi\left(\widehat{\theta}\left(\epsilon_{in}\right)\right)\right)-\nabla_{\theta}Q\left(\xi\left(\theta^{*}\right)\right) \\ = & \left[\nabla_{\theta\theta'}Q\left(\xi\left(\theta^{*}\right)\right)\right]'\left(\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right)+O\left(\left\|\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right\|^{2}\right). \end{aligned}$$

The second-order condition of the GMM minimization problem is guaranteed by nonsingularity of $\nabla_{\theta\theta'}Q\left(\xi\left(\theta^{*}\right)\right)$, which is often assumed. Then, the term involving $\left(\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right)$ does not vanish, and our error bound can be expressed in terms of $\left(\widehat{\theta}\left(\epsilon_{in}\right)-\theta^{*}\right)$, unlike in DFS. Thus, we can obtain a sharper error bound. See Appendix A for a complete proof.

It is worth noting that the theorems in DFS are not incorrect. We just have different sets of assumptions and theorems. However, we believe that all our assumptions (as well as their assumptions) are satisfied in most applications. Assumption (a) in Theorem 1 is made implicitly in DFS. Assumption (b) is common. Indeed, $\nabla_{\theta\theta'}Q\left(\xi\left(\theta\right)\right)$ is often assumed to be nonsingular at

[3] Kim and Park (2010) also assume the differentiability and find that the error bound in the estimate has the same order as the inner loop tolerance in large samples. We (and DFS) study small sample errors.

[4] DFS also assume this differentiability in the proof of Theorem 2.

the true parameter value of θ to guarantee asymptotic normality of the estimator (Newey and McFadden 1994). In Assumption (c), we impose a boundedness condition on $\frac{\partial \Gamma(\xi,\theta)}{\partial \xi'}$, while DFS impose a boundedness condition on $\frac{\partial Q}{\partial \xi'}$. We take into account that (6) is a practical stopping criterion, where $\Gamma\left(\xi\left(\theta,\epsilon_{in}\right),\theta\right)$ approximates $\nabla_\theta Q\left(\xi\left(\theta\right)\right)$ and $\Gamma\left(\xi,\theta\right)$ is differentiable with respect to ξ given θ. We believe that $\Gamma\left(\xi,\theta\right)$ is differentiable in almost all settings. This differentiability leads to a sharper error bound than that of DFS.

3.2 Numerical Errors in NFP/NT

In this subsection, we discuss an error bound for Newton's method. To derive a numerical property of NFP/NT, we adopt the following assumptions to invoke a version of the Kantorovich theorem. Let $D \subset \mathbb{R}^J$ be an open convex set and $\|A\|$ denote the matrix norm for a matrix A.

Assumption 1. *Given θ, (i) $s\left(\xi;\theta\right)$ is differentiable with respect to ξ, (ii) $\nabla_\xi s\left(\xi;\theta\right)|_{\xi=\xi^0_{NT}}$ is invertible, (iii) $s\left(\xi^0_{NT};\theta\right) \neq S$, (iv) $\left\|\left(\nabla_\xi s\left(\xi;\theta\right)|_{\xi=\xi^0_{NT}}\right)^{-1}\left(\nabla_\xi s\left(\xi;\theta\right)|_{\xi=\xi'} - \nabla_\xi s\left(\xi;\theta\right)|_{\xi=\xi''}\right)\right\| \leq \kappa\left\|\xi' - \xi''\right\|$ for all $\xi',\xi'' \in D$, (v) $\rho \equiv \kappa\eta < \frac{1}{2}$ for $\eta \equiv \left\|\left(\nabla_\xi s\left(\xi;\theta\right)|_{\xi=\xi^0_{NT}}\right)^{-1} s\left(\xi^0_{NT};\theta\right)\right\|$, and (vi)*

$$\left\{\xi \in \mathbb{R}^J : \left\|\xi^1_{NT} - \xi\right\| \leq \eta\frac{1-\sqrt{1-2\rho}}{1+\sqrt{1-2\rho}}\right\} \subset D.$$

From the definition of $s\left(\xi;\theta\right)$, Assumption 1(i) is satisfied. Lee and Seo (2015, Lemma 1) show that Assumption 1(ii) holds for any θ and initial guess ξ^0_{NT}. If Assumption 1(iii) were not true, the initial guess would be the solution. The remaining conditions are technical assumptions to deliver existence of the solution and convergence of the Newton iterates.

Lemma 3. *Suppose that Assumption 1 holds. Then,*

$$\left\|\xi_{NT}\left(\theta,\epsilon_{in}\right) - \xi\left(\theta\right)\right\| \leq \frac{\kappa\left(\theta\right)}{\sqrt{1-2\rho\left(\theta\right)}}\epsilon^2_{in}. \tag{11}$$

Lemma 3 indicates that the error bound of the numerical solution $\xi_{NT}\left(\theta,\epsilon_{in}\right)$ has an order of ϵ^2_{in}. This is smaller than that of the contraction mapping iterations, $O\left(\epsilon_{in}\right)$. This can be easily checked by comparing (3) with

$$\left\|\xi^{h+1}_{NT} - \xi^h_{NT}\right\| \leq \frac{\kappa}{2\sqrt{1-2\rho}}\left\|\xi^h_{NT} - \xi^{h-1}_{NT}\right\|^2, \tag{12}$$

where $\frac{\kappa}{2\sqrt{1-2\rho}}$ is not too large, which can be deduced from Yamamoto (1986). This implies that Newton's method converges quadratically to a local solution, whereas contraction mapping iterations converge linearly.

Moreover, we can achieve a tighter inner loop error with only a few more iterations, which makes the error propagation even smaller. Let ξ_{NT}^h be the first iterate satisfying the stopping rule (5), given the inner loop tolerance, ϵ_{in}. Then, by (12), ξ_{NT}^{h+1} achieves the error less than $\frac{\kappa}{2\sqrt{1-2\rho}}\epsilon_{in}^2$. Because $\frac{\kappa}{2\sqrt{1-2\rho}}$ is not expected to be very large in most applications, one more iteration can make the error much smaller. Roughly speaking, when ξ_{NT}^h achieves the error of 10^{-6}, ξ_{NT}^{h+1} or ξ_{NT}^{h+2} will achieve the error less than 10^{-12}.

Next, consider an error bound of the estimate computed using Newton's method. We define $\widehat{\theta}_{NT}(\epsilon_{in}) \equiv \widetilde{\theta}_{NT}(\epsilon_{in}, 0)$.

Corollary 4. *Suppose that the assumptions of Theorem 1 and Assumption 1 hold. Then,*

$$\left\|\widehat{\theta}_{NT}(\epsilon_{in}) - \theta^*\right\| \leq O\left(\frac{\kappa\left(\widehat{\theta}_{NT}(\epsilon_{in})\right)}{\sqrt{1-2\rho\left(\widehat{\theta}_{NT}(\epsilon_{in})\right)}}\epsilon_{in}^2\right).$$

Corollary 4 provides a smaller upper bound of the estimate error for NFP/NT than NFP/CTR and is easily obtained when the second term of the RHS of (7) is replaced with (11).

4 Monte Carlo Simulation

In the Monte Carlo simulation, we implement three approaches: MPEC, NFP/CTR, and NFP/NT. We generate synthetic data sets under the same setting used by DFS (we do not repeat their settings here for compactness).

We ran all experiments on a computer with a CPU Intel Xeon processor E3 1270 v2, 16GB of RAM, Windows 7 64 bit, MATLAB R2012a, and TOMLAB 7.9 equipped with KNITRO 8.0. We use the codes that DFS provide in *Econometrica* after some modifications. We use the interior point algorithm (for the outer loop) with algorithm option ALG=1 in KNITRO 8.0. We provide NFP with the analytic Jacobian/gradient and Hessian. We use 10^{-6} for the outer loop (optimality) tolerance.

For NFP/NT, we use a combination of contraction mapping and Newton iterations because Newton's method does not converge globally but contraction mapping iterations do. Our approach is similar to Rust's (1987) strategy in that the inner loop begins with the contraction mapping iterations and switches to Newton's method after the contraction mapping converges under a loose tolerance. However, the way we deal with numerical instability is different from Rust (2000).[5]

[5] Our theoretical results on convergence rate of numerical errors still apply to our implemention of modified Newton's method. Recall that we assume convergence. Conditional on convergence, the last several guesses are close

Table 1: Monte Carlo Results Varying the Inner Loop Tolerance

Inner loop tolerance	Approach	Runs conv.	CPU time (seconds)	Major iter.	Func. eval.	Grad/Hess. eval.	CTR. iter.	NT. iter.
10^{-3}	NFP/CTR	14%	92.9	26	114	52	530	
	NFP/NT	7%	91.6	26	88	54	118	32
10^{-6}	NFP/CTR	16%	47.1	11	65	24	700	
	NFP/NT	80%	37.1	10	26	23	53	42
10^{-9}	NFP/CTR	60%	46.9	10	31	22	991	
	NFP/NT	100%	33.1	9	13	21	23	48
10^{-12}	NFP/CTR	99%	45.3	9	14	21	1,147	
	NFP/NT	100%	33.2	9	13	21	36	53

Notes: We generate 20 data sets for each level of inner loop tolerance with $E\left[\beta_i^0\right] = 2$, $T = 50$, $J = 25$ and $n_s = 1,000$. We use five starting points in the estimation for each data set. The reported means are over 100 runs. The analytic Jacobian/gradient and the Hessian are provided to NFP and MPEC. Major iter. is the number of iterations in the outer loop. Func. eval. is the number of function evaluations in the outer loop. Grad/Hess. eval. is the number of gradient and Hessian evaluations in the outer loop. CTR. iter. is the number of contraction mapping iterations in the inner loop. NT. iter. is the number of Newton iterations in the inner loop.

There are issues with the numerical instability of NFP/NT. For example, Newton's method does not perform well if the initial guess is not close enough to the solution, or the derivative of $s(\xi;\theta)$ with respect to ξ in (4) may be numerically ill-posed, in which case an inaccurate Newton step may be obtained. We discuss how we handle these problems in detail in Appendix B.[6]

4.1 Results of Monte Carlo Experiments

In this subsection, we run a Monte Carlo experiment using synthetic data sets to examine how the level of inner loop tolerance affects convergence and numerical accuracy. We report the results in Tables 1 and 2.

We report the rate of runs that converged for each method in Table 1. The third column in the table shows that NFP/CTR does not converge in many cases for the loose tolerances of

to the solution and by the local convergence property of Newton's method, the original Netwon steps work well on these guesses. Our modifications come into play only if the original Newton step fails to find a reasonable direction. Thus when the inner loop converges, the termination condition of the original Newton method is satisfied.

[6] As Iskhakov et al. (2015) discuss, if a researcher uses MPEC (e.g., KNITRO) instead of NFP, an effort to customize the programming code may be reduced or avoided.

10^{-3} through 10^{-9}, as DFS also show. In particular, for 10^{-3}, only 14 runs out of 100 converged. However, for 10^{-12}, the algorithm converges in almost all runs. NFP/NT converges better than NFP/CTR under the same level of inner loop tolerance, except 10^{-3}. NFP/NT converges 80% for a tolerance of 10^{-6} and 100% for tighter tolerances.

We also report the numbers of iterations in the outer loop and in the inner loop in Table 1. For both NFP/NT and NFP/CTR, as the inner loop tolerance gets looser, the numbers of iterations in the outer loops (i.e., major iterations, function evaluations, and gradient and Hessian evaluations) increase, but the numbers of iterations in the inner loop (i.e., contraction mapping iterations and Newton iterations) decrease. This implies that a large inner loop error propagates into the outer loop and thus prevents the outer loop from converging, resulting in more iterations in the outer loop, which DFS also point out. For the inner loop tolerances of 10^{-12}, NFP/NT and NFP/CTR have almost the same numbers of iterations in the outer loop, but there is a drastic difference in the numbers of iterations in the inner loop. For example, with $\epsilon_{in} = 10^{-12}$, both NFP/NT and NFP/CTR have, on average, 9 major iterations in the outer loop, but NFP/NT has 89 iterations (36 contraction mapping iterations and 53 Newton's iterations) in the inner loop while NFP/CTR has $1,147$ contraction mapping iterations. The small number of iterations in the inner loop of NFP/NT is the main driver for speeding up the estimation procedure as in Iskhakov et al. (2015). Furthermore, as we tighten the inner loop tolerance from 10^{-9} to 10^{-12}, the total number of iterations in the inner loop until the outer loop converges increases by 18 (13 contraction mapping iterations and 5 Newton's iterations) for NFP/NT, while the number of contraction mapping iterations increases by 156 for NFP/CTR. Because the objective function is evaluated 13 times until convergence, NFP/NT needs only one or two more inner loop iterations for each objective function evaluation to achieve a tighter inner loop tolerance 10^{-12} from 10^{-9}. This is consistent with the quadratic convergence rate of Newton's method.

Next, we measure the numerical accuracy of NFP. For this, we need the theoretically exact estimate in the finite sample. Because the numerical error-free solution of the BLP GMM problem is not available, we assume that the MPEC estimates are the exact estimates. The optimality and feasibility tolerances for MPEC are set to a very tight level, 10^{-10}, to obtain an estimate as accurate as possible. DFS use 10^{-6} in their experiments. To simplify an order of magnitude of discrepancy, we take the logarithm (base 10) of the absolute deviations of the NFP estimates from the MPEC estimates.

Table 2 reports the mean and maximum of the log absolute deviations for each level of inner loop tolerance. Both the NFP and the MPEC estimates are the minimizers among the local minima for each of 20 data sets.[7] We find that for all data sets with convergence reported, the

[7] There is one case in which NFP and MPEC converge to different local minimizers, which we report as not

Table 2: Log deviation of NFP SD estimates from MPEC estimates

Inner loop tolerance	Approach	Datasets conv.	True values:	Log dev. of estimates from MPEC 0.7071	0.7071	0.7071	0.7071	0.4472	All
10^{-3}	NFP/CTR	30%	Avg	-9.34	-7.53	-7.40	-7.86	-7.35	-7.90
			Max	-5.94	-6.76	-7.09	-6.96	-6.97	-5.94
	NFP/NT	20%	Avg	-13.24	-8.16	-7.56	-7.89	-7.34	-8.84
			Max	-6.47	-7.38	-7.26	-7.55	-7.04	-6.47
10^{-6}	NFP/CTR	40%	Avg	-11.29	-8.22	-8.09	-8.62	-7.93	-8.83
			Max	-7.06	-7.39	-7.58	-7.88	-7.57	-7.06
	NFP/NT	100%	Avg	-8.46	-9.05	-8.75	-9.01	-8.57	-8.77
			Max	-6.18	-7.54	-7.96	-7.54	-7.34	-6.18
10^{-9}	NFP/CTR	85%	Avg	-8.85	-8.72	-8.62	-9.13	-8.79	-8.82
			Max	-6.92	-7.87	-8.12	-8.10	-8.09	-6.92
	NFP/NT	100%	Avg	-9.28	-9.39	-9.14	-9.72	-9.01	-9.31
			Max	-6.47	-7.87	-8.19	-7.81	-8.37	-6.47
10^{-12}	NFP/CTR	100%	Avg	-9.37	-9.63	-9.27	-9.81	-9.08	-9.43
			Max	-7.92	-8.45	-8.60	-8.43	-8.38	-7.92
	NFP/NT	100%	Avg	-9.55	-9.79	-9.32	-9.87	-9.10	-9.52
			Max	-7.20	-8.48	-8.52	-8.70	-8.37	-7.20

Notes: For each tolerance level, there are 20 data sets as described in Table 1. The local minimum with the lowest objective value is determined to be the global minimum for each data set. Absolute deviations between the estimates of NFP and MPEC are calculated, and the log with base 10 of the deviation is shown.

absolute deviations between the estimates of NFP/CTR and MPEC are, on average, less than the optimality tolerance, 10^{-6}. Furthermore, the maximum of the log deviations ranges from -5.94 to -7.92. Even for $\epsilon_{in} = 10^{-3}$, the maximum log deviation is -5.94, which is small. According to Theorem 1, the upper bound of the deviation of NFP/CTR has an order of 10^{-3} when $\epsilon_{in} = 10^{-3}$. We conjecture that there might exist a sharper upper bound than what Theorem 1 suggests or that actual deviations do not always achieve the upper bound.

It is worth comparing Table 2 with Tables I and II in DFS. First, the purpose of the latter tables is not to illustrate their error bound derived in their Theorem 3, but instead to illustrate consequences of failure to find local optima of the objective function due to a loose inner loop tolerance. Our purpose of the table is to illustrate the error bound of the estimate on convergence. Second, the numbers in the tables may appear inconsistent, in particular for NFP/CTR with a loose inner loop tolerance. For example, the fraction of convergence in our table is 30% for the tolerance of 10^{-3}, while that in Tables I and II in DFS is 0. The maximum absolute deviation for a tolerance of 10^{-3} is small in our table, whereas the difference (in the mean own-price elasticities) between loose (10^{-4}) and tight (10^{-14}) inner loop tolerances is large in their tables. One reason for these differences is that we supply the analytic gradient of the objective function in our experiment, and DFS do not in Table I. Because a numerical gradient generates an additional error, it explains the difference in the fractions of convergence. However, in Table II, DFS apply analytic derivatives but do not observe convergence. DFS use the pseudo-real cereal data set from Nevo (2000), which may be harder to optimize with a loose inner loop tolerance than the Monte Carlo data set that we use. Another reason for the difference is that we use the global minimum (i.e., the minimum among the local minima from the five starting points) for each data set, and they use all results from all the starting points, including the ones that did not converge.

As we discuss in Section 3, our analysis of the error bound in Corollary 2 assumes convergence. Tables 1 and 2 in DFS show that a loose inner loop tolerance of NFP/CTR sacrifices convergence. However, when convergence is achieved under a tight outer loop tolerance such as 10^{-6}, the NFP/CTR deviations from MPEC are very small even at the loose inner loop tolerances. Thus, it seems that the level of inner loop tolerance significantly affects convergence rather than accuracy of the estimate. DFS do not investigate the combination of a tight outer loop tolerance, a loose inner loop tolerance, and the optimization routine reporting convergence simply because this never happens in their examples in Tables I and II.

Corollary 4 predicts that NFP/NT produces smaller estimate errors, but Table 2 shows that the estimate errors of NFP/NT seem close to those of NFP/CTR. A potential reason may be related to an outer loop error because an outer loop error becomes dominant as an inner loop error gets

convergent.

smaller. Thus, the different inner loop errors seem to make no noticeable difference between the estimates of NFP/NT and NFP/CTR.

5 Conclusion

In this article, we study the numerical performance of NFP/CTR and NFP/NT in BLP's random coefficients logit demand model. We show the theoretical result that the upper bound on the numerical error in the estimate of NFP/CTR is sharper than that proposed by DFS under our assumptions. In an experiment with synthetic data, we demonstrate that NFP/CTR achieves reasonably accurate estimates on convergence even with a loose inner loop tolerance if a tight outer loop tolerance is applied. This implies that the error from the inner loop does not severely affect numerical accuracy of the estimate. However, a loose inner loop tolerance sacrifices convergence of NFP/CTR. As an alternative, we argue that NFP/NT can achieve a very tight inner loop tolerance rather easily. Therefore, NFP/NT performs well in numerical aspects such as convergence and accuracy.

A Appendix: Proofs

Proof of Theorem 1: For simple notations, let $\widetilde{\theta} = \widetilde{\theta}(\epsilon_{in}, \epsilon_{out})$ and write $\frac{\partial \Gamma(\xi,\theta)}{\partial \xi'}|_{\left(\xi\left(\widetilde{\theta}\right),\widetilde{\theta}\right)}$ instead of $\frac{\partial \Gamma(\xi,\theta)}{\partial \xi'}|_{(\xi,\theta)=\left(\xi\left(\widetilde{\theta}\right),\widetilde{\theta}\right)}$, and similarly for other derivatives. Let $\theta = (\theta_1, ..., \theta_K)'$ and, for $k = 1, ..., K$, $\Gamma_k(\xi, \theta)$ be the k-th component of $\Gamma(\xi, \theta)$.

For any $k = 1, ..., K$, compute

$$
\begin{aligned}
& \Gamma_k\left(\xi\left(\widetilde{\theta}, \epsilon_{in}\right), \widetilde{\theta}\right) \\
= \; & \Gamma_k\left(\xi\left(\widetilde{\theta}, \epsilon_{in}\right), \widetilde{\theta}\right) - \Gamma_k\left(\xi\left(\widetilde{\theta}\right), \widetilde{\theta}\right) + \Gamma_k\left(\xi\left(\widetilde{\theta}\right), \widetilde{\theta}\right) - \Gamma_k\left(\xi\left(\theta^*\right), \theta^*\right) \\
= \; & \Gamma_k\left(\xi\left(\widetilde{\theta}, \epsilon_{in}\right), \widetilde{\theta}\right) - \Gamma_k\left(\xi\left(\widetilde{\theta}\right), \widetilde{\theta}\right) + \frac{dQ\left(\xi\left(\theta\right)\right)}{d\theta_k}|_{\theta=\widetilde{\theta}} - \frac{dQ\left(\xi\left(\theta\right)\right)}{d\theta_k}|_{\theta=\theta^*} \\
= \; & \left[\frac{\partial \Gamma_k(\xi,\theta)}{\partial \xi}|_{(\xi,\theta)=\left(\xi\left(\widetilde{\theta}\right),\widetilde{\theta}\right)}\right]' \left[\xi\left(\widetilde{\theta}, \epsilon_{in}\right) - \xi\left(\widetilde{\theta}\right)\right] + O\left(\left\|\xi\left(\widetilde{\theta}, \epsilon_{in}\right) - \xi\left(\widetilde{\theta}\right)\right\|^2\right) \\
& + \left[\frac{d^2 Q\left(\xi\left(\theta\right)\right)}{d\theta d\theta_k}|_{\theta=\theta^*}\right]' \left[\widetilde{\theta} - \theta^*\right] + O\left(\left\|\widetilde{\theta} - \theta^*\right\|^2\right).
\end{aligned}
$$

The first and second equalities follow because $\Gamma_k\left(\xi\left(\theta^*\right), \theta^*\right) = \frac{dQ(\xi(\theta))}{d\theta_k}|_{\theta=\theta^*} = 0$. Stacking for all

$j = 1, ..., J$ and rearranging the terms, we obtain

$$\begin{aligned}
& \left[\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}\right]\left[\widetilde{\theta}-\theta^*\right]+O\left(\left\|\widetilde{\theta}-\theta^*\right\|^2\right)1_K \\
= \; & \Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right)-\left[\frac{\partial\Gamma\left(\xi,\theta\right)}{\partial\xi'}|_{(\xi,\theta)=(\xi(\widetilde{\theta}),\widetilde{\theta})}\right]\left[\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right] \\
& +O\left(\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\|^2\right)1_K,
\end{aligned}$$

where 1_K is the K-dimensional column vector of 1's. Since $\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}$ is nonsingular,

$$\begin{aligned}
& \left(\widetilde{\theta}-\theta^*\right)+O\left(\left\|\widetilde{\theta}-\theta^*\right\|^2\right)1_K \\
= \; & \left[\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}\right]^{-1}\Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right) \\
& -\left[\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}\right]^{-1}\left[\frac{\partial\Gamma\left(\xi,\theta\right)}{\partial\xi'}|_{(\xi,\theta)=(\xi(\widetilde{\theta}),\widetilde{\theta})}\right]\left[\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right] \\
& +O\left(\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\|^2\right)1_K.
\end{aligned}$$

Then, since $\frac{\partial\Gamma(\xi,\theta)}{\partial\xi'}|_{(\xi,\theta)=(\xi(\widetilde{\theta}),\widetilde{\theta})}$ is bounded,

$$\begin{aligned}
& \left\|\widetilde{\theta}-\theta^*\right\| \\
\leq \; & \left\|\left[\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}\right]^{-1}\right\|\left\|\Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right)\right\| \\
& +\left\|\left[\nabla_{\theta\theta'} Q\left(\xi\left(\theta\right)\right)|_{\theta=\theta^*}\right]^{-1}\right\|\left\|\frac{\partial\Gamma\left(\xi,\theta\right)}{\partial\xi'}|_{(\xi,\theta)=(\xi(\widetilde{\theta}),\widetilde{\theta})}\right\|\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\| \\
& +O\left(\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\|^2\right) \\
\leq \; & O\left(\left\|\Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right)\right\|\right)+O\left(\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\|\right).
\end{aligned}$$

Here, $O\left(\left\|\widetilde{\theta}-\theta^*\right\|^2\right)$ and $O\left(\left\|\xi\left(\widetilde{\theta},\epsilon_{in}\right)-\xi\left(\widetilde{\theta}\right)\right\|^2\right)$ are ignored because there are lower order terms. □

Proof of Corollaries 2 and 4: Let $\epsilon_{out} = 0$ in Theorem 1. Then,

$$\Gamma\left(\xi\left(\widetilde{\theta},\epsilon_{in}\right),\widetilde{\theta}\right)=0.$$

Also by (8) and (11), the required properties are derived. □

Proof of Lemma 3: We apply Yamamoto (1986, Lemma 2.4 and Corollary 3.1.1). The prerequisites of the lemma and the corollary are satisfied by Assumption 1. Thus, we deduce

$$\begin{aligned}\left\|\xi_{NT}^{h}-\xi\left(\theta\right)\right\| &\leq 2\left\|\xi_{NT}^{h+1}-\xi_{NT}^{h}\right\| \text{ and}\\ \left\|\xi_{NT}^{h+1}-\xi_{NT}^{h}\right\| &\leq \frac{\kappa}{2\sqrt{1-2\rho}}\left\|\xi_{NT}^{h}-\xi_{NT}^{h-1}\right\|^{2}.\end{aligned}$$

Let h be the first that satisfies (5). Then,

$$\begin{aligned}\left\|\xi_{NT}\left(\theta,\epsilon_{in}\right)-\xi\left(\theta\right)\right\| &\leq 2\left\|\xi_{NT}^{h+1}-\xi_{NT}^{h}\right\| \leq \frac{\kappa}{\sqrt{1-2\rho}}\left\|\xi_{NT}^{h}-\xi_{NT}^{h-1}\right\|^{2}\\ &\leq \frac{\kappa}{\sqrt{1-2\rho}}\epsilon_{in}^{2}.\end{aligned}$$

Observe that $1-2\rho>0$ and $\frac{\kappa}{\sqrt{1-2\rho}}$ is bounded (in fact, a constant) when ϵ_{in} changes. This completes the proof. □

B Appendix: Additional Techniques for Newton's Method

When we apply Newton's method for the inner fixed-point problems, computations may be unstable under some parameter values. As is known, numerically ill-posed equations may produce inaccurate solutions. We explain the techniques that we use to alleviate the numerical instability.

We introduce some notations. We omit the market index t assuming that there is only one market. We concentrate out the linear parameter as Nevo (2000) suggests. Then, the equation we need to solve in the inner loop is

$$S_j=\frac{1}{n_s}\sum_{i=1}^{n_s}\frac{\exp\left(\delta_j\right)\cdot\exp\left(\sum_{k=1}^{K}x_j^k\sigma_k v_i^k\right)}{1+\sum_{j'=1}^{J}\exp\left(\delta_{j'}\right)\cdot\exp\left(\sum_{k=1}^{K}x_{j'}^k\sigma_k v_i^k\right)}\equiv s_j\left(\delta;\sigma\right),\text{ for } j=1,2,\cdots,J.$$

Here, S_j and $s_j\left(\delta;\sigma\right)$ are the observed market share and the predicted market share function of product j, respectively. The variable $X_{jt}=\left(x_{jt}^1,...,x_{jt}^K\right)$ is a K-dimensional vector containing the product characteristics, and $\beta_i=(\beta_{i1},...,\beta_{iK})$ is assumed to be of the form $\beta_{ik}=\overline{\beta}_k+\sigma_k v_{ik}$ with $v_{ik}\sim N\left(0,1\right)$. Only $\sigma=(\sigma_1,...,\sigma_K)$ is to be estimated in the concentrated GMM problem, and $\overline{\beta}_k$ will be estimated later. For each product j, $X_j\beta_i$ can be expressed by the sum of its mean utility $\delta_j=\sum_{k=1}^{K}x_j^k\overline{\beta}_k$ and the individual i's utility $\sum_{k=1}^{K}x_j^k\sigma_k v_i^k$ deviated from the mean utility. For each given σ, we need to solve for $\delta=(\delta_1,...,\delta_J)$.

1. Adjust the scale of the unknowns to be solved: Instead of solving for δ directly, we solve for $\omega=(\omega_1,...,\omega_J)$ where $\omega_j=\frac{\exp(\delta_j)}{\alpha_j},\ j=1,2,\cdots,J$ given some constant vector $\alpha=(\alpha_1,\cdots,\alpha_J)$.

The equation is then,

$$S_j = \frac{1}{n_s}\sum_{i=1}^{n_s} \frac{\omega_j \cdot \exp\left(\sum_{k=1}^{K} x_j^k \sigma_k v_i^k\right) \cdot \alpha_j}{1 + \sum_{j'=1}^{J} \omega_{j'} \cdot \exp\left(\sum_{k=1}^{K} x_{j'}^k \sigma_k v_i^k\right) \cdot \alpha_{j'}} \equiv \widetilde{s}_j(\omega;\sigma), \qquad j = 1, 2, \cdots, J.$$

After solving for ω, it is easy to obtain δ. In our simulation, we choose $\alpha_j = S_j$ for each $j = 1, ..., J$. Then, even for a very small value of S_j, such as around 10^{-9}, ω_j is not very small and we expect ω_j and $\omega_{j'}$ to be of a similar magnitude, unlike $\exp(\delta_j)$ and $\exp(\delta_{j'})$.

2. Adjust the scale of the matrix $\nabla_\omega \widetilde{s}(\cdot)$ if its inversion is ill-posed: The Newton step $\Delta\omega$ is computed as follows:

$$\Delta\omega = \left[\nabla_\omega \widetilde{s}(\cdot)\right]^{-1}\left[S - \widetilde{s}(\cdot)\right],$$

where $S = (S_1, \cdots, S_J)$, $\widetilde{s}(\cdot) = (\widetilde{s}_1(\cdot), \cdots, \widetilde{s}_J(\cdot))$ and $\nabla_\omega \widetilde{s}(\cdot)$ is the matrix of the first partial derivatives of $\widetilde{s}(\cdot)$ with respect to ω. If elements of $\nabla_\omega \widetilde{s}(\cdot)$ are badly scaled, the computation may be very inaccurate. To handle the bad scale problem, we use the identity

$$\left[\nabla_\omega \widetilde{s}(\cdot)\right]^{-1}\left[S - \widetilde{s}(\cdot)\right] = B\left[A\nabla_\omega \widetilde{s}(\cdot)\, B\right]^{-1} A\left[S - \widetilde{s}(\cdot)\right]$$

for any invertible $(J \times J)$ matrices A and B. We set A to be the identity matrix and B a diagonal matrix whose (j, j) element is $\left(\partial \widetilde{s}_j(\cdot)/\partial\omega_j\right)^{-1}$. The reason for adjusting the scale of the diagonal elements is that the matrix $\nabla_\omega \widetilde{s}(\cdot)$ is diagonally dominant (Lee and Seo 2015, Lemma 1). We check the condition number of $\nabla_\omega \widetilde{s}(\cdot)$ to decide whether to apply the above identity. Roughly speaking, if the condition number is 10^m, we may lose up to m digits of accuracy. In our simulation, we apply the above identity if the condition number is greater than 10^{15}.

3. Use the contraction iteration, if computing $B\left[\nabla_\omega \widetilde{s}(\cdot)\, B\right]^{-1}\left[S - \widetilde{s}(\cdot)\right]$ is still numerically ill-posed: Adjusting the scale of the matrix $\nabla_\omega \widetilde{s}(\cdot)$ does help obtain numerical stability but does not guarantee it. Thus, if the condition number of $\nabla_\omega \widetilde{s}(\cdot)\, B$ is still greater than 10^{15}, the computation may be still unreliable. In this case, we iterate contraction mappings 10 times and then try the Newton step again.

4. Adopt a line search: Even if the condition number of $\nabla_\omega \widetilde{s}(\cdot)$ is smaller than 10^{15}, the Newton step may not produce descent improvement if the current iterate is not close enough to the solution. This is a well-known problem of the original Newton iteration (see for example Nocedal and Wright 2006). A line search is one of the popular choices to solve this problem. Even though there may be more efficient line search algorithm in the literature, we believe pursuing efficiency of a line search is beyond this paper. Instead, we do the following simple line search: If the next Newton iterate $\omega^{\tau+1} = \omega^\tau + \Delta\omega \in \mathbb{R}^J$ contains 0 or a negative element, we believe that the Newton step is too large to be stable because the solution vector ω satisfies $\omega_j > 0$ for all $j = 1, ..., J$. If so, we take

the next Newton step as follows:

$$\omega^\tau + a\Delta\omega$$

for some $a > 0$. We set

$$a = \tfrac{1}{2} / \max_j \frac{|\Delta\omega_j|}{\left|\omega_j^\tau\right|}$$

so that each component of $\omega^{a,\tau+1} \equiv \omega^\tau + a\Delta\omega$ is positive. But, since there is no guarantee that this process converges, we compute the contraction iterate, $\omega_j^{CTR,\tau+1} \equiv \omega_j^\tau S_j / \widetilde{s}_j\left(\omega^\tau; \sigma\right)$ as well. Then, between these two iterates, we choose the one producing a smaller change of ω: If $\left\|\omega^{a,\tau+1} - \omega^\tau\right\| < \left\|\omega^{CTR,\tau+1} - \omega^\tau\right\|$, $\omega^{a,\tau+1}$ is chosen as the next iterate. Otherwise, $\omega^{CTR,\tau+1}$ is chosen and 10 contraction iterations are applied since the Newton step seems unstable at the current iterate.

5. Initial guess for ω in the next outer loop iteration: The value of a new parameter candidate is not expected to move very far from the current value in the outer loop, and thus it is common to use the current solution ω_σ for the current parameter σ as the initial guess for the market share equations in the next outer loop iteration. However, the solution ω_σ may not be accurately found under some parameter values. (For example, it is not possible to compute the predicted market share accurately under large parameter values such as standard deviation parameter values being 100.) Then, in the next iteration of the outer loop, it is also likely that the solution $\omega_{\sigma_{new}}$, given a new parameter value σ_{new}, is hard to compute. Therefore, the objective function may produce different values even under the same parameter values because the inaccurate solution could depend on the initial guess for the market share equation. Then, unstable objective function values may make the optimization routine move to a wrong direction and, even worse, oscillate. To avoid this problem, if the inner loop does not converge at the current outer loop iteration, we set the initial guess for the next outer loop iteration to be a predetermined value, not the inaccurate solution of the current outer loop iteration. This way, the optimization routine may get out of a bad parameter value area quickly.

References

[1] Ackerberg, D., J. Geweke, and J. Hahn (2009): "Comments on "Convergence Properties of the Likelihood of Computed Dynamic Models"," *Econometrica*, 77(6), 2009-2017.

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

[3] Dubé, J. P., J. T. Fox, and C. L. Su (2012): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231-2267.

[4] Houde, J. (2012): "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, 102(5), 2147-2182.

[5] Iskhakov, F., J. Lee, J. Rust, B. Schjerning, and K. Seo (2015): "Constrained Optimization Approaches to Estimation of Structural Models: Comment," *Econometrica*, forthcoming.

[6] Kim, K. and M. Park (2010): "Approximation error in the nested fixed point algorithm for BLP model estimation," Working paper.

[7] Lee, J. and K. Seo (2015): "A Computationally Fast Estimator for Random Coefficients Logit Demand Models Using Aggregate Data," *RAND Journal of Economics*, 46(1), 86-102.

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

[9] Newey, W. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," In Engle, R. and D. McFadden, eds., *The Handbook of Econometrics.* North Holland, 1994.

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

[11] Patel, K. (2012): "How to Estimate Random Coefficients Demand Models," Essays in Industrial Organization, Doctoral dissertation, Northwestern University.

[12] Reynaerts, J., R. Varadhan, and J. Nash (2012): "Enhancing the Convergence Properties of the BLP (1995) Contraction Mapping," VIVES Discussion Paper 35.

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

[14] ——— (2000): "Nested Fixed Point Algorithm Documentation Manual: Version 6," Department of Economics, Yale University.

[15] Yamamoto, T. (1986): "A Method for Finding Sharp Error Bounds for Newton's Method Under the Kantorovich Assumptions," *Numerische Mathematik*, 49, 203-220.