



Asymptotic theory for differentiated products demand models with many markets[☆]



Joachim Freyberger

Department of Economics, University of Wisconsin, Madison, United States

ARTICLE INFO

Article history:

Received 16 April 2012

Received in revised form

12 October 2014

Accepted 20 October 2014

Available online 11 November 2014

JEL classification:

C13

C15

L13

Keywords:

Demand estimation

Differentiated products

Many markets

Bias correction

Adjusted standard errors

ABSTRACT

This paper develops asymptotic theory for differentiated product demand models with a large number of markets T . It takes into account that the predicted market shares are approximated by Monte Carlo integration with R draws and that the observed market shares are approximated from a sample of N consumers. The estimated parameters are \sqrt{T} consistent and asymptotically normal as long as R and N grow fast enough relative to T . Both approximations yield additional bias and variance terms in the asymptotic expansion. I propose a bias corrected estimator and a variance adjustment that takes the leading terms into account. Monte Carlo simulations show that these adjustments should be used in applications to avoid severe undercoverage caused by the approximation errors.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Discrete choice models have been widely used in the empirical industrial economics literature to estimate demand for differentiated products. In these models, consumers in market t can typically choose one of J_t products or an outside option. The consumers choose the product that maximizes their utility, which leads to an expression of the markets shares. The parameters of the utility function can then be estimated using observed markets shares and product characteristics.

The challenge of the asymptotic theory in these models is to deal with several approximation errors. First, in models with heterogeneous consumers, such as the model of Berry et al. (1995) (referred to as the BLP model), the market shares involve integrals over the distribution of random coefficients. When estimating the parameters of the model, these integrals cannot be calculated

[☆] I thank an associate editor and three referees for helpful comments which have greatly improved the results and the exposition of the paper. I am grateful to Ivan Canay, Joel Horowitz, and Elie Tamer for helpful discussions and suggestions. I have also received valuable feedback from Mike Abito, Steven Berry, Mark Chicu, Roland Eisenhuth, Amit Gandhi, Aviv Nevo, Henrique de Oliveira, and Ketan Patel.

E-mail address: jfreyberger@ssc.wisc.edu.

analytically and are usually approximated by Monte Carlo integration with R random draws from the distribution of the random coefficients.¹ Second, many data sets do not contain true market shares but approximations from a sample of N consumers. Consequently, asymptotic theory needs to allow for three sources of errors: the simulation error in approximating the shares predicted by the model, the sampling error in estimating the market shares, and the underlying model error.

The limiting distribution of the estimated parameters can be obtained by either letting the number of products, the number of markets, or both approach infinity. Since the asymptotic distribution of the estimator serves as an approximation of its unknown finite sample distribution, it depends on the particular data set which approximation is most suitable. While in some cases using an approximation where the number of products approaches infinity is appropriate (as in Berry et al., 1995), in other cases the number of markets is a lot larger than the number of products (e.g. Nevo, 2001). As shown in this paper the asymptotic properties of the estimator differ a lot depending on which approximation is

¹ Although the focus lies on the effect of using Monte Carlo integration to approximate integrals, polynomial-based quadrature rules are also discussed. See Remark 4 for further details.

used. It is therefore important that both approximations are well understood.

Berry et al. (2004) provide asymptotic theory for a large number of products in one market. In their paper all market shares go to 0 at the rate $1/J$ and a necessary condition for asymptotic normality is that J^2/R and J^2/N are bounded. In this case, they find that their estimator $\hat{\theta}$ of the parameter vector θ_0 satisfies

$$\sqrt{J}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_{GMM} + \lambda_1 V_{MC} + \lambda_2 V_{SM}),$$

where $\lambda_1 = \lim_{J,R \rightarrow \infty} J^2/R$ and $\lambda_2 = \lim_{J,N \rightarrow \infty} J^2/N$. Here V_{GMM} denotes the variance of the estimator when the integral is calculated exactly and the true market shares are observed. V_{MC} and V_{SM} are additional variance terms due to the approximation error of the integrals and the market shares, respectively. Hence, if J^2/R or J^2/N are bounded away from 0, the asymptotic distribution of the estimated parameter vector is centered at 0 but Monte Carlo integration and market share approximation leads to a larger variance.

This paper is concerned with the asymptotic theory for a small number of products, J_t , in a growing number of markets T , which is the natural choice in many applications. However, this setup has not been considered in the literature so far. In Nevo (2001), for example, there are 1124 markets and 24 products (see also among others Kim, 2004, and Villas-Boas, 2007). Furthermore, in a similar (but more general) class of models, Berry and Haile (2014) provide nonparametric identification results for a large number of markets and a fixed number of products. These results can serve as a basis for nonparametric or semiparametric estimation of the model. Before such a flexible estimation procedure is developed, it is interesting to know how the commonly used fully parametric estimators behave. I prove consistency and asymptotic normality for these cases where T approaches infinity and $J_t \leq \bar{J}$ where \bar{J} is fixed. The assumptions in the main part of the paper are stated for the BLP model, but the results in the appendix use higher level assumptions for a more general class of models. I find that the estimated parameters are \sqrt{T} consistent and asymptotically normal as long as \sqrt{T}/R and \sqrt{T}/N are bounded. In this case, $\hat{\theta}$ satisfies

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\tilde{\lambda}_1 \mu_1 + \tilde{\lambda}_2 \mu_2, V_{GMM})$$

where $\tilde{\lambda}_1 = \lim_{T,R \rightarrow \infty} \sqrt{T}/R$ and $\tilde{\lambda}_2 = \lim_{T,N \rightarrow \infty} \sqrt{T}/N$, again V_{GMM} is the variance of the estimator without approximation errors in the integrals and market shares, and μ_1 and μ_2 are constants. Hence, if \sqrt{T}/R or \sqrt{T}/N are bounded away from 0, Monte Carlo integration and market share approximation lead to an asymptotic distribution which is not centered at 0. The intuition for this result is that Monte Carlo integration (and similarly market share approximation) yields both additional bias as well as additional variance terms in the asymptotic expansion. The leading bias term is of order $O_p(\sqrt{T}/R)$, while the leading variance term is of order $O_p(1/\sqrt{R})$. Hence, the bias dominates the variance if R does not increase faster than T , and the variance does not affect the asymptotic distribution as long as $R \rightarrow \infty$. These results rely on using different draws to approximate the integral in different markets. If the same R draws are used in all markets one needs T/R to be bounded to obtain \sqrt{T} consistency, which means that more draws are needed to approximate each integral relative to the number of markets.²

² Remark 3 in Section 2.2 shows the differences in the asymptotic expansions when the same draws and different draws are used in each market. This section also highlights the differences between asymptotics in the number of markets and the number of products with and without different draws in different markets. See Remark 5 for details.

These results highlight that there are important differences between letting the number of products or the number of markets approach infinity. With a large number of products, it is important to correct the variance of the estimator due to the approximations. With a large number of markets and different draws in each market, the asymptotic distribution might not be centered at 0 and hence, a bias corrected estimator is needed. In both cases if R is too small, confidence intervals based on the usual asymptotic GMM distribution have the wrong size even asymptotically. Notice that when \sqrt{T}/R and \sqrt{T}/N converge to 0, the approximations do not affect the asymptotic distribution of the estimator. Contrary, in the setup where J approaches infinity, this result is only obtained when J^2/R and J^2/N converge to 0, which is a stronger requirement on the number of draws relative to the sample size.

The finite sample properties of the estimator depend on R and N due to both the additional bias and variance terms. I suggest an analytical bias correction which eliminates the leading bias term from the asymptotic distribution. I also show how one can easily incorporate the leading variance term when calculating standard errors. These two corrections greatly improve finite sample results. In particular, Monte Carlo simulations demonstrate that using a small number of simulation draws in comparison to the number of markets and using the usual GMM asymptotic distribution can yield distorted inference while the use of bias correction and adjusted standard errors leads to a considerably better performance.

These results might suggest that practitioners can simply use a very large number of draws and ignore Monte Carlo integration issues. Although this might be feasible depending on the model and computing resources available, in applications this is often not possible for several reasons. First, one does not know in advance how many draws suffice to obtain satisfactory results. As discussed in Section 4, the number of draws needed depends, among others, on the sample size, the number of random coefficients as well as unknown parameters, such as the variance of the random coefficients. Second, taking a very large number of draws is computationally very demanding because one needs to solve a complicated nonlinear optimization problem to estimate the parameters. The Monte Carlo results of the random coefficients logit model presented in Section 4 are based on a small number of products ($J = 4$) and six random coefficients to make the problem tractable. However, in the same setup as in Section 4 but with a sample size of $J = 24$ and $T = 1124$ (as in Nevo, 2001) it takes more than 24 h to minimize the objective function when $R = 2000$ and the starting values of the parameters are close to the true values.³ Since we are dealing with a nonlinear optimization problem one needs to use several different starting values in applications. Taking a smaller number of draws considerably speeds up calculations. Third, even when taking the same draws for each product and each random coefficient, the number of draws needed is $T \times R$. In the previous example this means that 2248,000 draws are used to calculate the shares and the draws have to be stored before optimizing the function. As a consequence, more than 20 GB of memory is needed to run the program which is used to do the simulations in this paper. Finally, in case one wants to integrate over empirical distributions of demographic characteristics, R is constrained by the number of people in the database for a certain market.

The implication for empirical work with Monte Carlo integration or approximated market shares is that bias corrections and variance adjustments should be used. If the number of draws and the number of consumers is sufficiently large, the bias correction is close to 0 and the corrected standard errors will be very close to the

³ Computational details, including a description of the processors used, are presented in Section 4.

usual GMM standard errors. If the number of draws or consumers is small, the approximations will affect the finite sample performance of the estimator, the usual GMM standard errors underestimate the true variance, and the estimates are biased. In this case the proposed corrections, which can be computed easily, considerably improve the finite sample performance. Nevertheless, the number of draws and consumers should be as large as possible, subject to computational constraints and data availability, in order to improve the precision of the initial estimator, which is used to calculate the bias correction.

The focus of this paper is on understanding the asymptotic theory when Monte Carlo integration is employed because this is the method which has almost exclusively been used in applications. Furthermore, an advantage of Monte Carlo integration is that it can easily be used to integrate over distributions with an unknown functional form such as demographic characteristics as in Nevo (2001). An interesting alternative is to use polynomial-based quadrature rules recently advocated by Skrainka and Judd (2011). These approximations are shown to perform well in simulations when integrating over a normal distribution. However, it is not clear how a distribution of demographic characteristics, which does not have a closed form expression, can be handled with quadrature rules unless a parametric distribution is assumed. Although the focus lies on Monte Carlo integration, the asymptotic expansions derived in this paper also provide insights into finite sample bias from polynomial-based quadrature rules.

The results in this paper are related to similar findings of Lee (1995) in simulated maximum likelihood estimation (SMLE) of discrete choice models. Lee's paper and my paper are both based on higher order expansions of the derivative of the objective function in which the simulation error of the integral is linearized. The first term in this expansion is the derivative without the simulation error, the second term leads to a second variance term, and the third term to the bias. Lee's paper is tailored to SML while mine is tailored to discrete choice random coefficient models. Lee (1995) also finds that the asymptotic distribution might not be centered at 0 if the number of draws is small relative to the sample size.⁴ Similar findings are obtained in nonlinear simulated least squares estimation by Laffont et al. (1995), where again the assumptions and proofs are specific to that setup. Simulated methods of moments estimators studied by McFadden (1989) and Pakes and Pollock (1989) differ from the previously discussed estimators because they do not lead to a bias term in the asymptotic expansion. Hahn and Newey (2004) and Arellano and Hahn (2007) propose bias corrections in nonlinear fixed effects panel data models where the bias is not due to a simulation error, but the incidental parameters problem and there is no variance adjustment.

Kristensen and Salanié (2013) deal with simulation based estimators in a general setup where the function to be approximated is the same for all t . They assume that an expansion of the estimator exists and that the remainder term is negligible. Under these assumptions, they propose bias corrections which eliminate the leading bias term and variance adjustments. These results do not directly apply to the BLP model because both the distribution of the random coefficients and the true market shares might vary over markets. However, in case the true market shares are observed and the distribution of random coefficients is the same in all markets, the expansions and the bias correction here are a special case of those in Kristensen and Salanié (2013). In this case, I still need to derive the actual expansion, which is needed to calculate the bias correction and the variance adjustment, and show that the remainder term is negligible under low level conditions. Doing so is

particularly challenging in discrete choice demand models because there is no closed form expression for the objective function and the resulting formulas are algebraically tedious. See Remark 2 for more details.

My results provide theoretical justifications for simulation results in recent studies by Skrainka and Judd (2011) and Skrainka (2012), who find among others finite sample bias and excessively tight standard errors when using Monte Carlo integration. I show in Monte Carlo simulations that my adjusted standard errors are not artificially tight and that the finite sample bias caused by the integral approximation can be eliminated by using the analytical bias correction. Other recent contributions to the literature on estimation of discrete choice demand models include Gandhi et al. (2013) and Armstrong (2013). In both papers $J \rightarrow \infty$. Gandhi et al. (2013) allow for interactions between the unobserved demand error and product characteristics, which affect both the identification arguments and estimation method. Armstrong (2013) discusses the validity of commonly used instruments in models with a large number of products under conditions on economic primitives. He shows that in several models using product characteristics as instruments for price yields inconsistent estimates and he shows how consistent estimates can be obtained. His findings on invalidity of instruments do not apply in my setting with a fixed number of products.

The remainder of this paper is organized as follows. The next section introduces the BLP model, provides low level assumptions, and states the main results. A more general class of models under higher level assumptions is discussed in the Appendix. Section 3 explains how the bias correction and the variance adjustment can be implemented. Finally, I demonstrate with Monte Carlo simulations that the proposed corrections considerably improve the finite sample performance.

2. Random coefficient logit model

Let $p_{jt} \in \mathbb{R}$ be the price of product j in market t . Define $x_{jt} = (p_{jt}, x_{jt}^{(2)}) \in \mathbb{R}^{d_x}$, where $x_{jt}^{(2)}$ are observed product characteristics other than price and may include a constant or product dummies. Let $\xi_{jt} \in \mathbb{R}$ be a product characteristic of product j in market t , which is observed by firms and consumers, but not by the econometrician. In the BLP model consumer i in market t chooses the product j which maximizes the utility

$$u_{ijt} = x'_{jt} \beta_{it} + \xi_{jt} + \epsilon_{ijt}, \quad (1)$$

where $\beta_{it} = \beta_0 + \Sigma_0 v_{it}$, $\beta_0 \in \mathbb{R}^{d_x}$, Σ_0 is a $d_x \times d_x$ matrix, and $v_{it} \in \mathbb{R}^{d_x}$ is a random vector with distribution function P_{0t} . Assuming that ϵ_{ijt} are independent and identically distributed (iid) extreme value random variables, the market share of product j in market t is given by

$$s_{jt} = \int \frac{\exp(x'_{jt} \beta_0 + \xi_{jt} + x'_{jt} \Sigma_0 v)}{1 + \sum_{k=1}^J \exp(x'_{kt} \beta_0 + \xi_{kt} + x'_{kt} \Sigma_0 v)} dP_{0t}(v). \quad (2)$$

I assume that P_{0t} is known to the econometrician. In many applications P_{0t} is assumed to be a standard normal distribution and Σ_0 is a diagonal matrix, but P_{0t} could also be a distribution of demographic characteristics which varies across markets. Let $x_t = (x_{1t}, \dots, x_{Jt})$, $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})$, $\theta = (\beta, \text{vec}(\Sigma))$, and define

$$v_{jt}(\theta, x_t, \xi_t, v) \equiv \frac{\exp(x'_{jt} \beta + \xi_{jt} + x'_{jt} \Sigma v)}{1 + \sum_{k=1}^J \exp(x'_{kt} \beta + \xi_{kt} + x'_{kt} \Sigma v)}.$$

⁴ See Remark 1 for further comments on differences and similarities of the proofs.

The function $v_{jt}(\theta, x_t, \xi_t, v)$ represents the market share of product j , predicted by the model, for a given parameters θ , product characteristics x_t and ξ_t , and consumer type v . Since for any market t the market shares are generated by integrating over all consumers, the shares predicted by the model are defined as

$$\sigma_{jt}(\theta, x_t, \xi_t, P_t) \equiv \int v_{jt}(\theta, x_t, \xi_t, v) dP_t(v) \tag{3}$$

and the true shares are $s_{jt} = \sigma_{jt}(\theta_0, x_t, \xi_t, P_{0t})$. Let $\sigma_t(\theta, x_t, \xi_t, P_t)$ be the $J_t \times 1$ vector with elements $\sigma_{jt}(\theta, x_t, \xi_t, P_t)$ and define $v_t(\theta, x_t, \xi_t, v)$ analogously.

Berry (1994) shows that for any $\theta \in \Theta, x_t \in \mathbb{R}^{d_x \times J_t}$, distribution P , and $s_{jt} \in (0, 1)$ there exists $\xi_t(\theta, P, s_t, x_t) \in \mathbb{R}^{J_t}$ such that for all j

$$s_{jt} = \int v_{jt}(\theta, x_t, \xi_t(\theta, P, s_t, x_t), v) dP(v).$$

In particular, one can solve for ξ_t in the system of J_t equations given in (2) or, in other words, $\sigma_t(\theta, x_t, \cdot, P_t)$ is invertible. I denote the inverse function by $\sigma_t^{-1}(\theta, x_t, \cdot, P_t)$.

Now assume that there are instruments $z_t \in \mathbb{R}^{J_t \times d_z}$ such that

$$E(z_t' \xi_t(\theta_0, P_{0t}, s_t, x_t)) = 0 \tag{4}$$

and that these moment conditions identify θ_0 . Then an estimator of θ_0 is

$$\begin{aligned} \tilde{\theta} \equiv \arg \min_{\theta \in \Theta} & \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{0t}, s_t, x_t) \right)' W_T \\ & \times \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{0t}, s_t, x_t) \right), \end{aligned}$$

where W_T is a weighting matrix which converges in probability to a positive definite matrix W . The estimator $\tilde{\theta}$ is infeasible in practice for two reasons. First, it is infeasible to calculate the integral in (2) analytically. Therefore, it is often approximated by R random draws from P_{0t} . In particular,

$$\begin{aligned} \int v_{jt}(\theta, x_t, \xi_t, v_{rt}) dP_0(v) & \approx \frac{1}{R} \sum_{i=1}^R v_{jt}(\theta, x_t, \xi_t, v_{rt}) \\ & = \int \frac{\exp(x_{jt}'\beta + \xi_{jt} + x_{jt}'\Sigma v)}{1 + \sum_{k=1}^J \exp(x_{kt}'\beta + \xi_{kt} + x_{kt}'\Sigma v)} dP_{Rt}(v), \end{aligned}$$

where v_{rt} are draws from the distribution P_{0t} , which may differ across markets, and P_{Rt} denotes the empirical distribution function. Second, often the true market shares s_t are not observed but are instead approximated from a sample of N consumers. That is, the observed market shares are $s_{jt}^N = \frac{1}{N} \sum_{n=1}^N s_{jnt}$, where $s_{jnt} \in \{0, 1\}$, and $E(s_{jnt}) = s_{jt}$.

Now define the estimator used in applications

$$\begin{aligned} \hat{\theta} \equiv \arg \min_{\theta \in \Theta} & \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{Rt}, s_t^N, x_t) \right)' W_T \\ & \times \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{Rt}, s_t^N, x_t) \right). \end{aligned}$$

2.1. Assumptions

I now provide primitive sufficient conditions for the asymptotic normality results discussed in the introduction to hold in the BLP model. A more general class of models under higher level assumptions is discussed in the Appendix. Let \bar{J} be a positive integer.

Assumption RC1. The data $\{z_t, x_t, s_t\}_{t=1}^T$ are independent across markets and $J_t \leq \bar{J}$.

Assumption RC2. v_{rt} are iid draws over r and independent over t from P_{0t} and independent of (x_t, z_t, s_t) , $\ln(T)/R \rightarrow 0$, and $E_{P_{0t}}(v^4) < \infty$.

Assumption RC3. Let $\tilde{\mathcal{P}} = \{P^1, \dots, P^k\}$ be a finite set of distributions. $v_t = g(a_t, w_t)$ where $g : \mathcal{A} \times \mathbb{R}^{d_w} \rightarrow \mathbb{R}^{d_x}$ is continuous in both arguments, a_t is a vector of constants and $\mathcal{A} \subset \mathbb{R}^{d_a}$ is a compact set, w_t is a random vector and the distribution of each element of w_t either (a) is in $\tilde{\mathcal{P}}$ and is the same for all t or (b) its support is a subset of $[M_l, M_u]$ for some constants M_l and M_u .

Assumption RC4. $s_{jtn} \in \{0, 1\}$ are iid draws over n with $E(s_{jtn}) = s_{jt}$, independent over t , independent of $(x_t, z_t, \xi_t) \mid s_t$, and $\ln(T)/N \rightarrow 0$.

Assumption RC5. There exists $\varepsilon > 0$ such that $\varepsilon \leq s_{jt} \leq 1 - \varepsilon$ for all $t = 1, \dots, T$ and $j = 0, \dots, J_t$ with $s_{0t} \equiv 1 - \sum_{j=1}^{J_t} s_{jt}$.

Assumption RC6. x_t is in a compact set. The parameter space Θ is compact and θ_0 is in the interior of Θ .

Assumption RC7. The instruments matrix has full rank, $E(z_{jt}^2) \leq M$ for some M and $E(z_{jt}^4) < \infty$ for all $l = 1, \dots, d_z$ and $j = 1, \dots, J_t$.

Assumption RC8. For all functions $h_j(z_t, x_t, s_t, P_t, J_t)$ with $E(h_j(z_t, x_t, s_t, P_{0t}, J_t)^2) \leq M$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} E(h_j(z_t, x_t, s_t, P_{0t}, J_t))$$

exists.

Assumption RC9. The parameter vector θ_0 is identified from the moment conditions in (4). In particular, for all $\delta > 0$, there exists $C(\delta) > 0$ such that

$$\begin{aligned} \lim_{T \rightarrow \infty} \Pr \left(\inf_{\theta \notin \mathcal{N}_{\theta_0}(\delta)} \left\| \frac{1}{T} \sum_{t=1}^T z_t' (\xi_t(\theta, P_{0t}, s_t, p_t) \right. \right. \\ \left. \left. - \xi_t(\theta_0, P_{0t}, s_t, p_t)) \right\| \geq C(\delta) \right) = 1, \end{aligned}$$

where $\| \cdot \|$ denoted the Euclidean norm, and the matrix

$$\Gamma \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \theta} E(z_t' \xi_t(\theta_0, P_{0t}, s_t, x_t))$$

has full rank.

Assumption RC1 states that the observed data is independent across markets and that the number of products in each market is bounded by \bar{J} . One way to allow for dependence across markets is to assume that $\xi_{jt} = \xi_j + \Delta \xi_{jt}$ and to estimate the model with product dummies as in Nevo (2001). In this way product dummies can capture the part of the product characteristics that do not vary across markets.⁵ **Assumption RC2** assumes that the draws used to

⁵ Since θ is finite dimensional, the number of different products across all markets needs to be finite in this case. Moreover, the identification assumption then requires that, asymptotically, all products appear in infinitely many markets. An important special case is when the same products are available in all markets. Also notice that with product dummies, parameters in front of observed product characteristics that do not vary across markets are not identified.

approximate the integral in each market are iid and independent of the observed data and that the number of draws diverge.⁶ Notice that I assume that different draws are used in different markets. This assumption does not lead to much larger computational costs compared to using the same draws in all markets, but it leads to a smaller variance of the estimator as shown in Remark 3. Moreover, the distribution of random coefficients needs to have four bounded moments. Assumption RC3 places additional restrictions on the distribution of random coefficients. If the distribution has compact support, this assumption does not restrict how it can vary across markets, which allows using demographics with compact support. It also allows that the distribution has unbounded support but in this case the way it varies over t is controlled by the finite dimensional parameter a_t . For example, one could model demographics with a normal or a log normal distribution where the mean and the variance change across markets. Clearly, also the often used normal distribution with an unknown mean and variance is permitted. Assumption RC4 says that the consumers are sampled randomly in each market and across markets and that the number of consumers grows with T . Assumptions RC5 and RC6 state that the market shares are bounded away from 0 and 1 (including the shares of the outside option s_{0t}) and that the product characteristics and parameters are in a compact set, respectively. This assumption is common in the simulation based estimation literature (for example McFadden (1989), Lee (1995), and Berry et al. (2004)). Assumption RC7 puts moment conditions on the instrument. Assumption RC8 is a limit condition which is needed because the data are not assumed to be identically distributed across market (in which case the assumption would follow immediately). The assumption holds for example if we assume that J_t and P_{0t} are random and iid across markets and that all other assumptions hold conditional on J_t and P_{0t} . The assumption can also hold for deterministic J_t and P_{0t} . It mainly rules out systematic changes of J_t or P_{0t} over t . Finally, Assumption RC9 contains standard global and local identification conditions. Nonparametric identification with many markets and a finite number of products is shown by Berry and Haile (2014).

2.2. Asymptotic properties of the estimator

The previous assumptions lead to the following theorem which characterizes the asymptotic expansion of the estimator.

Theorem 1. Assume that Assumptions RC1–RC9 hold. Then there exist $d_z \times d_z$ matrices Φ_1, Φ_2 , and Φ_3 , as well $d_z \times 1$ vectors $\bar{\mu}_1$ and $\bar{\mu}_2$ such that

$$\begin{aligned} \sqrt{T}(\hat{\theta} - \theta_0) &= \left((\Gamma'W\Gamma)^{-1} \Gamma'W + o_p(1) \right) \\ &\times \left(Q_{1T} + \frac{1}{\sqrt{R}}Q_{2T} + \frac{1}{\sqrt{N}}Q_{3T} + \frac{\sqrt{T}}{R}C_{1T} \right. \\ &\left. + \frac{\sqrt{T}}{N}C_{2T} + o_p\left(\frac{\sqrt{T}}{R}\right) + o_p\left(\frac{\sqrt{T}}{N}\right) \right) \end{aligned}$$

where

$$Q_{1T} \xrightarrow{d} N(0, \Phi_1), \quad Q_{2T} \xrightarrow{d} N(0, \Phi_2), \quad Q_{3T} \xrightarrow{d} N(0, \Phi_3),$$

and

$$C_{1T} \xrightarrow{p} \bar{\mu}_1, \quad C_{2T} \xrightarrow{p} \bar{\mu}_2.$$

Furthermore, Q_{1T}, Q_{2T} , and Q_{3T} are asymptotically independent.

⁶ I assume throughout that the draws are random although in practice pseudo random draws are used. The implicit assumption is that pseudo random draws lead to the same asymptotic properties. The Monte Carlo simulation results support this assumption.

The proof is in the Appendix. It verifies higher level assumptions of a general class of models. An immediate consequence of Theorem 1 is the following result.

Corollary 1. Assume that Assumptions RC1–RC9 hold. If $\lambda_1 = \lim_{T \rightarrow \infty} \frac{\sqrt{T}}{R} < \infty$ and $\lambda_2 = \lim_{T \rightarrow \infty} \frac{\sqrt{T}}{N} < \infty$, then

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left((\Gamma'W\Gamma)^{-1} \Gamma'W(\lambda_1 \bar{\mu}_1 + \lambda_2 \bar{\mu}_2), V_1 \right)$$

where

$$V_1 = (\Gamma'W\Gamma)^{-1} \Gamma'W\Phi_1W\Gamma(\Gamma'W\Gamma)^{-1}.$$

Theorem 1 shows that the use of Monte Carlo integration (as opposed to evaluating the integral exactly) leads to additional variance and additional bias terms. The leading variance term is of order $1/\sqrt{R}$ while the leading bias term is of order \sqrt{T}/R . Hence, if R grows more slowly than T , the leading bias term dominates the leading variance term, which may lead to an asymptotic distribution that is not centered at 0. As a consequence, if $\lambda_1 > 0$, confidence intervals based on the usual GMM asymptotic distribution have the wrong size asymptotically. If R grows faster than T , the leading variance term becomes dominating, but the first order asymptotic distribution is not affected by Monte Carlo integration. Similar arguments apply to the approximations of the market shares.

Remark 1. The proof of Theorem 1 is based on a third order expansion of the derivative of the objective function in which the simulation error of the integral is linearized. This is analogous to Lee (1995), who makes use of a similar approach and obtains similar results for simulated maximum likelihood estimation. I also make use of similar tools as Lee to show that the remainder term of the expansion is negligible.⁷ Despite sharing these similarities at a general level, deriving the actual expansion and showing that the remainder term is negligible is significantly different in both models. In discrete choice demand models these results are challenging to derive (and very particular to this setting) because there is no closed form expression for the objective function. Moreover, the Taylor expansion is algebraically tedious because it involves differentiating the inverse of a function from R^t to R^t . In the next subsection, I illustrate these steps for the special case where $J_t = 1$.

Remark 2. Kristensen and Salanié (2013) deal with a general class of simulation based estimators, they use very similar expansions, and they obtain similar conclusions in their Theorem 2. However, Kristensen and Salanié (2013) do not allow the objects to be approximated to change over markets (here s_t and P_{0t}) and they directly assume that a higher order expansion exists and that the remainder term is negligible (in their Assumption A4). I derive these results, which depend on the specific structure of the model. As a consequence, I obtain closed form expressions for the variance and the bias, which are the basis for the adjustments proposed in Section 2.4. If the true market shares are observed and if the distribution of random coefficients is the same in all markets, the expansions and the bias correction here are a special case of those in Kristensen and Salanié (2013). In this case I still have to derive the actual expansion and show that the remainder term is negligible, which is the major part of the proof. The latter could be achieved by verifying Assumption A4 in Kristensen and Salanié (2013) using my low level conditions, but this condition is stronger than necessary. I instead show that the remainder term is $o_p\left(\frac{\sqrt{T}}{R} + \frac{\sqrt{T}}{N}\right)$, which suffices to prove my results.

⁷ In particular, Lemmas A.2 and A.3 in the Appendix are very similar to Lemma A.2. and Proposition A.3. of Lee (1995), respectively.

Remark 3. It can also be shown that, under slightly different assumptions, if the distribution of random coefficients is the same in all markets and if one uses the same draws from P_{0t} in all markets then

$$\begin{aligned} \sqrt{T}(\hat{\theta} - \theta_0) &= \left((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1) \right) \\ &\times \left(Q_{1T} + \frac{\sqrt{T}}{\sqrt{R}}Q_{2T}^* + \frac{1}{\sqrt{N}}Q_{3T} + \frac{\sqrt{T}}{R}C_{1T} \right. \\ &\left. + \frac{\sqrt{T}}{N}C_{2T} + o_p\left(\frac{\sqrt{T}}{\sqrt{R}}\right) + o_p\left(\frac{\sqrt{T}}{N}\right) \right) \end{aligned}$$

where Q_{2T}^* converges to a normally distributed random variable with mean 0 as well. In this case one needs T/R to be bounded to obtain asymptotic normality, which is a stronger condition than the rate in Corollary 2, and the additional variance term dominates the additional bias term. Therefore, if $\tilde{\lambda} = \lim_{T \rightarrow \infty} \frac{\sqrt{T}}{\sqrt{R}} < \infty$ and for simplicity $N = \infty$, then

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_1 + \tilde{\lambda}^2 V_2)$$

where $V_1 = (\Gamma'W\Gamma)^{-1}\Gamma'W\Phi_1W\Gamma(\Gamma'W\Gamma)^{-1}$ and $V_2 = (\Gamma'W\Gamma)^{-1}\Gamma'W\Phi_2W\Gamma(\Gamma'W\Gamma)^{-1}$.

These results imply that using different draws in different markets is superior because the additional variance term is of smaller order while the bias is of the same order. This advantage of using different draws has been pointed out in other papers that deal with simulation based estimators, such as [McFadden \(1989\)](#) or [Kristensen and Salanié \(2013\)](#).

Remark 4. The expansions are very similar if one uses polynomial-based quadrature rules. However, the nodes are not random in these cases and hence, one cannot use laws of large numbers or central limit theorems to deal with the terms Q_{2T} and Q_{3T} . Since none of these terms has a mean of 0 with nonstochastic approximations, quadrature rules lead to an additional bias term only and one obtains asymptotic normality if the number of nodes grows fast enough relative to T . Quantifying the order of the bias is beyond the scope of this paper.

Remark 5. The main difference between the asymptotics in the number products ([Berry et al., 2004](#)) and the number of markets is that with many products all market shares are assumed to converge to 0, which leads to particular challenges for the asymptotic theory. [Berry et al. \(2004\)](#) assume that the same draws are used to approximate the predicted shares for all products and markets. Under this assumption, with many products $\frac{J^2}{R}$ needs to be bounded for \sqrt{J} consistency, while I require that $\frac{T}{R}$ needs to be bounded for \sqrt{T} consistency. The additional factor of J in the fraction appears because all market shares are assumed to converge to 0.

2.3. Intuition with $J_t = 1$

To get a better sense of where the additional bias and variance terms come from, I now present an intuitive outline of the asymptotic normality proof with $J_t = 1$. The intuition for the general case is very similar. To simplify the notation, I drop x_t as an argument from all functions. Since we are dealing with a GMM estimator, I show that the first order condition can be written as

$$\sqrt{T}(\hat{\theta} - \theta_0) = \left((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1) \right) \sqrt{T}G_T(\theta_0, P_R, s^N) \quad (5)$$

where

$$G_T(\theta_0, P_R, s^N) = \frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta_0, P_{Rt}, s_t^N).$$

Now consider $G_T(\theta_0, P_R, s^N)$ and write

$$\begin{aligned} G_T(\theta_0, P_R, s^N) &= G_T(\theta_0, P_0, s) + G_T(\theta_0, P_R, s) \\ &\quad - G_T(\theta_0, P_0, s) + G_T(\theta_0, P_R, s^N) - G_T(\theta_0, P_R, s). \end{aligned}$$

The first term would be the only term left if the integral was calculated exactly and if the true shares were observed. Since $E(z_t' \xi_t(\theta_0, P_{0t}, s_t)) = 0$, the assumptions imply that $\sqrt{T}G_T(\theta_0, P_0, s)$ converges to a normally distributed random variable. Next write

$$\begin{aligned} G_T(\theta_0, P_R, s) - G_T(\theta_0, P_0, s) &= \frac{1}{T} \sum_{t=1}^T z_t' (\xi_t(\theta_0, P_{Rt}, s_t) - \xi_t(\theta_0, P_{0t}, s_t)). \quad (6) \end{aligned}$$

This difference is not 0 because of the approximation error of the integral. To quantify the effect of the simulation error I use a Taylor expansion. I first rewrite $\xi_t(\theta_0, P_{Rt}, s_t)$ as a function of $\sigma_t(\theta, \xi_t, P_{0t})$ and $\xi_t(\theta_0, P_{0t}, s_t)$ as a function of $\sigma_t(\theta, \xi_t, P_{Rt})$, where σ_t is defined in Eq. (3), using invertibility of this predicted share function. Since $\xi_t = \xi_t(\theta_0, P_{0t}, s_t)$, we get $s_t = \sigma_t(\theta_0, \xi_t, P_{0t}) = \sigma_t(\theta_0, \xi_t(\theta_0, P_{Rt}, s_t), P_{Rt})$ and thus

$$\begin{aligned} \xi_t(\theta_0, P_{Rt}, s_t) - \xi_t(\theta_0, P_{0t}, s_t) &= \sigma_t^{-1}(\theta_0, \sigma_t(\theta_0, \xi_t, P_{0t}), P_{Rt}) - \sigma_t^{-1}(\theta_0, \sigma_t(\theta_0, \xi_t, P_{Rt}), P_{Rt}). \end{aligned}$$

With this expression, I use a Taylor approximation to show that

$$\begin{aligned} \xi_t(\theta_0, P_{Rt}, s_t) &= \xi_t(\theta_0, P_{0t}, s_t) - H_{0t}^{-1}e_{Rt} - \frac{1}{2}I_{0t}(e_{Rt})^2 \\ &\quad + H_{0t}^{-2} \left(\frac{\partial e_{Rt}}{\partial \xi} \right) e_{Rt} + o_p \left(\frac{\sqrt{T}}{R} \right) \quad (7) \end{aligned}$$

where H_{0t} and I_{0t} constants,

$$e_{Rt} \equiv \sigma_t(\theta, \xi_t, P_{Rt}) - \sigma_t(\theta, \xi_t, P_{0t}),$$

and

$$\frac{\partial e_{Rt}}{\partial \xi} \equiv \frac{\partial \sigma_t(\theta, \xi_t, P_{Rt})}{\partial \xi} - \frac{\partial \sigma_t(\theta, \xi_t, P_{0t})}{\partial \xi}.$$

The first three terms on the right hand side of Eq. (7) result from the Taylor expansion while the fourth term arises because P_{Rt} enters $\sigma_t^{-1}(\theta_0, \sigma_t(\theta_0, \xi_t, P_{Rt}), P_{Rt})$ both through σ_t and as the last argument. Plugging this expansion back into (6) gives

$$\begin{aligned} \sqrt{T}(G_T(\theta_0, P_R, s) - G_T(\theta_0, P_0, s)) &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' H_{0t}^{-1} e_{Rt} \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' \left(H_{0t}^{-2} \left(\frac{\partial e_{Rt}}{\partial \xi} \right) e_{Rt} - \frac{1}{2} I_{0t} (e_{Rt})^2 \right) + o_p \left(\frac{\sqrt{T}}{R} \right). \end{aligned}$$

The first term,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' H_{0t}^{-1} e_{Rt} &= \frac{1}{\sqrt{T}} \frac{1}{R} \sum_{t=1}^T \sum_{r=1}^R z_t' H_{0t}^{-1} \\ &\quad \times (v_{jt}(\theta, \xi_t, v_{rt}) - \sigma_t(\theta, \xi_t, P_{0t})), \end{aligned}$$

is $O_p\left(\frac{1}{\sqrt{R}}\right)$ and converges to a normally distributed random variable when multiplied by \sqrt{R} . The second term does not have mean zero and it can be shown that the assumptions imply that

$$\frac{R}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' \left(H_{0t}^{-2} \left(\frac{\partial e_{Rt}}{\partial \xi} \right) e_{Rt} - \frac{1}{2} I_{0t} (e_{Rt})^2 \right) \xrightarrow{p} \bar{\mu}.$$

Hence, it is $O_p\left(\frac{\sqrt{T}}{R}\right)$ and converges in probability to a constant when multiplied by $\frac{R}{\sqrt{T}}$. Thus,

$$\begin{aligned} & \sqrt{T} (G_T(\theta_0, P_R, s) - G_T(\theta_0, P_0, s)) \\ &= O_p\left(\frac{1}{\sqrt{R}}\right) + O_p\left(\frac{\sqrt{T}}{R}\right) + o_p\left(\frac{\sqrt{T}}{R}\right). \end{aligned}$$

Similarly, using that $\xi_t(\theta_0, P_{Rt}, s_t^N) = \sigma_t^{-1}(\theta_0, s_t^N, P_{Rt})$ and $\xi(\theta_0, P_{Rt}, s_t) = \sigma_t^{-1}(\theta_0, s_t, P_{Rt})$ a Taylor approximation gives

$$\begin{aligned} & \sqrt{T} (G_T(\theta_0, P_R, s^N) - G_T(\theta_0, P_R, s)) = \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' H_{0t}^{-1} (s_t^N - s_t) \\ & - \frac{1}{2} \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t' I_{0t} (s_t^N - s_t)^2 + o_p\left(\frac{\sqrt{T}}{N}\right). \end{aligned}$$

Again the assumptions imply that the first term on the right hand side converges to a normal distribution when multiplied by \sqrt{N} and the second term converges to constant when multiplied by $\frac{N}{\sqrt{T}}$.

Putting the previous results together yields the conclusions of **Theorem 1**. If R and N approach ∞ at a slower rate than T , the additional bias terms (second order terms in the Taylor expansion) dominate the additional variance terms (first order terms). Furthermore, the additional bias terms yield rate restrictions on R and N relative to T whereas the first order terms converge to 0 as $R \rightarrow \infty$ and $N \rightarrow \infty$ at any rate. In the following Section 1 show how the leading bias term can be removed and how and the additional variance can be taken into account when calculating standard errors.

2.4. Bias and variance corrections

This section shows how the leading bias terms, namely the $O_p\left(\frac{\sqrt{T}}{R}\right)$ term and the $O_p\left(\frac{\sqrt{T}}{N}\right)$ term in the expansion above, can be eliminated by using an analytic bias correction method. Similar methods have been suggested by Lee (1995), Arellano and Hahn (2007), and Kristensen and Salanié (2013) in related setups.⁸ Furthermore, the leading additional variance terms, in particular the $O_p\left(\frac{1}{R}\right)$ term and the $O_p\left(\frac{1}{N}\right)$ terms, can easily be taken into account when calculating standard errors.

2.4.1. Analytic bias correction

Define the bias adjusted estimator as

$$\hat{\theta}_A \equiv \hat{\theta} - \left(\hat{\Gamma}' W_T \hat{\Gamma}\right)^{-1} \hat{\Gamma}' W_T \left(\frac{1}{R} \hat{\mu}_1 + \frac{1}{N} \hat{\mu}_2\right)$$

where

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \theta} z_t' \xi_t(\hat{\theta}, P_{Rt}, s_t, x_t)$$

and expressions for $\hat{\mu}_1$ and $\hat{\mu}_2$, which are sample analogs of $\bar{\mu}_1$ and $\bar{\mu}_2$, are given in Section 3. Subtracting an estimator of $(\Gamma' W \Gamma)^{-1} \Gamma' W \left(\frac{1}{R} \bar{\mu}_1 + \frac{1}{N} \bar{\mu}_2\right)$ from $\hat{\theta}$ eliminates the leading bias term from the asymptotic expansion, which is established by the following theorem.

⁸ An alternative to an analytic bias correction is a jackknife bias correction as in Kristensen and Salanié (2013). However, this estimator is computationally costly because one would have to solve the optimization problem several times. Furthermore, this procedure increases the variance of the estimator. Similarly, the panel jackknife bias correction suggested by Hahn and Newey (2004) is computationally very demanding in this setting.

Theorem 2. Assume that Assumptions RC1–RC9 hold. Then

$$\begin{aligned} & \sqrt{T} (\hat{\theta}_A - \theta_0) = \left((\Gamma' W \Gamma)^{-1} \Gamma' W + o_p(1) \right) \\ & \times \left(Q_{1T} + \frac{1}{\sqrt{R}} Q_{2T} + \frac{1}{\sqrt{N}} Q_{3T} + o_p\left(\frac{\sqrt{T}}{R}\right) + o_p\left(\frac{\sqrt{T}}{N}\right) \right). \end{aligned}$$

As opposed to the results of **Theorem 1**, now $\sqrt{T} (\hat{\theta}_A - \theta_0) \xrightarrow{d} N(0, V_1)$ as long as $\frac{\sqrt{T}}{R}$ is bounded.

2.4.2. Variance correction

The variance of the estimator can be estimated by

$$\hat{V} = \left(\hat{\Gamma}' \hat{W} \hat{\Gamma}\right)^{-1} \hat{\Gamma}' \hat{W} \left(\hat{\Phi}_1 + \frac{1}{R} \hat{\Phi}_2 + \frac{1}{N} \hat{\Phi}_3\right) \hat{W} \hat{\Gamma} \left(\hat{\Gamma}' \hat{W} \hat{\Gamma}\right)^{-1}$$

where $\hat{\Phi}_1$, $\hat{\Phi}_2$, and $\hat{\Phi}_3$ are estimators of Φ_1 , Φ_2 , and Φ_3 provided in Section 3. In this way the variance of the $O_p\left(\frac{1}{R}\right)$ term and the $O_p\left(\frac{1}{N}\right)$ term is taken into account as well. In case a bias adjustment is used, $\hat{\theta}$ can be replaced by $\hat{\theta}_A$.

3. Bias and variance implementation in the BLP model

In this Section 1 explain that both the bias correction and the variance adjustment are easy to implement in the BLP model. In fact, most quantities needed to calculate these adjustments have already been calculated in other parts of the estimation procedure.⁹

First recall that when the distribution of the random coefficients is P_{Rt} , the market shares predicted by the model are

$$\sigma_{jt}(\theta, x_t, \xi_t, P_{Rt}) = \frac{1}{R} \sum_{r=1}^R v_{jt}(\theta, x_t, \xi_t, v_{rt})$$

where

$$v_{jt}(\theta, x_t, \xi_t, v) = \frac{\exp(x_{jt}' \beta + \xi_{jt} + x_{jt}' \Sigma v)}{1 + \sum_{k=1}^{J_t} \exp(x_{kt}' \beta + \xi_{kt} + x_{kt}' \Sigma v)}$$

Also $\sigma_t(\theta, x_t, \xi_t, P_{Rt})$ is the $J_t \times 1$ vector with elements $\sigma_{jt}(\theta, x_t, \xi_t, P_{Rt})$ and $v_t(\theta, x_t, \xi_t, v_{rt})$ is defined analogously. To calculate the variance of $\hat{\theta}$, even without any adjustments, one needs to calculate

$$\hat{\Gamma} = \frac{\partial}{\partial \theta} \frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\hat{\theta}, P_{Rt}, s_t^N, x_t).$$

This can be done by using the implicit function theorem which yields

$$\begin{aligned} \hat{\Gamma} &= -\frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \sigma_t(\hat{\theta}, x_t, \xi_t(\hat{\theta}, P_{Rt}, s_t^N, x_t), P_{Rt})}{\partial \xi_t} \right)^{-1} \\ & \times \frac{\partial \sigma_t(\hat{\theta}, x_t, \xi_t(\theta, P_{Rt}, s_t^N, x_t), P_{Rt})}{\partial \theta}. \end{aligned}$$

Let $\hat{\xi}_t = \xi_t(\hat{\theta}, P_{Rt}, s_t^N, x_t)$. Notice that $\hat{\xi}_t$ is needed in the estimation procedure when calculating the GMM objective function. When MPEC is used, $\hat{\xi}_t$ are parameters. Hence, $\hat{\Gamma}$ mainly involves

⁹ The code is available on the author's homepage: <http://www.ssc.wisc.edu/~jfreyberger>.

derivatives of the predicted shares, which are easy to calculate. For example,

$$\begin{aligned} \frac{\partial \sigma_t(\theta, x_t, \xi, P_{Rt})}{\partial \xi_t} &= \frac{1}{R} \sum_{r=1}^R \frac{\partial v_t(\theta, x_t, \xi, v_{rt})}{\partial \xi_t} \\ &= \frac{1}{R} \sum_{r=1}^R \text{diag}(v_t(\theta, x_t, \xi, v_{rt})) \\ &\quad - v_t(\theta, x_t, \xi, v_{rt}) v_t(\theta, x_t, \xi, v_{rt})' \end{aligned}$$

where $\text{diag}(v_t)$ is a diagonal matrix with v_t on the diagonal. Similarly, one can calculate the derivative of the shares with respect to θ . Since these derivatives are already needed to calculate the unadjusted standard errors, they have been discussed in detail by Nevo (2000). Next define

$$\hat{H}_t = \left(\frac{\partial \sigma_t(\hat{\theta}, x_t, \hat{\xi}_t, P_{Rt})}{\partial \xi} \right)$$

and

$$\hat{I}_{jt} = \sum_{k=1}^{J_t} \hat{H}_t^{-1} \left(\frac{\partial^2 \sigma_t(\hat{\theta}, x_t, \hat{\xi}_t, P_{Rt})}{\partial \xi \partial \xi_k} \right) \hat{H}_t^{-1} e_{jt} e_{kt}' \hat{H}_t^{-1}$$

where e_{jt} denotes the j th column of the $J_t \times J_t$ identity matrix. To calculate \hat{I}_{jt} , we need the second derivative of the predicted shares. It is easy to show that

$$\begin{aligned} \frac{\partial^2 \sigma_t(\hat{\theta}, x_t, \hat{\xi}_t, P_{Rt})}{\partial \xi \partial \xi_k} &= \frac{1}{R} \sum_{r=1}^R \text{diag} \left(\frac{\partial v_t(\theta, x_t, \xi_t, v_{rt})}{\partial \xi_k} \right) \\ &\quad - \left(\frac{\partial v_t(\theta, x_t, \xi_t, v_{rt})}{\partial \xi_k} \right) v_t(\theta, x_t, \xi_t, v_{rt})' \\ &\quad - v_t(\theta, x_t, \xi_t, v_{rt}) \left(\frac{\partial v_t(\theta, x_t, \xi_t, v_{rt})}{\partial \xi_k} \right)'. \end{aligned}$$

These are all the derivatives that we need to calculate. All adjustments are now simply functions of the previous quantities. So simplify the expression of the adjustments, define the demeaned version of v_t and its derivative by

$$\bar{v}_t(\theta, x_t, \xi_t, v) = v_t(\theta, x_t, \xi_t, v) - \sigma_t(\theta, x_t, \xi_t, P_{Rt})$$

and

$$\frac{\partial \bar{v}_t(\theta, x_t, \xi_t, v)}{\partial \xi} = \frac{\partial v_t(\theta, x_t, \xi_t, v)}{\partial \xi} - \frac{\partial \sigma_t(\theta, x_t, \xi_t, P_{Rt})}{\partial \xi}$$

as well as the demeaned individual purchasing decisions by

$$\bar{s}_{tm} = s_{tm} - \frac{1}{N} \sum_{n=1}^N s_{tm}.$$

Now recall that the variance of the estimator can be estimated by

$$\hat{V} = (\hat{r}' \hat{W} \hat{r})^{-1} \hat{r}' \hat{W} \left(\hat{\phi}_1 + \frac{1}{R} \hat{\phi}_2 + \frac{1}{N} \hat{\phi}_3 \right) \hat{W} \hat{r} (\hat{r}' \hat{W} \hat{r})^{-1}.$$

Here

$$\hat{\phi}_1 = \frac{1}{T} \sum_{t=1}^T z_t' \hat{\xi}_t \hat{\xi}_t' z_t,$$

$$\hat{\phi}_2 = \frac{1}{R} \frac{1}{T} \sum_{r=1}^R \sum_{t=1}^T z_t' \hat{H}_t^{-1} \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt}) \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt})' \hat{H}_t^{-1} z_t,$$

$$\hat{\phi}_3 = \frac{1}{N} \frac{1}{T} \sum_{n=1}^N \sum_{t=1}^T z_t' \hat{H}_t^{-1} \bar{s}_{tm} \bar{s}_{tm}' \hat{H}_t^{-1} z_t.$$

We also have all ingredients to calculate the bias correction. To do so define

$$\hat{S}_{1jt} = \frac{1}{R} \sum_{r=1}^R \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt})' \hat{I}_{jt} \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt}),$$

$$\hat{S}_{2jt} = \frac{1}{n} \sum_{n=1}^N \bar{s}_{tm}' \hat{I}_{jt} \bar{s}_{tm},$$

$$\hat{C}_t = \frac{1}{R} \sum_{r=1}^R \hat{H}_t^{-1} \frac{\partial \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt})}{\partial \xi} \hat{H}_t^{-1} \frac{\partial \bar{v}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_{rt})}{\partial \xi}.$$

All of the matrix are simply transformations of the first two derivatives of the predicted shares. The estimated bias terms are

$$\hat{\mu}_1 = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} z_{jt}' \left(e_j' \hat{C}_t - \frac{1}{2} \hat{S}_{1jt} \right)$$

and

$$\hat{\mu}_2 = -\frac{1}{2} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} z_{jt}' \hat{S}_{2jt}.$$

4. Monte Carlo simulations

In this section, I illustrate that the simulation error of the integral will affect the finite sample performance of the estimator because the usual GMM standard errors underestimate the true variance and the estimates are biased. I use the model described in Section 2. The setup for the Monte Carlo simulations is adapted from Dubé et al. (2012) with some changes to accommodate the asymptotics in the number of markets. A similar setup is also used by Skrainka and Judd (2011). The number of products is set to 4, I generate all shares using $N = 3000$ consumer, and I vary the number of markets, T , and draws, R . I include three additional product characteristics, which affect choices. One of the product characteristics does not vary across markets. These product characteristics, x_{kjt} for $k = 1, \dots, 4$, are distributed as $TN(0, 1)$, where TN denotes the standard normal distribution truncated at -2 and 2 and x_{1jt} does not vary over markets. There is also a constant term, $x_{0jt} = 1$, for all j . The unobserved product characteristics are $\xi_{jt} = \frac{1}{2}(\xi_j + \Delta \xi_{jt})$ where

$$\xi_j \sim TN(0, 1), \quad \Delta \xi_{jt} \sim TN(0, 1), \quad j = 1, \dots, 4, \quad t = 1, \dots, T.$$

Hence, also the unobserved product characteristics have a component which does not change over markets. I capture this part by using product dummies as parameters as in Nevo (2001).¹⁰ The price is

$$p_{jt} = \frac{1}{2} \left| 2 + 0.5 \xi_{jt} + e_{jt} + 1.1 \sum_{k=1}^4 x_{kjt} \right|$$

where $e_{jt} \sim TN(0, 1)$. There is a random coefficient on all product characteristics including price and the constant term. The random coefficient are distributed as follows

$$\begin{pmatrix} \beta_i^0 \\ \beta_i^1 \\ \beta_i^2 \\ \beta_i^3 \\ \beta_i^4 \\ \beta_i^p \end{pmatrix} \sim N \left(\begin{pmatrix} 3 \\ 1.5 \\ 1.5 \\ 1 \\ 0.5 \\ -2 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{pmatrix} \right).$$

¹⁰ See footnote 5 for a discussion on using product dummies as parameters.

Here β_i^k and β_i^p denote the random coefficients on x_{kjt} and p_{jt} , respectively.¹¹ Allowing for either correlated product characteristics or correlated random coefficients does not change the simulation results much. For each product j in market t , I generate 3 instruments

$$z_{ljt} = a_{ljt} + 0.25 \left(e_{jt} + 1.1 \sum_{k=1}^4 x_{kjt} \right),$$

$$l = 1, \dots, 3, j = 1, \dots, 4, t = 1, \dots, T,$$

where $a_{ljt} \sim U(0, 1)$. Next to z_{ljt} , x_{2jt} , x_{3jt} , and product dummies, I also use z_{ljt}^2 and z_{ljt}^3 for all l , x_{2jt}^2 , x_{3jt}^2 , x_{4jt}^2 , x_{2jt}^3 , x_{3jt}^3 , x_{4jt}^3 as well as $x_{2jt}z_{ljt}$, $x_{3jt}z_{ljt}$ and $x_{4jt}z_{ljt}$ as instruments.¹² Just as Su and Judd (2012) and Dubé et al. (2012) I use a reduced form process for the price and the instruments instead of generating them from a structural model. Consequently, the instruments here might be stronger than those commonly found in applications.

I make use of the Mathematical Programming with Equilibrium Constraints (MPEC) approach (see Su and Judd, 2012 and Dubé et al., 2012) to estimate the BLP model, where I use KNITRO (Version 8.1) to solve the constrained optimization problem. I supply the solver with analytic gradients for the objective function and the constraints but not with the Jacobian. I use the same tolerance levels for the constraints and the objective function as Dubé et al. (2012).^{13, 14} I also use a similar method as Dubé et al. (2012) to obtain 5 different starting values for each simulated data set. That is, I first draw values for the standard deviations of the random coefficients from a $U(0, 1)$ distribution. I then obtain the starting values for the means (and the other parameters for the MPEC procedure) with a two stage least squares procedure taking the variances as given. The starting values correspond to the minimum of the objective function (10) for a given value of the variances. As a consequence, all starting values are feasible. In the majority of the simulated data sets, even when $R = 50$, most starting values yield the same minimum of the objective function. For example, when $R = 50$ and $T = 400$ all five starting values yield the same minimum in over 60% of the simulated data sets. Moreover, when R increases it is more likely that all starting values converge to the same solution.¹⁵

Below I investigate the actual coverage rate of a 95% nominal confidence interval for $\beta^p = E(\beta_i^p) = -2$ using bias correction methods as well as standard errors with and without correcting

for the approximation errors. The usual GMM standard errors are estimated using

$$\hat{v}_1 = \left(\hat{\Gamma}' \hat{W} \hat{\Gamma} \right)^{-1} \hat{\Gamma}' \hat{W} \hat{\phi}_1 \hat{W} \hat{\Gamma} \left(\hat{\Gamma}' \hat{W} \hat{\Gamma} \right)^{-1}$$

and the adjusted standard errors are calculated using

$$\hat{v} = \left(\hat{\Gamma}' \hat{W} \hat{\Gamma} \right)^{-1} \hat{\Gamma}' \hat{W} \left(\hat{\phi}_1 + \frac{1}{R} \hat{\phi}_2 + \frac{1}{N} \hat{\phi}_3 \right) \hat{W} \hat{\Gamma} \left(\hat{\Gamma}' \hat{W} \hat{\Gamma} \right)^{-1}.$$

I also compare the mean lengths of the confidence intervals obtained from simulations with and without variance corrections. I simulate the data with 50,000 draws from the joint distribution of random coefficients. I treat these 50,000 draws as the true distribution and then take 50–800 demeaned random draws from this distribution depending on the setup. I make use of 50–800 markets. I use different draws to approximate the integral in different markets. Hence, with 100 markets and 800 draws I sample in total 80,000 times from the true distribution. The computational costs mainly depend on the number of draws that are used to evaluate each integral, which is 800 in this case. All coverage rates are based on 1000 Monte Carlo iterations, i.e. 1000 simulated data sets.

Table 1 shows that the number of draws affects the actual coverage rate of a nominal 95% confidence interval if the usual GMM asymptotic distribution is employed. For example with 800 markets the actual coverage rate is only 74.5% with 50 draws, while it increases to 91.1% with 800 draws. Using the bias corrected estimator leads to an improvement in these coverage rates. However, in general there is still a large difference between using a small number and large number of draws. For instance, with 800 markets and 50 draws one obtains a coverage rate of 85.3%, while 800 draws yield a coverage rate of 94.8%. The same holds when simulation adjusted standard errors but no bias adjustment is used. In case one uses both the analytical bias adjustment and the standard error adjustment, the coverage rate is very close to 95% even with a small number of draws. For example with 800 markets and 50 draws the coverage rate is 92.0%. Therefore, as mentioned by Skrainka and Judd (2011), the usual GMM standard errors are too tight when the number of draws is small, but the adjusted standard seems to be of the right magnitude.

The costs of the improved coverage rates are wider confidence intervals. Table 2 shows the mean length of confidence intervals with adjusted standard errors divided by the mean length with unadjusted standard errors. The means are taken over the 1000 simulated data sets. When the number of draws is 50 the adjusted confidence intervals are up to 25% larger. When the number of draws increases, this ratio decreases. With 800 draws, the difference is less than 10%. Thus, corrected standard errors only have a large effect on coverage rates and the lengths of the confidence intervals when the number of simulations R is small.

Table 3 shows the finite sample bias with and without bias correction. The finite sample bias of the GMM point estimates decreases as R increases and as T increases. Especially if T is small, the finite sample bias is still substantial even when R is large, because this nonlinear estimator is biased in finite samples even if $R = \infty$. The bias correction reduces the finite sample bias significantly. For the bias corrected point estimates, there is no significant difference in the bias when R is large and when R is small (especially when T is large), which suggest that the bias correction eliminates most of the bias due to the simulation error. The remaining bias is due to the nonlinear objective function.

Finally, Table 4 shows the ratio of the mean squared error of $\hat{\beta}^p$ when the same and different draws are used to approximate the integral in different markets. When the number of draws is small, the mean squared error with the same draws is up to 37.2% larger (with 400 markets). As the number of draws increases, this

¹¹ Normally distributed random coefficients imply that $P(\beta_i^p > 0) > 0$. Here $P(\beta_i^p > 0) = 0.00234$.

¹² As Dubé et al. (2012) and Skrainka and Judd (2011) I use standard powers of instruments and regressors. Orthogonal polynomial are an alternative choice. Notice that here all instruments are symmetrically distributed around 0 and thus, z_{ljt} and z_{ljt}^2 as well as z_{ljt}^3 and z_{ljt}^3 are uncorrelated.

¹³ The code of Dubé et al. (2012), which uses the same draws in each market, can be downloaded from Dubé's website: <http://faculty.chicagobooth.edu/jean-pierre.dube/research/MPECcode.html>.

¹⁴ I ran the programs on the Northwestern Social Sciences Computing Cluster, which uses 260 AMD Opteron CPU cores running at 2.8 GHz, with MATLAB R2012b. The programs use up to 6 GB of memory.

¹⁵ I adapted the first data generating process (DGP) from Dubé et al. (2012) by adding a constant term (i.e. 2) when generating the price. I do so because otherwise the instruments and the price are uncorrelated (although the product of two instruments is correlated with the price). Dubé et al. (2012) also use a second DGP where instruments and prices are correlated but where the variation in price is a lot larger (around 7 compared to 0.53 in my DGP). As a consequence, Dubé et al. (2012) get very large variations of market shares over markets and many market shares very close to 0. These differences in the DGPs might be the reason why Skrainka and Judd (2011), who use the second DGP, find many false local optima with Monte Carlo integration, whereas I do not experience this problem. Also notice that as opposed to Skrainka and Judd (2011), I use different draws to approximate the integrals in different markets.

Table 1
Coverage rates of 95% confidence intervals for β^p .

	50 draws	100 draws	200 draws	400 draws	800 draws
GMM point estimates and GMM standard errors					
50 markets	0.784	0.860	0.849	0.856	0.904
100 markets	0.786	0.839	0.889	0.890	0.885
200 markets	0.798	0.849	0.885	0.885	0.915
400 markets	0.797	0.851	0.896	0.898	0.909
800 markets	0.745	0.847	0.883	0.903	0.911
Bias corrected point estimates and GMM standard errors					
50 markets	0.833	0.908	0.893	0.893	0.929
100 markets	0.869	0.892	0.937	0.934	0.926
200 markets	0.889	0.909	0.940	0.927	0.949
400 markets	0.892	0.930	0.943	0.946	0.947
800 markets	0.853	0.909	0.932	0.951	0.948
GMM point estimates and adjusted standard errors					
50 markets	0.870	0.908	0.893	0.891	0.924
100 markets	0.865	0.891	0.925	0.928	0.919
200 markets	0.888	0.893	0.923	0.911	0.933
400 markets	0.865	0.902	0.921	0.924	0.933
800 markets	0.821	0.893	0.907	0.932	0.930
Bias corrected point estimates and adjusted standard errors					
50 markets	0.911	0.943	0.920	0.913	0.942
100 markets	0.924	0.939	0.957	0.959	0.937
200 markets	0.939	0.948	0.958	0.947	0.965
400 markets	0.943	0.945	0.960	0.963	0.963
800 markets	0.920	0.947	0.954	0.966	0.964

The nominal coverage rate is 0.95 and the number of Monte Carlo iterations is 1000. The true value of β^p is -2 . If the actual coverage rate is 95%, the standard error with 1,000 simulations is around 0.0069. If the actual coverage rate is 80%, the standard error increases to 0.0126.

Table 2
Ratio of mean length of unadjusted and adjusted confidence intervals for β^p .

	50 draws	100 draws	200 draws	400 draws	800 draws
50 markets	1.251	1.174	1.129	1.110	1.089
100 markets	1.239	1.179	1.134	1.112	1.093
200 markets	1.246	1.178	1.134	1.108	1.095
400 markets	1.244	1.183	1.140	1.113	1.097
800 markets	1.240	1.180	1.137	1.112	1.098

Table 3
Finite sample bias of $\hat{\beta}^p$.

	50 draws	100 draws	200 draws	400 draws	800 draws
GMM point estimates					
50 markets	-0.131	-0.110	-0.127	-0.129	-0.126
100 markets	-0.112	-0.101	-0.095	-0.097	-0.100
200 markets	-0.091	-0.083	-0.080	-0.081	-0.085
400 markets	-0.087	-0.070	-0.061	-0.065	-0.062
800 markets	-0.078	-0.060	-0.056	-0.054	-0.052
Bias corrected point estimates					
50 markets	-0.076	-0.065	-0.087	-0.092	-0.091
100 markets	-0.055	-0.052	-0.052	-0.057	-0.063
200 markets	-0.022	-0.032	-0.036	-0.040	-0.045
400 markets	-0.019	-0.014	-0.014	-0.021	-0.021
800 markets	-0.013	-0.006	-0.009	-0.010	-0.010

Table 4
Ratio of mean squared error of $\hat{\beta}^p$ with same and different draws.

	50 draws	100 draws	200 draws	400 draws
50 markets	0.937	1.134	1.022	1.031
100 markets	1.099	0.995	1.116	1.121
200 markets	1.363	1.162	1.160	0.977
400 markets	1.372	1.203	1.204	1.056

ratio converges to 1. Table A1 in the supplementary appendix (see Appendix A) presents additional simulation results when the same draws are used, which show that the adjustments are

also important in this case. The unadjusted confidence intervals generally have a lower coverage rate than the ones obtained with different draws, especially when R is small relative to T . The adjustments again yield coverage rates close to 95% in all cases. When the same draws are used the adjustments are more costly. In particular, the ratio of the mean length of the adjusted and unadjusted confidence intervals now depends on T and R (as opposed to only R with different draws) and adjusted confidence intervals are up to 72.4% larger. As a consequence, adjusted confidence intervals are larger when the same draws are used compared to using different draws. For example, when $T = 400$ and $R = 50$, the mean length of adjusted confidence intervals with

the same draws is 24.5% larger. These results are all in line with the theoretical findings in this paper and verify that different draws should be used to calculate the integrals in different markets.

Clearly, many parameter choices drive the results in this Monte Carlo study. For example, a low variance of the error term (relative to the variance of the product characteristics) or strong instruments yield more precise estimates for a given number of markets. Furthermore, the effect of the number of draws depends on the variance of the random coefficients relative to the variance of the product characteristics. A high variance of the random coefficient implies that we need many draws to eliminate the effect of the second term of the asymptotic expansion. It is therefore hard to give a general guideline of how many draws (or how many markets) suffice to obtain satisfactory results. The simulation results, however, demonstrate that practitioners should use the bias correction and the adjusted standard errors in application. If the number of draws is sufficiently large, the bias correction is small and the adjusted standard errors will be very close to the GMM standard errors. If the number of draws is small, the simulation error will affect the finite sample performance of the estimator and using the usual GMM asymptotic distribution yields biased estimators and underestimation of the true variance. Since a large number of draws improves the precision of the initial estimator, which is in turn used to calculate the bias correction, the number of draws should be as large as possible, subject to computational constraints and data availability.

5. Conclusion

This paper develops asymptotic theory for estimated parameters in differentiated product demand systems with a small number of products and a large number of markets T . The asymptotic theory takes into account that the predicted market shares are approximated by Monte Carlo integration with R draws and that the observed market shares are approximated from a sample of N consumers. Both approximations affect the asymptotic distribution, because they both lead to a bias and a variance term in the asymptotic expansion of the estimator. I show that when R and N do not increase faster than the number of markets, the bias terms dominate the variance terms. In this case, the asymptotic distribution might not be centered at 0 and standard confidence intervals do not have the right size, even asymptotically. These findings differ from the setup with a large number of products where the variance term always dominates.

I propose both bias corrections and variance adjustments in order to take the approximation errors into account. I then demonstrate with Monte Carlo simulations that these adjustments, which are easy to compute, should be used in applications. In particular, I show that if the number of draws and the number of consumers are sufficiently large, the bias correction is small and the adjusted standard errors will be very close to the GMM standard errors. However, if the number of draws or the number of consumers is small, the approximations will affect the finite sample performance of the estimator and using the usual GMM asymptotic distribution yields biased estimators and underestimation of the true variance. As a consequence the coverage rate of confidence intervals can be significantly below the nominal rate. The estimates and confidence intervals with the adjustments do not suffer from these issues.

Appendix A. Useful lemmas

I use the following lemmas to prove the main results. The proofs are in the supplementary appendix (see Appendix A).

Lemma A.1. Let $f(x, \theta, v) : \mathcal{X} \times \Theta \times \mathbb{R}^{d_v} \rightarrow [-M_1, M_2] \subset \mathbb{R}$ be a continuously differentiable function in all arguments where \mathcal{X} is a compact subset of \mathbb{R}^{d_x} and Θ is a compact subset of \mathbb{R}^{d_θ} . Let v_{1t}, \dots, v_{Rt} be iid draws from $P_t \in \mathcal{P}$. Let $x_t \in \mathcal{X}$ denote the (random) data. Let $\hat{\mathcal{P}} = \{P^1, \dots, P^k\}$ be a finite set of distributions. Assume that

- (i) $\ln(T)/R(T) \rightarrow 0$ as $T \rightarrow \infty$,
- (ii) v_{rt} and x_t are independent, and
- (iii) Any random variable v_t with distribution P_t satisfies the following: $v_t = g(a_t, w_t)$ where $g : \mathcal{A} \times \mathbb{R}^{d_w} \rightarrow \mathbb{R}^{d_v}$ is continuous in both arguments, a_t is a vector of constants and $\mathcal{A} \subset \mathbb{R}^{d_a}$ is a compact set, w_t is a random vector and the distribution of each element of w_t either (a) is in $\hat{\mathcal{P}}$ and is the same for all t or (b) the support is a subset of $[M_l, M_u]$ for some constants M_l and M_u .

Then

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \left| \frac{1}{R} \sum_{r=1}^R f(x_t, \theta, v_{rt}) - \int f(x_t, \theta, v) dP_t(v) \right| \xrightarrow{p} 0$$

as $T \rightarrow \infty$.

Lemma A.2. Let Y_1, Y_2, \dots be independent random variables with mean 0. Let K be an even integer. Then

$$E \left(\left| \sum_{r=1}^R Y_r \right|^K \right) \leq A_K R^{(K/2-1)} \sum_{r=1}^R E (|Y_r|^K)$$

where A_K is a universal constant depending only on K .

Lemma A.3. Suppose that $(v_1^{(t)}, \dots, v_R^{(t)}, x_t)$ with $t = 1, \dots, T$ are random vectors such that $v_r^{(t)}$ is iid across r and independent of x_t . Let

$$y_{Rt} = s(x_t) \left[\frac{1}{R} \sum_{r=1}^R p(v_r^{(t)}, x_t) \right]^{m_1} \left[\frac{1}{R} \sum_{r=1}^R q(v_r^{(t)}, x_t) \right]^{m_2}$$

where m_1 and m_2 are nonnegative integers and s, p and q are measurable functions such that

$$E(p(v_r^{(t)}, x_t) | x_t) = E(q(v_r^{(t)}, x_t) | x_t) = 0$$

for all t . Also assume that if $m_1 > 0$ and $m_2 > 0$ for some a and b satisfying $\frac{1}{a} + \frac{1}{b} = 1$ it holds that for some finite M

$$E(|s(x_t)|^{2a} p(v_r^{(t)}, x_t)^{2am_1}) \leq M$$

and

$$E(|s(x_t)|^{2b} q(v_r^{(t)}, x_t)^{2bm_2}) \leq M.$$

If $m_1 > 0$ and $m_2 = 0$ assume instead that

$$E(|s(x_t)|^2 p(v_r^{(t)}, x_t)^{2m_1}) \leq M.$$

Then

$$\frac{1}{T} \sum_{t=1}^T |y_{Rt}| = O_p(R^{-(m_1+m_2)/2}).$$

Lemma A.4. Suppose $F: \mathbb{R}^J \rightarrow \mathbb{R}$ is a $k + 1$ times continuously differentiable function on an open convex set $S \subseteq \mathbb{R}^J$. For any multi-index $\alpha \in \mathbb{N}_+^J$, let $|\alpha| = \alpha_1 + \dots + \alpha_j$ and $\alpha! = \alpha_1! \dots \alpha_j!$. Furthermore, for $x \in \mathbb{R}^J$ let

$$x^{|\alpha|} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_j^{\alpha_j}$$

and

$$\frac{\partial^{|\alpha|} F(x)}{\partial x^{|\alpha|}} = \frac{\partial^{|\alpha|} F(x)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_j^{\alpha_j}}$$

If $a \in S$ and $a + h \in S$, then for some $c \in (0, 1)$

$$F(a + h) = \sum_{\alpha \in \mathbb{N}_+^J: |\alpha| \leq k} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} F(x)}{\partial x^{|\alpha|}} \Big|_{x=a} h^{|\alpha|} \right) + \sum_{\alpha \in \mathbb{N}_+^J: |\alpha| = k+1} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} F(x)}{\partial x^{|\alpha|}} \Big|_{x=a+ch} \right) h^{|\alpha|}$$

Appendix B. A general model

In this section I present a general model, high level assumptions, and consistency and asymptotic normality results. The BLP is a special case of this general setup and I prove Theorem 1 by verifying the assumptions in this section. I make use of the following notation. I denote the norm of a $a \times b$ matrix A by $\|A\| = \text{tr}(A'A)^{1/2}$ and I denote a neighborhood of a vector $a_0 \in \mathbb{R}^{d_a}$ by

$$\mathcal{N}_{a_0}(\varepsilon) \equiv \{a \in \mathbb{R}^{d_a} : \|a - a_0\| \leq \varepsilon\}.$$

Let $x_t \in \mathbb{R}^{d_x}$, $\xi_t \in \mathbb{R}^{d_\xi}$, and $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ be defined as in Section 2. Suppose there exists a d_v dimensional random vector v with distribution function P_{0t} , which may change over markets. I assumed that P_{0t} is known, but $P_t \in \mathcal{P}$ where \mathcal{P} is a space of probability distributions, which is restricted in the assumptions that follow. For given product characteristics (x_t, ξ_t) , a parameter value θ , and a distribution P_t , the $J_t \times 1$ vector of market shares in market t predicted by the model is denoted by $\sigma_t(\theta, x_t, \xi_t, P_t)$. Define the function $v_t : \mathbb{R}^{d_x} \times \Theta \times \mathbb{R}^{d_v} \rightarrow \mathbb{R}^{J_t}$. I assume that the predicted shares have the form

$$\sigma_t(\theta, x_t, \xi_t, P_t) = \int v_t(\theta, x_t, \xi_t, v) dP_t(v), \quad \forall t = 1, \dots, T.$$

The $J_t \times 1$ vector of true market shares in market t is denoted by s_t and I assume that

$$s_t = \int v_t(\theta_0, x_t, \xi_t, v) dP_{0t}(v), \quad \forall t = 1, \dots, T \tag{8}$$

for some θ_0 and $P_{0t} \in \mathcal{P}$ which is analogous to the discussion in Section 2 but without specific functional form or distributional assumptions. In particular, in Section 2

$$v_{jt}(\theta, x_t, \xi_t, v) = \frac{\exp(x'_{jt} \beta + \xi_{jt} + x'_{jt} \Sigma v)}{1 + \sum_{k=1}^{J_t} \exp(x'_{kt} \beta + \xi_{kt} + x'_{kt} \Sigma v)}$$

A different functional form for v_{jt} arises for example if ϵ_{ijt} in (1) is not extreme value distributed.

To simplify notation, I suppress x_t and refer to $v_t(\theta, x_t, \xi_t, v)$ as $v_t(\theta, \xi_t, v)$ and to $\sigma_t(\theta, x_t, \xi_t, P_t)$ as $\sigma_t(\theta, \xi_t, P_t)$. The j th element of $\sigma_t(\theta, \xi_t, P_t)$ is denoted by $\sigma_{jt}(\theta, \xi_t, P_t)$. To state the moment conditions and the estimator of θ_0 , I make the following assumptions.

Assumption A.1. For any pair $(P, \theta) \in \mathcal{P} \times \Theta$ and vector $(m_t, s_t) \in \mathcal{M} \times (0, 1)^{J_t}$ there exists a unique solution ξ to $s_t - \sigma_t(\theta, \xi, P) = 0$. This solution is denoted by $\xi_t(\theta, P, s_t, x_t)$ and abbreviated by $\xi_t(\theta, P, s_t)$.

Next define the function

$$G_t(\theta, P_t) \equiv E(z_t' \xi_t(\theta, P_t, s_t)),$$

and the sample moment

$$G_T(\theta, P, s) \equiv \frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_t, s_t)$$

where $z_t \in \mathbb{R}^{J_t \times d_z}$ is a vector of instrumental variables. The moment conditions of the model are

$$E(z_t' \xi_t(\theta_0, P_{0t}, s_t)) = 0. \tag{9}$$

The function $\sigma_t(\theta, \xi_t, P_{0t})$ involves an integral, which is approximated using the empirical probability measure P_{Rt} of an iid sample v_{1t}, \dots, v_{Rt} from P_{0t} . We may also not observe the true market shares s_t but instead an approximation from a sample of N consumers. That is, for $n = 1, \dots, N$ and $t = 1, \dots, T$, we observe $s_{tn} \in \{0, 1\}^{J_t}$ with $E(s_{tn}) = s_t$ and $\sum_{j=1}^{J_t} s_{jtn} \leq 1$. Define the observed approximated market shares by $s_t^N \equiv \frac{1}{N} \sum_{n=1}^N s_{tn}$.

I assume that the number of draws, R , and the number of consumers, N , are functions of T and all limits are taken as $T \rightarrow \infty$. Now, define the estimator

$$\hat{\theta} \equiv \arg \min_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{Rt}, s_t^N) \right)' \times W_T \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta, P_{Rt}, s_t^N) \right). \tag{10}$$

The estimator $\hat{\theta}$ is the one used in practice and its asymptotic properties are analyzed here.

B.1. Consistency

I first provide general assumptions, which are sufficient conditions for consistency.

Assumption A.2. The market shares are generated by (8) and moment conditions (9) hold.

Assumption A.3. $J_t \leq \bar{J}$ for some integer \bar{J} .

Assumption A.4. The data $\{z_t, x_t, p_t, s_t\}_{t=1}^T$ are independent across markets.

Assumption A.5. The matrix $\frac{1}{T} \sum_{t=1}^T z_t' z_t$ has full rank and is stochastically bounded, i.e. for all $\varepsilon > 0$ there exists an $M(\varepsilon)$ such that $\Pr \left(\left\| \frac{1}{T} \sum_{t=1}^T z_t' z_t \right\| > M(\varepsilon) \right) < \varepsilon$.

Assumption A.6. $E(\xi_{jt}(\theta_0, P_{0t}, s_t)^2 | z_t) \leq M$ for all $t = 1, \dots, T$ and $j = 1, \dots, J_t$.

Assumption A.7. As $T \rightarrow \infty$

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \max_{1 \leq j \leq J_t} |\sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt}) - \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})| \xrightarrow{P} 0$$

and

$$\max_{1 \leq t \leq T} \max_{1 \leq j \leq J_t} |s_{jt}^N - s_{jt}| \xrightarrow{P} 0.$$

Assumption A.8. For all $\delta > 0$, there exists $C(\delta) > 0$ such that for all $t = 1, \dots, T$,

$$\lim_{T \rightarrow \infty} \Pr \left(\inf_{\theta \in \Theta} \left\{ \inf_{\xi \notin \mathcal{N}_{\xi_t(\theta, P_{Rt}, s_t^N)}(\delta)} \left\| \sigma_t(\theta, \xi, P_{Rt}) - \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt}) \right\| \right\} > C(\delta) \right) = 1.$$

Assumption A.9. For all $\delta > 0$, there exists $C(\delta) > 0$ such that

$$\lim_{T \rightarrow \infty} \Pr \left(\inf_{\theta \notin \mathcal{N}_{\theta_0}(\delta)} \|G_T(\theta, P_0, s) - G_T(\theta_0, P_0, s)\| \geq C(\delta) \right) = 1.$$

I assume that the number of products in each market is bounded in Assumption A.3 and that we observe independent markets in Assumption A.4. Assumptions A.5 and A.6 state that the matrix of instruments has full rank, is stochastically bounded, and that the unobserved product characteristics have a bounded second conditional moment.

Assumptions A.7 and A.8 are needed because in the proof it is required that

$$\sup_{\theta \in \Theta} \|G_T(\theta, P_R, s^N) - G_T(\theta, P_0, s)\| = o_p(1)$$

$$= \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T z_t' (\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)) \right\| = o_p(1).$$

A sufficient condition for this to hold is that $\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)$ converges to 0 in probability uniformly over θ and t . Since there is no closed form expression for ξ_t , I assume instead that the market shares generated by the model, with the true and the approximated distribution, as well as the true and approximated market shares are uniformly close (Assumption A.7) and that this would be violated if $\xi_t(\theta, P_{0t}, s_t)$ was not close to $\xi_t(\theta, P_{Rt}, s_t^N)$ (Assumption A.8). Assumption A.8 says that asymptotically the ξ that sets the model's predictions for shares equal to the actual shares can be distinguished from other values of ξ . Assumption A.7 implies that $R(T) \rightarrow \infty$ and $N(T) \rightarrow \infty$ as $T \rightarrow \infty$. All assumptions can be verified under low level condition in the BLP model as explained in Section 2.1. Finally, Assumption A.9 states that θ_0 is identified from the moment conditions.

These assumptions imply the following theorem.

Theorem A.1. Suppose that Assumptions A.1–A.9 hold. Then $\hat{\theta} \xrightarrow{p} \theta_0$ as $T \rightarrow \infty$.

Proof. The proof consists of two parts. In the first part I show that an estimator defined as any sequence that satisfies $\|G_T(\hat{\theta}, P_0, s)\| = \inf_{\theta \in \Theta} \|G_T(\theta, P_0, s)\| + o_p(1)$ is a consistent estimator for θ . To do so notice that for any $\delta > 0$,

$$\Pr \left(\|\hat{\theta} - \theta_0\| \geq \delta \right)$$

$$= \Pr \left(\|\hat{\theta} - \theta_0\| \geq \delta, \|G_T(\theta_0, P_0, s) - G_T(\hat{\theta}, P_0, s)\| \geq C(\delta) \right)$$

$$+ \Pr \left(\|\hat{\theta} - \theta_0\| \geq \delta, \|G_T(\theta_0, P_0, s) - G_T(\hat{\theta}, P_0, s)\| < C(\delta) \right)$$

$$\leq \Pr \left(\|G_T(\theta_0, P_0, s) - G_T(\hat{\theta}, P_0, s)\| \geq C(\delta) \right)$$

$$+ \Pr \left(\inf_{\theta \notin \mathcal{N}_{\theta_0}(\delta)} \|G_T(\theta, P_0, s) - G_T(\theta_0, P_0, s)\| < C(\delta) \right).$$

The second term on the right hand side converges to 0 by Assumption A.9. For the first term notice that Assumptions A.4–A.6, and Kolmogorov's first law of large numbers implies that $\|G_T(\theta_0, P_0, s)\| = o_p(1)$. Thus

$$\|G_T(\theta_0, P_0, s) - G_T(\hat{\theta}, P_0, s)\| \leq 2\|G_T(\theta_0, P_0, s)\| + o_p(1) = o_p(1).$$

Therefore, $\hat{\theta} \xrightarrow{p} \theta_0$.

In the second part I show that $\sup_{\theta \in \Theta} \|G_T(\theta, P_R, s^N) - G_T(\theta, P_0, s)\|$ converges to 0 in probability. This then implies by the triangle inequality that for any sequence $\theta_T \in \Theta$

$$\|G_T(\theta_T, P_R, s^N) - \|G_T(\theta_T, P_0, s)\| \leq \|G_T(\theta_T, P_R, s^N) - G_T(\theta_T, P_0, s)\| = o_p(1).$$

Let $\tilde{\theta} = \arg \min_{\theta \in \Theta} \|G_T(\theta, P_0, s)\|$. Then $\hat{\theta}$ satisfies $\|G_T(\hat{\theta}, P_0, s)\| = \inf_{\theta \in \Theta} \|G_T(\theta, P_0, s)\| + o_p(1)$ because

$$0 \leq \|G_T(\hat{\theta}, P_0, s)\| - \inf_{\theta \in \Theta} \|G_T(\theta, P_0, s)\| = \|G_T(\hat{\theta}, P_0, s)\| - \|G_T(\tilde{\theta}, P_0, s)\| = \|G_T(\hat{\theta}, P_R, s^N)\| - \|G_T(\tilde{\theta}, P_0, s)\| + o_p(1) \leq \|G_T(\tilde{\theta}, P_R, s^N)\| - \|G_T(\tilde{\theta}, P_0, s)\| + o_p(1) = o_p(1).$$

Hence, by the first part, proving $\sup_{\theta \in \Theta} \|G_T(\theta, P_R, s^N) - G_T(\theta, P_0, s)\| = o_p(1)$ is sufficient for consistency. Let Z be the $\sum_{t=1}^T J_t \times d_z$ matrix of instruments. By the Cauchy–Schwarz inequality,

$$\|G_T(\theta, P_R, s^N) - G_T(\theta, P_0, s)\|^2 = \frac{1}{T^2} \|Z'(\xi(\theta, P_R, s^N) - \xi(\theta, P_0, s))\|^2 \leq \frac{1}{T} \|Z'Z\| \times \frac{1}{T} \|\xi(\theta, P_R, s^N) - \xi(\theta, P_0, s)\|^2.$$

Since $\frac{1}{T} \|Z'Z\| = O_p(1)$ by Assumption A.5, it suffices to prove that

$$\sup_{\theta \in \Theta} \frac{1}{T} \|\xi(\theta, P_R, s^N) - \xi(\theta, P_0, s)\|^2 = \sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \|\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)\|^2 = o_p(1).$$

By Assumption A.7 we have

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \|\sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt}) - \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt})\| = \sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \|\sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t}) - \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt}) + s_t^N - s_t\| = o_p(1). \tag{11}$$

This then implies that $\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \|\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)\|^2 = o_p(1)$ because by Assumption A.8, if instead $\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \|\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)\|^2 > \delta$, then

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \|\sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt}) - \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt})\| > C(\delta)$$

with probability approaching 1, which contradicts (11). \square

B.2. Asymptotic normality

Next I present sufficient conditions for the asymptotic normality results. To do so define

$$\varepsilon_{0rt} \equiv v_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), v_{rt}) - \int v_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), v) dP_{0t}(v)$$

and

$$d\varepsilon_{0rt} \equiv \frac{\partial v_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), v_{rt})}{\partial \xi} - \int \frac{\partial v_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), v)}{\partial \xi} dP_{0t}(v).$$

Also let e_{jt} denote the j th column of the $J_t \times J_t$ identity matrix and define

$$H_{0t} \equiv \frac{\partial \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}), P_{0t})}{\partial \xi}$$

$$I_{0jt} \equiv \sum_{k=1}^{J_t} H_{0t}^{-1} K_{0kt} H_{0t}^{-1} e_{jt} e_{kt}' H_{0t}^{-1},$$

and $K_{0kt} \equiv \frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi_k \partial \xi_k}$.

For any function $h(v, x)$ denote by $E_t^*(h(v, x))$ the expectation with respect to P_{0t} given the data.

Assumption B.1. θ_0 is an interior point of Θ .

Assumption B.2. For all $P \in \mathcal{P}$, the function $G_T(\theta, P, s)$ is differentiable at θ_0 . Define the derivative matrix $\Gamma_t \equiv \frac{\partial G_t(\theta_0, P_{0t})}{\partial \theta} \cdot \frac{1}{T} \sum_{t=1}^T \Gamma_t$ converges to a matrix, Γ , of full rank.

Assumption B.3. $v_t(\theta, \xi_t, v)$ is four times continuously differentiable in ξ_t for all $\xi_t \in \mathbb{R}^{J_t}$, $\theta \in \Theta$, and $v \in \mathbb{R}^{d_v}$. $v_t(\theta, \xi_t, v)$ and its first four derivatives with respect to ξ_t are bounded, continuous, and differentiable with respect to θ for all $\xi_t \in \mathbb{R}^{J_t}$, $\theta \in \Theta$, and $v \in \mathbb{R}^{d_v}$.

Assumption B.4. $s_{jtn} \in \{0, 1\}$ are iid draws with $E(s_{jtn}) = s_{jt}$, the draws are independent of (p_t, z_t, ξ_t) conditional on s_t , and $\ln(T)/N \rightarrow 0$.

Assumption B.5. As $T \rightarrow \infty$

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \max_{1 \leq j, k \leq J_t} \left| \frac{\partial \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt})}{\partial \xi_k} - \frac{\partial \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_k} \right| \xrightarrow{p} 0,$$

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \max_{1 \leq j, k, l \leq J_t} \left| \frac{\partial^2 \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt})}{\partial \xi_k \partial \xi_l} - \frac{\partial^2 \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_k \partial \xi_l} \right| \xrightarrow{p} 0$$

$$\sup_{\theta \in \Theta} \max_{1 \leq t \leq T} \max_{1 \leq j, k, l, m \leq J_t} \left| \frac{\partial^3 \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{Rt})}{\partial \xi_k \partial \xi_l \partial \xi_m} - \frac{\partial^3 \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_k \partial \xi_l \partial \xi_m} \right| \xrightarrow{p} 0.$$

Assumption B.6. v_{rt} is iid across r , independent across t , and independent of (p_t, z_t, ξ_t) .

Assumption B.7. The random variables in x_t have bounded support and $E(z_{jt}^4) \leq M$ for some constant M and for all $l = 1, \dots, d^2$, $t = 1, \dots, T$ and $j = 1, \dots, J_t$.

Assumption B.8. The absolute value of each element of H_{0t}^{-1} is bounded above by some constant M in a neighborhood of θ_0 for all $t = 1, \dots, T$.

Assumption B.9. There exists a $J_t \times d_\theta$ matrix $H_{1t}(v)$ and a $d_\theta \times d_\theta$ matrix $H_2(v)$ such that for all $j = 1, \dots, J_t$,

$$\left| \frac{\partial v_t(\theta, \xi_t, v)}{\partial \theta} \right| \leq H_{1t}(v), \quad \left| \frac{\partial^2 v_{jt}(\theta, \xi_t, v)}{\partial \theta \partial \xi_t'} \right| \leq H_{1t}(v),$$

$$\left| \frac{\partial^2 v_{jt}(\theta, \xi_t, v)}{\partial \theta \partial \theta'} \right| \leq H_2(v),$$

and each element of $H_{1t}(v)$ and $H_2(v)$ has 4 bounded absolute moments with respect to P_{0t} . The inequalities are understood

element by element and the elements of $H_{1t}(v)$ do not depend on t .

Assumption B.10. Assume that there are positive definite matrices Φ_1, Φ_2 , and Φ_3 such that

$$\lim_{T \rightarrow \infty} E \left(\frac{1}{T} \sum_{t=1}^T z_t' \xi_t(\theta_0, P_{0t}, s_t) \xi_t(\theta_0, P_{0t}, s_t)' z_t \right) = \Phi_1,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{Var} (z_t' H_{0t}^{-1} \varepsilon_{0rt}) = \Phi_2,$$

and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{Var} (z_t' H_{0t}^{-1} (s_{nt} - s_t)) = \Phi_3.$$

Assumption B.11. Assume that the following limit exists

$$\bar{\mu}_1 \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} E \left(z_{jt}' \left(e_{jt}' H_{0t}^{-1} E_t^* (d\varepsilon_{0rt} H_{0t}^{-1} \varepsilon_{0rt}) - \frac{1}{2} E_t^* (\varepsilon_{0rt}' I_{0jt} \varepsilon_{0rt}) \right) \right)$$

and

$$\bar{\mu}_2 \equiv \lim_{T \rightarrow \infty} -\frac{1}{2} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} E (z_{jt}' (\text{diag}(I_{0jt})' s_t - s_t' I_{0jt} s_t)).$$

The first two assumptions are standard. **Assumption B.3** assumes smoothness of the function generating the market shares and can easily be checked once more structure is imposed. It rules out the pure characteristics model, where v_{jt} is an indicator function. **Assumption B.4** states that the observed shares are generated from a random sample of N consumers. **Assumption B.5** is very similar to **Assumption A.7**. **Assumption B.6** states that the draws, v_{rt} , are iid and differ across markets. **Assumption B.7** states that the prices and product characteristics have bounded support and that the instruments have four bounded moments. **Assumption B.8** states that the elements of the inverse of a matrix are bounded. This assumption generally holds under the assumptions in this paper if the inverse exists. **Assumption B.9** is easily verified once more structure on the function generating the market shares is imposed. **Assumptions B.10** and **B.11** place restrictions on the data generating process. They guarantee the applicability of central limit theorems and the convergence of the bias terms in the asymptotic expansion, respectively. Both assumptions are implied by the other assumptions if the distributions of the data and the random coefficients are the same in all markets.

These assumptions lead to the main theorem of this section.

Theorem A.2. Assume that **Assumptions A.1–A.9** and **B.1–B.11** hold. Then

$$\sqrt{T} (\hat{\theta} - \theta_0) = \left(\left(\Gamma' W \Gamma \right)^{-1} \Gamma' W + o_p(1) \right) \times \left(Q_{1T} + \frac{1}{\sqrt{R}} Q_{2T} + \frac{1}{\sqrt{N}} Q_{3T} + \frac{\sqrt{T}}{R} C_{1T} + \frac{\sqrt{T}}{N} C_{2T} + o_p \left(\frac{\sqrt{T}}{R} \right) + o_p \left(\frac{\sqrt{T}}{N} \right) \right)$$

where $Q_{1T} \xrightarrow{d} N(0, \Phi_1)$, $Q_{2T} \xrightarrow{d} N(0, \Phi_2)$, $Q_{3T} \xrightarrow{d} N(0, \Phi_3)$, $C_{1T} \xrightarrow{p} \bar{\mu}_1$, and $C_{2T} \xrightarrow{p} \bar{\mu}_2$. Furthermore, Q_{1T} , Q_{2T} , and Q_{3T} are asymptotically independent.

The proof is given after a corollary, which is an immediate consequence of Theorem A.2.

Corollary 2. Assume that Assumptions A.1–A.9 and B.1–B.11 hold. If $\lambda_1 = \lim_{T \rightarrow \infty} \frac{\sqrt{T}}{R} < \infty$ and $\lambda_2 = \lim_{T \rightarrow \infty} \frac{\sqrt{T}}{N} < \infty$, then

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left((\Gamma'W\Gamma)^{-1}\Gamma'W(\lambda_1\bar{\mu}_1 + \lambda_2\bar{\mu}_2), V_1\right)$$

where

$$V_1 = (\Gamma'W\Gamma)^{-1}\Gamma'W\Phi_1W\Gamma(\Gamma'W\Gamma)^{-1}.$$

Proof of Theorem A.2. The objective is to minimize $G_T(\theta, P_R, s^N)'W_T G_T(\theta, P_R, s^N)$. The first order condition is

$$\left(\frac{\partial}{\partial \theta} G_T(\hat{\theta}, P_R, s^N)'\right) W_T G_T(\hat{\theta}, P_R, s^N) = 0.$$

Define

$$D_T(\theta, P_R, s^N) \equiv \frac{\partial}{\partial \theta} G_T(\theta, P_R, s^N).$$

The mean value theorem implies that a first order expansion of $G_T(\hat{\theta}, P_R, s^N)$ around $\theta = \theta_0$ yields

$$D_T(\hat{\theta}, P_R, s^N)'W_T \left(G_T(\theta_0, P_R, s^N) + D_T(\tilde{\theta}, P_R, s^N)(\hat{\theta} - \theta_0)\right) = 0$$

where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}$.¹⁶ Thus

$$\begin{aligned} \sqrt{T}(\hat{\theta} - \theta_0) &= \left(D_T(\hat{\theta}, P_R, s^N)'W_T D_T(\tilde{\theta}, P_R, s^N)\right)^{-1} \\ &\quad \times D_T(\hat{\theta}, P_R, s^N)'W_T \sqrt{T}G_T(\theta_0, P_R, s^N). \end{aligned}$$

The proof now consists of two parts. First I show that

$$\begin{aligned} \sqrt{T}G_T(\theta_0, P_R, s^N) &= \sqrt{T}G_T(\theta_0, P_0, s) + O_p\left(\frac{1}{\sqrt{R}}\right) + O_p\left(\frac{\sqrt{T}}{R}\right) \\ &\quad + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{N}\right) + o_p\left(\frac{\sqrt{T}}{R}\right) + o_p\left(\frac{\sqrt{T}}{N}\right) \end{aligned}$$

and I derive an expressions for the terms on the right hand side. Next I prove that for any consistent estimator $\hat{\theta}$ of θ it holds that $\hat{D}_T(\hat{\theta}, P_R, s^N)$ converges to Γ in probability. Combining these results yields the conclusion of the theorem.

We have

$$\begin{aligned} \sqrt{T}G_T(\theta_0, P_R, s^N) &= \sqrt{T}G_T(\theta_0, P_0, s) + \sqrt{T}(G_T(\theta_0, P_R, s) \\ &\quad - G_T(\theta_0, P_0, s)) + \sqrt{T}(G_T(\theta_0, P_R, s^N) - G_T(\theta_0, P_R, s)). \end{aligned} \quad (12)$$

The first term is $O_p(1)$ and belongs to the GMM objective function without simulation error. By Assumptions A.4 and B.7–B.10 and Lyapunov's CLT it converges to a normally distributed random variable.¹⁷ Next consider

$$\begin{aligned} G_T(\theta_0, P_R, s) - G_T(\theta_0, P_0, s) &= \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^J z'_{jt}(\xi_{jt}(\theta_0, P_{Rt}, s_t) - \xi_{jt}(\theta_0, P_{0t}, s_t)) \end{aligned}$$

¹⁶ The notation is a bit simplified because $\tilde{\theta}$ differs for each element of $D_T(\hat{\theta}, P_R, s^N)$. See Lemma A.4.

¹⁷ Since $\frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} = -\left(\frac{\partial \sigma_t(\xi_t(\theta, P_{0t}, s_t), \theta, P_{0t})}{\partial \xi_t} \frac{\partial \sigma_t(\xi_t(\theta, P_{0t}, s_t), \theta, P_{0t})}{\partial \theta}\right)^{-1} \frac{\partial \sigma_t(\xi_t(\theta, P_{0t}, s_t), \theta, P_{0t})}{\partial \theta}$, Assumptions B.7–B.9 guarantee the existence of $2 + \delta$ moments, which is needed for Lyapunov's CLT to apply.

$$\begin{aligned} &= \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^J z'_{jt}(\sigma_{jt}^{-1}(\theta_0, s_t, P_{Rt}) \\ &\quad - \sigma_{jt}^{-1}(\theta_0, \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{Rt}), P_{Rt})). \end{aligned}$$

To derive an expression for this sum, first let $F : \mathbb{R}^J \rightarrow \mathbb{R}^J$ be a three times continuously invertible function. The inverse function is defined by $F^{-1} : \mathbb{R}^J \rightarrow \mathbb{R}^J$. Then by Lemma A.4 for any $s_1, s_0 \in \mathbb{R}^J$ there exists a $c_j \in (0, 1)$ such that

$$\begin{aligned} F_j^{-1}(s_1) &= F_j^{-1}(s_0) + \frac{\partial F_j^{-1}(s)}{\partial s} \Big|_{s=s_0} (s_1 - s_0) \\ &\quad + \frac{1}{2} (s_1 - s_0)' \frac{\partial^2 F_j^{-1}(s)}{\partial s' \partial s} \Big|_{s=s_0} (s_1 - s_0) \\ &\quad + \sum_{\alpha \in \mathbb{N}_+^J; |\alpha|=3} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} F_j^{-1}(s)}{\partial s^{|\alpha|}} \Big|_{s=s_0+c_j(s_1-s_0)} \right) (s_1 - s_0)^{|\alpha|} \end{aligned}$$

where F_j^{-1} denotes the j th element of vector F^{-1} and α is a J -dimensional multi-index. Next I derive expressions for the first two derivatives of the inverse function. Since $F^{-1}(F(x)) = x$,

$$\frac{\partial F^{-1}(F(x))}{\partial x} = I \Leftrightarrow \frac{\partial F^{-1}(F(x))}{\partial F(x)} \frac{\partial F(x)}{\partial x} = I$$

and with $s = F(x)$ we get $\frac{\partial F^{-1}(s)}{\partial s} = \left(\frac{\partial F(x)}{\partial x}\right)^{-1}$. The j th row of this

matrix is $\frac{\partial F_j^{-1}(s)}{\partial s}$ which implies that

$$\frac{\partial F_j^{-1}(s)}{\partial s} = e'_j \left(\frac{\partial F(x)}{\partial x}\right)^{-1} \quad \text{and} \quad \frac{\partial F_j^{-1}(s)}{\partial s'} = \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j,$$

where e_j is a $J \times 1$ vector of zeros with a 1 at the j th element. Next notice that

$$\frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial x_i} = \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial F(x)} \frac{\partial F(x)}{\partial x_i},$$

which implies that

$$\begin{aligned} \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial x} &= \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial F(x)} \frac{\partial F(x)}{\partial x} \\ \text{or} \quad \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial F(x)} &= \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial x} \left(\frac{\partial F(x)}{\partial x}\right)^{-1}. \end{aligned}$$

Finally

$$\begin{aligned} \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial x_i} &= \frac{\partial \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j}{\partial x_i} \\ &= -\left(\frac{\partial F(x)}{\partial x'}\right)^{-1} \left(\frac{\partial^2 F(x)}{\partial x' \partial x_i}\right) \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j. \end{aligned}$$

It now follows that

$$\begin{aligned} \frac{\partial^2 F_j^{-1}(F(x))}{\partial F(x)' \partial F(x)} &= -\left(\left(\frac{\partial F(x)}{\partial x'}\right)^{-1} \frac{\partial^2 F(x)}{\partial x' \partial x_1} \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j\right. \\ &\quad \left. \dots \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} \frac{\partial^2 F(x)}{\partial x' \partial x_j} \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j\right) \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} \\ &= -\sum_{k=1}^J \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} \frac{\partial^2 F(x)}{\partial x' \partial x_k} \left(\frac{\partial F(x)}{\partial x'}\right)^{-1} e_j e'_k \left(\frac{\partial F(x)}{\partial x'}\right)^{-1}. \end{aligned}$$

Now let $e_{Rt} \equiv \sigma_t(\theta_0, \xi_t, P_{Rt}) - \sigma_t(\theta_0, \xi_t, P_{0t})$. Then since $\xi_{jt}(\theta_0, P_{Rt}, s_t) = \sigma_{jt}^{-1}(\theta_0, s_t, P_{Rt})$ and $\xi_{jt}(\theta_0, P_{0t}, s_t) = \sigma_{jt}^{-1}(\theta_0, s_t, P_{0t})$,

$(\theta_0, P_{0t}, s_t), P_{Rt})$ the previous expansion implies that there exists $c_{jt} \in (0, 1)$ such that

$$\begin{aligned} \xi_{jt}(\theta_0, P_{Rt}, s_t) &= \xi_{jt}(\theta_0, P_{0t}, s_t) - e'_{jt} H_{0t}^{-1} e_{Rt} + e'_{jt} H_{0t}^{-1} \frac{\partial e_{Rt}}{\partial \xi} H_{0t}^{-1} e_{Rt} \\ &\quad - \frac{1}{2} e'_{Rt} \left(\sum_{k=1}^{J_t} H_{0t}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi \partial \xi_k} \right) H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1} \right) \\ &\quad \times e_{Rt} + error_{jt} \end{aligned}$$

where e_{jt} is the j th column of the $J_t \times J_t$ identity matrix,

$$H_{0t} = \frac{\partial \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi},$$

$$H_{Rt} = \frac{\partial \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{Rt})}{\partial \xi},$$

$$\begin{aligned} error_{jt} &= e'_{jt} (H_{Rt}^{-1} - H_{0t}^{-1}) (H_{Rt} - H_{0t}) H_{0t}^{-1} e_{Rt} \\ &\quad - \frac{1}{2} e'_{Rt} \left(\sum_{k=1}^{J_t} \left(H_{Rt}^{-1} \left(\frac{\partial^2 \sigma_t(\xi_t(\theta_0, P_{0t}, s_t), \theta_0, P_{Rt})}{\partial \xi \partial \xi_k} \right) H_{Rt}^{-1} e_{jt} e'_{kt} H_{Rt}^{-1} \right. \right. \\ &\quad \left. \left. - H_{0t}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi \partial \xi_k} \right) H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1} \right) \right) e_{Rt} \\ &\quad + \sum_{\alpha \in \mathbb{N}_{+}^{J_t}: |\alpha|=3} \frac{1}{\alpha!} \\ &\quad \times \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_t s_t + (1-c_t) \sigma_t(\theta_0, \xi(\theta_0, P_{0t}, s_t), P_{Rt})} \right) e_{Rt}^{|\alpha|}. \end{aligned}$$

Now define

$$K_{0kt} \equiv \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi \partial \xi_k} \right) \quad \text{and}$$

$$I_{0jt} \equiv \sum_{k=1}^{J_t} H_{0t}^{-1} K_{0kt} H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1}.$$

Then

$$\begin{aligned} \xi_{jt}(\theta_0, P_{Rt}, s_t) &= \xi_{jt}(\theta_0, P_{0t}, s_t) - e'_{jt} H_{0t}^{-1} e_{Rt} + e'_{jt} H_{0t}^{-1} \frac{\partial e_{Rt}}{\partial \xi} H_{0t}^{-1} e_{Rt} \\ &\quad - \frac{1}{2} e'_{Rt} I_{0jt} e_{Rt} + error_{jt}, \end{aligned}$$

It now follows that $\sqrt{T} (G_T(\theta_0, P_R, s) - G_T(\theta_0, P_0, s))$ equals

$$\begin{aligned} &-\frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t H_{0t}^{-1} e_{Rt} + \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j \\ &\quad \times \left(e'_{jt} H_{0t}^{-1} \frac{\partial e_{Rt}}{\partial \xi} H_{0t}^{-1} e_{Rt} - \frac{1}{2} e'_{Rt} I_{0jt} e_{Rt} \right) + \frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t error_{jt}. \end{aligned}$$

Now we have three terms. The first term is $O_p\left(\frac{1}{\sqrt{R}}\right)$ and converges to a normally distributed random variable when multiplied by \sqrt{R} by Assumptions B.3, B.6–B.8 and B.10 and Lyapunov's CLT for triangular arrays. This normal is asymptotically independent of the asymptotic distribution of $G_T(\theta_0, P_0, s)$ by Assumption B.6.

The second term does not have mean 0 because

$$\begin{aligned} E &\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j \left(e'_{jt} H_{0t}^{-1} \left(\frac{\partial e_{Rt}}{\partial \xi} \right) H_{0t}^{-1} e_{Rt} - \frac{1}{2} e'_{Rt} I_{0jt} e_{Rt} \right) \right) \\ &= \frac{\sqrt{T}}{R} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} E \\ &\quad \times \left(z'_j \left(e'_{jt} H_{0t}^{-1} E_t^* (d\varepsilon_{0rt} H_{0t}^{-1} \varepsilon_{0rt}) - \frac{1}{2} E_t^* (e'_{0rt} I_{0jt} \varepsilon_{0rt}) \right) \right). \end{aligned}$$

By Assumption B.11

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{J_t} E &\left(z'_j \left(e'_{jt} H_{0t}^{-1} E_t^* (d\varepsilon_{0rt} H_{0t}^{-1} \varepsilon_{0rt}) \right. \right. \\ &\quad \left. \left. - \frac{1}{2} E_t^* (e'_{0rt} I_{0jt} \varepsilon_{0rt}) \right) \right) = \bar{\mu}_1. \end{aligned}$$

By Assumptions B.6–B.8 and B.11, Chebyshev's law of large numbers (LLN), and Lyapunov's CLT

$$\frac{R}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j \left(e'_{jt} H_{0t}^{-1} \left(\frac{\partial e_{Rt}}{\partial \xi} \right) H_{0t}^{-1} e_{Rt} - \frac{1}{2} e'_{Rt} I_{0jt} e_{Rt} \right) \xrightarrow{p} \bar{\mu}_1,$$

which implies that the second term times $\frac{R}{\sqrt{T}}$ converges in probability to a constant.

Next I will prove that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t error_{jt} = o_p\left(\frac{\sqrt{T}}{R}\right).$$

Recall that

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j error_{jt} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j e'_{jt} (H_{Rt}^{-1} - H_{0t}^{-1}) \left(\frac{\partial e_{Rt}}{\partial \xi} \right) H_{0t}^{-1} e_{Rt} \\ &\quad - \frac{1}{2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j e'_{jt} e_{Rt} \\ &\quad \times \left(\sum_{k=1}^{J_t} \left(H_{Rt}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{Rt})}{\partial \xi \partial \xi_k} \right) H_{Rt}^{-1} e_{jt} e'_{kt} H_{Rt}^{-1} \right. \right. \\ &\quad \left. \left. - H_{0t}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi \partial \xi_k} \right) H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1} \right) \right) e_{Rt} \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j \sum_{\alpha \in \mathbb{N}_{+}^{J_t}: |\alpha|=3} \frac{1}{\alpha!} \\ &\quad \times \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_t s_t + (1-c_t) \sigma_t(\theta_0, \xi(\theta_0, P_{0t}, s_t), P_{Rt})} \right) e_{Rt}^{|\alpha|}. \end{aligned}$$

By Assumptions B.5 and B.8, $H_{Rt}^{-1} = H_{0t}^{-1} + o_p(1)$ where each element of the $J_t \times J_t$ $o_p(1)$ term does not depend on t . Similarly

$$\begin{aligned} &\sum_{k=1}^{J_t} H_{Rt}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{Rt})}{\partial \xi \partial \xi_k} \right) H_{Rt}^{-1} e_{jt} e'_{kt} H_{Rt}^{-1} \\ &= \sum_{k=1}^{J_t} H_{0t}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, s_t), P_{0t})}{\partial \xi \partial \xi_k} \right) \\ &\quad \times H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1} + o_p(1), \end{aligned}$$

where each element of the $J_t \times J_t$ $o_p(1)$ term does not depend on t . It now follows from Assumptions B.6–B.8 as well as Lemma A.3 that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_j e'_{jt} (H_{Rt}^{-1} - H_{0t}^{-1}) \left(\frac{\partial e_{Rt}}{\partial \xi} \right) H_{0t}^{-1} e_{Rt} = o_p\left(\frac{\sqrt{T}}{R}\right)$$

and

$$\begin{aligned} & \frac{1}{2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_{jt} e'_{Rt} \\ & \times \left(\sum_{k=1}^{J_t} \left(H_{Rt}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, S_t), P_{Rt})}{\partial \xi \partial \xi_k} \right) H_{Rt}^{-1} e_{jt} e'_{kt} H_{Rt}^{-1} \right. \right. \\ & \left. \left. - H_{0t}^{-1} \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}, S_t), P_{0t})}{\partial \xi \partial \xi_k} \right) H_{0t}^{-1} e_{jt} e'_{kt} H_{0t}^{-1} \right) \right) e_{Rt} \\ & = o_p \left(\frac{\sqrt{T}}{R} \right). \end{aligned}$$

Finally consider

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t \sum_{\alpha \in \mathbb{N}_+^t: |\alpha|=3} \frac{1}{\alpha!} \\ & \times \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})\sigma_t(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})} \right) e_{Rt}^{|\alpha|}. \end{aligned}$$

Similarly, we can write the third partial derivatives of the inverse function as a function of partial derivatives of $\sigma_t(\theta_0, \xi, P_{0t})$ evaluated at $\tilde{\xi}_t$ that satisfies

$$\begin{aligned} \sigma_{jt}(\theta_0, \tilde{\xi}_t, P_{Rt}) &= \sigma_{jt}(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt}) \\ &+ c_{jt}(s_{jt} - \sigma_{jt}(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})). \end{aligned}$$

Since $\max_{1 \leq t \leq T} |s_{jt} - \sigma_{jt}(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})| = o_p(1)$ also $\max_{1 \leq t \leq T} |s_{jt} - \sigma_{jt}(\theta_0, \tilde{\xi}_t, P_{Rt})| = o_p(1)$ which is equivalent to

$$\max_{1 \leq t \leq T} \left| \sigma_{jt}(\theta_0, \xi(\theta_0, P_{Rt}, S_t), P_{Rt}) - \sigma_{jt}(\theta_0, \tilde{\xi}_t, P_{Rt}) \right| = o_p(1).$$

This implies by Assumption A.8 that $\max_{1 \leq t \leq T} \left| \xi_{jt}(\theta_0, P_{Rt}, S_t) - \tilde{\xi}_{jt} \right| = o_p(1)$, which in turn means that $\max_{1 \leq t \leq T} \left| \xi_{jt}(\theta_0, P_{0t}, S_t) - \tilde{\xi}_{jt} \right| = o_p(1)$. Next, it is easy to verify that

$$\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})\sigma_t(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})}$$

is a function of $\left(\frac{\partial \sigma_t(\theta_0, \xi, P_{Rt})}{\partial \xi} \Big|_{\xi_t = \tilde{\xi}_t} \right)^{-1}$ and as well as partial derivatives of $\sigma_t(\theta_0, \xi, P_{Rt})$ up to order 3 evaluated at $\tilde{\xi}_t$. But since $\max_{1 \leq t \leq T} \left| \xi(\theta_0, P_{0t}, S_t) - \tilde{\xi}_t \right| = o_p(1)$, it follows from Assumptions B.3 and B.5 that

$$\frac{\partial^{|\alpha|} \sigma_t^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})\sigma_t(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})} = Q_{t,0,|\alpha|} + o_p(1)$$

where $Q_{t,0,|\alpha|}$ is bounded for each t and $|\alpha|$ and the $o_p(1)$ term does not depend on t . As a consequence, it follows from Lemma A.3 that

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_{jt} \sum_{\alpha \in \mathbb{N}_+^t: |\alpha|=3} \frac{1}{\alpha!} \\ & \times \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})\sigma_t(\theta_0, \xi(\theta_0, P_{0t}, S_t), P_{Rt})} \right) e_{Rt}^{|\alpha|} \\ & = o_p \left(\frac{\sqrt{T}}{R} \right). \end{aligned}$$

Next consider the last term of (12), namely

$$\begin{aligned} & \sqrt{T} (G_T(\theta_0, P_R, s^N) - G_T(\theta_0, P_R, s)) \\ & = \frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t (\xi_t(\theta_0, P_{Rt}, s_t^N) - \xi(\theta_0, P_{Rt}, s_t)) \\ & = \frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t (\sigma_t^{-1}(\theta_0, s_t^N, P_{Rt}) - \sigma_t^{-1}(\theta_0, s_t, P_{Rt})) \\ & = \frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t \tilde{H}_{Rt}^{-1} (s_t^N - s_t) \\ & \quad - \frac{1}{2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_{jt} (s_t^N - s_t)' \tilde{I}_{Rjt} (s_t^N - s_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_{jt} \\ & \quad \times \left(\sum_{\alpha \in \mathbb{N}_+^t: |\alpha|=3} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})s_t^N} \right) \right) \\ & \quad \times (s_t^N - s_t)^{|\alpha|}, \end{aligned}$$

where $c_{jt} \in (0, 1)$, $\tilde{I}_{Rjt} \equiv \sum_{k=1}^{J_t} \tilde{H}_{Rt}^{-1} \tilde{K}_{Rkt} \tilde{H}_{Rt}^{-1} e_{jt} e'_{kt} \tilde{H}_{Rt}^{-1}$ with

$$\tilde{H}_{Rt} = \frac{\partial \sigma_t(\theta_0, \xi(\theta_0, P_{Rt}, S_t), P_{Rt})}{\partial \xi},$$

$$\text{and } \tilde{K}_{Rkt} \equiv \left(\frac{\partial^2 \sigma_t(\theta_0, \xi_t(\theta_0, P_{Rt}, S_t), P_{Rt})}{\partial \xi \partial \xi_k} \right).$$

Similar arguments as before yield $\tilde{H}_{Rt} = H_{0t} + o_p(1)$ and $\tilde{I}_{Rjt} = I_{0jt} + o_p(1)$, where the $o_p(1)$ terms do not depend on t or j . Moreover

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T z'_t H_{0t}^{-1} (s_t^N - s_t) = \frac{1}{\sqrt{T}} \frac{1}{N} \sum_{t=1}^T \sum_{n=1}^N z'_t H_{0t}^{-1} (s_{tn} - s_t).$$

Multiplied by \sqrt{N} this term converges to a normal distribution by Assumptions B.3, B.4, B.7, B.8 and B.10 and Lyapunov's CLT for triangular arrays. Moreover, arguments as before show that

$$\frac{N}{\sqrt{T}} \left(-\frac{1}{2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_{jt} (s_t^N - s_t)' I_{0jt} (s_t^N - s_t) \right) \xrightarrow{p} \bar{\mu}_2.$$

Also, just as before,

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j=1}^{J_t} z'_t \left(\sum_{\alpha \in \mathbb{N}_+^t: |\alpha|=3} \frac{1}{\alpha!} \right. \\ & \times \left. \left(\frac{\partial^{|\alpha|} \sigma_{jt}^{-1}(\theta_0, s, P_{Rt})}{\partial s^{|\alpha|}} \Big|_{s=c_{jt}s_t + (1-c_{jt})s_t^N} \right) (s_t^N - s_t)^{|\alpha|} \right) \\ & = o_p \left(\frac{\sqrt{T}}{N} \right). \end{aligned}$$

Finally consider

$$D_T(\hat{\theta}, P_R, s^N) = \frac{1}{T} \sum_{t=1}^T z'_t \frac{\partial \xi_t(\hat{\theta}, P_{Rt}, s_t^N)}{\partial \theta}.$$

To show that for any consistent estimator, $\check{\theta}$, of $\theta D_T(\check{\theta}, P_R, s^N) \xrightarrow{p} \Gamma$ it is sufficient to show (by Assumptions B.2 and B.9) that $D_T(\theta, P_R, s^N) - \frac{1}{T} \sum_{t=1}^T \frac{\partial G_t(\theta, P_{0t})}{\partial \theta} \xrightarrow{p} 0$ uniformly over θ in a neighborhood of θ_0 . This then implies that

$$\begin{aligned} & \left(D_T(\hat{\theta}, P_R, s^N)' W_T D_T(\check{\theta}, P_R, s^N) \right)^{-1} D_T(\hat{\theta}, P_R, s^N)' W_T \\ & \xrightarrow{p} (\Gamma' W \Gamma)^{-1} \Gamma' W. \end{aligned}$$

Notice that

$$D_T(\theta, P_R, s^N) - \frac{1}{T} \sum_{t=1}^T \frac{\partial G_T(\theta, P_{0t})}{\partial \theta} = D_T(\theta, P_R, s^N) - \frac{\partial G_T(\theta, P_0, s)}{\partial \theta} + \frac{1}{T} \sum_{t=1}^T z_t' \frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} - E \left(z_t' \frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} \right)$$

and that

$$\frac{1}{T} \sum_{t=1}^T z_t' \frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} - E \left(z_t' \frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} \right) \xrightarrow{p} 0$$

uniformly over θ by Assumptions A.4, B.8 and B.9 and the ULLN of Andrews (1987).¹⁸

Next rewrite

$$D_T(\theta, P_R, s^N) - \frac{\partial G_T(\theta, P_0, s)}{\partial \theta} = \frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \xi_t(\theta, P_{Rt}, s_t^N)}{\partial \theta} - \frac{\partial \xi_t(\theta, P_{0t}, s_t)}{\partial \theta} \right) = -\frac{1}{T} \sum_{t=1}^T z_t' \left(\left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \xi_t} \right)^{-1} \times \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \theta} - \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right),$$

where $\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta}$ is the derivatives with respect to the first element only. Similar as before

$$\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \xi_t} = \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} + o_p(1)$$

in a neighborhood of θ_0 , where the $o_p(1)$ term does not depend θ or t . Now by Assumption B.8

$$\left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \xi_t} \right)^{-1} = \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} + o_p(1),$$

and the $o_p(1)$ term does not depend θ or t . By Hölder's inequality and Assumptions A.4, B.7 and B.9 is $o_p(1)$. Hence,

$$D_T(\theta, P_R) - \frac{\partial G_T(\theta, P_0, s)}{\partial \theta} = -\frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} \times \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \theta} - \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right) - \frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1}$$

$$\times \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} - \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right) + o_p(1).$$

It follows from Assumptions A.7, A.8 and B.9 that for all j

$$\left| \frac{\partial \sigma_{jt}(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \theta'} - \frac{\partial \sigma_{jt}(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta'} \right| \leq \frac{1}{R} \sum_{r=1}^R |H'_{1t}(v_{rt})| |\xi_t(\theta, P_{Rt}, s_t^N) - \xi_t(\theta, P_{0t}, s_t)| = o_p(1)$$

uniformly over θ and t . Thus, by Hölder's inequality and Assumptions A.4, B.7 and B.9

$$\frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} \times \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{Rt}, s_t^N), P_{Rt})}{\partial \theta} - \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right)$$

is $o_p(1)$ uniformly over θ and t . Finally

$$\frac{1}{T} \sum_{t=1}^T z_t' \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} \times \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} - \frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right) = \frac{1}{T} \sum_{t=1}^T \frac{1}{R} \sum_{r=1}^R z_t' \left(\frac{\partial \sigma_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi_t} \right)^{-1} \times \left(\frac{\partial v_t(\theta, \xi_t(\theta, P_{0t}, s_t), v_{rt})}{\partial \theta} - E_t^* \left(\frac{\partial v_t(\theta, \xi_t(\theta, P_{0t}, s_t), P_{0t})}{\partial \theta} \right) \right)$$

with is $o_p(1)$ uniformly over θ by the uniform law of large numbers of Andrews (1987) and Assumptions A.4 and B.6–B.9. Hence $\sup_{\theta \in \Theta} \left| D_T(\theta, P_R) - \frac{\partial G_T(\theta, P_0, s)}{\partial \theta} \right| = o_p(1)$. \square

B.3. Bias correction

Theorem 2 also holds in the general case under Assumptions A.1–A.9 and B.1–B.11, where expressions for $\hat{\mu}_1$ and $\hat{\mu}_2$ are given in Section 3. In the last part of the proof of Theorem A.2, I show that for any consistent estimator, $\hat{\theta}$, of θ , it holds that $\frac{1}{T} \sum_{t=1}^T z_t' \frac{\partial \xi_t(\hat{\theta}, P_{Rt}, s_t^N)}{\partial \theta} \xrightarrow{p} \Gamma$. Analogous arguments can be used here to show $\hat{\mu}_1 \xrightarrow{p} \mu_1$ as $T \rightarrow \infty$ and $\hat{\mu}_2 \xrightarrow{p} \mu_2$ as $T \rightarrow \infty$.

Appendix C. Proofs of Theorems 1 and 2

I will first prove that the assumptions imply that $\xi_t(\theta, P_{0t}, s_t, x_t)$ is an element of a bounded subset of \mathbb{R}^k and then use this result to verify the assumptions of Theorem A.2. First decompose $w_t = (w_t^1, w_t^2)$ where $w_t^1 \sim Q_t$ has compact support and $w_t^2 = w^2 \sim F$ with $F \in \hat{\mathcal{P}} \times \dots \times \hat{\mathcal{P}}$. Let \mathcal{X} be the support of x_t . Now assume that $\xi_t(\theta, P_{0t}, s_t, x_t)$ is not an element of a bounded subset of \mathbb{R}^k . Then, there exists a sequence $\{s^k, x^k, \theta^k\}$, $k = 1, 2, \dots$ with $\varepsilon \leq s_j^k \leq 1 - \varepsilon$, $j = 0, \dots, J_t$, $\theta \in \Theta$, and $x^k \in \mathcal{X}$ such that for some j , $\xi_j^k \equiv \xi_j(\theta^k, P_{0t_k}, s^k, x^k) \rightarrow \infty$ or $\xi_j^k \rightarrow -\infty$ and

$$s_j^k = \int v_{jt}(\theta^k, x^k, \xi^k, g(a^k, w_1, w_2)) dQ_{t_k}(w_1) dF(w_2).$$

¹⁸ Assumptions A.4, B.8 and B.9 can be used to verify Assumptions A1, A2, and A5 of Andrews (1987).

Next define

$$s_j(\xi^k) \equiv \sup_{\theta \in \Theta} \sup_{x \in \mathcal{X}} \sup_{w_1 \in \text{supp}(Q_t)} \sup_{a \in \mathcal{A}} \int v_{kt}(x, \xi^k, \theta, g(a, w_1, w_2)) dF(w_2).$$

If $\xi_j^k \rightarrow \infty$, then $s^nk_0 \leq s_0(\xi^k) \rightarrow 0$, which is a contradiction. Similarly if $\xi_j^k \rightarrow -\infty$, then $s_j^k \rightarrow 0$.

I now verify the assumptions of Theorem A.2. Assumption A.1 holds in the BLP model as shown by Berry (1994). The result also holds with approximated predicted shares and observed shares, as long as the observed shares are positive with probability one, which is guaranteed by Assumption RC4. Assumptions A.2 and A.3 clearly hold. Assumption A.4 is assumed in Assumption RC1. Assumption A.5 is implied by Assumption RC7 and Assumption A.6 holds because $\xi_t(\theta_0, P_{0t}, s_t)$ is bounded. The first part of Assumption A.7 follows from Lemma A.1 (where $x_t = (x_t, \xi_t(\theta, P_{0t}, s_t, x_t))$ and $f(x, \theta, v) = v_{jt}(\theta, x, \xi, v)$). The conditions of the lemma are implied by Assumptions RC2–RC6.

The second part holds because by the Bernstein inequalities there exists a constant c such that

$$\Pr \left(\max_t \max_j |s_{jt}^N - s_{jt}| > \varepsilon \right) \leq \sum_{t=1}^T \sum_{j=1}^J \Pr (|s_t^N - s_t| > \varepsilon) \leq 2\bar{J} \sum_{t=1}^T \exp(-cN).$$

To see why Assumption A.8 holds, take ξ_t such that $\|\xi_t - \xi_t(\theta, P_{Rt}, s_t^N)\| \geq J\delta$. Assume without loss of generality that $|\xi_{1t} - \xi_{1t}(\theta, P_{Rt}, s_t^N)| \geq |\xi_{jt} - \xi_{jt}(\theta, P_{Rt}, s_t^N)|$ for all j . Also assume that $\xi_{1t} - \xi_{1t}(\theta, P_{Rt}, s_t^N) \geq \delta$ but similar arguments apply to $\xi_{1t} - \xi_{1t}(\theta, P_{Rt}, s_t^N) \leq -\delta$. Let $\xi_t + \delta = (\xi_{1t} + \delta, \dots, \xi_{jt} + \delta)$ and notice that for all $\delta > 0$

$$\exp(\delta) \int v_1(\theta, x_t, \xi_t, v) dP(v) > \int v_1(\theta, x_t, \xi_t + \delta, v) dP(v) > \int v_1(\theta, x_t, \xi_t, v) dP(v)$$

and

$$\int v_1(\theta, x_t, \xi_t, v) dP(v) \geq \int v_1(\theta, x_t, \xi_t(\theta, P, s_t) + \delta, v) dP(v) > \int v_1(\theta, x_t, \xi_t(\theta, P, s_t), v) dP(v).$$

Also notice that by Assumptions RC2 and RC4–RC6 there exists $\eta > 0$ and a compact set $\mathcal{V}_t \subset \mathbb{R}^{d_x}$, such that with probability approaching 1, $\eta < v_{jt}(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) < 1 - \eta$ for all j and $v \in \mathcal{V}_t$, and $\int_{\mathcal{V}_t} dP_{Rt}(v) > p > 0$.

Next define $\delta_0 = \min \left\{ \delta, \frac{1}{4} \ln \left(\frac{1}{1-\eta} \right) \right\}$ and $C(\delta) = p\eta(1 - \exp(2\delta_0)(1 - \eta))\delta_0 > 0$. The previous arguments imply that for some $\bar{\delta} \in (0, \delta_0)$ it holds that with probability approaching 1

$$\begin{aligned} & \int v_1(\theta, x_t, \xi_t, v) dP_{Rt}(v) \\ & - \int v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) dP_{Rt}(v) \\ & \geq \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t, v) dP_{Rt}(v) \\ & - \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) dP_{Rt}(v) \end{aligned}$$

$$\begin{aligned} & \geq \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N) + \delta_0, v) dP_{Rt}(v) \\ & - \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) dP_{Rt}(v) \\ & \geq \left(\int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N) + \bar{\delta}, v) dP_{Rt}(v) \right. \\ & \quad \left. - \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N) + \bar{\delta}, v) \right. \\ & \quad \left. \times \sum_{k=1}^J v_k(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N) + \bar{\delta}, v) dP_{Rt}(v) \right) \delta_0 \\ & \geq \left(\int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) dP_{Rt}(v) \right. \\ & \quad \left. - \exp(2\delta_0) \int_{\mathcal{V}_t} v_1(\theta, x_t, \xi_t(\theta, P_{Rt}, s_t^N), v) \right. \\ & \quad \left. \times (1 - \eta) dP_{Rt}(v) \right) \delta_0 \\ & \geq \int_{\mathcal{V}_t} dP_{Rt}(v) \eta \left(1 - \exp(2\delta_0)(1 - \eta) \right) \delta_0 \\ & > p\eta \left(1 - \exp(2\delta_0)(1 - \eta) \right) \delta_0. \end{aligned}$$

This proves that Assumption A.8 holds. Assumption A.9 is assumed in Assumption RC9.

Assumption B.1 is directly assumed. Differentiability in Assumption B.2 holds. Full rank is assumed in Assumption RC9 and the limit result holds by Assumption RC8 (the moment conditions in Assumptions RC5–RC7 guarantee that the moment condition in Assumption RC8 holds). Assumption B.3 holds since the derivatives are bounded. Assumptions B.4, B.6 and B.7 are easily interpretable and directly assumed. Assumption B.5 holds by Lemma A.1, just like Assumption A.7. By the contraction mapping property of the BLP model, $\frac{\partial \sigma_t(\theta, \xi(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi}$ is invertible (see Dubé et al., 2012). Thus, the determinant is not equal to 0 for all θ, s_t, x_t . By Assumptions RC5 and RC6 and continuity of the determinant in all arguments, it follows that the determinant is bounded away from 0. Since the elements of $\frac{\partial \sigma_t(\theta, \xi(\theta, P_{0t}, s_t), P_{0t})}{\partial \xi}$ are bounded, it follows that Assumption B.8 holds. Assumption B.9 holds with $H(v) = Cv$ for some constant C . Existence of four moments is assumed in Assumption RC2. Just like Assumption RC9, Assumptions B.10 and B.11 follow from Assumption RC8.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2014.10.009>.

References

Andrews, D.W., 1987. Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* 55 (6), 1465–1471.
 Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: Some recent developments. In: Blundell, R., Newey, W., Persson, T. (Eds.), *Advances in Economics and Econometrics*, vol. 3. Cambridge University Press.
 Armstrong, T., 2013. Large market asymptotics for differentiated product demand estimators with economic models of supply. Working Paper.
 Berry, S., 1994. Estimating discrete-choice models of product differentiation. *The RAND J. Economics* 25 (2), 242–262.
 Berry, S., Haile, P.A., 2014. Identification in differentiated products markets using market level data. *Econometrica* 82 (5), 1749–1797.
 Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. *Econometrica* 63 (4), 841–890.
 Berry, S., Linton, O.B., Pakes, A., 2004. Limit theory for estimating the parameters of differentiated product demand systems. *Rev. Econ. Stud.* 71, 613–654.

- Dubé, J.P., Fox, J.T., Su, C.L., 2012. Improving the numerical performance of BLP static and dynamic discrete choice random coefficients demand estimation. *Econometrica* 80 (5), 2231–2267.
- Gandhi, A., Kim, K., Petrin, A., 2013. Identification and estimation in discrete choice demand models when endogenous variables interact with the error. Working Paper.
- Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panelmodels. *Econometrica* 72 (4), 1295–1319.
- Kim, D., 2004. Estimation of the effects of new brands on incumbents—profits and consumer welfare: The U.S. processed cheese market case. *Rev. Indus. Org.* 25, 275–293.
- Kristensen, D., Salanié, B., 2013. Higher-order properties for approximate estimators. Working Paper.
- Laffont, J.-J., Ossard, H., Vuong, Q., 1995. Econometrics of first price auctions. *Econometrica* 63 (4), 953–980.
- Lee, L., 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory* 11 (3), 437–483.
- McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57 (5), 995–1026.
- Nevo, A., 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. *J. Econom. Manag. Strategy* 9 (4), 513–548.
- Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69 (2), 307–342.
- Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57 (5), 1027–1057.
- Skrainka, B., 2012. A large scale study of the small sample performance of random coefficient models of demand. Working Paper.
- Skrainka, B., Judd, K., 2011. High performance quadrature rules: how numerical integration affects a popular model of product differentiation. Working Paper.
- Su, C., Judd, K., 2012. Constrained optimization approaches to estimation of structural models. *Econometrica* 80 (5), 2213–2230.
- Villas-Boas, S., 2007. Vertical relationships between manufacturers and retailers: Inference with limited data. *Rev. Econom. Stud.* 74, 625–652.