Approximation error in the nested fixed point algorithm for BLP model estimation

Kyoo il Kim and Minjung Park*
University of Minnesota
February 2009, Updated April 2010

Abstract

In this note we show that an approximation error in the inner loop of the nested fixed point algorithm for BLP model estimation affects the outer loop GMM estimation with the same order of magnitude. Therefore the approximation error does not propagate into the GMM estimation.

Keywords: Approximation error, BLP, Nested fixed point algorithm

1 Introduction

We consider a standard demand model of differentiated products with random coefficients following Berry, Levinsohn, and Pakes (1995, 2004, BLP hereafter), Nevo (2001), and Goolsbee and Petrin (2004). The market share equation for a product $j$ among $J$ alternative products (excluding one outside good) at market $t$ with the random coefficient logit model is (e.g.)

$$s_j(x_t, p_t, \xi_t; \theta) = \int_{\beta} \frac{\exp(\beta_0 + x_{jt}' \beta_x - \beta_p p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta_0 + x_{kt}' \beta_x - \beta_p p_{kt} + \xi_{kt})} dF(\beta; \theta)$$

where $x_t = (x_{1,t}', ..., x_{J,t}')'$ is a vector of observable characteristics of products, $p_t = (p_{1,t}, ..., p_{J,t})'$ is a vector of prices, $\xi_t = (\xi_{1,t}, ..., \xi_{J,t})'$ is a vector of market and product specific demand shocks (unobservable to econometricians), $\beta = (\beta_0, \beta_x', \beta_p)'$ are random coefficients distributed as the cumulative distribution function $F(\beta; \theta)$, and $\theta$ is the parameter that determines the distribution of random coefficients (means and standard deviations). We assume there are $T$ independent markets (e.g., across different regions or times).

In BLP, the parameter $\theta$ is of our interest. The estimation proceeds in two steps (or iterative steps). We first compute $\xi_t(\theta)$ using the contraction mapping for each given value of $\theta$ as proposed in BLP. Let $S_t$ be the vector of observed market shares and $s^{-1}(S_t, \theta)$ be the inverse mapping of the

*Correspondence: kyookim@umn.edu, mpark@umn.edu, Address: Department of Economics, University of Minnesota, 4-101 Hanson Hall, 1925 Fourth Street South, Minneapolis, MN 55455
market share equation (1). Then, we can write
\[
\xi_t(\theta) = s^{-1}(S_t, \theta) \tag{2}
\]
and with an abuse of notation, we further write
\[
\xi_{jt}(\theta) = s^{-1}_j(S_t, \theta) \text{ for } j = 1, \ldots, J. \tag{3}
\]

Then we estimate \( \theta \) using GMM based on the moment function\(^2\)
\[
Q(\xi(\theta)) = \hat{m}_j(\xi(\theta))'W\hat{\bar{m}}_j(\xi(\theta)) \tag{4}
\]
where \( \bar{z}_{jt} \)'s are valid instruments including demand shifters and cost shifters to deal with the endogeneity of prices that are possibly correlated with the unobserved demand shocks.

Define \( \hat{m}_j(\xi(\theta)) = \frac{1}{T} \sum_{j=1}^{J} E[m_j(\xi_{jt}(\theta))] \).\(^3\) Further let \( \xi(\theta) = (\xi'_1(\theta), \ldots, \xi'_T(\theta))' \) and \( z_t = (z'_{1t}, \ldots, z'_{Jt})' \). Then we write the population criterion function for the GMM estimation with a weight matrix \( W \) as\(^4\)
\[
Q(\xi(\theta)) = \hat{m}_j(\xi(\theta))'W\hat{\bar{m}}_j(\xi(\theta)) \tag{5}
\]
so that it can accommodate the case that \( J \rightarrow \infty \), which will be more akin to the asymptotics in BLP and Berry, Linton, and Pakes (2004).

We also write the sample analogue of the population criterion function for the estimation as
\[
\tilde{Q}_{JT}(\xi(\theta)) = \hat{m}_j(\xi(\theta))'\tilde{W}_{JT}\hat{\bar{m}}_j(\xi(\theta)) \tag{6}
\]
where \( \hat{m}_j(\xi(\theta)) = \frac{1}{T} \sum_{j=1}^{J} \sum_{t=1}^{T} m_j(\xi_{jt}(\theta)) \) and \( \tilde{W}_{JT} \rightarrow_p W \).

Recall that in BLP the GMM estimation using the sample criterion function (5) requires the inner loop step to find the fixed point of the inverse mapping in (2) for each trial value of \( \theta \). From this we call the GMM estimator in BLP the nested fixed point (NFP) estimator. In particular, we obtain \( \xi_t(\theta) \) by iterating the contraction mapping
\[
\xi_t^{H+1} = \xi_t^H + \varkappa (\log S_t - \log s^H(\xi^H_t, \theta)) \text{ for } t = 1, \ldots, T \tag{7}
\]
where \( \varkappa \in (0, 1) \) is a tuning parameter and \( H \) denotes the number of iterations. Note that in practice, the convergence in (6) cannot be exact. In other words, we stop the iteration when \( \xi_t^{H+1} \) is sufficiently

\(^1\)So we are abstract from the sampling error of the observed market shares and the simulation error if one approximates the integral in (1) using a simulation method.

\(^2\)One can view this as a set of unconditional moment conditions implied by the conditional moment condition
\[
E[\xi_{jt}(\theta_0)|Z_{jt}] = 0 \text{ and } z_{jt} \text{ consists of functions of instruments } Z_{jt}.
\]

\(^3\)Therefore we assume \( \xi_{jt} \) and \( \bar{z}_{jt} \) are identically distributed across \( t = 1, \ldots, T \) but allow them to be non-identically distributed across \( j = 1, \ldots, J \). This can be relaxed further but it will not change the key results of this paper.

\(^4\)In our asymptotics we fix \( J \) and let \( T \rightarrow \infty \), we can also use \( Q(\xi(\theta)) = \hat{m}(\xi(\theta))'W\hat{\bar{m}}(\xi(\theta)) \) with an appropriate weight matrix \( W \) where \( \hat{m}(\xi(\theta)) = E[m_1(\xi_{1t}(\theta)), \ldots, m_J(\xi_{Jt}(\theta))]' \) and \( \hat{\bar{m}}_T(\xi(\theta)) = \hat{m}_T(\xi(\theta))'\tilde{W}_T\hat{\bar{m}}_T(\xi(\theta)) \) where \( \hat{m}_T = \frac{1}{T} \sum_{t=1}^{T} m_1(\xi_{1t}(\theta)), \ldots, m_J(\xi_{Jt}(\theta))' \) and \( \tilde{W}_T \rightarrow_p W \) to gain the efficiency of the GMM estimator. We have obtained the same conclusion in this framework too.
close to $\xi_H$, allowed by a tolerance level. To be precise, we impose a stopping rule that the iterative procedure stops if
\[
\{||\xi_H^t - \xi_H^{t+1}||^2 / J\}^{1/2} \leq \epsilon
\]
is satisfied where we may arbitrarily pick $\epsilon$ as (e.g.) $10^{-6}$ or $10^{-12}$.

Naturally researchers often use a lenient tolerance level if the NFP contraction mapping is slow. It is a plausible question to ask then how the approximation error due to this tolerance level may affect the outer loop GMM estimation. Clearly the inner loop approximation error would have some effects on the GMM estimator. We are more interested in the order of the effect.

As a closely related problem, Fernández-Villaverde, Rubio-Ramírez, and Santos (2006, FRS) consider a likelihood based estimation of economic models that are solved via approximation methods (e.g., first order or second order approximation of an optimal policy function). Their question is “What are the effects on statistical inference of using an approximated likelihood instead of the exact likelihood?” Their main conclusion is “second order approximation errors in the policy function have first order effects on the likelihood function.” In a recent paper, Ackerberg, Geweke, and Hahn (2009) show that the impact of approximation error is not as large as one might conclude from FRS. In the pseudo maximum likelihood estimation framework, they establish that the effects of approximation error on classical inference of the parameter are of the same order as the approximation error itself.

In this note we pose a similar question and show that an approximation error in the inner loop of the NFP algorithm for BLP model estimation creates the same magnitude of bias in the outer loop estimation. Our finding is in line with Ackerberg, Geweke, and Hahn (2009). This means that the inner loop approximation error is neither accumulated nor magnified in the outer loop estimation. We also derive a condition under which the approximation error does not contribute to the asymptotic distribution of the GMM estimator. We believe these results would be of great interest to many applied IO researchers since estimation of the BLP model via the NFP algorithm is very commonly used in the IO literature and an approximation error is inevitable in the procedure.

2 The effects of inner loop approximation errors on the outer loop

We note that the outer loop GMM estimation is based on the approximated population criterion function instead of (4),
\[
Q(\xi^H(\theta)) = \bar{m}_J(\xi^H(\theta))'W\bar{m}_J(\xi^H(\theta)).
\]
Then we introduce the following notation:
\[
\tilde{\theta}^H = \argmin_{\theta \in \Theta} Q_{JT}(\xi^H(\theta)) = \argmin_{\theta \in \Theta} \hat{m}_{JT}(\xi^H(\theta))'\tilde{W}_{JT}\hat{m}_{JT}(\xi^H(\theta)),
\]
\[
\theta^H = \argmin_{\theta \in \Theta} Q(\xi^H(\theta)),
\]
\[
(7) \quad \theta_0 = \argmin_{\theta \in \Theta} Q(\xi(\theta)).
\]

\footnote{A recent working paper by Dubé, Fox, and Su (2008) poses a similar question. But their focus and approach are different from ours.}
\( \hat{\theta}^H \) is the GMM estimator that a researcher will obtain in estimation of the BLP model via the NFP algorithm, \( \theta^H \) is the pseudo true value assuming the contraction mapping is stopped at the iteration \( H = H(\epsilon) \) induced by the tolerance level \( \epsilon \), and \( \theta_0 \) is the true parameter value without approximation error (i.e., the inverse mapping is exact). We let

\[
\xi_t(\theta) = \xi_t^{H(0)}(\theta)
\]

when there is no approximation error in the contraction mapping. Let \( \nabla_{\theta} \) and \( \nabla_{\theta_0} \) denote the first and the second order derivative with respect to \( \theta \). Without loss of generality, we define the approximation error as

\[
(8) \quad \max_{t=0,1,2} \left\{ \mathbb{E}[\sup_{\theta \in \Theta} \| (\nabla_{\theta})^T \xi_t^{H(\epsilon)}(\theta) - (\nabla_{\theta})^T \xi_t(\theta) \|^2 / J] \right\}^{1/2} \leq \eta(\epsilon)
\]

and call \( \eta(\epsilon) \) the approximation error in the inner loop.

In our study to focus on the approximation error of the inner loop fixed point algorithm we will be abstract from the sampling error of the observed market shares (i.e., we treat the observed shares as the true shares \(^6\) and also be abstract from the simulation error if one uses simulation draws to approximate the integral (with respect to the random coefficients) in (1). Therefore our analysis will be based on a standard GMM asymptotics rather than the asymptotics of Berry, Linton, and Pakes (2004) that address these two issues, sampling error and simulation error. On the other hand Berry, Linton, and Pakes (2004) do not consider the approximation error of the NFP algorithm. Therefore our study complements Berry, Linton, and Pakes (2004). One may combine their framework with ours to fully address these three sources of errors in the asymptotics.

We start with imposing standard regularity conditions for the consistency and the asymptotic normality of the GMM estimator without the approximation error (e.g., Theorem 2.6 and Theorem 3.4 in Newey and McFadden (1994)). Since we have the specific form of the moment function as (3), we impose the regularity conditions on \( \xi_t(\theta) \).

**Condition 1** (Consistency). Suppose that \( \{ (z_{j,t}, \xi_{j,t}) \}_{j=1,...,J} \) are i.i.d across \( t \), \( \hat{W}_{JT} \rightarrow_p W \), and (i) \( W \) is positive semi-definite and \( W \hat{m}_J(\xi(\theta)) = 0 \) only if \( \theta = \theta_0 \); (ii) \( \theta_0 \in \Theta \), which is compact; (iii) \( \xi_t(\theta) \) is continuous at each \( \theta \in \Theta \) with probability one; (iv) \( \mathbb{E}[\| z_t \|] / \sqrt{J} < \infty \) and \( \mathbb{E}[\sup_{\theta \in \Theta} \| \xi_t(\theta) \| / \sqrt{J}] < \infty \).

Under Condition 1 we have \( \hat{\theta}^{H(0)} \rightarrow_p \theta_0 \) where \( \hat{\theta}^{H(0)} \) denotes the GMM estimator without the approximation error in the inner loop.

**Condition 2** (Asymptotic normality). (i) \( \theta_0 \) is in the interior of \( \Theta \); (ii) \( \xi_t(\theta) \) is continuously differentiable in a neighborhood \( \Theta_0 \) of \( \theta_0 \) with probability approaching one; (iii) \( \hat{m}_J(\xi(\theta_0)) = 0 \), \( \mathbb{E}[\| z_t \|^2 / J] < \infty \), and \( \mathbb{E}[\| \xi_t(\theta_0) \|^2 / J] < \infty \); (iv) \( \mathbb{E}[\sup_{\theta \in \Theta_0} \| \nabla_{\theta} \xi_t(\theta) \|^2 / J] < \infty \); (v) \( \nabla_{\theta} \hat{m}_J(\xi(\theta_0))^TW \nabla_{\theta} \hat{m}_J(\xi(\theta_0)) \) is nonsingular.

\(^6\)The sampling error does not affect the asymptotics results as long as the number of consumers that generate the shares is large enough (see BLP).
Under Condition 2 we have the $\sqrt{JT}$-asymptotic normality of $\hat{\theta}^{H(0)}$.

Now we consider the consistency of the GMM estimator with the approximation error. Regarding the consistency of the estimator, what matters is the distance from the estimator $\hat{\theta}^{H}$ to the true value i.e.,
\[ \| \hat{\theta}^{H} - \theta_0 \|. \]

First note that from a set of standard regularity conditions for the GMM estimation treating $Q(\xi^{H}(\theta))$ as the true criterion function for $H$ large enough (i.e., $\epsilon$ is small enough), one can obtain
\[ \hat{\theta}^{H} = \theta^{H} + O_p(1/\sqrt{JT}). \]

We impose the following regularity conditions for this purpose.

**Condition 3** (Consistency of pseudo true value estimator). For $H = H(\epsilon)$ large enough (i.e., $\epsilon$ is small enough) (i) $W\hat{m}_{J}(\xi^{H}(\theta)) = 0$ only if $\theta = \theta^{H}$; (ii) $\theta^{H} \in \Theta$, which is compact; (iii) $\xi_{t}^{H}(\theta)$ is continuous at each $\theta \in \Theta$ with probability one; (iv) $E[||z_{t}||/\sqrt{J}] < \infty$ and $E[\sup_{\theta \in \Theta}||\xi^{H}_{t}(\theta)||/\sqrt{J}] < \infty$.

Under Condition 3 we have $\hat{\theta}^{H} \rightarrow_{P} \theta^{H}$.

**Condition 4** (Asymptotic normality of pseudo true value estimator). For $H = H(\epsilon)$ large enough (i.e., $\epsilon$ is small enough) (i) $\theta^{H}$ is in the interior of $\Theta$; (ii) $\xi_{t}^{H}(\theta)$ is continuously differentiable in a neighborhood $\Theta_{H}$ of $\theta^{H}$ with probability approaching one; (iii) $\hat{m}_{J}(\xi^{H}(\theta^{H})) = 0$, $E[||z_{t}||^{2}/J] < \infty$, and $E[||\xi_{t}^{H}(\theta^{H})||^{2}/J] < \infty$; (iv) $E[\sup_{\theta \in \Theta} ||(\nabla_{\theta}^{2}\xi^{H}_{t}(\theta))||^{2}/J] < \infty$, $l = 1, 2$; (v) $\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta^{H}))^{\prime}W\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta^{H}))$ is nonsingular.\(^7\)

Under Condition 3 and 4 we have (9) and therefore, we can write
\[ \| \hat{\theta}^{H} - \theta_0 \| \leq \| \hat{\theta}^{H} - \theta^{H} \| + \| \theta^{H} - \theta_0 \| = O_p(1/\sqrt{JT}) + \| \theta^{H} - \theta_0 \| \]
for $H$ sufficiently large.

Then we can focus on the distance between the pseudo true value and the true value. From the mean-value expansion of the FOC for the pseudo true value around $\theta_0$, we have
\[ 0 = \nabla_{\theta}Q(\xi^{H}(\theta^{H})) = \nabla_{\theta}Q(\xi^{H}(\theta_0)) + \nabla_{\theta\theta}Q(\xi^{H}(\hat{\theta}^{H}))(\theta^{H} - \theta_0) \]
\[ = \nabla_{\theta}Q(\xi^{H}(\theta_0)) + \nabla_{\theta\theta}Q(\xi(\theta_0))(\theta^{H} - \theta_0) + [\nabla_{\theta\theta}Q(\xi^{H}(\hat{\theta}^{H})) - \nabla_{\theta\theta}Q(\xi(\theta_0))](\theta^{H} - \theta_0) \]
\[ = \nabla_{\theta}Q(\xi^{H}(\theta_0)) + \nabla_{\theta\theta}Q(\xi(\theta_0))(\theta^{H} - \theta_0) \]
\[ + [\nabla_{\theta\theta}Q(\xi^{H}(\hat{\theta}^{H})) - \nabla_{\theta\theta}Q(\xi(\hat{\theta}^{H}))](\theta^{H} - \theta_0) + [\nabla_{\theta\theta}Q(\xi^{H}(\hat{\theta}^{H})) - \nabla_{\theta\theta}Q(\xi(\theta_0))](\theta^{H} - \theta_0) \]

\(^7\)We can further show this instead of assuming it as follows. Because $\nabla_{\theta}\hat{m}_{J}(\xi(\theta))$ is continuous and $\Theta$ is compact, Condition 2 (v) implies that $\nabla_{\theta}\hat{m}_{J}(\xi(\theta))^{\prime}W\nabla_{\theta}\hat{m}_{J}(\xi(\theta))$ is still nonsingular in a small neighborhood of $\theta_0$. Also we can take $H = H(\epsilon)$ large enough (i.e., $\epsilon$ is small enough) such that $\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta))^{\prime}W\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta))$ is arbitrary close to $\nabla_{\theta}\hat{m}_{J}(\xi(\theta))^{\prime}W\nabla_{\theta}\hat{m}_{J}(\xi(\theta))$ in this small neighborhood of $\theta_0$. Therefore when $\theta^{H}$ is close enough to $\theta_0$ (i.e., the approximation error is small enough), $\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta^{H}))^{\prime}W\nabla_{\theta}\hat{m}_{J}(\xi^{H}(\theta^{H}))$ becomes nonsingular.
where $\tilde{\theta}^H$ lies between $\theta^H$ and $\theta_0$. We show $\nabla_{\theta\theta} Q(\xi^H(\tilde{\theta}^H)) - \nabla_{\theta\theta} Q(\xi(\tilde{\theta}^H))$ is bounded by the approximation error $\eta(\epsilon)$. Consider for all $l,l' \in \{1, \ldots, \dim(\theta)\}$

\[
\nabla_{\theta_l\theta_{l'}} Q(\xi^H(\tilde{\theta}^H)) - \nabla_{\theta_l\theta_{l'}} Q(\xi(\tilde{\theta}^H)) \\
= 2 \left\{ \nabla_{\theta_l} \bar{m}_J(\xi^H(\theta))|_{\theta=\tilde{\theta}^H} \right\}' W \left\{ \nabla_{\theta_{l'}} \bar{m}_J(\xi^H(\theta))|_{\theta=\tilde{\theta}^H} \right\}' - 2 \left\{ \nabla_{\theta_l} \bar{m}_J(\xi(\theta))|_{\theta=\tilde{\theta}^H} \right\}' W \left\{ \nabla_{\theta_{l'}} \bar{m}_J(\xi(\theta))|_{\theta=\tilde{\theta}^H} \right\}' \\
+ 2 \left\{ \nabla_{\theta_l\theta_{l'}} \bar{m}_J(\xi^H(\theta))|_{\theta=\tilde{\theta}^H} \right\}' W \bar{m}_J(\xi^H(\tilde{\theta}^H)) - 2 \left\{ \nabla_{\theta_l\theta_{l'}} \bar{m}_J(\xi(\theta))|_{\theta=\tilde{\theta}^H} \right\}' W \bar{m}_J(\xi(\tilde{\theta}^H))
\]

and therefore the triangle inequality, the Cauchy-Schwarz inequality, Condition 4, and the definition of the approximation error in (8) imply

\[
||\nabla_{\theta\theta} Q(\xi^H(\tilde{\theta}^H)) - \nabla_{\theta\theta} Q(\xi(\tilde{\theta}^H))|| \\
\leq C_1 \left\{ E[\sup_{\theta \in \Theta} ||\xi^H_1(\theta) - \xi_1(\theta)||^2 / J] \right\}^{1/2} + C_2 \left\{ E[\sup_{\theta \in \Theta} ||\nabla_\theta \xi^H(\theta) - \nabla_\theta \xi(\theta)||^2 / J] \right\}^{1/2} \\
+ C_3 \left\{ E[\sup_{\theta \in \Theta} ||(\nabla_\theta)^2 \xi^H_1(\theta) - (\nabla_\theta)^2 \xi_1(\theta)||^2 / J] \right\}^{1/2}
\]

\[
\leq C \eta(\epsilon)
\]

for some constants $C, C_1, C_2,$ and $C_3$.

Then because $\nabla_{\theta\theta} Q(\xi(\theta))$ is nonsingular and $\nabla_{\theta\theta} Q(\xi(\theta))$ is continuous in $\theta$ in the neighborhood of $\theta_0$ by Condition 2, from (11) and (12) we obtain

\[
||\theta^H - \theta_0|| = O \left( ||\nabla_{\theta} Q(\xi^H(\theta_0))|| \right).
\]

Further note that

\[
\nabla_{\theta} Q(\xi^H(\theta_0)) = \nabla_{\theta} Q(\xi^H(\theta_0)) - \nabla_{\theta} Q(\xi(\theta_0)) \\
= 2 \left\{ \nabla_{\theta} \bar{m}_J(\xi^H(\theta))|_{\theta=\theta_0} \right\}' W \bar{m}_J(\xi^H(\theta_0)) - 2 \left\{ \nabla_{\theta} \bar{m}_J(\xi(\theta))|_{\theta=\theta_0} \right\}' W \bar{m}_J(\xi(\theta_0)) \\
= 2 \left\{ \nabla_{\theta} \bar{m}_J(\xi^H(\theta))|_{\theta=\theta_0} - \nabla_{\theta} \bar{m}_J(\xi(\theta))|_{\theta=\theta_0} \right\}' W \bar{m}_J(\xi(\theta_0)) \\
+ 2 \left\{ \nabla_{\theta} \bar{m}_J(\xi^H(\theta))|_{\theta=\theta_0} \right\}' W (\bar{m}_J(\xi^H(\theta_0)) - \bar{m}_J(\xi(\theta_0))) \\
= 0 + 2 \left\{ \nabla_{\theta} \bar{m}_J(\xi^H(\theta))|_{\theta=\theta_0} \right\}' W \cdot \frac{1}{J} \sum_{j=1}^{J} E \left[ (\xi^H_{j,t}(\theta_0) - \xi_{j,t}(\theta_0)) z_{j,t} \right]
\]

where the first equality holds by $\nabla_{\theta} Q(\xi(\theta_0)) = 0$ (i.e., the FOC of the true objective function (7)) and the last equality holds by the moment condition, $\bar{m}_J(\xi(\theta_0)) = 0$. It follows that by the Cauchy-Schwarz inequality and the triangle inequality,

\[
||\nabla_{\theta} Q(\xi^H(\theta_0))|| \leq C \left\{ E[||\xi^H_1(\theta_0) - \xi_1(\theta_0)||^2 / J] \right\}^{1/2} \left\{ E[||z_t||^2 / J] \right\}^{1/2}
\]

for some positive constant $C$. We, therefore, conclude that

\[
||\theta^H - \theta_0|| = O \left( ||\nabla_{\theta} Q(\xi^H(\theta_0))|| \right) = O(\eta(\epsilon)).
\]

by (8), (13), and $E[||z_t||^2 / J] < \infty$. Combining this and (10) we summarize the key result here.
Theorem 1. Suppose Conditions 1-4 hold. Then

\[ \left\| \hat{\theta}^H - \theta_0 \right\| = O_p(1/\sqrt{JT}) + O(\eta(\epsilon)). \]

The key finding is that the outer loop GMM estimator has the same magnitude of bias as the approximation error in the inner loop. This means if a researcher chooses the tolerance level equal to \(10^{-6}\) in the inner loop of the fixed point algorithm, then it creates the bias of the order of \(10^{-6}\) in the outer loop GMM estimator.

This also implies that for a valid inference (e.g., asymptotic normal distribution approximation), we need\(^8\)

\[ \sqrt{JT} \eta(\epsilon) \to 0 \text{ so that } \sqrt{JT} (\hat{\theta}^H - \theta_0) = \sqrt{JT} \left( \hat{\theta}^H - \theta^H \right) + o(1) = O_p(1) \]

i.e., with a larger sample size, we should use more strict convergence criteria in the inner loop.

3 Conclusion

We study the effects of approximation errors in the inner loop contraction mapping on statistical inference of the outer loop GMM estimator in the context of BLP model. We show that the inner loop approximation error will create the same magnitude of bias in the outer loop estimation. Therefore numerical error from the inner loop does not propagate into the outer loop.

---

\(^8\)Following Gouriéroux and Monfort (1991) and Hajvassiliou and Ruud (1994) (as suggested in Ackerberg, Geweke, and Hahn (2009)) we can further show that under \(\sqrt{JT} \eta(\epsilon) \to 0\) the asymptotic distribution of \(\sqrt{JT} (\hat{\theta}^H - \theta_0)\) converges to that of \(\sqrt{JT} (\hat{\theta}^{H(0)} - \theta_0)\).
References


