



On Asymptotic Properties of the Parameters of Differentiated Product Demand and Supply Systems When Demographically-Categorized Purchasing Pattern Data are Available

明城, 聡
金澤, 雄一郎

(Citation)

神戸大学経済学研究科 Discussion Paper, 1009

(Issue Date)

2010

(Resource Type)

technical report

(Version)

Version of Record

(URL)

<https://hdl.handle.net/20.500.14094/81002302>



On Asymptotic Properties of the Parameters of Differentiated
Product Demand and Supply Systems When
Demographically-Categorized Purchasing Pattern Data are Available

Satoshi MYOJO¹

Kobe University

Yuichiro KANAZAWA²

University of Tsukuba

June, 2010

Abstract

In this paper, we derive asymptotic theorems for the Petrin (2002) extension of the Berry, Levinsohn, and Pakes (BLP, 1995) framework to estimate demand-supply models with micro moments. The micro moments contain the information relating the consumer demographics to the characteristics of the products they purchase. With additional assumptions, the extended estimator is shown to be CAN and more efficient than the BLP estimator. We discuss the conditions under which these asymptotic theorems hold for the random coefficient logit model. We implement extensive simulation studies and confirm the benefit of the micro moments in estimating the random coefficient logit model.

¹This research is supported in part by the Grant-in-Aid for Scientific Research (B) 22730246 from the Japan Society for the Promotion of Science. Please address correspondence to: Satoshi Myojo, Faculty of Economics, Kobe University, 2-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo, 657-8501, Japan. E-mail:myojo@econ.kobe-u.ac.jp

²This research is also supported in part by the Grant-in-Aid for Scientific Research (C)(2) 12680310, (C)(2) 16510103, and (B) 20310081 from the Japan Society for the Promotion of Science.

1 Introduction

Some recent empirical studies in the industrial organization and marketing extend the framework proposed by BLP (1995, henceforth BLP (1995)) and try to integrate the information on consumer demographics to the utility functions in order to make their models more realistic and convincing. For example, Nevo’s examination on price competition in the ready-to-eat cereal industry (Nevo 2001) uses individual’s income, age and a dummy variable indicating the individual has a child or not in the utility function. Sudhir (2001) includes household’s income to model the U.S. automobile demand in the study of competitive interactions among firms in different market segments. The background behind these is that public sources of information such as Current Population Survey (CPS) and Integrated Public Use Microdata Series (IPUMS) are widely available. Those sources give us information on the joint distribution of the U.S. household’s demographics such as income, age of household’s head, and family size.

In the analysis of the U.S. automobile market, Petrin (2002) goes further and tries to link demographics of new-vehicle purchasers to the vehicles they purchased. Specifically, given a purchasing pattern such as “buying a minivan,” he proposes to match the model-predicted average consumer’s demographics with the average consumer’s demographics from Consumer Expenditure Survey (CEX) automobile supplement in the GMM estimation. Petrin’s framework presupposes the market information on the population average, which is readily accessible through public sources.³ He claims that “the extra information plays the same role as consumer-level data, allowing estimated substitution patterns and (thus) welfare to directly reflect demographic-driven differences in tastes for observed characteristics (page, 706, lines 22-25).” His intention, it seems, is to reduce the bias associated with “a heavy dependence on the idiosyncratic logit “taste” error” (page 707, lines 5-6). If so, his contention that a source of his idea is from Imbens and Lancaster (1994) is unfortunate, because Imbens and Lancaster use micro moments to improve the efficiency.⁴

Petrin adds the set of functions of the expected value of consumer’ demographics given specific product characteristics consumers choose (e.g., expected family size of households that purchased minivans) as additional moments in the GMM estimation, where the original moment conditions used in BLP (1995) are orthogonal conditions of the unobserved quality ξ_j and the unobserved cost shifter ω_j with the corresponding instrumental variables z_j^d and z_j^c for product j . To evaluate the additional moments, individuals are sampled from the population. So Petrin’s additional moments are sample average *over individuals*, while

³Berry, Levinsohn, and Pakes (2004), on the other hand, uses detailed consumer-level data, which include not only individuals’ choices but also the choices they would have made had their first choice products not been available. Although the proposed method should improve the out-of-sample model’s prediction, it requires proprietary consumer-level data, which are not readily available to researchers, as the authors themselves acknowledged in the paper: the CAMIP data “are generally not available to researchers outside of the company” (page 79, line 30).

⁴The efficiency argument in the Imbens and Lancaster’s (1994) estimation is basically supported by that of maximum likelihood estimate (MLE), but this is not the case for Petrin’s approach; BLP framework does not use any distributional assumption on the product-level error terms (ξ_j and ω_j) other than mean independent condition and thus the functional form of the score functions are unknown.

BLP moments are *over products*.

It should be noted that these new moments are subject to the simulation and sampling errors in the BLP estimation. This is because the expectations of consumer demographics are evaluated conditional on the product characteristics $(\mathbf{X}, \boldsymbol{\xi})$, where the $\boldsymbol{\xi}$ includes the simulation and the sampling errors induced through the BLP's contraction mapping. In addition, the additional market information itself contain another type of sampling error. This is because the additional market information is typically an estimate for the population average demographics obtained from the sample of consumers (e.g., CEX sample) and this is separate from the one from which the observed market share \mathbf{s}^n is calculated. This error also affect the evaluation of the new moments. In summary each of the four errors (the simulation error, the sampling error in the observed market shares, the sampling error *induced when researcher evaluates the additional moments*, and the sampling error *in the additional information itself*) as well as the stochastic nature of the product characteristics will affect the evaluation of the additional moments. The estimator proposed by Petrin appears to assume that we are able to control the impacts from the first four errors. Moreover, it is not apparent if Petrin samples another set of individuals to evaluate additional moments, independent of those used to simulate the market shares of products. Unfortunately, Petrin (2002) does not provide any asymptotic theorems for the estimator.

We write this paper to generalize the GMM estimator extended by Petrin (2002) and provide the conditions under which this estimator not only has the CAN properties, but is more efficient than the original BLP estimator. We assume the econometrician samples two sets of individuals independent of each other, one to simulate the market share of products and the other to evaluate the additional moments, in order to avoid intractable correlations between the two sets of individuals. We also assume the given additional information on demographics of consumers are calculated from the sample independent of these two samples. We follow the rigorous work of Berry, Linton, and Pakes (2004) (hereinafter, BLP (2004)) in which the authors presented the asymptotic theorems applicable to the random coefficient logit models of demand in BLP (1995). Then we implement extensive simulation studies and confirm the benefit of the micro moments in estimating the random coefficient logit model.

This paper is organized as follows. In section 2, we operationalize the Petrin's extension to the BLP framework which utilizes the additional micro moments and define the sampling and simulation errors in the GMM objective function. In section 3, we provide assumptions for these errors and the structure of the product space to follow and then give the outline of the proofs of the asymptotic theorems for the extension. In section 4, we derive rates at which the numbers of two distinct samples (one to calculate the observed market shares and the other to compute the additional information data) and the number of simulation draws must grow relative to the number of products in the market to guarantee our asymptotic theorems to hold for the random coefficient logit model. Results from the extensive Monte Carlo experiments are presented in section 5. Finally, in section 6, we give concluding remarks and briefly discuss the case where the two samples, one is used for simulating market shares of products and other is from which the additional

information is derived, are correlated. Detail of the proofs are given in appendix.

2 System of Demand and Supply with Micro Moments

In this section, we give precise definition to the product space, refocusing the estimation procedure of BLP framework in combining the demand and the supply side moment conditions, and construct the additional moment conditions which relate consumer demographics to the characteristics of products they purchase. Since our approach extends BLP (2004), notations and the most of definitions are kept as identical as possible to those in BLP (2004).

2.1 Demand Side Model

The discrete choice differentiated product demand model formulates that the utility of consumer i for product j is a function of demand side parameters θ_d , observed product characteristics \mathbf{x}_j , unobserved (by the econometrician) product characteristics ξ_j , and random consumer tastes ν_{ij} . Given the product characteristics (\mathbf{x}_j, ξ_j) for all (J) products marketed, consumers either buy one of the products or choose the “outside” good. Each consumer makes the choice to maximize his/her utility. Different consumers assign different utility to the same choices because their tastes are different. The tastes follow the distribution P^0 .

Although most product characteristics are not correlated with the unobserved product characteristics $\xi_j \in \mathfrak{R}$, $j = 1, \dots, J$, some of them (e.g., price) are. We denote the vector of observed product characteristics $\mathbf{x}_j = (\mathbf{x}'_{1j}, \mathbf{x}'_{2j})'$ where $\mathbf{x}_{1j} \in \mathfrak{R}^{K_1}$ are exogenous and not correlated with ξ_j , while $\mathbf{x}_{2j} \in \mathfrak{R}^{K_2}$ are endogenous and correlated with ξ_j . We assume the set of exogenous product characteristics (\mathbf{x}_{1j}, ξ_j) , $j = 1, \dots, J$ are random sample of product characteristics of size J from the underlying population of product characteristics. Thus, (\mathbf{x}_{1j}, ξ_j) are assumed independent across j , while \mathbf{x}_{2j} are not in general across j since they are endogenously determined in the market as functions of others' and its own product characteristics. The ξ_j 's are assumed to be mean independent of $\mathbf{X}_1 = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1J})'$ and to have a finite conditional variance as

$$(1) \quad \text{E}[\xi_j | \mathbf{X}_1] = 0 \quad \text{and} \quad \sup_{1 \leq j \leq J} \text{E}[\xi_j^2 | \mathbf{x}_{1j}] < \infty$$

with probability one. The set of observed product characteristics for all the products is denoted by $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_J)'$.

The conditional purchase probability σ_{ij} of product j is a map from consumer i 's tastes $\nu_i \in \mathfrak{R}^v$, a demand side parameter vector $\theta_d \in \Theta_d$, and the set of characteristics of all products $(\mathbf{X}, \boldsymbol{\xi})$, and is thus denoted as $\sigma_{ij}(\mathbf{X}, \boldsymbol{\xi}, \nu_i; \theta_d)$. BLP (1995) framework generates the vector of market shares, $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \theta_d, P)$, by aggregating over the individual choice probability with the distribution P of the consumer tastes ν_i as

$$\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \theta_d, P) = \int \sigma_{ij}(\mathbf{X}, \boldsymbol{\xi}, \nu_i; \theta_d) dP(\nu_i)$$

where P is typically the empirical distribution of the tastes from a random sample drawn from P^0 . Note that these market shares are still random variables due to the stochastic nature of the product characteristics \mathbf{X} and $\boldsymbol{\xi}$. If we evaluate $\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ at $(\boldsymbol{\theta}_d^0, P^0)$, where $\boldsymbol{\theta}_d^0$ is the true value, we have the “conditionally true” market shares \mathbf{s}^0 given the product characteristics $(\mathbf{X}, \boldsymbol{\xi})$ in the population, i.e., $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d^0, P^0) \equiv \mathbf{s}^0$.

Equation in the form of $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \mathbf{s}$ can, in theory, be solved for $\boldsymbol{\xi}$ as a function of $(\mathbf{X}, \boldsymbol{\theta}_d, \mathbf{s}, P)$. BLP (1995) provides general conditions under which there is a unique solution for

$$(2) \quad \mathbf{s} - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \mathbf{0}$$

for every $(\mathbf{X}, \boldsymbol{\theta}_d, \mathbf{s}, P) \in \mathcal{X} \times \Theta_d \times \mathcal{S}_J \times \mathcal{P}$, where \mathcal{X} is a space for the product characteristics \mathbf{X} , and \mathcal{P} is a family of probability measures. If we solve (2) at any $(\boldsymbol{\theta}_d, \mathbf{s}, P) \neq (\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$, the independence assumption for the resulting $\xi_j(\mathbf{X}, \boldsymbol{\theta}_d, \mathbf{s}, P)$ no longer holds because the two factors deciding the ξ_j —the market share s_j and the endogenous product characteristics \mathbf{x}_{2j} for product j —are endogenously determined through the market equilibrium (e.g., Nash in prices or quantities) as a function of the product characteristics not only of its own but also of its competitors. However, if we solve the identity $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d^0, P^0) = \mathbf{s}^0$ with respect to $\boldsymbol{\xi}$ under the conditions to guarantee the uniqueness of the $\boldsymbol{\xi}$ in (2), we are able to retrieve the original $\xi_j(\mathbf{X}, \boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$ which we assume are independent across j .

2.2 Supply Side Model

In this paper we take into account supply side moment condition unlike BLP (2004). The framework is based on BLP (1995). Here, we give the model and define notations.

The supply side model formulates the pricing equations for the J products marketed. We assume an oligopolistic market where a finite number of suppliers provide multiple products. Suppliers ($m = 1, \dots, F$) are maximizers of the profit from the combination of products they produce. By assuming the Bertrand-Nash pricing for supplier’s strategy, the first order condition for the product j of the manufacturer m is

$$\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) + \sum_{l \in \mathcal{J}_m} (p_l - c_l) \partial \sigma_l(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial p_j = 0 \quad \text{for } j \in \mathcal{J}_m,$$

where \mathcal{J}_m denotes the set of products provided by the manufacturer m , and these p_j and c_j are respectively the price and the marginal cost of the product j . This equation can be expressed in matrix form

$$(3) \quad \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) + \boldsymbol{\Delta}(\mathbf{p} - \mathbf{c}) = \mathbf{0}$$

where $\boldsymbol{\Delta}$ is the $J \times J$ non-singular gradient matrix of $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ with respect to \mathbf{p} whose (j, k) element is defined by

$$(4) \quad \Delta_{jk} = \begin{cases} \partial \sigma_k(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial p_j, & \text{if the products } j \text{ and } k \text{ are produced by the same firm;} \\ 0, & \text{otherwise.} \end{cases}$$

We define the marginal cost c_j as a function of the observed cost shifters \mathbf{w}_j and the unobserved (by the econometrician) cost shifters ω_j as

$$(5) \quad g(c_j) = \mathbf{w}'_j \boldsymbol{\theta}_c + \omega_j$$

where $g(\cdot)$ is a monotonic function and $\boldsymbol{\theta}_c \in \Theta_c$ is a cost side parameter vector. While the choice of $g(\cdot)$ depends on application, we assume $g(\cdot)$ is continuously differentiable with a finite derivative for all realizable values of cost. Suppose that the observed cost shifters \mathbf{w}_j consist of exogenous $\mathbf{w}_{1j} \in \Re^{L_1}$ as well as endogenous $\mathbf{w}_{2j} \in \Re^{L_2}$, and thus we write $\mathbf{w}_j = (\mathbf{w}'_{1j}, \mathbf{w}'_{2j})'$ and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_J)'$. The exogenous cost shifters include not only the cost variables determined outside the market under consideration (e.g. factor price), but also the product design characteristics suppliers cannot immediately change in response to consumer's demand. The cost variables determined at the market equilibrium (e.g. production scale) are treated as endogenous cost shifters. As in the formulation of (\mathbf{x}_{1j}, ξ_j) on the demand side, we assume the set of exogenous cost shifters $(\mathbf{w}_{1j}, \omega_j)$ is a random sample of cost shifters from the underlying population of cost shifters. Thus $(\mathbf{w}_{1j}, \omega_j)$ are assumed to be independent across j , while \mathbf{w}_{2j} are in general not independent across j as they are determined in the market as functions of cost shifters of other products. Similar to the demand side unobservables, the unobserved cost shifters ω_j are assumed to be mean independent of the exogenous cost shifters $\mathbf{W}_1 = (\mathbf{w}_{11}, \dots, \mathbf{w}_{1J})'$, and satisfy with probability one,

$$(6) \quad \mathbb{E}[\omega_j | \mathbf{W}_1] = 0, \quad \text{and} \quad \sup_{1 \leq j \leq J} \mathbb{E}[\omega_j^2 | \mathbf{w}_{1j}] < \infty.$$

Define $\mathbf{g}(\mathbf{x}) \equiv (g(x_1), \dots, g(x_J))$. Solving the first order condition (3) with respect to \mathbf{c} and substituting for (5) give the vector of the unobserved cost shifters

$$(7) \quad \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}, P) = \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\mathbf{X}, \boldsymbol{\theta}_d, \mathbf{s}, P), \boldsymbol{\theta}_d, P)) - \mathbf{W}\boldsymbol{\theta}_c,$$

where

$$(8) \quad \mathbf{m}_g \equiv -\boldsymbol{\Delta}^{-1} \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$$

represents the vector of the profit margins for all the products in the market. Hereafter, we suppress the dependence of ξ_j and ω_j on \mathbf{X} and \mathbf{W} to express $\xi_j(\boldsymbol{\theta}_d, \mathbf{s}, P)$ and $\omega_j(\boldsymbol{\theta}, \mathbf{s}, P)$ respectively for notational simplicity. Notice that the parameter vector $\boldsymbol{\theta}$ in $\boldsymbol{\omega}$ contains both the demand and supply side parameters, i.e., $\boldsymbol{\theta} = (\boldsymbol{\theta}'_d, \boldsymbol{\theta}'_c)'$. Since the profit margin $m_{g_j}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ for product j is determined not only by its unobserved product characteristics ξ_j , but by those of the other products on the market, these ω_j are in general dependent across j when $(\boldsymbol{\theta}, \mathbf{s}, P) \neq (\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)$. However, when (7) is evaluated at $(\boldsymbol{\theta}, \mathbf{s}, P) = (\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)$, we are able to recover the original $\omega_j, j = 1, \dots, J$, and they are assumed independent across j .

2.3 GMM Estimation with Micro Moments

Let us define the $J \times M_1$ demand side instrument matrix $\mathbf{Z}_d = (\mathbf{z}_1^d, \dots, \mathbf{z}_J^d)'$ whose components \mathbf{z}_j^d can be written as $\mathbf{z}_j^d(\mathbf{x}_{11}, \dots, \mathbf{x}_{1J}) \in \mathfrak{R}^{M_1}$, where $\mathbf{z}_j^d(\cdot) : \mathfrak{R}^{K_1 \times J} \rightarrow \mathfrak{R}^{M_1}$ for $j = 1, \dots, J$. It should be noted that the demand side instruments \mathbf{z}_j^d for product j are assumed to be a function of the exogenous characteristics not only of its own, but of the other products in the market. This is because the instruments by definition must correlate with the product characteristics \mathbf{x}_{2j} , and these endogenous variables \mathbf{x}_{2j} (e.g. price) are determined by both its own and its competitors' product characteristics.

Similar to the demand side, we define the $J \times M_2$ supply side instrumental variables $\mathbf{Z}_c = (\mathbf{z}_1^c, \dots, \mathbf{z}_J^c)'$ as a function of the exogenous cost shifters $(\mathbf{w}_{11}, \dots, \mathbf{w}_{1J})$ of all the products. Here, $\mathbf{z}_j^c(\mathbf{w}_{11}, \dots, \mathbf{w}_{1J}) \in \mathfrak{R}^{M_2}$ and $\mathbf{z}_j^c(\cdot) : \mathfrak{R}^{L_1 \times J} \rightarrow \mathfrak{R}^{M_2}$ for $j = 1, \dots, J$.

Assume for moment, that we know the underlying taste distribution of P^0 and that we are able to observe the true market share \mathbf{s}^0 . Considering stochastic nature of the product characteristics \mathbf{X}_1 and $\boldsymbol{\xi}$, we set forth the demand side restriction as

$$(9) \quad \mathbb{E}_{\mathbf{x}_1, \boldsymbol{\xi}} [\mathbf{z}_j^d \boldsymbol{\xi}_j(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] = \mathbf{0}$$

at $\boldsymbol{\theta}_d = \boldsymbol{\theta}_d^0$ where the expectation is taken with respect not only to $\boldsymbol{\xi}$, but also to \mathbf{X}_1 . Supply side restriction we use is

$$(10) \quad \mathbb{E}_{\mathbf{w}_1, \omega} [\mathbf{z}_j^c \omega_j(\boldsymbol{\theta}, \mathbf{s}^0, P^0)] = \mathbf{0}$$

at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. The BLP(1995) framework uses the orthogonal conditions between the unobserved product characteristics $(\boldsymbol{\xi}_j, \omega_j)$ and the exogenous instrumental variables $(\mathbf{z}_j^d, \mathbf{z}_j^c)$ as moment conditions to obtain the GMM estimate of the parameter $\boldsymbol{\theta}$. The sample moments for the demand and supply systems are

$$(11) \quad \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) = \begin{pmatrix} \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) \\ \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \end{pmatrix} = \begin{pmatrix} \sum_j \mathbf{z}_j^d \boldsymbol{\xi}_j(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) / J \\ \sum_j \mathbf{z}_j^c \omega_j(\boldsymbol{\theta}, \mathbf{s}^0, P^0) / J \end{pmatrix}.$$

For some markets, market summaries are publicly available such as average demographics of consumers who purchased a specific type of products, even if their detailed individual-level data such as their purchasing histories are not. In the U.S. automobile market, for instance, we can obtain the data on the median income of consumers who purchased domestic, European, or Japanese vehicles from publications such as the *Ward's Motor Vehicle Facts & Figures*.

We now operationalize the idea given by Petrin (2002), which extends the BLP (1995) framework by adding moment conditions constructed from the market summary data. First we define some words and notations. *Discriminating attribute* is the product characteristic or the product attribute that enables consumers to discriminate some products from others. When we say consumer i chooses discriminating attribute q , this means that consumer chooses a product from a group of products whose characteristic or attribute

have discriminating attribute q . Discriminating attribute q is assumed to be a function of observed product characteristics \mathbf{X} . An automobile attribute “import” is one such discriminating attribute. When we say a consumer chooses this attribute, what we mean is that the consumer purchases an imported vehicle. Similarly, “minivan” and “costing between \$20,000 to \$30,000” are examples of the discriminating attribute. On the other hand, unobservable consumer’s proximity to a dealership is a function of ξ only and may not be regarded as a discriminating attribute as defined. We consider a finite number of discriminating attributes ($q = 1, \dots, N_p$) and denote a set of all the products that have attribute q as \mathcal{Q}_q . We assume the market share of products with discriminating attribute is positive (i.e., $\Pr[C_i \in \mathcal{Q}_q | \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] > 0$, where C_i denotes the choice of randomly sampled consumer i).

We next consider expectation of consumer’s demographics conditional on a specific discriminating attribute. Suppose that the consumer i ’s demographics can be decomposed into observable and unobservable components $\boldsymbol{\nu}_i = (\boldsymbol{\nu}_i^{obs}, \boldsymbol{\nu}_i^{unobs})$. The joint densities of $\boldsymbol{\nu}_i$ and $\boldsymbol{\nu}_i^{obs}$ are respectively denoted as $P^0(d\boldsymbol{\nu}_i)$ and $P^0(d\boldsymbol{\nu}_i^{obs})$. Observable demographic variables such as age, family size, or, income, is already numerical, but for other demographics such as household with children, belonging to a certain age group, choice of residential area, can be numerically expressed using indicators. We denote this numerically represented D dimensional demographics as $\boldsymbol{\nu}_i^{obs} = (\nu_{i1}^{obs}, \dots, \nu_{iD}^{obs})'$. We assume that the joint density of demographics $\boldsymbol{\nu}_i^{obs}$ is of bounded support. The consumer i ’s d -th observed demographic ν_{id}^{obs} , $d = 1, \dots, D$ is averaged over all consumers choosing discriminating attribute q in the population to obtain the conditional expectation $\eta_{dq}^0 = \mathbb{E}[\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \mathbf{X}, \xi(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)]$. An example of this conditional expectation would be the expected value of income of consumers in the population P^0 who purchased imported vehicles. We assume η_{dq}^0 has a finite mean and variance for all J , i.e., $\mathbb{E}_{\mathbf{x}, \xi}[\eta_{dq}^0] < \infty$ and $\text{V}_{\mathbf{x}, \xi}[\eta_{dq}^0] < \infty$ for $d = 1, \dots, D, q = 1, \dots, N_p$.

Let $\Pr[d\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)]$ be the conditional density of consumer i ’s demographics ν_{id}^{obs} given his/her choice of discriminating attribute q and product characteristics $(\mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0))$. Since the conditional expectation η_{dq}^0 can be written as

$$\begin{aligned}
(12) \quad & \mathbb{E}[\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] \\
&= \int \nu_{id}^{obs} \Pr[d\nu_{id}^{obs} | C_i \in \mathcal{Q}_q, \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] \\
&= \frac{\int \nu_{id}^{obs} \Pr[C_i \in \mathcal{Q}_q | \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \nu_{id}^{obs}] P^0(d\nu_{id}^{obs})}{\Pr[C_i \in \mathcal{Q}_q | \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)]} \\
&= \frac{\int \nu_{id}^{obs} \Pr[C_i \in \mathcal{Q}_q | \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\nu}_i] P^0(d\boldsymbol{\nu}_i)}{\Pr[C_i \in \mathcal{Q}_q | \mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)]} \\
&= \int \nu_{id}^{obs} \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{ij}(\mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\nu}_i; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \xi(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} P^0(d\boldsymbol{\nu}_i),
\end{aligned}$$

we can form an identity, which is the basis for additional moment conditions

$$(13) \quad \eta_{dq}^0 - \int \nu_{id}^{obs} \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{ij}(\mathbf{X}, \xi(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\nu}_i; \boldsymbol{\theta}_d^0)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \xi(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)} P^0(d\boldsymbol{\nu}_i) \equiv 0$$

for $q = 1, \dots, N_p, d = 1, \dots, D$.

Although P^0 is so far assumed known, we typically are not able to calculate the second term on the left-hand side of (13) analytically and will have to approximate it by using the empirical distribution P^T of i.i.d. sample $\boldsymbol{\nu}_t, t = 1, \dots, T$ from the underlying distribution P^0 . The corresponding sample moments $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$ (a on the shoulder stands for additional) are

$$(14) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = \boldsymbol{\eta}^0 - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$$

where

$$(15) \quad \boldsymbol{\eta}^0 = (\eta_{11}^0, \dots, \eta_{1N_p}^0, \dots, \eta_{D1}^0, \dots, \eta_{DN_p}^0)', \quad \boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \begin{pmatrix} \frac{\sum_{j \in \mathcal{Q}_1} \sigma_{tj}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_1} \sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \\ \vdots \\ \frac{\sum_{j \in \mathcal{Q}_{N_p}} \sigma_{tj}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_{N_p}} \sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \end{pmatrix}.$$

The symbol \otimes denotes the Kronecker product. The quantity $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ is the consumer t 's model-calculated purchasing probability of products with discriminating attribute q relative to the model-calculated market share of the same products. Note that these additional moments are again conditional on product characteristics $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0))$, and thus depend on the indices J and T .

We use the set of the three moments, two from (11) and from (14) as

$$(16) \quad \mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = \begin{pmatrix} \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) \\ \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \\ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) \end{pmatrix}$$

to estimate $\boldsymbol{\theta}$ in theory. As pointed out in BLP (2004), we have two issues when evaluating $\|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|$. First, we assume P^0 is known so far, we typically are not able to calculate $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ analytically and have to approximate it by a simulator, say $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$, where P^R is the empirical measure of i.i.d. sample $\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_R$ from the underlying distribution P^0 , and the sample is independent of the sample $\boldsymbol{\nu}_t, t = 1, \dots, T$ in (14) for evaluating the additional moments. Simulated market shares are then given by

$$(17) \quad \sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \int \sigma_{ij}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_i; \boldsymbol{\theta}_d) dP^R(\boldsymbol{\nu}_i) \equiv \frac{1}{R} \sum_{r=1}^R \sigma_{rj}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d).$$

Second, we are not necessarily able to observe the true market shares \mathbf{s}^0 . Instead, the vector of given observed market shares, \mathbf{s}^n , are typically constructed from n i.i.d. draws from the population of consumers, and hence is not equal to the population value \mathbf{s}^0 in general. The observed market share of product j is

$$(18) \quad s_j^n = \frac{1}{n} \sum_{i=1}^n 1(C_i = j),$$

where the indicator variable $1(C_i = j)$ takes one if $C_i = j$ and zero otherwise. Since C_i denotes the choice of randomly sampled consumer i , they are i.i.d. across i .

We substitute $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ given as a solution of $\mathbf{s}^n - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \mathbf{0}$ for (11) to obtain

$$(19) \quad \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) = J^{-1} \sum_{j=1}^J \mathbf{z}_j^d \xi_j(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R).$$

Furthermore, substituting $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) = (\omega_1(\boldsymbol{\theta}, \mathbf{s}^n, P^R), \dots, \omega_J(\boldsymbol{\theta}, \mathbf{s}^n, P^R))'$ obtained from evaluating (7) at $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ and $P = P^R$ for (11) gives

$$(20) \quad \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R) = J^{-1} \sum_{j=1}^J \mathbf{z}_j^c \omega_j(\boldsymbol{\theta}, \mathbf{s}^n, P^R).$$

In addition, we have another issue when evaluating the additional moments in (14). In general, we do not know the conditional expectation of demographics η_{dq}^0 , instead, we have its estimate η_{dq}^N from independent sources, which is typically estimated from the sample of N consumers. The sample counterparts we can calculate for the additional moments are thus

$$(21) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) = \boldsymbol{\eta}^N - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$$

for $\boldsymbol{\theta}_d \in \Theta_d$. As a result, the actual sample-based objective function we minimize in the GMM estimation is the sum of norm of $\mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$, $\mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R)$, and $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$, that is, the norm of

$$(22) \quad \mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) = \begin{pmatrix} \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) \\ \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R) \\ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) \end{pmatrix}.$$

Notice that the first two moments \mathbf{G}_J^d and \mathbf{G}_J^c in (22) are sample moments averaged over products $j = 1, \dots, J$, while the third moment $\mathbf{G}_{J,T}^a$ is averaged over consumers $t = 1, \dots, T$. Note also that in the expression $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$, there exist five distinct randomness: one from the draws of the product characteristics $(\mathbf{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$, two from the sampling processes not controlled by the econometrician of consumers for \mathbf{s}^n and $\boldsymbol{\eta}^N$, two from the empirical distributions P^R and P^T employed by the econometrician. The impact of these randomness on the estimate of $\boldsymbol{\theta}$ are decided by the relative size of the sample— J , n , N , R and T . Now we are going to operationalize the sampling and the simulation errors in the following.

2.4 The sampling and simulation errors

The sampling error, $\boldsymbol{\epsilon}^n$, is defined as the difference between the observed market shares \mathbf{s}^n and the true market share \mathbf{s}^0 . Specifically, its component ϵ_j^n for the product j is

$$(23) \quad \epsilon_j^n \equiv s_j^n - s_j^0 = \frac{1}{n} \sum_{i=1}^n \{1(C_i = j) - s_j^0\} = \frac{1}{n} \sum_{i=1}^n \epsilon_{ji}$$

for $j = 1, \dots, J$, where $\epsilon_{ji} \equiv 1(C_i = j) - s_j^0, i = 1, \dots, n$ are the differences of the sampled consumer's choice from the population market share (s_j^0) of the same choice and are assumed independent across i .

Note that from (2), for any $\theta_d \in \Theta_d$, the unique solutions ξ for $\mathbf{s}^n - \sigma(\mathbf{X}, \xi, \theta_d, P^R) = \mathbf{0}$ and $\mathbf{s}^0 - \sigma(\mathbf{X}, \xi, \theta_d, P^0) = \mathbf{0}$ are written as $\xi(\theta_d, \mathbf{s}^n, P^R)$ and $\xi(\theta_d, \mathbf{s}^0, P^0)$ respectively. So, substituting these ξ s back into $\sigma(\mathbf{X}, \xi, \theta_d, P^R)$ and $\sigma(\mathbf{X}, \xi, \theta_d, P^0)$ retrieves \mathbf{s}^n and \mathbf{s}^0 respectively, or $\mathbf{s}^n = \sigma(\mathbf{X}, \xi(\theta_d, \mathbf{s}^n, P^R), \theta_d, P^R)$ and $\mathbf{s}^0 = \sigma(\mathbf{X}, \xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)$ for any $\theta_d \in \Theta_d$. Similarly, if we evaluate (2) with the observed (true) market share \mathbf{s}^n (\mathbf{s}^0) and the underlying (empirical) population P^0 (P^R) of consumers, the resulting $\xi(\theta_d, \mathbf{s}^n, P^0)$ ($\xi(\theta_d, \mathbf{s}^0, P^R)$) satisfies $\mathbf{s}^n = \sigma(\mathbf{X}, \xi(\theta_d, \mathbf{s}^n, P^0), \theta_d, P^0)$ ($\mathbf{s}^0 = \sigma(\mathbf{X}, \xi(\theta_d, \mathbf{s}^0, P^R), \theta_d, P^R)$) for all $\theta_d \in \Theta_d$. These facts are used to define the simulation errors below.

The simulation process generates the simulation error $\epsilon^R(\theta_d)$, which is for any θ_d the difference between the simulated market shares in (17) from the P^R and those from the P^0 . The simulation error ϵ_j^R for product j with sample of R consumers is defined as follows.

$$(24) \quad \begin{aligned} \epsilon_j^R(\theta_d) &\equiv \sigma_j(\mathbf{X}, \xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R) - \sigma_j(\mathbf{X}, \xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0) \\ &= \frac{1}{R} \sum_{r=1}^R \epsilon_{jr}^*(\mathbf{X}, \xi(\theta_d, \mathbf{s}^0, P^0), \theta_d) \end{aligned}$$

for $j = 1, \dots, J$, where $\epsilon_{jr}^*(\mathbf{X}, \xi, \theta_d) = \sigma_{rj}(\mathbf{X}, \xi, \nu_r; \theta_d) - \sigma_j(\mathbf{X}, \xi, \theta_d, P^0), r = 1, \dots, R$ are independent across r conditional on (\mathbf{X}, ξ) by the simulating process.

We also assume N independent consumer draws with their purchasing histories are used to construct the additional information $\boldsymbol{\eta}^N = (\eta_{11}^N, \dots, \eta_{1N_p}^N, \dots, \eta_{D1}^N, \dots, \eta_{DN_p}^N)'$ and define the sampling error $\boldsymbol{\epsilon}^N$ in the additional information $\boldsymbol{\eta}^N$ itself as follows.

$$(25) \quad \boldsymbol{\epsilon}^N \equiv \boldsymbol{\eta}^N - \boldsymbol{\eta}^0 = \frac{1}{N} \sum_{i'=1}^N \boldsymbol{\epsilon}_{i'}^\#.$$

In short, we assume here that $\boldsymbol{\eta}^N$ is the average of N conditionally independent random variables given the set of product characteristics (\mathbf{X}, ξ) of all products.

Since we use the sample of T draws of consumer to evaluate the additional moments, this also induces the sampling error in $\mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ in (21). Note that quantities n and N are normally *beyond the control* of the econometrician. On the other hand quantities R and T are both *chosen by* the econometrician.

2.5 Metrics, Neighborhoods, and Notations

The metrics, neighborhoods, notations are kept as identical as possible to those in BLP (2004). We work with the product space $\Theta \times \mathcal{S}_J \times \mathcal{P}$. The parameter space Θ is a compact subset of \Re^K and we use the Euclidean metric on Θ , $\rho_E(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|$. The space for the market share vector \mathbf{s} is $J + 1$ dimensional unit simplex \mathcal{S}_J , $\mathcal{S}_J = \{(s_0, \dots, s_J)' \mid 0 < s_j < 1 \text{ for } j = 0, \dots, J, \text{ and } \sum_{j=0}^J s_j = 1\}$. Since the market

share s_j tends to shrink as the number J of the products on the market increases, we need to make sure the speed at which the s_j converges to the true share s_j^0 be faster than the speed at which s_j^0 converges to zero. Hence, we use the metric $\rho_{s^0}(\mathbf{s}, \mathbf{s}^*) = \max_{0 \leq j \leq J} |(s_j - s_j^*)/s_j^0|$ on \mathcal{S}_J .

The \mathcal{P} is the set of probability measures of consumer's tastes. The L_∞ metric $\rho_P(P, P^*) = \sup_{B \in \mathcal{B}} |P(B) - P^*(B)|$ is adopted on \mathcal{P} , where \mathcal{B} is the class of all Borel sets on \mathfrak{R}^v , where v is the dimension of $\boldsymbol{\nu}_i$ in the purchasing probability. This metric measures the distance between the empirical distribution P^R and the underlying distribution P^0 of $\boldsymbol{\nu}_i$.

Since the dimension of the unobserved product characteristics $\boldsymbol{\xi}$ increases as the number J of products increases, element by element convergence of $\boldsymbol{\xi}$ to $\boldsymbol{\xi}^*$ does not necessarily guarantee that $\|\boldsymbol{\xi} - \boldsymbol{\xi}^*\| = o_p(1)$. What we need is the convergence of the unobserved product characteristics $\boldsymbol{\xi}$ as a vector to another vector $\boldsymbol{\xi}^*$, not an element by element convergence. Hence we use the averaged Euclidean metric $\rho_\xi(\boldsymbol{\xi}, \boldsymbol{\xi}^*) = J^{-1} \|\boldsymbol{\xi} - \boldsymbol{\xi}^*\|^2 = J^{-1} \sum_{j=1}^J (\xi_j - \xi_j^*)^2$, which allows for the possibility that a finite number of elements in $\boldsymbol{\xi}$ do not converge to the corresponding elements in $\boldsymbol{\xi}^*$.

With these metrics, we define the δ neighborhoods for $\boldsymbol{\theta}^0$, \mathbf{s}^0 , and P^0 respectively as $\mathcal{N}_{\boldsymbol{\theta}^0}(\delta) = \{\boldsymbol{\theta} : \rho_E(\boldsymbol{\theta}, \boldsymbol{\theta}^0) \leq \delta\}$, $\mathcal{N}_{\mathbf{s}^0}(\delta) = \{\mathbf{s} : \rho_s(\mathbf{s}, \mathbf{s}^0) \leq \delta\}$, and $\mathcal{N}_{P^0}(\delta) = \{P : \rho_P(P, P^0) \leq \delta\}$. Also for each $\boldsymbol{\theta}$, the δ neighborhood of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ is defined by $\mathcal{N}_{\boldsymbol{\xi}^0}(\boldsymbol{\theta}_d; \delta) = \{\boldsymbol{\xi} : \rho_\xi(\boldsymbol{\xi}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)) \leq \delta\}$.

The notation we use for the Euclidean norm of any $m \times n$ matrix \mathbf{A} is $\|\mathbf{A}\| = \{\text{tr}(\mathbf{A}'\mathbf{A})\}^{1/2}$. We use the $O_p(\cdot)$ and $o_p(\cdot)$ notation of Mann and Wald (1944) to denote the stochastic order of magnitude. When applied to vectors and matrices, they measure element by element magnitude. If \mathbf{x} is a $k \times 1$ vector, $\text{diag}[\mathbf{x}]$ denotes a $k \times k$ diagonal matrix with the element of \mathbf{x} along its principle diagonal.

3 Asymptotic Properties of the GMM estimator

In this section, we derive the asymptotic theorems for the GMM estimator $\hat{\boldsymbol{\theta}}$ which minimizes the norm of $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ in (22). With some additional assumptions, we extend theorems in BLP (2004) to show that the suggested estimator has CAN properties. The proofs are in Appendix.

3.1 Consistency

The consistency argument is established by showing that

(1-i) the estimator $\tilde{\boldsymbol{\theta}}$ defined as any sequence that satisfies $\|\mathbf{G}_{J,T}(\tilde{\boldsymbol{\theta}}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| = \inf_{\boldsymbol{\theta} \in \Theta} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| + o_p(1)$ is consistent for $\boldsymbol{\theta}^0$, and

(1-ii) $\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| = o_p(1)$.

A consequence of (1-ii) is that $\|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)\|$ and $\|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|$ have the same asymptotic distribution uniformly in $\boldsymbol{\theta}$, and thus the estimator $\hat{\boldsymbol{\theta}}$ which minimizes the former is very close to the $\tilde{\boldsymbol{\theta}}$ that minimizes the latter. Therefore $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}^0$ from (1-i).

In what follows, we explain the roles of assumptions play to obtain the consistency as we present them. Assumptions A1–A9 govern the limiting behavior of the random components both in the demand, supply and additional moments. They include assumptions A1–A6 in BLP (2004) on the demand side.

Assumptions A1 are on various errors. In Assumption A1(a), we assume the observed market shares s_j^n for product j are multinomial random variables averaged over the n sampled consumers ($i = 1, \dots, n$). Assumption A1(b) guarantees that the simulation error $\epsilon_{j_r}^*$ in (24) relative to the number R of the simulation draws is of the same order as the sampling error ϵ_{ji} relative to the number n of the sample. With assumption A1(c), η_{dq}^N is unbiased conditional on $(\mathbf{X}, \boldsymbol{\xi})$ and $N^{1/2}$ consistent for the true η_{dq}^0 . These assumptions are used to control the magnitudes of the respective errors. Note that A1(a) and (c) are assumptions on the consumer behaviors because \mathbf{s}^n and \mathbf{s}^0 are the results of actual consumers' choices, and the consumers are assumed to be able to observe the true unobserved product characteristics, $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$. As a result, for A1(a) and (c) we can condition on \mathbf{X} and on $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$, but not on a general $\boldsymbol{\xi}$ when evaluating the moments of the difference $\mathbf{s}^n - \mathbf{s}^0$. On the other hand, A1(b) is an assumption on consumer behaviors from the econometrician's point of view because $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ and $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$, both of which are model-calculated shares, are the devices the econometrician uses and s/he is not able to observe the unobserved product characteristics, true or otherwise. As a result, we need to condition on general unobserved $\boldsymbol{\xi}$ along with on the \mathbf{X} . Formally,

Assumption A1 (a) *Given the set of product characteristics $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$, the difference $\mathbf{s}^n - \mathbf{s}^0$ between the observed market share \mathbf{s}^n and the “conditionally” true market share \mathbf{s}^0 has conditional mean $E_{\epsilon|\mathbf{x}, \boldsymbol{\xi}}[\mathbf{s}^n - \mathbf{s}^0 | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = \mathbf{0}$ with the conditional variance-covariance matrix $\mathbf{V}_2 = E_{\epsilon|\mathbf{x}, \boldsymbol{\xi}}[(\mathbf{s}^n - \mathbf{s}^0)(\mathbf{s}^n - \mathbf{s}^0)' | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = (\text{diag}[\mathbf{s}^0] - \mathbf{s}^0 \mathbf{s}^{0'})/n$.*

(b) *For each $\boldsymbol{\theta}_d$, given the set of product characteristics $(\mathbf{X}, \boldsymbol{\xi})$, the difference $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ has conditional mean $E_{\epsilon^*|\mathbf{x}, \boldsymbol{\xi}}[\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) | \mathbf{X}, \boldsymbol{\xi}] = \mathbf{0}$ with the conditional variance-covariance matrix $\mathbf{V}_3 = E_{\epsilon^*|\mathbf{x}, \boldsymbol{\xi}}[\{\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)\} \cdot \{\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)\}' | \mathbf{X}, \boldsymbol{\xi}]$ whose order of magnitude relative to R is the same as that of \mathbf{V}_2 relative to n or, $R \cdot O(\mathbf{V}_3) = n \cdot O(\mathbf{V}_2)$.*

(c) *For all observed consumer's demographics $d = 1, \dots, D$ and for all discriminating attributes $q = 1, \dots, N_p$, the sampling error $\eta_{dq}^N - \eta_{dq}^0$ has conditional mean $E_{\epsilon^\#|\mathbf{x}, \boldsymbol{\xi}}[\eta_{dq}^N - \eta_{dq}^0 | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = 0$ with the conditional variance $\mathbf{V}_4 = V_{\epsilon^\#|\mathbf{x}, \boldsymbol{\xi}}[\eta_{dq}^N - \eta_{dq}^0 | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)]$ whose order of magnitude is $1/N$.*

Assumption A2 is a smoothness or regularity condition for the share function. In A2(a), we first assume the model-calculated market share $\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ for product j does not abruptly change as the unobserved product quality ξ_k for product k changes. We further assume the $\mathbf{H} = \partial \boldsymbol{\sigma} / \partial \boldsymbol{\xi}'$ is invertible, and this means one can measure the change in unobserved product quality $\partial \xi_j$ for product j ($j = 1, \dots, J$) associated with the change in the model-calculated market share $\partial \sigma_k$ for product k ($k = 1, \dots, J$). Assumption A2(b) stipulates how the model-calculated market share $\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ for product j is affected by the changes in unobserved product quality for product k . It is positively affected by the improvement of its own unobserved

quality, but adversely influenced by those of the other products. Assumptions A2(a) and (b) are sufficient for the existence of a unique solution $\boldsymbol{\xi}$ to (2) for every $(\boldsymbol{\theta}_d, \mathbf{s}, P)$ (See appendix in Berry (1994) for detail).

It looks as if we need a similar setup for the supply side unobserved cost shifter ω_j relative to the model-calculated market share σ_k . This is not so, however, because as clearly seen in (7), the $\omega_j(\boldsymbol{\theta}, \mathbf{s}, P)$ can be obtained as a function of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P)$ with the observed (p_j, \mathbf{w}_j) and the given parameters $(\boldsymbol{\theta}_d, \boldsymbol{\theta}_c)$ once we decide to choose on which (\mathbf{s}, P) it is evaluated. This enables the characteristics of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P)$ to transmit to $\omega_j(\boldsymbol{\theta}, \mathbf{s}, P)$ if there exists a profit margin $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P), \boldsymbol{\theta}_d, P)$ in (8) that is at least locally smooth with respect to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P)$ along with smoothness in $g(\cdot)$. Assumption A2(c) guarantees the existence of $\boldsymbol{\Delta}^{-1}$, which in turn guarantees the existence of $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P), \boldsymbol{\theta}_d, P)$ in (8). We place local smoothness of $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P), \boldsymbol{\theta}_d, P)$ relative to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P)$ in the form to appear in assumption A7. As for smoothness of $g(\cdot)$, we reiterate that the single argument function $g(\cdot)$ is monotonic and continuously differentiable with finite derivative for all realizable values of cost. We choose not to include this in the assumptions simply because this does not rise to the same level as the other assumptions are. Therefore,

Assumption A2 (a) For every finite J , for all $\boldsymbol{\theta}_d \in \Theta_d$, and for all P in a neighborhood of P^0 , $\partial\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\xi_k$ exists, and is continuously differentiable both in $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d$. The matrix $\mathbf{H}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \partial\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\boldsymbol{\xi}'$ is invertible for all J .

(b) For every $(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, $\partial\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\xi_j > 0$ for $j = 1, \dots, J$, and $\partial\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\xi_k < 0$ for $k, j = 1, \dots, J, k \neq j$.

(c) For every finite J , for all $\boldsymbol{\theta}_d \in \Theta_d$, and for all P in a neighborhood of P^0 , $\partial\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial p_k$ exists for $j, k = 1, \dots, J$, and the matrix $\boldsymbol{\Delta}$ whose (j, k) element is defined in (4) is invertible for all J and continuously differentiable both in $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d$.

In cases we consider here, the number J of the products in the market increases. This means that each component of the “conditionally” true market share \mathbf{s}^0 and also of the theoretical market share $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ generally approaches to zero as J grows large. Assumptions A3(a),(b) guarantee that \mathbf{s}^n and $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ respectively converge to \mathbf{s}^0 and $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ faster than the speed at which each component of \mathbf{s}^0 and of $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ converges to zero.

Assumption A3 The observed market shares \mathbf{s}^n are consistent with respect to \mathbf{s}^0 , i.e., for any $\delta > 0$,

$$(a) \quad \rho_{\mathbf{s}^0}(\mathbf{s}^n, \mathbf{s}^0) = \max_{0 \leq j \leq J} |(s_j^n - s_j^0)/s_j^0| = o_p(1).$$

Similarly, the simulated market shares $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ are consistent with respect to $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ uniformly over $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d \in \Theta_d$, i.e., for any $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d \in \Theta_d$,

$$(b) \quad \rho_{\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R), \boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) = \max_{0 \leq j \leq J} \left| \frac{\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}{\sigma_j(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)} \right| = o_p(1).$$

Assumption A4 is on instrumental variables. Throughout the paper, we treat the product characteristics \mathbf{x}_{1j} as exogenous and so do the demand side instruments \mathbf{z}_j^d . We impose in A4(a) stochastic boundedness and uniformly integrability on \mathbf{z}_j^d . In assumption A4(b), the same restrictions are imposed on the supply side instruments \mathbf{z}_j^c .

Assumption A4 (a) *The demand side instrumental variables are such that the matrix $\mathbf{Z}'_d \mathbf{Z}_d / J$ is stochastically bounded, i.e., for all $\epsilon > 0$ there exists an M_ϵ such that $\Pr[\|\mathbf{Z}'_d \mathbf{Z}_d / J\| > M_\epsilon] < \epsilon$. Moreover, we suppose $\|\mathbf{Z}'_d \mathbf{Z}_d / J\|$ is uniformly integrable in J , i.e., $\lim_{\alpha \rightarrow \infty} \sup_J \int \|\mathbf{Z}'_d \mathbf{Z}_d / J\| \mathbb{1}\{\|\mathbf{Z}'_d \mathbf{Z}_d / J\| > \alpha\} dP_{\mathbf{x}_1}(\mathbf{X}_1) = 0$ where $P_{\mathbf{x}_1}(\cdot)$ is the joint distribution of \mathbf{X}_1 .*

(b) *The supply side instrumental variables are such that the matrix $\mathbf{Z}'_c \mathbf{Z}_c / J$ is stochastically bounded and uniformly integrable in J .*

Assumption A5 is a condition that bounds $\|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)\|$ away from $\|\mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\|$ (which converges to zero in probability) over $\boldsymbol{\theta}$ outside of a neighborhood of $\boldsymbol{\theta}^0$.

Assumption A5 *For all $\delta > 0$, there exists $C(\delta)$ such that*

$$\lim_{J \rightarrow \infty} \Pr \left[\inf_{\boldsymbol{\theta} \notin \mathcal{N}_{\boldsymbol{\theta}^0}(\delta)} \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\| \geq C(\delta) \right] = 1.$$

For all $\boldsymbol{\theta}_d$, the value of $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ that satisfies the equation $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) = \mathbf{s}^0$ is assumed unique. Since the sum of the market shares including that of the outside good s^0_0 , is unity, this $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ also satisfies $\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) / \sigma_0(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) = \mathbf{s}^0 / s^0_0$. Define a function $\boldsymbol{\tau}_J(\cdot) : \mathbb{R}^J \rightarrow \mathbb{R}^J$ such that $\boldsymbol{\tau}_J(\mathbf{s}) = (\log(s_1/s_0), \dots, \log(s_J/s_0))$. Then, the relation is equivalent to saying that $\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) = \boldsymbol{\tau}_J(\mathbf{s}^0) = \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0))$ at $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ for all $\boldsymbol{\theta}_d$. Assumption A6 guarantees that any $\boldsymbol{\xi}$ outside the δ neighborhood of the $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ cannot make $\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0))$ close to $\boldsymbol{\tau}_J(\mathbf{s}^0)$ within the range of $C(\delta)$ in terms of the averaged Euclidean distance with probability tending to one. The choice of this metric is necessary because we need to allow for the fact that the dimension of the model-calculated market share $\boldsymbol{\sigma}$ increases as the number J of products increases. The particular form of $\boldsymbol{\tau}_J$ makes this assumption easier to verify for logit-like demand models.

Assumption A6 *For all $\delta > 0$, there exists $C(\delta)$ such that*

$$\lim_{J \rightarrow \infty} \Pr \left[\inf_{\boldsymbol{\theta}_d \in \Theta_d} \inf_{\boldsymbol{\xi} \notin \mathcal{N}_{\boldsymbol{\xi}^0}(\boldsymbol{\theta}_d, \delta)} J^{-\frac{1}{2}} \|\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0))\| > C(\delta) \right] = 1.$$

The following assumption A7 is one that we additionally impose on the profit margin for the vector of products, because we incorporate the supply side as well. In assumption A7, we assume the profit margins $J^{-\frac{1}{2}} \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}, P), \boldsymbol{\theta}_d, P)$ have stochastically equicontinuity-like characteristics in $(\boldsymbol{\xi}, P)$ at $(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), P^0)$ for any $\boldsymbol{\theta}_d \in \Theta_d$. As seen in the consistency proof of BLP (2004), $\Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) \notin \mathcal{N}_{\boldsymbol{\xi}^0}(\boldsymbol{\theta}_d, \delta)] \rightarrow 0$ and $\Pr[P^R \notin \mathcal{N}_{P^0}(\delta)] \rightarrow 0$ for $\delta > 0$ as J grows large. With these convergence in probability results along with assumption A7, we are able to show the averaged Euclidean distance between $\mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ and $\mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ is close uniformly in $\boldsymbol{\theta}_d \in \Theta_d$. We should note that assumption A7 is not stochastic equicontinuity as normally defined because the dimension of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ grows large as J grows, though $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ converges to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ in probability in averaged Euclidean metric.⁵

⁵One more comment on the behavior of the dimension increasing $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$. It should be noted that when evaluated at the true parameter value $\boldsymbol{\theta}_d^0$ as J increases, say, from 100 to 500, the first 100 elements of $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$ at $J = 500$ must be

Assumption A7 For all $\delta > 0$ and for any $\boldsymbol{\theta}_d \in \Theta_d$,

$$\lim_{J \rightarrow \infty} \Pr \left[\sup_{(\boldsymbol{\xi}, P) \in \mathcal{N}_{\boldsymbol{\xi}^0}(\boldsymbol{\theta}_d; \delta) \times \mathcal{N}_{P^0}(\delta)} J^{-\frac{1}{2}} \|\mathbf{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| > \delta \right] = 0.$$

In assumption A8, we assume an asymptotic property the discriminating attributes $q, q = 1, \dots, N_p$ must obey. We guarantee non-zero aggregate market share for the products with discriminating attribute q when the number of products J grows large. With this assumption, the additional moment defined in (14) has finite variance at $\boldsymbol{\theta}_d = \boldsymbol{\theta}_d^0$.

Assumption A8 For all discriminating attributes $q = 1, \dots, N_p$, $\{\sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\}^{-2}$ has a finite mean and variance for every J .

Assumption A9 is on $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, the model-calculated purchasing probabilities of consumer t of products with discriminating attribute q relative to the model-calculated market share of the same products t relative to the population P . We assume that the average absolute distance between $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ and $\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ converges to zero in probability within the δ neighborhood of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ for any $\boldsymbol{\theta}_d \in \Theta_d$. This assumption will be used to bring the sample analogue of the additional moments, $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ close enough to $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^N)$ for any $\boldsymbol{\theta}_d$.

Assumption A9 For any $\boldsymbol{\theta}_d \in \Theta_d$, and for all $\delta > 0$,

$$\lim_{J, T \rightarrow \infty} \Pr \left[\sup_{(\boldsymbol{\xi}, P) \in \mathcal{N}_{\boldsymbol{\xi}^0}(\boldsymbol{\theta}_d; \delta) \times \mathcal{N}_{P^0}(\delta)} T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| > \delta \right] = 0,$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = (\boldsymbol{\psi}_1(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \boldsymbol{\psi}_T(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))'$.

Now we are ready to state the consistency of the Petrin estimator with the additional moments:

Theorem 1 (Consistency of $\hat{\boldsymbol{\theta}}$) Suppose that A1–A9 hold for some $n(J, T), R(J, T)$, and N , all of which grow infinitely as J and T grow infinitely. Then, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^0$.

3.2 Asymptotic Normality

To establish asymptotic normality, we first approximate $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ in (22) by $\mathcal{G}_{J,T}(\boldsymbol{\theta}) = (\mathcal{G}_J(\boldsymbol{\theta})', \mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d)')'$ within δ neighborhood of $\boldsymbol{\theta}^0$, where $\mathcal{G}_{J,T}(\boldsymbol{\theta})$ is $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$ plus the terms associated with sampling and simulation errors. Then, we show that

$$(2\text{-i}) \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq \delta_{J,T}} \left\| \begin{array}{l} J^{\frac{1}{2}} [\mathcal{G}_J(\boldsymbol{\theta}) - \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^n, P^R)] \\ T^{\frac{1}{2}} [\mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)] \end{array} \right\| \xrightarrow{p} 0 \text{ when } \delta_{J,T} \rightarrow 0, \text{ and}$$

(2-ii) an estimator that minimizes $\|\mathcal{G}_{J,T}(\boldsymbol{\theta})\|$ over $\boldsymbol{\theta} \in \Theta$; (1) is asymptotically normal at the rate $J^{\frac{1}{2}}$ assuming T goes to infinity faster than J , and (2) has a variance-covariance matrix which is the sum of equal to the all 100 elements of $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$ at $J = 100$. This fact does not hold in general when evaluated at $\boldsymbol{\theta}_d \neq \boldsymbol{\theta}_d^0$. For instance there is no guarantee that the first 100 elements of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ at $J = 500$ are equal to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ at $J = 100$.

three mutually uncorrelated terms (one resulting from randomness in the draws on exogenous variables $(\mathbf{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$, one from sampling errors ϵ_j^n , and one from simulation error $\epsilon_j^R(\boldsymbol{\theta}_d)$).

Given consistency, a consequence of (2-i) is that the estimator obtained from minimizing $\|\mathbf{G}_{J,T}(\boldsymbol{\theta})\|$, has the same limiting distribution as our estimator that minimizes $\|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)\|$.

As in BLP (2004), we first decompose the unobserved quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ into three random terms—the unobserved quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$, the term generated from the sampling error ϵ^n , and the term generated from the simulation error $\epsilon^R(\boldsymbol{\theta}_d)$. This allows us to express the demand side moment $\mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ in (19) as the sum of the three conditionally independent terms as

$$\begin{aligned}
(26) \quad \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) &= J^{-1} \mathbf{Z}'_d \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) \\
&= J^{-1} \mathbf{Z}'_d \left[\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) + \{ \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R) \} + \{ \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) \} \right] \\
&= \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) + J^{-1} \mathbf{Z}'_d \{ \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \epsilon^n - \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \epsilon^R(\boldsymbol{\theta}_d) \}
\end{aligned}$$

where $\bar{\boldsymbol{\xi}} \equiv (\bar{\xi}_1, \dots, \bar{\xi}_J)$ is a set of $J \times 1$ vectors of the values between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$, and so is $\underline{\boldsymbol{\xi}} \equiv (\underline{\xi}_1, \dots, \underline{\xi}_J)$ between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ with the notation

$$\mathbf{H}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P) = \begin{pmatrix} \frac{\partial \sigma_1}{\partial \xi_1} \Big|_{\bar{\xi}_1} & \cdots & \frac{\partial \sigma_1}{\partial \xi_J} \Big|_{\bar{\xi}_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J}{\partial \xi_1} \Big|_{\bar{\xi}_J} & \cdots & \frac{\partial \sigma_J}{\partial \xi_J} \Big|_{\bar{\xi}_J} \end{pmatrix} \quad \text{and} \quad \mathbf{H}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P) = \begin{pmatrix} \frac{\partial \sigma_1}{\partial \xi_1} \Big|_{\underline{\xi}_1} & \cdots & \frac{\partial \sigma_1}{\partial \xi_J} \Big|_{\underline{\xi}_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J}{\partial \xi_1} \Big|_{\underline{\xi}_J} & \cdots & \frac{\partial \sigma_J}{\partial \xi_J} \Big|_{\underline{\xi}_J} \end{pmatrix}.$$

Using the similar decomposition of the cost side unobservable $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R)$, the cost side moment $\mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R)$ in (20) can be expressed as

$$\begin{aligned}
(27) \quad \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R) &= J^{-1} \mathbf{Z}'_c \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) \\
&= J^{-1} \mathbf{Z}'_c \left[\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0) + \{ \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R) \} + \{ \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \} \right] \\
&= \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \\
&\quad + J^{-1} \mathbf{Z}'_c \{ \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \} \\
&\quad - J^{-1} \mathbf{Z}'_c \mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \epsilon^n \\
&\quad + J^{-1} \mathbf{Z}'_c \mathbf{L}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \epsilon^R(\boldsymbol{\theta}_d)
\end{aligned}$$

where $\bar{\boldsymbol{\xi}}$ is between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$, and so is $\underline{\boldsymbol{\xi}}$ between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ with the notation $\mathbf{L}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \text{diag}[\dot{g}(p_1 - m_{g_1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)), \dots, \dot{g}(p_J - m_{g_J}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))]$ and $\mathbf{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \partial \mathbf{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial \boldsymbol{\xi}'$. Actually, $J \times J$ matrices $\mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)$ and $\mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)$ contain $\bar{\xi}_1, \dots, \bar{\xi}_J$ in its 1st to the J th rows, all of which can be distinct, but we here suppress this fact for notational simplicity.

As for the additional moments, we rewrite (21) in the following form:

$$(28) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) \\ = \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \left\{ \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right. \\ \left. + \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\ddagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right\} + \boldsymbol{\epsilon}^N$$

where $\boldsymbol{\xi}^\dagger \equiv (\boldsymbol{\xi}_1^\dagger, \dots, \boldsymbol{\xi}_J^\dagger)$ is a set of intermediate vectors between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$, and so are $\boldsymbol{\xi}^\ddagger \equiv (\boldsymbol{\xi}_1^\ddagger, \dots, \boldsymbol{\xi}_J^\ddagger)$ between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$ with the notation $\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \equiv \partial \boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial \boldsymbol{\xi}'$.

We approximate $\mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$, $\mathbf{G}_J^c(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$, and $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ within the neighborhood of $\boldsymbol{\theta}^0$ respectively by the following functions.

$$(29) \quad \begin{pmatrix} \mathbf{G}_J^d(\boldsymbol{\theta}_d) \\ \mathbf{G}_J^c(\boldsymbol{\theta}) \\ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d) \end{pmatrix} = \begin{pmatrix} \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) + J^{-1} \mathbf{Z}'_d \mathbf{H}_0^{-1} \{ \boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \} \\ \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - J^{-1} \mathbf{Z}'_c \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \{ \boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \} \\ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \{ \boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \} + \boldsymbol{\epsilon}^N \end{pmatrix}.$$

where $\mathbf{H}_0 = \mathbf{H}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$, $\mathbf{L}_0 = \mathbf{L}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$, $\mathbf{M}_0 = \mathbf{M}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$ and $\boldsymbol{\Upsilon}_t^0 \equiv \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Let $\mathcal{G}_J(\boldsymbol{\theta}) \equiv (\mathcal{G}_J^d(\boldsymbol{\theta}_d)', \mathcal{G}_J^c(\boldsymbol{\theta})')'$ and $\mathcal{G}_{J,T}(\boldsymbol{\theta}) \equiv (\mathcal{G}_J(\boldsymbol{\theta})', \mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d)')'$. The first terms on the right hand side of (29) are the sample moments evaluated at $(\mathbf{s}, P, \boldsymbol{\eta}) = (\mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$ as in (16) and thus contains neither the sampling nor simulation errors, while the remaining terms are approximations to the differences between $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ and $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$. Each term in (29) obviously has zero expectation at the true parameter values under assumptions A1 because of the orthogonality conditions of the demand, supply and additional moments. This property will transmit to the estimator that minimizes the norm of (29).

Note that the three components in $\mathcal{G}_{J,T}(\boldsymbol{\theta})$ —those involving $\boldsymbol{\epsilon}^n$, $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$, and $\boldsymbol{\epsilon}^N$ —are *not* jointly independent because they all include the product characteristics \mathbf{X} as well as the unobserved product quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)$, both of which are random. However they are *uncorrelated if evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$* since $\boldsymbol{\epsilon}^n$, $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$, and $\boldsymbol{\epsilon}^N$ are generated by the distinct sampling processes conditional on $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$ as in assumption A1(a), (b) and (c). These facts together enable us to calculate the asymptotic variance-covariance matrix of $(J^{1/2} \mathcal{G}_J(\boldsymbol{\theta}^0), T^{1/2} \mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d))$ as the sum of the variance-covariance matrices, each derived from these separate components in $\mathcal{G}_{J,T}(\boldsymbol{\theta}^0)$. To make this even more clearly, let us first define

$$\begin{aligned} (\mathbf{a}_1^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \mathbf{a}_J^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) &\equiv \mathbf{Z}'_d \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \\ (\mathbf{a}_1^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \mathbf{a}_J^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) &\equiv -\mathbf{Z}'_c \mathbf{L}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \mathbf{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \\ (\mathbf{a}_1^a(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \mathbf{a}_J^a(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) &\equiv -\sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \end{aligned}$$

and further define $\mathbf{Y}_{JT_i}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ and $\mathbf{Y}_{JT_r}^*(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ using $\mathbf{a}_j^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, $\mathbf{a}_j^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, and $\mathbf{a}_j^a(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ as

$$\mathbf{Y}_{JT_i}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \equiv \begin{pmatrix} \frac{1}{nJ^{1/2}} \sum_{j=1}^J \mathbf{a}_j^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{ji} \\ \frac{1}{nJ^{1/2}} \sum_{j=1}^J \mathbf{a}_j^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{ji} \\ \frac{1}{nT^{1/2}} \sum_{j=1}^J \mathbf{a}_j^a(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{ji} \end{pmatrix}, \quad \mathbf{Y}_{JT_r}^*(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \equiv \begin{pmatrix} \frac{1}{RJ^{1/2}} \sum_{j=1}^J \mathbf{a}_j^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{jr}^*(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d) \\ \frac{1}{RJ^{1/2}} \sum_{j=1}^J \mathbf{a}_j^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{jr}^*(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d) \\ \frac{1}{RT^{1/2}} \sum_{j=1}^J \mathbf{a}_j^a(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \epsilon_{jr}^*(\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d) \end{pmatrix}.$$

Then, we can conveniently re-express (29) with the associated size indices $J^{1/2}$ and $T^{1/2}$ as the sum of the four terms involving the stochastic exogenous variables $(\mathbf{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$, the sampling error ϵ_j^n , the simulation error $\epsilon_j^R(\boldsymbol{\theta}_d)$, and the sampling error $\epsilon_{i'}^\#$ as

$$(30) \quad \begin{pmatrix} J^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) \\ T^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d) \end{pmatrix} = \begin{pmatrix} J^{\frac{1}{2}} \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \\ T^{\frac{1}{2}} \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) \end{pmatrix} + \sum_{i=1}^n \mathbf{Y}_{JT_i} + \sum_{r=1}^R \mathbf{Y}_{JT_r}^* + \sum_{i'=1}^N \begin{pmatrix} \mathbf{0} \\ T^{\frac{1}{2}} \epsilon_{i'}^\# / N \end{pmatrix}.$$

The four terms on the right hand side of (30) are separable only when evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. To establish (2-ii), therefore, we apply Theorem 3.3 in Pakes and Pollard (1989) in which each of the four terms on the right hand side of (30) are asymptotically normal when J and T simultaneously grow large.

Assumptions B1, B2 and B3 have essentially the same roles as the conditions (v), (ii) and (iii) respectively in Theorem 3.3 of Pakes and Pollard (1989). Assumption B1 is on the true parameter $\boldsymbol{\theta}^0$. Assumption B2 is the differentiability condition (with respect to $\boldsymbol{\theta}$) for the expectation of $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$. Given B2, assumption B3 implies that $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$ can be approximated by $\boldsymbol{\Gamma}_{J,T}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + \mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$ near $\boldsymbol{\theta}^0$, where $\boldsymbol{\Gamma}_{J,T}$ is the first-order derivative of $\mathbb{E}[\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)]$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$.

Assumption B1 $\boldsymbol{\theta}^0$ is an interior point of Θ .

Assumption B2 For all $\boldsymbol{\theta}$ in some $\delta > 0$ neighborhood of $\boldsymbol{\theta}^0$,

$$\mathbb{E}[\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] = \begin{pmatrix} \mathbb{E}_{\mathbf{x}_1, \xi}[\mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] \\ \mathbb{E}_{\mathbf{w}_1, \omega}[\mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)] \\ \mathbb{E}_{\mathbf{x}, \xi, \nu}[\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] \end{pmatrix} = \boldsymbol{\Gamma}_{J,T}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\|)$$

uniformly in J and T . The matrix $\boldsymbol{\Gamma}_{J,T} = (\boldsymbol{\Gamma}_J^{d'}, \boldsymbol{\Gamma}_J^{c'}, \boldsymbol{\Gamma}_{J,T}^{a'})' \rightarrow (\boldsymbol{\Gamma}^{d'}, \boldsymbol{\Gamma}^{c'}, \boldsymbol{\Gamma}^{a'})'$ as $J, T \rightarrow \infty$, where $\boldsymbol{\Gamma}_{J,T}$ has full column rank.

Assumption B3 For all sequences of positive numbers $\delta_{J,T}$ such that $\delta_{J,T} \rightarrow 0$,

$$(a) \quad \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| \leq \delta_{J,T}} \left\| \left\| J^{\frac{1}{2}} \left\{ \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) - \mathbb{E}_{\mathbf{x}_1, \xi}[\mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)] \right\} \right. \right. \\ \left. \left. - J^{\frac{1}{2}} \left\{ \mathbf{G}_J^d(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) - \mathbb{E}_{\mathbf{x}_1, \xi}[\mathbf{G}_J^d(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] \right\} \right\| \right\| = o_p(1)$$

$$(b) \quad \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq \delta_{J,T}} \left\| \left\| J^{\frac{1}{2}} \left\{ \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbb{E}_{\mathbf{w}_1, \omega}[\mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)] \right\} \right. \right. \\ \left. \left. - J^{\frac{1}{2}} \left\{ \mathbf{G}_J^c(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0) - \mathbb{E}_{\mathbf{w}_1, \omega}[\mathbf{G}_J^c(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)] \right\} \right\| \right\| = o_p(1).$$

$$(c) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \left\| T^{\frac{1}{2}} \{ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbb{E}[\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] \} \right. \\ \left. - T^{\frac{1}{2}} \{ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbb{E}[\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] \} \right\| = o_p(1).$$

Assumptions B4(a)–(d) determine the magnitude of the four components on the right hand side of (30), while assumptions B4(e)–(h) are the Lyapunov conditions to establish the central limit theorem.

Assumption B4 *The following finite positive definite matrices Φ_1 , Φ_2 , Φ_3 , and Φ_4^a exist.*

$$(a) \quad \lim_{J \rightarrow \infty} \mathbb{V} \begin{bmatrix} J^{\frac{1}{2}} \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \\ T^{\frac{1}{2}} \mathbf{G}_{J,T}^a(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) \end{bmatrix} = \Phi_1 = \begin{pmatrix} \Phi_1 & \Phi_1^{12} \\ \Phi_1^{12'} & \Phi_1^a \end{pmatrix},$$

$$(b) \quad \lim_{n, J \rightarrow \infty} n \mathbb{V}[\mathbf{Y}_{JTi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)] = \Phi_2 = \begin{pmatrix} \Phi_2 & \Phi_2^{12} \\ \Phi_2^{12'} & \Phi_2^a \end{pmatrix},$$

$$(c) \quad \lim_{R, J \rightarrow \infty} R \mathbb{V}_{\epsilon^*, x, \xi}[\mathbf{Y}_{JTr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)] = \Phi_3 = \begin{pmatrix} \Phi_3 & \Phi_3^{12} \\ \Phi_3^{12'} & \Phi_3^a \end{pmatrix},$$

$$(d) \quad \lim_{J, T, N \rightarrow \infty} N \mathbb{V}[T^{\frac{1}{2}} \boldsymbol{\epsilon}_{i'}^\# / N] = \Phi_4^a$$

The following Lyapunov conditions hold.

$$(e) \quad \sum_{(j,t)=(1,1)}^{(J,T)} \mathbb{E} \left[\left\| \begin{array}{c} z_j^d \xi_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) / J^{\frac{1}{2}} \\ z_j^c \omega_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) / J^{\frac{1}{2}} \\ (\boldsymbol{\eta}^0 - \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)) / T^{\frac{1}{2}} \end{array} \right\|^{2+\delta} \right] = o(1),$$

$$(f) \quad n \mathbb{E}[\|\mathbf{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\|^{2+\delta}] = o(1),$$

$$(g) \quad R \mathbb{E}[\|\mathbf{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\|^{2+\delta}] = o(1),$$

$$(h) \quad N \mathbb{E}[\|T^{\frac{1}{2}} \boldsymbol{\epsilon}_{i'}^\# / N\|^{2+\delta}] = o(1)$$

for some $\delta > 0$.

Assumptions B5(a)–(h) are conditions that enable us to control the differences between $(J^{1/2} \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^n, P^R), T^{1/2} \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N))$ and $(J^{1/2} \mathcal{G}_J(\boldsymbol{\theta}), T^{1/2} \mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d))$ within the shrinking neighborhood of $(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Specifically, in B5(a)–(d), (f) and (g), we assume those differences have stochastic equicontinuity-like characteristics at $(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = (\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. The assumptions B5(a) and (b) are respectively on the sampling and the simulation errors in the demand side moments, while B5(c)–(e) are those for the supply side moments and B5(f)–(h) are for the additional moments.

Assumption B5 *For all sequences of positive numbers $\delta_{J,T}$ such that $\delta_{J,T} \rightarrow 0$ as $J, T \rightarrow \infty$, we assume for demand side,*

$$(a) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, P) \in \{N_{\epsilon^0}(\theta_d^0; \delta_{J,T})\}^J \times N_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_d \{ \mathbf{H}^{-1}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P) - \mathbf{H}_0^{-1} \} \boldsymbol{\epsilon}^n \right\| = o_p(1),$$

$$(b) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^J \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_d \{ \mathbf{H}^{-1}(\xi_1, \theta_d, P) \epsilon^R(\theta_d) - \mathbf{H}_0^{-1} \epsilon^R(\theta_d^0) \} \right\| = o_p(1).$$

For supply side,

$$(c) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \right. \\ \left. \times \{ \mathbf{L}(\xi_1, \theta_d, P) \mathbf{M}(\xi_1, \theta_d, P) \mathbf{H}^{-1}(\xi_2, \theta_d, P) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \} \epsilon^n \right\| = o_p(1),$$

$$(d) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \right. \\ \left. \times \{ \mathbf{L}(\xi_1, \theta_d, P) \mathbf{M}(\xi_1, \theta_d, P) \mathbf{H}^{-1}(\xi_2, \theta_d, P) \epsilon^R(\theta_d) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \epsilon^R(\theta_d^0) \} \right\| = o_p(1),$$

$$(e) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{P \in \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \{ \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P) \right. \\ \left. - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)) \} \right\| = o_p(1).$$

And for the additional moments,

$$(f) \quad \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| T^{-\frac{1}{2}} \sum_{t=1}^T \left[\Upsilon_t(\xi_1, \theta_d, P) \mathbf{H}^{-1}(\xi_2, \theta_d, P) \epsilon^n \right. \right. \\ \left. \left. - \Upsilon_t(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0) \mathbf{H}^{-1}(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0) \epsilon^n \right] \right\| = o_p(1),$$

$$(g) \quad \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| T^{-\frac{1}{2}} \sum_{t=1}^T \left[\Upsilon_t(\xi_1, \theta_d, P) \mathbf{H}^{-1}(\xi_2, \theta_d, P) \epsilon^R(\theta_d) \right. \right. \\ \left. \left. - \Upsilon_t(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0) \mathbf{H}^{-1}(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0) \epsilon^R(\theta_d^0) \right] \right\| = o_p(1),$$

$$(h) \quad \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} T^{\frac{1}{2}} \sum_{j \in \mathcal{Q}_d} \epsilon_j^R(\theta_d) = o_p(1).$$

The quantity $\xi_1 = (\xi_{11}, \dots, \xi_{1J})$ and $\xi_2 = (\xi_{21}, \dots, \xi_{2J})$ are respectively a set of distinct J vectors, each vector corresponds to each row of $J \times J$ matrices $\mathbf{L}(\xi, \theta_d, P)$, $\mathbf{M}(\xi, \theta_d, P)$ and $\mathbf{H}^{-1}(\xi, \theta_d, P)$. The set $\{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_{J,T})\}^J$ indicates J sets of the $\delta_{J,T}$ neighborhood of $\xi(\theta_d^0, \mathbf{s}^0, P^0)$.

With these conditions we are ready to state asymptotic normality for the Petrin (2002) extension.

Theorem 2 (Asymptotic Normality of $\hat{\theta}$) Suppose that A1–A9 and B1–B5 hold for some increasing $n(J, T), R(J, T)$ such that $J/T \rightarrow 0$ as $J \rightarrow \infty, T \rightarrow \infty$ and $N \rightarrow \infty$. Then, the estimator $\hat{\theta}$ that minimizes $\|\mathbf{G}_{J,T}(\theta, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)\|$ is asymptotically normal at the rate of $J^{\frac{1}{2}}$:

$$J^{\frac{1}{2}}(\hat{\theta} - \theta^0) \overset{w}{\rightsquigarrow} N(\mathbf{0}, \mathbf{V})$$

with $\mathbf{V} = (\boldsymbol{\Gamma}'\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{a'}\boldsymbol{\Gamma}^a)^{-1} \boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{a'}\boldsymbol{\Gamma}^a)^{-1}$ where $\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \boldsymbol{\Phi}_3$.

Remark 1 When $J/T \rightarrow c > 0$ where c is constant, the variance-covariance matrix \mathbf{V} of Theorem 2 becomes

$$(31) \quad \mathbf{V} = (\boldsymbol{\Gamma}'\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{a'}\boldsymbol{\Gamma}^a)^{-1} (\boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma} + 2c^{1/2}\boldsymbol{\Gamma}'\boldsymbol{\Phi}^{12}\boldsymbol{\Gamma}^a + c\boldsymbol{\Gamma}^{a'}\boldsymbol{\Phi}^a\boldsymbol{\Gamma}^a) (\boldsymbol{\Gamma}'\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^{a'}\boldsymbol{\Gamma}^a)^{-1}$$

where $\Phi^{12} = \Phi_1^{12} + \Phi_2^{12} + \Phi_3^{12}$ is the off-diagonal component of the asymptotic variance covariance matrix of the $(J^{1/2}\mathcal{G}_J(\theta^0)', T^{1/2}\mathcal{G}_{J,T}^a(\theta_d^0)')$ and $\Phi^a = \Phi_1^a + \Phi_2^a + \Phi_3^a + \Phi_4^a$ is the asymptotic variance-covariance matrix of $T^{1/2}\mathcal{G}_{J,T}^a(\theta_d^0)$.

Remark 2 If the sampling error ϵ^n and the simulation error ϵ^R are negligibly small, the off-diagonal matrix Φ^{12} is also close to zero, and the asymptotic variance-covariance matrix V in (31) becomes

$$V \approx (\Gamma'\Gamma + \Gamma^{a'}\Gamma^a)^{-1}(\Gamma'\Phi\Gamma + c\Gamma^{a'}\Phi^a\Gamma^a)(\Gamma'\Gamma + \Gamma^{a'}\Gamma^a)^{-1}.$$

In this case, even when $J, T \rightarrow \infty$ but $J/T \rightarrow c > 0$, we can improve the efficiency of $\hat{\theta}$ by using the optimal weight matrix to the GMM objective function. With the weight matrix, we minimize $\|\mathbf{G}_{J,T}(\theta, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)\|^2 = \mathbf{G}_{J,T}(\theta, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)'\mathbf{W}_{J,T}\mathbf{G}_{J,T}(\theta, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ where $\mathbf{W}_{J,T} = \text{diag}(\mathbf{A}_J, \mathbf{A}_{J,T}^a)$ is diagonal and non-stochastic matrix. The asymptotic variance-covariance matrix of $\hat{\theta}$ corresponding to this objective function is thus

$$\tilde{V} \approx (\Gamma'\mathbf{A}_J\Gamma + \Gamma^{a'}\mathbf{A}_{J,T}^a\Gamma^a)^{-1}(\Gamma'\mathbf{A}_J\Phi\mathbf{A}_J\Gamma + c\Gamma^{a'}\mathbf{A}_{J,T}^a\Phi^a\mathbf{A}_{J,T}^a\Gamma^a)(\Gamma'\mathbf{A}_J\Gamma + \Gamma^{a'}\mathbf{A}_{J,T}^a\Gamma^a)^{-1}.$$

Obviously, $\mathbf{A}_J = \Phi^{-1}$ and $\mathbf{A}_{J,T}^a = (c\Phi^a)^{-1}$ are optimal and the \tilde{V} becomes

$$\tilde{V} = (\Gamma'\Phi^{-1}\Gamma + c^{-1}\Gamma^{a'}(\Phi^a)^{-1}\Gamma^a)^{-1}.$$

Relative to the asymptotic variance-covariance matrix $V^* = (\Gamma'\Phi\Gamma)^{-1}$ of the GMM estimator without the additional moments, $\tilde{V} < V^*$. In general, however, making the simulation error negligibly small may not be tenable given the computational burden, while ignoring the sampling error may be justified if sufficiently accurate market share data are available.

4 An Example of the Random Coefficient Logit Model of Demand

In estimating the demand model with the simple logit specification, BLP (2004) showed that, if the market shares of all the products stochastically go to zero at the rate of $1/J$, the assumptions in the consistency and the asymptotic normality are satisfied so long as n grows faster than J and J^2 respectively. Notice that the logit model has the closed-form solution for the equation (2) and thus do not incur the simulation error in the model, rendering the consideration of R unnecessary.

In what follows, we consider the random coefficient logit model of demand. As discussed in BLP (1995), this model has useful properties when product characteristics and consumers' taste are multi-dimensionally distributed and the nature of competition among products is complex. Unfortunately, the random coefficient logit model has no closed-form solution for (2) and for the inverse of $\mathbf{H}(\boldsymbol{\xi}, \theta_d, P)$ in assumption A2(a). Thus, our examination has to rely on its stochastic approximation.

Without loss of generality, we assume a random coefficient logit model with one random coefficient:

$$(32) \quad u_{ij} = \delta_j + \theta_x^u \nu_i^x x_j + \nu_{ij} \quad \text{with} \quad \delta_j = \theta_p p_j + \theta_x x_j + \xi_j$$

where ν_i^x represents consumer i 's random preference on the characteristic x_j relative to the price. The parameter θ_x^u indicates the magnitude of this preference, and when $\theta_x^u = 0$, the model reduces to the simple logit model. Provided that ν_{ij} 's are i.i.d. extreme value, the probability σ_{ij} that consumer i with preference ν_i^x chooses product j is given by

$$(33) \quad \sigma_{ij}(\boldsymbol{\xi}, \nu_i^x; \boldsymbol{\theta}_d) = \frac{\exp(\delta_j + \theta_x^u \nu_i^x x_j)}{1 + \sum_{k=1}^J \exp(\delta_k + \theta_x^u \nu_i^x x_k)}.$$

The market share of product j is obtained by integrating (33) in terms of ν_i^x over the population P^0 . We simulate it with a random sample of R individuals as

$$(34) \quad \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \equiv \frac{1}{R} \sum_{r=1}^R \sigma_{rj}(\boldsymbol{\xi}, \nu_r^x; \boldsymbol{\theta}_d) = \frac{1}{R} \sum_{r=1}^R \frac{\exp(\delta_j + \theta_x^u \nu_r^x x_j)}{1 + \sum_{k=1}^J \exp(\delta_k + \theta_x^u \nu_r^x x_k)}$$

If we assume that $\delta_j + \theta_x^u \nu_r^x x_j$ is stochastically bounded, the order of magnitudes of the individual's choice probability $\sigma_{rj}(\boldsymbol{\xi}, \nu_r^x; \boldsymbol{\theta}_d)$ and its average $\sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ are both of $1/J$. In the following, we put forward a condition on the magnitude of the individual choice probability. Although the condition makes individual's behavior restrictive, this treatment allows us to calculate the rates of n , R , N , and T relative to J , at which the random coefficient logit model follows our asymptotic theorems.

Condition S(a) For all consumer r with the demographics ν_r^x , and for all possible value of the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi})$, there exists positive finite constants \underline{c} and \bar{c} such that with probability one

$$(35) \quad \frac{\underline{c}}{J} \leq \inf_{\boldsymbol{\theta}_d \in \Theta_d} \sigma_{rj}(\boldsymbol{\xi}, \nu_r^x; \boldsymbol{\theta}_d) \leq \sup_{\boldsymbol{\theta}_d \in \Theta_d} \sigma_{rj}(\boldsymbol{\xi}, \nu_r^x; \boldsymbol{\theta}_d) \leq \frac{\bar{c}}{J}, \quad j = 0, 1, \dots, J.$$

(b) The constant \bar{c} further satisfies the relationship $\bar{c} J_m < J$ for each firm $m = 1, \dots, F$, where J_m is the number of products firm m produces in the markets.

With condition S(a), the individual choice probability $\sigma_{rj}(\boldsymbol{\xi}, \nu_r^x; \boldsymbol{\theta}_d)$ and its inverse are respectively $O_p(1/J)$ and $O_p(J)$. Obviously, this condition is sufficient for s_j^0 to be $O_p(1/J)$ for $j = 1, \dots, J$ because substituting $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ and integrating both sides of the inequality over the population P^0 immediately leads to $s_j = O_p(1/J)$. We assume our two sets of consumer draws, or $\nu_r^x, r = 1, \dots, R$ and $\nu_t^x, t = 1, \dots, T$, satisfy this condition. By condition S(b), we exclude the event that the aggregate market share for any of firms dominates in the market, i.e. $\sum_{j \in \mathcal{J}_m} s_j^0 \leq \sum_{j \in \mathcal{J}_m} \bar{c}/J = \bar{c} J_m / J < 1$ at any given J . This guarantees that the inverse of the aggregate market share for the other firms' products and the outside good, is finite and thus its order of magnitude is one, i.e., $1/(1 - \sum_{j \in \mathcal{J}_m} s_j^0) = O_p(1)$.

As stated above, the random coefficient logit model has no closed-form solution to the inverse of \boldsymbol{H} .

However, under condition S, we can approximate it by

$$(36) \quad \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) + \frac{1}{\sigma_0(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)} (1 + O_p(1/J)) \mathbf{i}\mathbf{i}',$$

where $\boldsymbol{\Sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \text{diag}(\sigma_1(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \sigma_J(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))$. In the appendix of BLP (2004, pp.651-652), an approximation essentially same as this was used to show that, even when we use the random coefficient logit model, the limiting behavior of the residual term on the sampling error in the demand side moment (26) is fundamentally similar to that for the logit model. As a result, the random coefficient logit model requires the same rate J^2 for n relative to J as the logit model to guarantee the GMM estimator to follow asymptotically normal. As for the number R of simulation draws, they presumed that symmetric arguments hold. Furthermore, we can show that the argument above apply to our supply side specification too.

Now we will examine a case where we have at our disposal additional moment conditions on demographically-categorized purchasing information. We suppose that we are now interested in estimating the parameter θ_x^u in (32) more accurately by using the information on consumers who choose specific sets of discriminating attributes in products. Denote the set of products having this attribute by \mathcal{Q} . Hereinafter, assume that we have a consistent estimate η^N , which was constructed from N independent consumer draws (not by researcher) from the population P^0 , separate from the n independent draws (again not by researcher) from P^0 for calculating the observed market share, with the expectation η^0 of ν_i^x conditional on the individual choosing a product in \mathcal{Q} . Given η^N , we will draw T individuals, independent of R simulation draws of individuals, from the population P^0 to construct an additional moment,

$$(37) \quad G_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) = \eta^N - \frac{1}{T} \sum_{t=1}^T \nu_t^x \psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$$

where $\psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \sum_{j \in \mathcal{Q}} \sigma_{tj}(\boldsymbol{\xi}, \nu_t^x, \boldsymbol{\theta}_d) / \sum_{j \in \mathcal{Q}} \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$.

The limiting behavior of the market shares, both observed and model-calculated, are assumed in A3. Assumptions A3(a) and (b) control the way in which \mathbf{s}^n and $\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ approach to the true market share \mathbf{s}^0 and $\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ respectively. To guarantee assumption A3 to hold, we require conditions on the growth rates of n and R relative to J as well as on the limiting behavior of the true market share \mathbf{s}^0 . BLP (2004) showed that $\Pr[\rho_{s^0}(\mathbf{s}^n, \mathbf{s}^0) > \delta] = J \mathbb{E}_{x,\xi}[\exp(-\delta^2 O_p(n/J))]$ for any $\delta > 0$ under assumption A1(a) and the condition $s_j^0 = O_p(1/J)$. This means that required rate of convergence of n is $J^{1+\epsilon}/n \rightarrow 0$ for any $\epsilon > 0$. Similarly, required rate of convergence of R for A3(b) is $J^{1+\epsilon}/R \rightarrow 0$.

To guarantee assumption A5, it is sufficient that the first order derivative matrix of $\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)$ in terms of $\boldsymbol{\theta} \in \Theta$ is of full column rank, since then for all $\delta > 0$, there exist C such that

$$\begin{aligned} \inf_{\boldsymbol{\theta} \notin \mathcal{N}_{\boldsymbol{\theta}^0}(\delta)} \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\| &= \inf_{\boldsymbol{\theta} \notin \mathcal{N}_{\boldsymbol{\theta}^0}(\delta)} \left\| \frac{\partial \mathbf{G}_J(\boldsymbol{\theta}^*, \mathbf{s}^0, P^0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \right\| \\ &\geq \inf_{\boldsymbol{\theta} \notin \mathcal{N}_{\boldsymbol{\theta}^0}(\delta)} C \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| = C\delta \end{aligned}$$

in probability tending to one as $J \rightarrow \infty$. In the following, we examine what it means to have $\partial \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'$ being of full-column rank. We should note that the demand side moment contains only the vector of demand parameters, $\boldsymbol{\theta}_d$, while that for cost side contains both of demand and cost side parameter vectors, $\boldsymbol{\theta}_d$ and $\boldsymbol{\theta}_c$. This means that the matrix $\partial \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'$ takes the following form

$$\frac{\partial \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} \partial \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_d & \mathbf{0} \\ \partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_d & \partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_c \end{pmatrix}.$$

This matrix is full-column rank if the components $\partial \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_d$ and $\partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_c$ are respectively of full-column rank, regardless of the value of $\partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_d$. Moreover, we know that $\partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_c = -J^{-1} \mathbf{Z}'_c \mathbf{W}$ by the definition of the cost side moment in section 2.3 and the assumed linear dependence of $\boldsymbol{\omega}$ on \mathbf{W} in (7). By properly choosing the cost side instruments \mathbf{Z}_c and cost shifter \mathbf{W} , we can construct $\partial \mathbf{G}_J^c(\boldsymbol{\theta}, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_c$ to be of full-column rank for all J . Therefore we only need to check $\partial \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)/\partial \boldsymbol{\theta}'_d$:

$$\frac{\partial \mathbf{G}_J^d(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)}{\partial \boldsymbol{\theta}'_d} = J^{-1} \mathbf{Z}'_d \frac{\partial \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)}{\partial \boldsymbol{\theta}'_d} = -J^{-1} \mathbf{Z}'_d \mathbf{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)}{\partial \boldsymbol{\theta}'_d}$$

where $\partial \boldsymbol{\sigma}(\cdot)/\partial \boldsymbol{\xi}' \cdot \partial \boldsymbol{\xi}/\partial \boldsymbol{\theta}'_d + \partial \boldsymbol{\sigma}(\cdot)/\partial \boldsymbol{\theta}'_d = \mathbf{0}$ from the implicit function theorem. Unfortunately, the full-column rankness of this matrix under the random coefficient logit specification cannot be checked analytically because of the existence of the inverse of \mathbf{H} . Thus we have to rely on numerical computations on a case-by-case basis. Assumption A6 requires similar argument. Assumption A7, on the other hand, can be verified using (36) and condition S(b) after tedious calculations.

For assumption A8, we assume the number of products in \mathcal{Q} increases as fast as the number of products in the market, which guarantees both of $\sum_{j \in \mathcal{Q}} \sigma_j$ and $1/\sum_{j \in \mathcal{Q}} \sigma_j$ to be $O_p(1)$ under condition S(a).

Since the quantity within the probability statement in assumption A9 is bounded from above as

$$(38) \quad \begin{aligned} & T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| \\ & \leq T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)\| \\ & \quad + T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| \end{aligned}$$

where $\boldsymbol{\Psi} = (\psi_1, \dots, \psi_T)'$ is a $T \times 1$ matrix, we separately evaluate the two terms on the right hand side of (38). The square of the first term is bounded by

$$\begin{aligned} & T^{-1} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)\|^2 \\ & = T^{-1} \left\| \frac{\partial \boldsymbol{\Psi}(\boldsymbol{\xi}^*, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} (\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)) \right\|^2 \\ & \leq \left(\frac{J}{T} \right) \left\| \frac{\partial \boldsymbol{\Psi}(\boldsymbol{\xi}^*, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \right\|^2 \cdot J^{-1} \|\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)\|^2 \end{aligned}$$

where ξ^* is between $\xi(\theta_d, \mathbf{s}^n, P^R)$ and $\xi(\theta_d, \mathbf{s}^0, P^0)$. Since we know that $J^{-1} \|\xi(\theta_d, \mathbf{s}^n, P^R) - \xi(\theta_d, \mathbf{s}^0, P^0)\|^2 = o_p(1)$ under assumptions A3 and A6, it remains to show that $\|\partial \Psi(\xi^*, \theta_d, P^R) / \partial \xi'\|^2 = O_p(T/J)$ to guarantee this whole term to be $o_p(1)$. We obtain the j th element of Υ_t as

$$(39) \quad \{\Upsilon_t(\xi, \theta_d, P)\}_j = \frac{\partial \psi_t(\xi, \theta_d, P)}{\partial \xi_j} \\ = \frac{\sigma_{tj}(1\{j \in \mathcal{Q}\} - \sum_{k \in \mathcal{Q}} \sigma_{tk})}{\sum_{k \in \mathcal{Q}} \sigma_k} - \frac{\sum_{k \in \mathcal{Q}} \sigma_{tk}}{\sum_{k \in \mathcal{Q}} \sigma_k} \cdot \frac{1\{j \in \mathcal{Q}\} \int \sigma_{rj} dP - \sum_{k \in \mathcal{Q}} \int \sigma_{rj} \sigma_{rk} dP}{\sum_{k \in \mathcal{Q}} \sigma_k}$$

where $\sigma_{rj} = \sigma_{rj}(\xi, \nu_r^x, \theta_d)$, $\sigma_{tj} = \sigma_{tj}(\xi, \nu_t^x, \theta_d)$ and $\sigma_j = \sigma_j(\xi, \theta_d, P)$. Under condition S(a), both of σ_{rj} and σ_j are $O_p(1/J)$, and $\sum_{j \in \mathcal{Q}} \sigma_j$ and $1/\sum_{j \in \mathcal{Q}} \sigma_j$ are both $O_p(1)$. Thus, we have $\partial \psi_t(\xi, \theta_d, P) / \partial \xi_j = O_p(1/J)$, and so $\|\partial \Psi(\xi^*, \theta_d, P) / \partial \xi'\|^2 = \sum_{j=1}^J \sum_{t=1}^T (\partial \psi_t(\xi^*, \theta_d, P) / \partial \xi_j)^2 = J \cdot T \cdot O_p(1/J)^2 = O_p(T/J)$.

The square of the second term on the right hand side of (38) is

$$T^{-1} \|\Psi(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R) - \Psi(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)\|^2 \\ = T^{-1} \sum_{t=1}^T \{\psi_t(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R) - \psi_t(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)\}^2 \\ = T^{-1} \sum_{t=1}^T \left\{ \frac{\sum_{j \in \mathcal{Q}} \sigma_{tj}(\xi(\theta_d, \mathbf{s}^0, P^0), \nu_t^x, \theta_d)}{\sum_{j \in \mathcal{Q}} \sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R)} - \frac{\sum_{j \in \mathcal{Q}} \sigma_{tj}(\xi(\theta_d, \mathbf{s}^0, P^0), \nu_t^x, \theta_d)}{\sum_{j \in \mathcal{Q}} \sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)} \right\}^2 \\ = \left\{ \frac{\sum_{j \in \mathcal{Q}} \{\sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R) - \sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)\}}{\sum_{j \in \mathcal{Q}} \sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^R) \cdot \sum_{j \in \mathcal{Q}} \sigma_j(\xi(\theta_d, \mathbf{s}^0, P^0), \theta_d, P^0)} \right\}^2 T^{-1} \sum_{t=1}^T \left\{ \sum_{j \in \mathcal{Q}} \sigma_{tj}(\xi(\theta_d, \mathbf{s}^0, P^0), \nu_t^x, \theta_d) \right\}^2 \\ = \left\{ \frac{\sum_{j \in \mathcal{Q}} O_p(1/R^{1/2} J^{1/2})}{\sum_{j \in \mathcal{Q}} O_p(1/J) \cdot \sum_{j \in \mathcal{Q}} O_p(1/J)} \right\}^2 T^{-1} \sum_{t=1}^T \left\{ \sum_{j \in \mathcal{Q}} O_p(1/J) \right\}^2 = O_p(J/R)$$

under assumption A1(b) and condition S(a). Therefore, R is required to grow faster than J .

We next move to assumptions in Theorem 2. We start with assumption B4 because B1 through B3 can be easily verified. In assumption B4(b), we need to keep the variance of $\sum_{i=1}^n \mathbf{Y}_{JT_i}^0$ in terms of the sampling error bounded. To accomplish this for the random coefficient logit model of demand, BLP (2004) showed that n and R are necessary to grow at the rate of J^2 . We focus on those on the additional moments. Let us denote the component of $\mathbf{Y}_{JT_i}^0 \equiv \mathbf{Y}_{JT_i}(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0)$ that corresponds to the additional moments as $Y_{JT_i}^{a0} = \sum_{j=1}^J a_j^{a0} \epsilon_{ji} / nT^{1/2}$, and abbreviate $\sigma_{tj}^0 = \sigma_{tj}(\xi(\theta_d^0, \mathbf{s}^0, P^0), \nu_t^x, \theta_d^0)$ and $\sigma_j^0 = \sigma_j(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0)$, then

$$(40) \quad a_j^{a0} \equiv a_j^a(\xi(\theta_d^0, \mathbf{s}^0, P^0), \theta_d^0, P^0) = \{-\sum_{t=1}^T \nu_t^x \Upsilon_t^0 \mathbf{H}_0^{-1}\}_j \\ = -\frac{\sum_{t=1}^T \nu_t^x \sigma_{tj}^0 (1\{j \in \mathcal{Q}\} - \sum_{l \in \mathcal{Q}} \sigma_{tl}^0)}{\sum_{l \in \mathcal{Q}} s_l^0} \cdot \frac{1}{s_j^0} - \frac{\sum_{t=1}^T \nu_t^x \sigma_{t0}^0 \sum_{l \in \mathcal{Q}} \sigma_{tl}^0}{\sum_{l \in \mathcal{Q}} s_l^0} \cdot \frac{1}{s_0^0} (1 + O_p(1/J)) \\ = \alpha(1 + \beta_j + O_p(1/J))$$

where

$$\alpha = -\frac{\sum_{t=1}^T \nu_t^x \sigma_{t0}^0 \sum_{l \in \mathcal{Q}} \sigma_{tl}^0}{\sum_{l \in \mathcal{Q}} s_l^0} \cdot \frac{1}{s_0^0}, \quad \beta_j = -\frac{\sum_{t=1}^T \nu_t^x \sigma_{tj}^0 (1\{j \in \mathcal{Q}\} - \sum_{l \in \mathcal{Q}} \sigma_{tl}^0)}{\sum_{t=1}^T \nu_t^x \sigma_{t0}^0 \sum_{l \in \mathcal{Q}} \sigma_{tl}^0} \cdot \frac{s_0^0}{s_j^0}$$

The α and β are respectively $O_p(T)$ and $O_p(1)$ under condition S(a). Using this $a_j^{a_0}$ we have further,

$$\begin{aligned}
& \sum_{j=1}^J (a_j^{a_0})^2 s_j^0 - (\sum_{j=1}^J a_j^{a_0} s_j^0)^2 \\
&= \sum_{j=1}^J \alpha^2 (1 + \beta_j + O_p(1/J))^2 s_j^0 - \{\sum_{j=1}^J \alpha (1 + \beta_j + O_p(1/J)) s_j^0\}^2 \\
&= \alpha^2 \left[s_0^0 (1 - s_0^0) (1 + O_p(1/J))^2 + 2(\sum_{j=1}^J \beta_j s_j^0) s_0^0 (1 + O_p(1/J)) + \sum_{j=1}^J \beta_j^2 s_j^0 - (\sum_{j=1}^J \beta_j s_j^0)^2 \right] \\
&\leq \alpha^2 \left[s_0^0 (1 - s_0^0) (1 + O_p(1/J))^2 + 2 \max_j |\beta_j| \cdot (\sum_{j=1}^J s_j^0) s_0^0 (1 + O_p(1/J)) + \max_j |\beta_j|^2 \cdot \sum_{j=1}^J s_j^0 \right] \\
&= \alpha^2 (1 - s_0^0) \left[s_0^0 (1 + O_p(1/J))^2 + 2 \max_j |\beta_j| s_0^0 (1 + O_p(1/J)) + \max_j |\beta_j|^2 \right] \\
&= O_p(T)^2 (1 - O_p(1/J)) \left[O_p(1/J) (1 + O_p(1/J))^2 + 2 O_p(1) O_p(1/J) (1 + O_p(1/J)) + O_p(1)^2 \right] \\
&= O_p(T^2).
\end{aligned}$$

where we substitute $\alpha = O_p(T)$ and $\beta_j = O_p(1)$. Therefore the variance of $\sum_{i=1}^n Y_{JT_i}^{a_0}$ is

$$\begin{aligned}
V_{\epsilon, \nu, x, \xi} [\sum_{i=1}^n Y_{JT_i}^{a_0}] &= \sum_{i=1}^n E_{\epsilon, \nu, x, \xi} [(1/n^2 T) (\sum_{j=1}^J a_j^{a_0} \epsilon_{ji})^2] \\
&= (1/nT) E_{\nu, x, \xi} \left[\sum_{j=1}^J (a_j^{a_0})^2 E_{\epsilon_{|x, \xi}} [\epsilon_{ji}^2 | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] \right. \\
&\quad \left. + \sum_{j \neq k} a_j^{a_0} a_k^{a_0} E_{\epsilon_{|x, \xi}} [\epsilon_{ji} \epsilon_{ki} | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] \right] \\
&= (1/nT) E_{\nu, x, \xi} \left[\sum_{j=1}^J (a_j^{a_0})^2 s_j^0 (1 - s_j^0) - \sum_{j \neq k} a_j^{a_0} a_k^{a_0} s_j^0 s_k^0 \right] \\
&= (1/nT) E_{\nu, x, \xi} \left[\sum_{j=1}^J (a_j^{a_0})^2 s_j^0 - (\sum_{j=1}^J a_j^{a_0} s_j^0)^2 \right] \\
&= (1/nT) E_{\nu, x, \xi} [O_p(T^2)] \\
&= E_{\nu, x, \xi} [O_p(T/n)].
\end{aligned}$$

To keep this variance bounded, n is needed to grow at the same rate as T . Similar calculation holds for assumption B4(c) and derives that R is required to grow at the same rate as T .

In assumption B4(d), we need to bound the variance of the residual term in the additional moment $T^{1/2} \mathcal{G}_{J,T}^a(\boldsymbol{\theta}_d^0)$ that corresponds to the sampling error in the additional information. The variance is

$$N V_{\epsilon^{\#}, x, \xi} [T^{1/2} N^{-1} \epsilon_i^{\#}] = E_{x, \xi} [V_{\epsilon^{\#} | x, \xi} [T^{1/2} (\eta^N - \eta^0) | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)]] = E_{x, \xi} [O_p(T/N)].$$

Thus we require the sample size T of consumer draws in constructing the additional moment in (37) to grow slower than the sample size N used for constructing the additional information η^N for B4(d) to hold.

Assumption B4(f) gives the Lyapunov condition the residual term $\sum_{i=1}^n Y_{JT_i}^{a_0}$ in the additional moment must follow. Since $a_j^{a_0}$ in (40) is $O_p(T)$ under condition S(a), we obtain

$$\begin{aligned}
n E_{\epsilon, \nu, x, \xi} [|Y_{JT_i}^{a_0}|^{2+\delta}] &= \frac{1}{n^{1+\delta} T^{(2+\delta)/2}} E_{\epsilon, \nu, x, \xi} [| \sum_{j=1}^J a_j^{a_0} \epsilon_{ji} |^{2+\delta}] \\
&\leq \frac{1}{n^{1+\delta} T^{(2+\delta)/2}} E_{\nu, x, \xi} [2^{2+\delta} \max_j |a_j^{a_0}|^{2+\delta}] \\
&= E_{\nu, x, \xi} [O_p(n^{-(1+\delta)} T^{(2+\delta)/2})].
\end{aligned}$$

Substituting $n = O(T^k)$ and solving $(2 + \delta)/2 - k(1 + \delta) < 0$ gives $k > 1$ for any $\delta > 0$, which means that T needs to grow slower than n . By similar argument for assumption B4(g) and B4(h), R is required to grow faster than T , while T needs to grow slower than N .

For assumption B5, we focus on those on the additional moments, B5(f) to B5(h). To have assumption B5(f), it is sufficient to show that both of the norm of $T^{1/2} \sum_{t=1}^T \Upsilon_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n$ and $T^{1/2} \sum_{t=1}^T \Upsilon_{t0} \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n$ are respectively $o_p(1)$. We abbreviate $\sigma_j^R = \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ and $\sigma_j^T = \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^T)$ and approximate the j th element of $T^{-1} \sum_{t=1}^T \Upsilon_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ by using \mathbf{H}^{-1} in (36) and $\partial \psi_t / \partial \xi_j$ in (39) as follows.

$$\begin{aligned}
(41) \quad & \left\{ T^{-1} \sum_{t=1}^T \Upsilon_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \right\}_j \\
&= T^{-1} \sum_{t=1}^T \sum_{l=1}^J \frac{\partial \psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)}{\partial \xi_l} H_{lj}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \\
&= \left[\frac{T^{-1} \sum_{t=1}^T (\sum_{l \in \mathcal{Q}} \sigma_{tl}) \sigma_{t0}}{\sum_{l \in \mathcal{Q}} \sigma_l^R} - \frac{(\sum_{l \in \mathcal{Q}} \sigma_l^T) R^{-1} \sum_{r=1}^R (\sum_{l \in \mathcal{Q}} \sigma_{rl}) \sigma_{r0}}{(\sum_{l \in \mathcal{Q}} \sigma_l^R)^2} \right] \cdot \frac{1}{\sigma_0^R} (1 + O_p(1/J)) \\
&+ \left[\frac{\sigma_j^T \cdot 1\{j \in \mathcal{Q}\} - T^{-1} \sum_{t=1}^T (\sum_{l \in \mathcal{Q}} \sigma_{tl}) \sigma_{tj}}{\sum_{l \in \mathcal{Q}} \sigma_l^R} \right. \\
&\quad \left. - \frac{(\sum_{l \in \mathcal{Q}} \sigma_l^T) \{\sigma_j^R \cdot 1\{j \in \mathcal{Q}\} - R^{-1} \sum_{r=1}^R (\sum_{l \in \mathcal{Q}} \sigma_{rl}) \sigma_{rj}\}}{(\sum_{l \in \mathcal{Q}} \sigma_l^R)^2} \right] \cdot \frac{1}{\sigma_j^R}.
\end{aligned}$$

Therefore, we use this equality to obtain

$$\begin{aligned}
& \left\| T^{-1} \sum_{t=1}^T \Upsilon_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right\| \\
&= \left| \sum_{j=1}^J \left\{ T^{-1} \sum_{t=1}^T \Upsilon_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \right\}_j \boldsymbol{\epsilon}_j^n \right| \\
&= \left| \left\{ \frac{T^{-1} \sum_{t=1}^T (\sum_{l \in \mathcal{Q}} \sigma_{tl}) \sigma_{t0}}{\sum_{l \in \mathcal{Q}} \sigma_l^R} - \frac{(\sum_{l \in \mathcal{Q}} \sigma_l^T) R^{-1} \sum_{r=1}^R (\sum_{l \in \mathcal{Q}} \sigma_{rl}) \sigma_{r0}}{(\sum_{l \in \mathcal{Q}} \sigma_l^R)^2} \right\} \frac{\sum_{j=1}^J (s_j^n - s_j^0)}{\sigma_0^R} (1 + O_p(1/J)) \right. \\
&\quad \left. + \sum_{j=1}^J \left\{ \frac{\sigma_j^T \cdot 1\{j \in \mathcal{Q}\} - T^{-1} \sum_{t=1}^T (\sum_{l \in \mathcal{Q}} \sigma_{tl}) \sigma_{tj}}{\sum_{l \in \mathcal{Q}} \sigma_l^R} \right. \right. \\
&\quad \left. \left. - \frac{(\sum_{l \in \mathcal{Q}} \sigma_l^T) \{\sigma_j^R \cdot 1\{j \in \mathcal{Q}\} - R^{-1} \sum_{r=1}^R (\sum_{l \in \mathcal{Q}} \sigma_{rl}) \sigma_{rj}\}}{(\sum_{l \in \mathcal{Q}} \sigma_l^R)^2} \right\} \cdot \frac{s_j^n - s_j^0}{\sigma_j^R} \right| \\
&= \left| \left\{ \frac{T^{-1} \sum_{t=1}^T O_p(1) \cdot O_p(1/J)}{O_p(1)} - \frac{O_p(1) R^{-1} \sum_{r=1}^R O_p(1) \cdot O_p(1/J)}{O_p(1)^2} \right\} \frac{\sum_{j=1}^J O_p(1/\sqrt{nJ})}{O_p(1/J)} (1 + O_p(1/J)) \right. \\
&\quad \left. + \sum_{j=1}^J \left\{ \frac{O_p(1/J) O_p(1) - T^{-1} \sum_{t=1}^T O_p(1) \cdot O_p(1/J)}{O_p(1)} \right. \right. \\
&\quad \left. \left. - \frac{O_p(1) \{O_p(1/J) O_p(1) - R^{-1} \sum_{r=1}^R O_p(1) \cdot O_p(1/J)\}}{O_p(1)^2} \right\} \cdot \frac{O_p(1/\sqrt{nJ})}{O_p(1/J)} \right| \\
&= O_p(\sqrt{J/n}).
\end{aligned}$$

Similarly, we obtain $\|T^{-1} \sum_{t=1}^T \Upsilon_{t0} \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n\| = O_p(\sqrt{J/n})$ using (36) and (39) evaluated at $(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) =$

$(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Hence,

$$\begin{aligned}
& \left\| T^{-1/2} \sum_{t=1}^T \{ \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \} \boldsymbol{\epsilon}^n \right\| \\
& \leq T^{1/2} \left\| T^{-1} \sum_{t=1}^T \{ \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right\| + T^{1/2} \left\| T^{-1} \sum_{t=1}^T \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \} \boldsymbol{\epsilon}^n \right\| \\
& = T^{1/2} O_p \left(\sqrt{J/n} \right) + T^{1/2} O_p \left(\sqrt{J/n} \right) \\
& = O_p \left(\sqrt{TJ/n} \right).
\end{aligned}$$

Therefore, random drawing of T individuals has to be done so that TJ grows slower than n . As for assumption B5(g), through a quite similar calculation as the calculation for assumption B5(f), we can show that the number R of simulation draws needs to grow faster than TJ .

We can easily see that assumption B5(h) requires R to grow faster than TJ as follows.

$$\sqrt{T} \sum_{j \in \mathcal{Q}} \epsilon_j^R(\boldsymbol{\theta}_d) = \sqrt{T} \sum_{j \in \mathcal{Q}} (\sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - s_j^0) = \sqrt{T} \sum_{j \in \mathcal{Q}} O_p \left(1/\sqrt{JR} \right) = O_p \left(\sqrt{TJ/R} \right).$$

In summary, for the random coefficient logit model, the estimator with the additional moment has consistency in Theorem 1 so long as n and R grow faster than J . The asymptotic normality in Theorem 2, on the other hand, requires that n and R to grow faster than J^2 and TJ , and N to grow faster than T . If we assume T grows at the rate of $J^{1+\epsilon_1}$ for $\epsilon_1 > 0$, a slightly faster than J , Theorem 2 requires n and R to grow faster than $J^{2+\epsilon_1}$ and N to grow faster than $J^{1+\epsilon_1}$. Table 1 lists the required growth rates of n , R , and N in the original BLP (1995) framework relative to the Petrin (2002) extension with the additional moments in terms of J . The table shows that Petrin (2002) extension requires slightly faster growth rates of n and R for the asymptotic normality to hold than those of the original BLP (1995) framework.

5 Monte Carlo Experiments

In this section, we evaluate Theorems 1 and 2 in section 3 through a series of Monte Carlo simulations for a version of random coefficient logit model of demand in the presence of oligopolistic suppliers when additional demographically categorized purchasing pattern data are available. We start with the system of

Table 1: Growth rate of n , R , and N relative to J necessary when T grows at $J^{1+\epsilon_1}$. Let $\epsilon, \epsilon_1, \epsilon_2 > 0$.

	Logit Model		Random Coefficient Logit Model			
	BLP (1995)		BLP (1995)		Petrin (2002)	
	Consistency	Normality	Consistency	Normality	Consistency	Normality
n	$J^{1+\epsilon}$	$J^{2+\epsilon}$	$J^{1+\epsilon}$	$J^{2+\epsilon}$	$J^{1+\epsilon}$	$J^{2+\epsilon_1+\epsilon_2}$
R	–	–	$J^{1+\epsilon}$	$J^{2+\epsilon}$	$J^{1+\epsilon}$	$J^{2+\epsilon_1+\epsilon_2}$
T	–	–	–	–	any rate	$J^{1+\epsilon_1}$
N	–	–	–	–	any rate	$J^{1+\epsilon_1+\epsilon_2}$

demand and supply of BLP (1995) in tables 2 and 3. In table 2 we consider cases where only the simulation errors are involved, and in table 3 we additionally assume that the sampling errors are present in the market share. We then proceed to the Petrin (2002) extension to the BLP (1995) framework with the additional moment condition in tables 4, 5, and 6. In table 4 we consider cases where no sampling errors exist in the additional information itself, and in table 5 we move to cases where the additional information itself contains another set of sampling errors. Table 6 serves dual purposes, in that it numerically verifies the asymptotic variances presented in Theorem 2 in the presence of all five errors. It also shows the potential benefit of the Petrin (2002) extension relative to the framework of BLP (1995) in the most realistic case. Since the simulation errors are under the control of the econometrician but reducing the simulation errors greatly increases computational burden, the econometrician is inclined to accept some degree of the simulation errors. Therefore we only consider cases where the simulation errors are present throughout these Monte Carlo repetitions. On the other hand, presence of the sampling errors in the observed share does not pose computational burden, and we consider them only in tables 3 and 6.

Throughout in this section, utility of consumer i for product j is

$$(42) \quad u_{ij} = -\alpha p_j + \beta x_j \nu_i^x + \xi_j + \nu_{ij}$$

where the unobserved quality ξ_j and the exogenous product characteristics x_j are respectively random draws from $N(0, 1)$ and $N(1, 1)$. These and other random draws employed in this section are all independent. The price of product p_j is, on the other hand, endogenously determined in the market. The ν_i^x is the consumer's random taste for x_j and distributed $N(0, 1)$. The ν_{ij} 's are i.i.d. extreme value draws. We set the demand side parameters as $\alpha = 1.0$ and $\beta = 1.0$. The market share σ_j is calculated by

$$(43) \quad \sigma_j = \int \frac{\exp(-\alpha p_j + \beta x_j \nu_i^x + \xi_j)}{1 + \sum_{l=1}^J \exp(-\alpha p_l + \beta x_l \nu_i^x + \xi_l)} P(d\nu_i^x).$$

The true market share s_j^0 is obtained by evaluating (43) with the underlying distribution P^0 of ν_i^x . We draw 10,000 consumers from $N(0, 1)$. They constitute the population.

For the supply side, we assume there exist five oligopolistic suppliers in the market, each producing the same number $J/5$ of products. These suppliers are assumed to have the same cost function

$$(44) \quad c_j = x_j \gamma + \omega_j$$

where the unobserved cost shifter ω_j is a random draw from $N(0, 1)$. For the cost side parameter, we set $\gamma = 1.5$. The true market share s_j^0 and the price p_j are determined at the equilibrium, and thus the values of p_j are obtained by solving $\mathbf{f}(\mathbf{p}) = \mathbf{c} - \mathbf{p} - \mathbf{\Delta}^{-1} \boldsymbol{\sigma} = \mathbf{0}$, that is, J dimensional nonlinear simultaneous equations, which is solved by an iterative Newton-Raphson algorithm.

We first estimate a version of the system of demand and supply of the BLP (1995) framework given in (43) and (44). We construct the three instruments from x_j — x_j itself, the company average of x_j , and the

average of x_j over the other companies. Table 2 shows the results for the mean of the estimated parameter values α , β , γ and the associated simulated standard errors for 100 Monte Carlo repetitions when the observed market shares have *no sampling errors*, i.e, the market shares are calculated from the population of 10,000 consumers. Each column corresponds to the number $J = 10, 25, 50, 100$ of products, while each row corresponds to the number $R = 10, 50, 100, 10J, J^2$ of consumer draws. In parentheses are the simulated standard errors—the standard errors of the estimated parameters across the repetitions.

In the table, we observe the simulated standard errors of parameters decrease as J increases. For J fixed, increasing R also contributes the reduction of the standard errors. Throughout, the standard error for β is much larger than those for α and for γ for the same pair of (R, J) . The β is harder to estimate because the consumer's taste for the product characteristics x_j is randomly altered by the ν_i^x and as such the information regarding the corresponding coefficient β is much harder to extract from the orthogonality condition between the unobserved quality ξ_j and the product characteristics x_j . In particular, when the number R of simulation draws is small at 10, the estimated β is found upwardly biased.

Table 3 shows the results when the observed market share s_j^n additionally contains the sampling error while the number of the simulation draws of consumers is set at $R = 100$. We construct the observed market share s_j^n from a multinomial sample of size n with the category probabilities (s_0^0, \dots, s_j^0) . When n is not large enough, some products are not purchased. Then we remove these products in estimating parameters. We observe that, the larger the n is, the smaller the simulated standard error is for any fixed J .

We next estimate the system of demand and supply given in (43) and (44) by the Petrin (2002) extension. We suppose that the information is available on (a) the expected value of ν_i^x over consumers who choose products priced higher than the average price; and (b) the expected value of ν_i^x over consumers who choose products with x_j greater than the average of x_j . So the additional moments are

$$(45) \quad \eta_1^0 = E[\nu_i^x | C_i \in \mathcal{Q}\{p_j \geq \bar{p}\}, x, \xi], \quad \eta_2^0 = E[\nu_i^x | C_i \in \mathcal{Q}\{x_j \geq \bar{x}\}, x, \xi]$$

where $\mathcal{Q}\{p_j \geq \bar{p}\}$ and $\mathcal{Q}\{x_j \geq \bar{x}\}$ represent respectively the set of products priced higher than the average \bar{p} , and the set of products whose characteristic x is larger than the average \bar{x} .

Table 4 is the results for cases where we know the expected values in (45) exactly and no sampling errors exist in the additional information. We draw T consumers from the population separately from the n and R consumers and then calculate the conditional average of ν_i^x by using their purchasing probabilities to calculate the additional moments. Here, we use the true market share s_j^0 as the observed market share ($n = 10,000$) and fix $R = 100$. This way, the effect of the additional moments on the accuracy of the estimates is more transparent.

The result indicates that information in the additional moment reduce the standard error of the random coefficient β considerably when the number T of consumer draws is large enough. For instance, when $J = 50, T = 1,000$, the standard error of β with the additional moments decreases to 0.137 in table 4

Table 2: Monte Carlo Results for the BLP (1995) Framework, 100 Repetitions, with No Sampling Error ($n = 10000$).

# of Consumer Draws (R)	$\alpha(1,0)$			$\beta(1,0)$			$\gamma(1,5)$					
	# of products (J) 10	# of products (J) 25	# of products (J) 50	# of Consumer Draws (R) 10	# of Consumer Draws (R) 25	# of Consumer Draws (R) 50	# of Consumer Draws (R) 10	# of Consumer Draws (R) 25	# of Consumer Draws (R) 50	# of products (J) 100		
10	0.974 (0.266)	0.953 (0.174)	0.952 (0.138)	0.934 (0.134)	1.303 (1.207)	1.385 (1.172)	1.223 (0.909)	1.177 (0.760)	1.558 (0.388)	1.543 (0.265)	1.546 (0.191)	1.518 (0.176)
50	0.974 (0.166)	0.990 (0.110)	0.989 (0.079)	0.971 (0.060)	0.957 (0.702)	0.983 (0.539)	0.958 (0.406)	0.936 (0.306)	1.595 (0.316)	1.609 (0.164)	1.602 (0.121)	1.574 (0.089)
100	0.982 (0.156)	0.997 (0.123)	0.989 (0.058)	0.979 (0.045)	0.909 (0.749)	0.981 (0.692)	0.912 (0.363)	0.935 (0.274)	1.583 (0.246)	1.613 (0.164)	1.605 (0.101)	1.582 (0.071)
10J	0.982 (0.156)	0.993 (0.099)	0.994 (0.056)	0.982 (0.036)	0.909 (0.749)	0.919 (0.543)	0.887 (0.347)	0.900 (0.238)	1.583 (0.246)	1.614 (0.158)	1.610 (0.097)	1.586 (0.073)
J^2	0.982 (0.156)	0.988 (0.093)	0.992 (0.055)	0.982 (0.035)	0.909 (0.749)	0.930 (0.605)	0.886 (0.325)	0.896 (0.240)	1.583 (0.246)	1.610 (0.156)	1.608 (0.098)	1.587 (0.073)

Standard error across repetitions in the parenthesis.

Table 3: Monte Carlo Results for the BLP (1995) Framework, 100 Repetitions, $R = 100$.

# of Consumer Draws (n)	$\alpha(1,0)$			$\beta(1,0)$			$\gamma(1,5)$					
	# of products (J)			# of Consumer Draws (n)			# of products (J)			# of Consumer Draws (n)		
	10	25	50	100	10	25	50	100	10	25	50	100
500	0.978 (0.180)	0.978 (0.235)	0.891 (0.107)	0.857 (0.082)	1.004 (0.824)	1.206 (1.348)	1.029 (0.476)	1.209 (0.457)	1.495 (0.274)	1.471 (0.189)	1.362 (0.178)	1.276 (0.134)
1000	0.987 (0.160)	0.988 (0.186)	0.935 (0.088)	0.918 (0.072)	0.972 (0.829)	1.108 (1.066)	1.000 (0.505)	1.115 (0.398)	1.528 (0.241)	1.529 (0.174)	1.458 (0.134)	1.396 (0.105)
2000	0.980 (0.164)	0.991 (0.134)	0.961 (0.078)	0.959 (0.058)	0.938 (0.787)	1.005 (0.698)	0.977 (0.454)	1.055 (0.328)	1.536 (0.241)	1.554 (0.161)	1.520 (0.110)	1.484 (0.084)
10J	0.917 (0.194)	0.925 (0.155)	0.891 (0.107)	0.918 (0.072)	1.054 (0.913)	1.290 (1.483)	1.029 (0.476)	1.115 (0.398)	1.329 (0.365)	1.377 (0.228)	1.362 (0.178)	1.396 (0.105)
J^2	0.917 (0.194)	0.974 (0.134)	0.963 (0.086)	0.984 (0.046)	1.054 (0.913)	1.127 (1.206)	0.978 (0.557)	0.945 (0.267)	1.329 (0.365)	1.493 (0.186)	1.520 (0.124)	1.570 (0.067)
∞	0.982 (0.156)	0.997 (0.123)	0.989 (0.068)	0.979 (0.045)	0.909 (0.749)	0.981 (0.692)	0.912 (0.363)	0.935 (0.274)	1.583 (0.246)	1.613 (0.164)	1.605 (0.101)	1.582 (0.071)

Standard error across repetitions stands in the parentheses.

from 0.363, which is the value without the additional moments in table 2 ($R = 100$ row, $J = 50$ column). Furthermore, when $J = 50$, if we change the size T of the sample to evaluate the additional moments from $T = 1000$ to $T = 2500(J^2)$, the standard error of β declines from 0.137 to 0.125. Similarly, when $J = 100$, increasing $T = 1000$ to $T = 10000(J^2)$ reduces the standard error of β from 0.134 to 0.087. On the other hand, when the number T is small, the standard error of β can increase rather than decrease. For example, the standard error of β at $T = 50$ and $J = 50$ increases to 0.392 in table 4 from 0.363 in table 2. These results show that the number T of consumer draws to evaluate the additional moments plays an important role in increasing the accuracy of β .

It should be noted that the additional moments have very limited influences on the standard errors of α and no influences on the standard errors of γ for any value for T . This is because the additional information is on the consumer's taste ν_i^x and contains little information on α and no information on γ .⁶

We then consider cases where the additional information itself contains another set of the sampling errors. Drawing N consumers from the population independent of the aforementioned T , n , and R consumers, we use the following estimators of η^N instead of η^0 ,

$$(46) \quad \eta_1^N = \sum_{i'=1}^N \frac{\nu_{i'}^x \cdot 1\{C_{i'} \in \mathcal{Q}\{p_j \geq \bar{p}\}\}}{N_p}, \quad \eta_2^N = \sum_{i'=1}^N \frac{\nu_{i'}^x \cdot 1\{C_{i'} \in \mathcal{Q}\{x_j \geq \bar{x}\}\}}{N_x}.$$

where $N_p = \sum_{i'=1}^N 1\{C_{i'} \in \mathcal{Q}\{p_j \geq \bar{p}\}\}$ and $N_x = \sum_{i'=1}^N 1\{C_{i'} \in \mathcal{Q}\{x_j \geq \bar{x}\}\}$ are respectively the number of consumers who choose products priced higher than the average price and the number of consumers who choose products whose characteristic x greater than the average product characteristic \bar{x} . These estimators are unbiased for η^0 conditional on x and ξ .

Table 5 shows the result for this case. The standard errors of β decreases as the size N of consumer draws to construct the additional information in (46). For instance, at $J = 50$, when we increase from $N = 1000$ to $N = \infty$ (that is, the population of 10000), the standard error of β decreases from 0.171 to 0.137. Similarly, when $J = 100$, increasing from $N = 1000$ to $N = \infty$ (10000) reduces the standard error of β from 0.169 to 0.134. These results show that the number N of consumer draws used for constructing the additional information also plays an important role in improving the accuracy of β . Again, the additional moments have very limited influences on the standard errors of α and no influences on the standard errors of γ in any value for N for the reasons aforementioned.

In concluding this section, we evaluate the asymptotic normality in Theorem 2 when we allow for all of the errors in the estimation. For $J = 25, R = 2000, n = 2000, N = 2000, T = 500$ fixed, we implement 1000 Monte Carlo repetitions of the estimation of α , β , and γ , and then we calculate their averages and simulated standard errors. We also numerically calculated the asymptotic variances of the GMM estimates of α , β , and γ in Theorem 2 in the following manner. For each simulated data, we calculate the moment conditions and their derivatives in terms of parameters (the parameters are fixed at true values). Then,

⁶The first order derivatives of the additional moments in terms of α are almost zero, while that for γ is zero.

Table 4: Monte Carlo Results for the Petrin (2002) Extension, 100 Repetitions, $n = 10000$, $N = 10000$, $R = 100$.

# of Consumer	$\alpha(1,0)$			$\beta(1,0)$			$\gamma(1,5)$			
	# of products (J)	# of Consumer	T	# of products (J)	# of Consumer	T	# of products (J)	# of Consumer	T	
10	10 25 50 100	10	10	10 25 50 100	10	10	10 25 50 100	10	10	
	0.985 (0.139) 0.978 (0.100) 0.989 (0.071) 0.993 (0.061)	0.930 (0.568) 1.039 (0.683) 0.954 (0.469) 0.999 (0.530)	10	0.930 (0.568) 1.039 (0.683) 0.954 (0.469) 0.999 (0.530)	10	1.630 (0.229) 1.594 (0.168) 1.620 (0.110) 1.607 (0.085)	10	1.630 (0.229) 1.594 (0.168) 1.620 (0.110) 1.607 (0.085)	10	1.630 (0.229) 1.594 (0.168) 1.620 (0.110) 1.607 (0.085)
50	1.007 (0.126) 0.985 (0.089) 0.989 (0.067) 0.993 (0.055)	0.978 (0.411) 0.999 (0.356) 0.978 (0.392) 0.958 (0.316)	50	0.978 (0.411) 0.999 (0.356) 0.978 (0.392) 0.958 (0.316)	50	1.648 (0.236) 1.605 (0.163) 1.621 (0.115) 1.608 (0.080)	50	1.648 (0.236) 1.605 (0.163) 1.621 (0.115) 1.608 (0.080)	50	1.648 (0.236) 1.605 (0.163) 1.621 (0.115) 1.608 (0.080)
100	1.019 (0.135) 0.988 (0.084) 0.997 (0.066) 0.996 (0.057)	0.974 (0.336) 0.991 (0.284) 0.953 (0.317) 0.933 (0.249)	100	0.974 (0.336) 0.991 (0.284) 0.953 (0.317) 0.933 (0.249)	100	1.677 (0.250) 1.610 (0.159) 1.629 (0.107) 1.610 (0.083)	100	1.677 (0.250) 1.610 (0.159) 1.629 (0.107) 1.610 (0.083)	100	1.677 (0.250) 1.610 (0.159) 1.629 (0.107) 1.610 (0.083)
500	1.017 (0.122) 0.988 (0.075) 0.996 (0.062) 1.008 (0.057)	0.991 (0.271) 0.961 (0.227) 0.981 (0.169) 0.958 (0.148)	500	0.991 (0.271) 0.961 (0.227) 0.981 (0.169) 0.958 (0.148)	500	1.676 (0.241) 1.617 (0.134) 1.620 (0.089) 1.615 (0.083)	500	1.676 (0.241) 1.617 (0.134) 1.620 (0.089) 1.615 (0.083)	500	1.676 (0.241) 1.617 (0.134) 1.620 (0.089) 1.615 (0.083)
1000	1.025 (0.133) 0.982 (0.072) 0.992 (0.062) 1.002 (0.054)	0.989 (0.234) 0.929 (0.134) 0.956 (0.137) 0.967 (0.134)	1000	0.989 (0.234) 0.929 (0.134) 0.956 (0.137) 0.967 (0.134)	1000	1.682 (0.238) 1.614 (0.139) 1.617 (0.097) 1.610 (0.087)	1000	1.682 (0.238) 1.614 (0.139) 1.617 (0.097) 1.610 (0.087)	1000	1.682 (0.238) 1.614 (0.139) 1.617 (0.097) 1.610 (0.087)
10J	1.019 (0.135) 0.983 (0.078) 0.996 (0.062) 1.002 (0.054)	0.974 (0.336) 0.967 (0.233) 0.981 (0.169) 0.967 (0.134)	10J	0.974 (0.336) 0.967 (0.233) 0.981 (0.169) 0.967 (0.134)	10J	1.677 (0.250) 1.612 (0.143) 1.620 (0.089) 1.610 (0.087)	10J	1.677 (0.250) 1.612 (0.143) 1.620 (0.089) 1.610 (0.087)	10J	1.677 (0.250) 1.612 (0.143) 1.620 (0.089) 1.610 (0.087)
J ²	1.019 (0.135) 0.992 (0.079) 0.996 (0.056) 0.999 (0.062)	0.974 (0.336) 0.959 (0.184) 0.954 (0.125) 0.955 (0.087)	J ²	0.974 (0.336) 0.959 (0.184) 0.954 (0.125) 0.955 (0.087)	J ²	1.677 (0.250) 1.620 (0.142) 1.621 (0.092) 1.606 (0.087)	J ²	1.677 (0.250) 1.620 (0.142) 1.621 (0.092) 1.606 (0.087)	J ²	1.677 (0.250) 1.620 (0.142) 1.621 (0.092) 1.606 (0.087)

Standard error across repetitions in the parentheses.

Table 5: Monte Carlo Results for the Petrin (2002) Extension, 100 Repetitions, $n = 10000$, $R = 100$, $T = 1000$.

# of Consumer N	$\alpha(1.0)$			$\beta(1.0)$			$\gamma(1.5)$					
	10	25	50	10	25	50	10	25	50	100		
500	1.023 (0.138)	0.995 (0.079)	0.991 (0.061)	1.004 (0.054)	0.980 (0.274)	0.970 (0.241)	0.950 (0.195)	0.998 (0.216)	1.679 (0.241)	1.624 (0.138)	1.617 (0.096)	1.611 (0.080)
1000	1.011 (0.125)	0.991 (0.075)	0.998 (0.061)	0.999 (0.054)	0.974 (0.240)	0.949 (0.185)	0.953 (0.171)	0.956 (0.169)	1.673 (0.246)	1.619 (0.135)	1.624 (0.093)	1.608 (0.084)
2000	1.023 (0.136)	0.989 (0.075)	0.995 (0.060)	1.002 (0.052)	0.994 (0.254)	0.967 (0.199)	0.946 (0.145)	0.967 (0.167)	1.681 (0.238)	1.619 (0.141)	1.621 (0.096)	1.609 (0.081)
10J	1.023 (0.140)	0.985 (0.081)	0.991 (0.061)	0.999 (0.054)	1.022 (0.435)	0.953 (0.283)	0.950 (0.195)	0.956 (0.169)	1.675 (0.253)	1.613 (0.141)	1.617 (0.096)	1.608 (0.084)
J^2	1.023 (0.140)	0.987 (0.065)	0.986 (0.058)	0.994 (0.051)	1.022 (0.435)	0.926 (0.210)	0.944 (0.145)	0.955 (0.127)	1.675 (0.253)	1.619 (0.136)	1.613 (0.092)	1.603 (0.086)
∞	1.025 (0.133)	0.982 (0.072)	0.992 (0.062)	1.002 (0.054)	0.989 (0.234)	0.929 (0.134)	0.956 (0.137)	0.967 (0.134)	1.682 (0.238)	1.614 (0.139)	1.617 (0.097)	1.610 (0.087)

Standard error across repetitions in the parentheses.

by averaging resulting values over the simulated data, we obtain the estimate for the expected value of $\mathbf{\Gamma}_{J,T}$ in assumption B2 and $\mathbf{\Phi}$ respectively to estimate the asymptotic variance of the parameters. Similar calculations are implemented also for the original BLP (1995) framework in which the asymptotic variance matrix is $(\mathbf{\Gamma}'\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\mathbf{\Phi}\mathbf{\Gamma}(\mathbf{\Gamma}'\mathbf{\Gamma})^{-1}$. Table 6 shows the result. The simulated standard errors of estimates seem

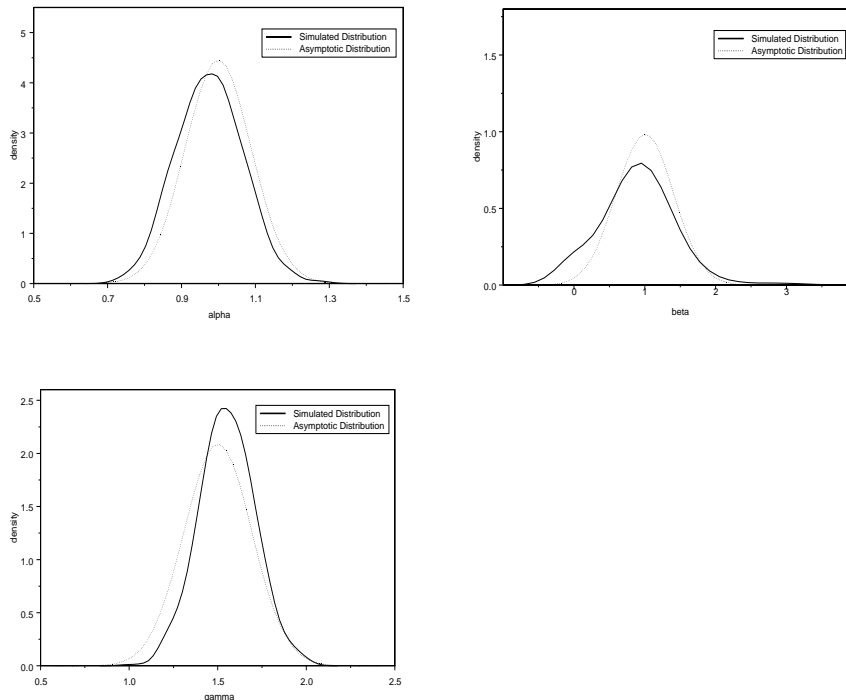
Table 6: Simulated and Estimated Standard Errors, $J = 25, n = 2000, R = 2000, N = 2000, T = 500$.

		α (1.0)	β (1.0)	γ (1.5)
BLP (1995) Framework	Mean	0.976	0.900	1.552
	Monte Carlo Std. Error	0.090	0.533	0.157
	Asymptotic Std. Error	0.088	0.393	0.186
Petrin (2002) Extension	Mean	0.996	1.022	1.570
	Monte Carlo Std. Error	0.077	0.254	0.149
	Asymptotic Std. Error	0.074	0.221	0.184

to be consistent with the asymptotic standard errors except those of β in BLP (1995) framework. It seems that difficulty in estimating correct β is even more pronounced for the BLP (1995) framework. This table shows the potential benefit of additional information in improving the accuracy of the random coefficient estimate.

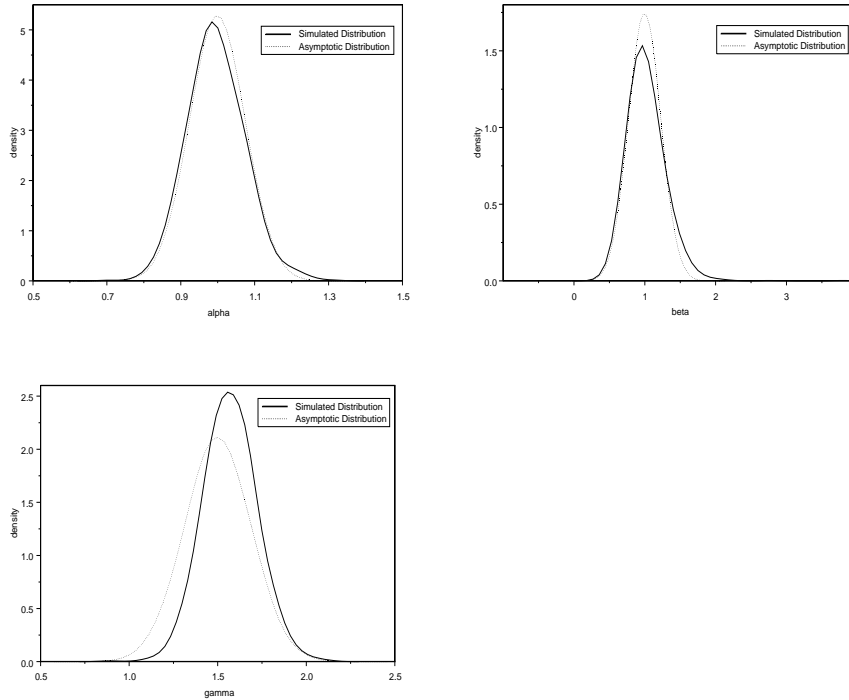
We make density estimates for the estimated parameters from the 1000 estimates used in table 6. (To make these plots, we use the density-plot command in the S-plus package with default options.) The solid

Figure 1: Kernel Density Estimate of Parameters, BLP (1995) Framework, $J=25, n=2000, R=2000$.



lines in Figures 1 and 2 show the densities of the estimated parameters, while the dotted lines show their asymptotic distributions using the true parameter values and the associated asymptotic variances in Table 6

Figure 2: Kernel Density Estimate of Parameters, the Petrin (2002) Extension, $J=25$, $n=2000$, $R=2000$, $T=500$, $N=2000$.



as mean and variance. From these plots, we observe that the simulated distributions of the estimates for the demand parameters α and β seem to improve significantly by the additional moments, while we also observe that the additional moments do not contribute at all in estimating the supply side parameter γ .

6 Conclusion and Discussion

In this paper, we generalize the GMM estimator extended by Petrin (2002) and provide the conditions under which this estimator not only has the CAN properties, but also is more efficient than the original BLP (1995) estimator. We sample two sets of individuals independent of each other, one to simulate the market share of products and the other to evaluate the additional moments, in order to avoid intractable correlations between these two sets of individuals. We also assume that the additional information on demographics of consumers are constructed from the sample independent of these two samples. With some additional assumptions, the suggested estimator is shown to have the CAN properties and to be more efficient than the BLP (1995) estimator.

We do not believe that the independent-source requirement is so restrictive or unrealistic. For instance, in analysing the U.S. automobile market we could sample individuals from the IPUMS-CPS to simulate the market shares of products, while the additional market information can be obtained from sources independent of the IPUMS-CPS such as J.D. Power and Associates.

In implementation, Petrin (2002) used the CEX to approximate the empirical distribution of demograph-

ics and he also used the CEX automobile supplement to link demographics of purchasers of new vehicles to the vehicles they purchase. We are not certain how the CEX automobile supplement was compiled, but the information of U.S. household purchasing patterns and demographics of purchasers of new vehicles may have been originated from the same Interview Survey of the CEX (Chapter 16, Consumer Expenditures and Income, The US Bureau of Labor Statistics, April, 2007). Therefore, the sample (P^R) used to simulate the market share of products and the sample used to obtain the additional market information may have been highly correlated. If so, the simulation error ϵ^R and the sampling error ϵ^N may have been correlated and Petrin's treatment of his data may not have satisfied our independent-sources requirement.

If this is to be the case, there is no guarantee that the estimator is asymptotically normal although its consistency remains valid.⁷ Asymptotic normality may not hold because the key part of the proof of the asymptotic normality relies on the fact that the three error terms (ϵ^n , ϵ^R , and ϵ^N) enter into the moment condition $\mathcal{G}_{J,T}(\theta)$ in additively separable and linear way, and if they are correlated, the sum of these terms in general does not weakly converge to the normal distribution even when each of them does so.

A Proofs

Proof of (1-i)

We will show (1-i) by using Theorem 3.1 of Pakes and Pollard (1989) which gives a sufficient condition under which an optimization estimator can be consistent for the true parameter value. The theorem guarantees that an estimator $\tilde{\theta}$ that satisfies $\|\mathcal{G}_{J,T}(\tilde{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| = \inf_{\theta \in \Theta} \|\mathcal{G}_{J,T}(\theta, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| + o_p(1)$ is consistent for θ^0 if

$$(1-i-a) \quad \mathcal{G}_{J,T}(\theta^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1), \text{ and}$$

$$(1-i-b) \quad \sup_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} \|\mathcal{G}_{J,T}(\theta, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|^{-1} = O_p(1) \quad \text{for each } \delta > 0.$$

(1-i-a)

We show (i-a) by applying the Bernoulli's weak law of large numbers to each row of $\mathcal{G}_{J,T}(\theta^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = (\mathbf{G}_J^d(\theta_d^0, \mathbf{s}^0, P^0)')$, $\mathbf{G}_J^c(\theta^0, \mathbf{s}^0, P^0)'$, $\mathbf{G}_{J,T}(\theta_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)'$. We illustrate how this can be done using the demand side sample moments. The m -th element of the demand side sample moments $\mathbf{G}_J^d(\theta^0, \mathbf{s}^0, P^0)$ is the average of $z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ over j where $z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ are not independent across j due to the interdependence of z_{jm}^d — $z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ are just conditionally independent given \mathbf{X}_1 . The Bernoulli's weak law of large numbers does not require independence nor identical distributedness among the $z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$, but requires the variance of $J^{-1} \sum_{j=1}^J z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ to converge to zero as J goes to infinity. Since z_{jm}^d are functions of \mathbf{X}_1 and the conditional expectation of $\xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ given \mathbf{X}_1 is zero in (1), the expectation and variance of $J^{-1} \sum_{j=1}^J z_{jm}^d \xi_j(\theta_d^0, \mathbf{s}^0, P^0)$ are respectively 0

⁷Consistency remains valid because our consistency proof of the GMM estimator with the additional moments does not require additional assumption even when the two error terms (ϵ^R and ϵ^N) are correlated. In the proof, (1-i) is only concerned with the behavior of $\mathcal{G}_{J,T}$ around the true value of market shares, demographics from population distribution, and additional information without errors. To establish (1-ii), we do not require independent-source assumption either because the relevant part $\sup_{\theta_d \in \Theta_d} \|\mathcal{G}_{J,T}^a(\theta_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathcal{G}_{J,T}^a(\theta_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|$ can be shown $o_p(1)$ only using A1(c) and A9.

and $\mathbb{E}_{\mathbf{x}_1}[J^{-2} \sum_{j=1}^J (z_{jm}^d)^2 \mathbb{E}_{\xi_j}[\xi_j^2(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) | \mathbf{X}_1]]$. Since the conditional variance of ξ_j is bounded in (1) by some constant $M > 0$ or $\mathbb{E}_{\xi_j}[\xi_j^2(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) | \mathbf{X}_1] < M$ with probability one, we have $J^{-2} \sum_{j=1}^J (z_{jm}^d)^2 \mathbb{E}_{\xi_j}[\xi_j^2(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) | \mathbf{X}_1] \leq (1/J)(\sum_{j=1}^J (z_{jm}^d)^2/J)M$. We know that $\sum_{j=1}^J (z_{jm}^d)^2/J$ is $O_p(1)$ and uniformly integrable by A4(a). Uniform integrability guarantees that the order of magnitude does not change after taking expectation, and this enable us to claim $\mathbb{E}_{\mathbf{x}_1}[\sum_{j=1}^J (z_{jm}^d)^2/J] = O(1)$. Hence $\mathbb{V}_{\mathbf{x}_1, \xi}[J^{-1} \sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = \mathbb{E}_{\mathbf{x}_1}[J^{-2} \sum_{j=1}^J (z_{jm}^d)^2 \mathbb{E}_{\xi_j}[\xi_j^2(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) | \mathbf{X}_1]] \leq (M/J) \mathbb{E}_{\mathbf{x}_1}[\sum_{j=1}^J (z_{jm}^d)^2/J] = (M/J) \cdot O(1) \rightarrow 0$ as $J \rightarrow \infty$. Bernoulli's weak law of large numbers ensures that the m -th element of $\mathbf{G}_{J,T}^d(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$ converges to 0 in probability, i.e., $\lim_{J \rightarrow \infty} \Pr[\{|\mathbf{G}_{J,T}^d(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)\}_m| > \epsilon] = \lim_{J \rightarrow \infty} \Pr[\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)/J > \epsilon] \leq \lim_{J \rightarrow \infty} \mathbb{V}_{\mathbf{x}_1, \xi}[\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)/J]/\epsilon^2 \leq \lim_{J \rightarrow \infty} (M/J) \cdot O(1)/\epsilon^2 = 0$. Thus $\|\mathbf{G}_{J,T}^d(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)\| = o_p(1)$. Similarly, we can show that the supply side moments $\mathbf{G}_{J,T}^c(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)$ converge to $\mathbb{E}_{\mathbf{w}_1, \omega}[\mathbf{G}_{J,T}^c(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = 0$ in probability by (6) and A4(b).

We denote the element of the additional moments $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}, P, \boldsymbol{\eta})$ corresponding to consumer's demographics d and discriminating attribute q as $\{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}, P, \boldsymbol{\eta})\}_{d,q}$. By the definition of η_{dq}^0 in (13), the expectation of $\{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}$ is zero, while the variance can be rewritten as follows.

$$\begin{aligned} & \mathbb{V}[\{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}] \\ &= \mathbb{E}_{\mathbf{x}, \xi} [\mathbb{V}_{\nu | \mathbf{x}, \xi} [\{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}]] \\ &= \frac{1}{T} \mathbb{E}_{\mathbf{x}, \xi} \left[\mathbb{E}_{\nu | \mathbf{x}, \xi} \left[\left\{ \nu_{td}^{obs} \sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0) \right\}^2 \right] / \left\{ \sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \right\}^2 \right] - \frac{1}{T} \mathbb{E}_{\mathbf{x}, \xi} \left[(\eta_{dq}^0)^2 \right] \right] \\ &\leq (1/T) \mathbb{E}_{\mathbf{x}, \xi} \left[\mathbb{E}_{\nu | \mathbf{x}, \xi} \left[(\nu_{td}^{obs})^2 \right] / \left\{ \sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \right\}^2 \right] - (1/T) \mathbb{E}_{\mathbf{x}, \xi} \left[(\eta_{dq}^0)^2 \right] \end{aligned}$$

where we abbreviate the vector $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$ in the conditional expectation for notational simplicity. Assumption A8 guarantees that $\mathbb{E}_{\mathbf{x}, \xi}[1/\{\sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\}^2] = O(1)$. Since the support of consumer's demographics distribution is assumed bounded, its second moment is finite, i.e., $\mathbb{E}_{\nu | \mathbf{x}, \xi}[(\nu_{td}^{obs})^2 | \mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0)] = \mathbb{E}_{\nu}[(\nu_{td}^{obs})^2] \leq M$ for some constant $M < \infty$. Moreover, we assume $\mathbb{E}_{\mathbf{x}, \xi}[(\eta_{dq}^0)^2] = O(1)$. Therefore, the variance of the additional moment is $\mathbb{V}[\{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}] \leq (1/T) \mathbb{E}_{\mathbf{x}, \xi}[M/\{\sum_{j \in \mathcal{Q}_q} \sigma_j(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\}^2] - (1/T) \mathbb{E}_{\mathbf{x}, \xi}[(\eta_{dq}^0)^2] \leq O(1/T) + O(1/T) = o(1)$. Thus the Bernoulli's law of large number ensures that $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)$ as $T \rightarrow \infty$ (and hence $J \rightarrow \infty$).

(1-i-b)

For every $(\epsilon, \delta) > (0, 0)$ and any positive function of δ , $C(\delta)$, the following relationship holds in general.

$$\begin{aligned} (B.1) \quad & \left\{ \inf_{\theta \notin \mathcal{N}_{\rho}(\delta)} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) \right\} \\ & \subset \left\{ \inf_{\theta \notin \mathcal{N}_{\rho}(\delta)} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| + \|\mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) \right\} \\ & \subset \left\{ \inf_{\theta \notin \mathcal{N}_{\rho}(\delta)} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) - \frac{\epsilon}{2} \right\} \cup \left\{ \|\mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq \frac{\epsilon}{2} \right\}. \end{aligned}$$

Taking probabilities and rearranging both sides of (B.1) give

$$(B.2) \quad \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho}(\delta)} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) - \frac{\epsilon}{2} \right]$$

$$\geq \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) \right] - \Pr \left[\|\mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq \frac{\epsilon}{2} \right].$$

For the second term on the right hand side of (B.2), since $\mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)$, for any $\epsilon > 0$, there exist $J_1(\epsilon)$ and $T_1(\epsilon)$ such that when $J > J_1$ and $T > T_1$

$$(B.3) \quad \Pr \left[\|\mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq \epsilon/2 \right] \leq \epsilon/2.$$

From assumption A5, for the ϵ and for any $\delta > 0$, there exist $C_2(\delta)$ and $J_2(\epsilon, \delta)$ such that when $J > J_2$

$$\Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\|^2 < C_2(\delta) \right] < \frac{\epsilon}{2}.$$

Thus, when $J > J_2$, we have

$$\begin{aligned} & \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|^2 < C_2(\delta) \right] \\ &= \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \{ \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\|^2 + \|\mathbf{G}_{J,T}^a(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\|^2 \} < C_2(\delta) \right] \\ &\leq \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) - \mathbf{G}_J(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0)\|^2 < C_2(\delta) \right] \leq \epsilon/2. \end{aligned}$$

Therefore, by setting $C(\delta) = C_2(\delta)^{1/2}$, for the first term on the right hand side of (B.2), we have

$$(B.4) \quad \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbf{G}_{J,T}(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) \right] \geq 1 - \frac{\epsilon}{2}.$$

By substituting (B.3) and (B.4) for (B.2), for $J > \max(J_1, J_2)$ and $T > T_1$,

$$\Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \geq C(\delta) - \epsilon/2 \right] \geq 1 - \frac{\epsilon}{2} - \frac{\epsilon}{2} = 1 - \epsilon.$$

Then we have $\limsup_{J,T} \Pr \left[\inf_{\theta \notin \mathcal{N}_{\rho^0(\delta)}} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| > C^*(\epsilon, \delta) \right] \geq 1 - \epsilon$ for $C^*(\epsilon, \delta) = C(\delta) - \epsilon/2$ and hence (1-i-b) is shown.

Proof of (1-ii)

We will show $\sup_{\theta \in \Theta} \|\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| = o_p(1)$. From the definitions of $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ and $\mathbf{G}_{J,T}(\boldsymbol{\theta}, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)$, we have

$$\begin{aligned} & \sup_{\theta \in \Theta} \|\mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0)\|^2 \\ &\leq \sup_{\theta_d \in \Theta_d} \|J^{-1} \mathbf{Z}'_d \{ \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) \}\|^2 + \sup_{\theta \in \Theta} \|J^{-1} \mathbf{Z}'_c \{ \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \}\|^2 \\ &\quad + \sup_{\theta_d \in \Theta_d} \|\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \\ &\leq J^{-1} \|\mathbf{Z}'_d \mathbf{Z}_d\| \times \sup_{\theta_d \in \Theta_d} J^{-1} \|\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)\|^2 + J^{-1} \|\mathbf{Z}'_c \mathbf{Z}_c\| \times \sup_{\theta \in \Theta} J^{-1} \|\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0)\|^2 \\ &\quad + \sup_{\theta_d \in \Theta_d} \|\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \end{aligned}$$

where the terms $\|\mathbf{Z}'_d \mathbf{Z}_d\|/J$ and $\|\mathbf{Z}'_c \mathbf{Z}_c\|/J$ are respectively $O_p(1)$ by assumptions A4(a) and A4(b). Thus it remains to show that

$$(B.5) \quad \sup_{\theta_d \in \Theta_d} J^{-1} \|\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)\|^2 = o_p(1),$$

$$(B.6) \quad \sup_{\theta \in \Theta} J^{-1} \|\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0)\|^2 = o_p(1),$$

$$(B.7) \quad \sup_{\theta_d \in \Theta_d} \|\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| = o_p(1).$$

Since the demand side condition (B.5) is established in BLP (2004) by A3 and A6, we will work on (B.6) and (B.7).

Proof of (B.6)

The Glivenko-Cantelli theorem gives $\Pr[P^R \notin \mathcal{N}_{P^0}(\delta)] \rightarrow 0$ for $\delta > 0$ as $R \rightarrow \infty$. With (B.5) already established, we have $\Pr[(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), P^R) \in \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d; \delta) \times \mathcal{N}_{P^0}(\delta)] \rightarrow 1$ for given $\delta > 0$. Thus assumption A7 guarantees that the differences in the profit margin behave uniformly over $\boldsymbol{\theta}_d \in \Theta_d$ as

$$(B.8) \quad \sup_{\theta_d \in \Theta_d} J^{-\frac{1}{2}} \|\mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| = o_p(1).$$

Since $\dot{g}(\cdot)$ is assumed finite for all realizable values of cost, we derive (B.6) by using (B.8) in the following inequality with the definition of $\omega_j(\boldsymbol{\theta}, \mathbf{s}, P)$ in (7).

$$\begin{aligned} & \sup_{\theta \in \Theta} J^{-1} \|\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0)\|^2 \\ &= \sup_{\theta_d \in \Theta_d} J^{-1} \sum_{j=1}^J \left\{ g(p_j - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)) - g(p_j - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \right\}^2 \\ &= \sup_{\theta_d \in \Theta_d} J^{-1} \sum_{j=1}^J \left[\dot{g}(p_j - \overline{m}_{g_j}) \left\{ m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right\} \right]^2 \\ &\leq \sup_{\theta_d \in \Theta_d} \sup_{1 \leq j \leq J} |\dot{g}(p_j - \overline{m}_{g_j})|^2 \cdot \sup_{\theta_d \in \Theta_d} J^{-1} \sum_{j=1}^J \left\{ m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right\}^2 \\ &= \sup_{\theta_d \in \Theta_d} \sup_{1 \leq j \leq J} |\dot{g}(p_j - \overline{m}_{g_j})|^2 \cdot \sup_{\theta_d \in \Theta_d} J^{-1} \|\mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\|^2 \\ &= o_p(1) \end{aligned}$$

where \overline{m}_{g_j} are between $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ and $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$. We should note that the difference between $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R)$ and $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0)$ includes only the demand side parameters $\boldsymbol{\theta}_d$ because of the linear dependence of $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}, P)$ on the supply side parameters $\boldsymbol{\theta}_c$ as seen in (7).

Proof of (B.7)

We show (B.7) as follows.

$$\begin{aligned} & \sup_{\theta_d \in \Theta_d} \|\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\| \\ &= \sup_{\theta_d \in \Theta_d} \|\boldsymbol{\eta}^N - T^{-1} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R), \boldsymbol{\theta}_d, P^R) \\ & \quad - \{\boldsymbol{\eta}^0 - T^{-1} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0), \boldsymbol{\theta}_d, P^0)\| \end{aligned}$$

$$\begin{aligned}
&\leq \|\boldsymbol{\eta}^N - \boldsymbol{\eta}^0\| + \sup_{\boldsymbol{\theta}_d \in \Theta_d} \left\| T^{-1} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\} \right\| \\
&= \|\boldsymbol{\eta}^N - \boldsymbol{\eta}^0\| + \sup_{\boldsymbol{\theta}_d \in \Theta_d} T^{-1} \|(\boldsymbol{\nu}^{obs})' \{\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\}\| \\
&\leq \|\boldsymbol{\eta}^N - \boldsymbol{\eta}^0\| + T^{-1/2} \|\boldsymbol{\nu}^{obs}\| \cdot \sup_{\boldsymbol{\theta}_d \in \Theta_d} T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\|, \\
&= O_p(N^{-1/2}) + O_p(1) \cdot o_p(1) = o_p(1)
\end{aligned}$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = (\boldsymbol{\psi}_1(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \dots, \boldsymbol{\psi}_T(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))'$ and $\boldsymbol{\nu}^{obs} = (\boldsymbol{\nu}_1^{obs}, \dots, \boldsymbol{\nu}_T^{obs})'$. In the last equality above, $\|\boldsymbol{\eta}^N - \boldsymbol{\eta}^0\| = O_p(N^{-1/2})$ comes from A1(c), and $T^{-1/2} \|\boldsymbol{\nu}^{obs}\| = O_p(1)$ is because the observed consumer demographics $\boldsymbol{\nu}_t^{obs}$ are assumed bounded. The $o_p(1)$ term follows from the next inequality with assumption A9:

$$\begin{aligned}
&\Pr \left[\sup_{\boldsymbol{\theta}_d \in \Theta_d} T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| > \delta \right] \\
&\leq \Pr \left[\sup_{\boldsymbol{\theta}_d \in \Theta_d} \sup_{(\boldsymbol{\xi}, P) \in \mathcal{N}_{\xi^0(\boldsymbol{\theta}_d, \delta)} \times \mathcal{N}_{P^0(\delta)}} T^{-1/2} \|\boldsymbol{\Psi}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\| > \delta \right] \\
&\quad + \Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) \notin \mathcal{N}_{\xi^0(\boldsymbol{\theta}_d, \delta)}] + \Pr[P^R \notin \mathcal{N}_{P^0(\delta)}] \\
&\rightarrow 0.
\end{aligned}$$

Derivation of (26), (27), and (28)

Using the Taylor series approximation of $\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ up to the first order, BLP (2004) showed

$$(B.9) \quad \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R) = \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n,$$

$$(B.10) \quad \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0) = -\mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d).$$

These expressions allow us to derive the demand side moments $\mathbf{G}_j^d(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R)$ in (26). The cost side derivation is performed using the demand side unobservables. Since $g(\cdot)$ is assumed to be continuously differentiable, the j -th element of $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R)$ can be written by the mean value theorem as

$$\begin{aligned}
\omega_j(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \omega_j(\boldsymbol{\theta}, \mathbf{s}^0, P^R) &= g(p_j - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R)) - g(p_j - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R)) \\
&= -\dot{g}(p_j - m_{g_j}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)) \frac{\partial m_{g_j}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)\}.
\end{aligned}$$

Substituting (B.9) for the above equations obtains the vector form expression

$$(B.11) \quad \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R) = -\mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n.$$

Similarly, we rewrite $\boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0)$ using (B.10) as follows.

$$\begin{aligned}
(B.12) \quad \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \mathbf{s}^0, P^0) &= \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R)) - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
&= \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
&\quad + \mathbf{L}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d).
\end{aligned}$$

These calculations in (B.11) and (B.12) lead us to the cost side moments $\mathbf{G}_j^c(\boldsymbol{\theta}, \mathbf{s}^n, P^R)$ in (27).

The additional moments $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ is rewritten as follows.

$$(B.13) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) \\ = \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) + \{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R, \boldsymbol{\eta}^N)\} \\ + \{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^N)\} + \{\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)\}.$$

Using (B.9) we express the second term on the right hand side of (B.13) as follows.

$$(B.14) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R, \boldsymbol{\eta}^N) \\ = \boldsymbol{\eta}^N - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \left\{ \boldsymbol{\eta}^N - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R) \right\} \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R)\} \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R)) \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n.$$

Similarly, with (B.10), the third term in (B.13) is

$$(B.15) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^N) \\ = \boldsymbol{\eta}^N - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \left\{ \boldsymbol{\eta}^N - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right\} \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\} \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \left\{ \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right. \\ \left. + \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0)) \right\} \\ = -\frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \left\{ \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right. \\ \left. - \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right\}.$$

The fourth term in (B.13) is

$$(B.16) \quad \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^N) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) = \boldsymbol{\eta}^N - \boldsymbol{\eta}^0.$$

Substituting (B.14), (B.15) and (B.16) for (B.13) obtains (28).

Proof of (2-i)

Suggested argument will be established by showing that for any $\delta_J \rightarrow 0$,

$$(B.17) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \left\| \mathcal{G}_J^d(\theta_d) - \mathbf{G}_J^d(\theta_d, \mathbf{s}^n, P^R) \right\| = o_p(1),$$

$$(B.18) \quad \sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \left\| \mathcal{G}_J^c(\theta) - \mathbf{G}_J^c(\theta, \mathbf{s}^n, P^R) \right\| = o_p(1),$$

$$(B.19) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} T^{\frac{1}{2}} \left\| \mathcal{G}_{J,T}^a(\theta_d) - \mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) \right\| = o_p(1).$$

Since (B.17) is shown in BLP (2004) under assumptions B5(a)(b), we focus on (B.18) and (B.19).

From (27) and (29), we know that

$$\begin{aligned} & \left\| J^{\frac{1}{2}} [\mathcal{G}_J^c(\theta) - \mathbf{G}_J^c(\theta, \mathbf{s}^n, P^R)] \right\| \\ &= \left\| -J^{-\frac{1}{2}} \mathbf{Z}'_c \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\} \right. \\ & \quad - J^{-\frac{1}{2}} \mathbf{Z}'_c \left[\mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \right. \\ & \quad \left. \left. - \mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n + \mathbf{L}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right] \right\| \\ & \leq \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left[\mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \mathbf{g}(\mathbf{p} - \mathbf{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \right] \right\| \\ & \quad + \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left\{ \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} - \mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \right\} \boldsymbol{\epsilon}^n \right\| \\ & \quad + \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left\{ \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \mathbf{L}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right\} \right\|. \end{aligned}$$

We show the three terms on the right-hand side of the inequality above are respectively $o_p(1)$ within the $\delta_{J,T}$ neighborhood of $\boldsymbol{\theta}_d^0$. We know the first term to be $o_p(1)$ by B5(e). The second term is shown $o_p(1)$ by using B5(c) as follows.

$$\begin{aligned} & \Pr \left[\sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left\{ \mathbf{L}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \right\} \boldsymbol{\epsilon}^n \right\| > c \right] \\ & \leq \Pr \left[\sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left\{ \mathbf{L}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P) \mathbf{M}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P) \mathbf{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \right\} \boldsymbol{\epsilon}^n \right\| > c \right] \\ & \quad + \Pr[(\bar{\boldsymbol{\xi}}_1, \dots, \bar{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[(\bar{\boldsymbol{\xi}}_1, \dots, \bar{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \\ & \rightarrow 0. \end{aligned}$$

Similarly, for the third term, we obtain by assumption B5(d)

$$\begin{aligned} & \Pr \left[\sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \left\{ \mathbf{L}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{M}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \right\} \right\| > c \right] \\ & \leq \Pr \left[\sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left\| J^{-\frac{1}{2}} \mathbf{Z}'_c \right. \right. \\ & \quad \left. \left. \times \left\{ \mathbf{L}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P) \mathbf{M}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P) \mathbf{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) - \mathbf{L}_0 \mathbf{M}_0 \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \right\} \right\| > c \right] \\ & \quad + \Pr[(\underline{\boldsymbol{\xi}}_1, \dots, \underline{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[(\underline{\boldsymbol{\xi}}_1, \dots, \underline{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \\ & \rightarrow 0. \end{aligned}$$

Thus, we obtain (B.18).

For the element of the additional moments $\mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N)$ which corresponds to consumer demographics d and discriminating attribute q , we have

(B.20)

$$\begin{aligned}
& \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{\frac{1}{2}} \left\{ \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d) - \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^n, P^R, \boldsymbol{\eta}^N) \right\}_{d,q} \right| \\
&= \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{\frac{1}{2}} \left[\frac{1}{T} \sum_{t=1}^T \nu_{td}^{obs} \{ \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \} \right. \right. \\
&\quad \left. \left. - \frac{1}{T} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \{ \boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \} \right. \right. \right. \\
&\quad \left. \left. \left. - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n + \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right] \right] \right| \\
&\leq \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \{ \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \} \right| \\
&\quad + \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right] \right| \\
&\quad + \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right] \right|
\end{aligned}$$

where $\boldsymbol{\Upsilon}_{tq}$ is the q th row vector of $\boldsymbol{\Upsilon}_t$. Thus, it is sufficient to show that the three terms on the right-hand side of the inequality above are respectively $o_p(1)$ or,

$$(B.21) \quad \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{td}^{obs} \{ \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \} \right| = o_p(1),$$

$$(B.22) \quad \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right] \right| = o_p(1),$$

$$(B.23) \quad \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right] \right| = o_p(1).$$

We obtain (B.21) as

$$\begin{aligned}
& \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \{ \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \} \right| \\
&= \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \left\{ \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)} - \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} \right\} \right| \\
&= \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \left\{ \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} \right\} \right. \\
&\quad \left. \times \frac{\sum_{j \in \mathcal{Q}_q} \{ \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) - \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R) \}}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)} \right| \\
&= \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1/2} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \frac{\sum_{j \in \mathcal{Q}_q} \{ -\epsilon_j^R(\boldsymbol{\theta}_d) \}}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)} \right| \\
&\leq \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \sup_{\|\boldsymbol{\theta}_d - \boldsymbol{\theta}_d^0\| < \delta_{J,T}} \left| \frac{\sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^R)} \right|
\end{aligned}$$

$$\begin{aligned}
&= \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| \frac{\sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \{\sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) + \epsilon_j^R(\boldsymbol{\theta}_d)\}} \right| \\
&= \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| \frac{\sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} s_j^0 + \sum_{j \in \mathcal{Q}_q} \epsilon_j^R(\boldsymbol{\theta}_d)} \right| \\
&= \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| \frac{\left(\sum_{j \in \mathcal{Q}_q} s_j^0\right)^{-1} \sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d)}{1 + \left(\sum_{j \in \mathcal{Q}_q} s_j^0\right)^{-1} T^{-1/2} \sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d)} \right| \\
&= \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| \frac{O_p(1) \cdot o_p(1)}{1 + O_p(1) \cdot T^{-1/2} o_p(1)} \right| \\
&= \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \cdot o_p(1) \\
&= o_p(1)
\end{aligned}$$

where we use assumption A8 for $(\sum_{j \in \mathcal{Q}_q} s_j^0)^{-1} = O_p(1)$ and assumption B5(h) for $\sum_{j \in \mathcal{Q}_q} T^{1/2} \epsilon_j^R(\boldsymbol{\theta}_d) = o_p(1)$. For the last equality above, we use a weak law of large number that ensures $\sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} |T^{-1} \sum_{t=1}^T \nu_{td}^{obs} \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d, P^0)| \xrightarrow{p} |\eta_{dq}^0| = O_p(1)$ where $\eta_{dq}^0 = O_p(1)$. For (B.22), we have

$$\begin{aligned}
&\Pr \left[\sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \nu_{td}^{obs} \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right] \right| > c \right] \\
&\leq \Pr \left[\max_t |\nu_{td}^{obs}| \cdot \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right] \right| > c \right] \\
&\leq \Pr \left[\max_t |\nu_{td}^{obs}| \cdot \sup_{\|\theta_d - \theta_d^0\| < \delta_{J,T}} \sup_{(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, P) \in \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \left| T^{-\frac{1}{2}} \sum_{t=1}^T \left[\boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \mathbf{H}_0^{-1} \boldsymbol{\epsilon}^n \right. \right. \right. \\
&\quad \left. \left. \left. - \boldsymbol{\Upsilon}_{tq}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P^R) \mathbf{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P^R) \boldsymbol{\epsilon}^n \right] \right| > c \right] \\
&\quad + \Pr[\boldsymbol{\xi}^\dagger \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[\bar{\boldsymbol{\xi}} \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \\
&= o(1)
\end{aligned}$$

where we use assumption B5(f) and $\max_t |\nu_{td}^{obs}| < M(\text{constant})$ as well as the facts $\Pr[\boldsymbol{\xi}^\dagger \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] \rightarrow 0$, $\Pr[\bar{\boldsymbol{\xi}} \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_{J,T})\}^J] \rightarrow 0$, and $\Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \rightarrow 0$. We also obtain (B.23) by the argument similar to the argument for (B.22) with assumption B5(g).

Proof of (2-ii)

To show that the asymptotic normality of the estimator $\check{\boldsymbol{\theta}}$ that minimizes the norm of $\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})$, we use a version of Theorem 3.3 in Pakes and Pollard (1989), which gives asymptotic normality to the estimator indexed by two distinct indices. From the theorem, if we can show the following five conditions,

- (i) $\|\boldsymbol{\mathcal{G}}_{J,T}(\check{\boldsymbol{\theta}})\| = o_p(J^{-\frac{1}{2}}) + o_p(T^{-\frac{1}{2}}) + \inf_{\boldsymbol{\theta}} \|\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})\|$;
- (ii) $E[\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})]$ is differentiable at $\boldsymbol{\theta}^0$ with a derivative matrix $\boldsymbol{\Gamma}_{J,T} = (\boldsymbol{\Gamma}'_J, \boldsymbol{\Gamma}_{J,T}^a)'$ of full rank where $\boldsymbol{\Gamma}_{J,T}$ converges to $(\boldsymbol{\Gamma}', \boldsymbol{\Gamma}^{a'})'$ as $J, T \rightarrow \infty$;

(iii) for every sequence $\{\delta_{J,T}\}$ of positive numbers that converges to zero as J, T goes to infinity,

$$(a) \quad \sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} \frac{\|\mathcal{G}_J(\theta) - \mathbb{E}[\mathcal{G}_J(\theta)] - \mathcal{G}_J(\theta^0)\|}{J^{-\frac{1}{2}} + \|\mathcal{G}_J(\theta)\| + \|\mathbb{E}[\mathcal{G}_J(\theta)]\|} = o_p(1);$$

$$(b) \quad \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \frac{\|\mathcal{G}_{J,T}^a(\theta_d) - \mathbb{E}[\mathcal{G}_{J,T}^a(\theta_d)] - \mathcal{G}_{J,T}^a(\theta_d^0)\|}{T^{-\frac{1}{2}} + \|\mathcal{G}_{J,T}^a(\theta_d)\| + \|\mathbb{E}[\mathcal{G}_{J,T}^a(\theta_d)]\|} = o_p(1);$$

(iv)

$$\begin{pmatrix} J^{\frac{1}{2}} \mathcal{G}_J(\theta^0) \\ T^{\frac{1}{2}} \mathcal{G}_{J,T}^a(\theta_d^0) \end{pmatrix} \overset{w}{\rightsquigarrow} N \left(\mathbf{0}, \begin{pmatrix} \Phi & \Phi^{12} \\ \Phi^{12'} & \Phi^a \end{pmatrix} \right);$$

(v) θ^0 is an interior point of Θ ,

(vi) The size indices T and J go to infinity so as to $J/T \rightarrow c \geq 0$,

then, we have $J^{1/2} \tilde{\theta} \overset{w}{\rightsquigarrow} N(\mathbf{0}, \mathbf{V})$ where $\mathbf{V} = (\mathbf{\Gamma}'\mathbf{\Gamma} + \mathbf{\Gamma}^{a'}\mathbf{\Gamma}^a)^{-1}(\mathbf{\Gamma}'\Phi\mathbf{\Gamma} + 2c^{1/2}\mathbf{\Gamma}'\Phi^{12}\mathbf{\Gamma}^a + c\mathbf{\Gamma}^{a'}\Phi^a\mathbf{\Gamma}^a)(\mathbf{\Gamma}'\mathbf{\Gamma} + \mathbf{\Gamma}^{a'}\mathbf{\Gamma}^a)^{-1}$.

Our estimator satisfies (i) by definition. Since the three random variables ϵ_{ji} , ϵ_{jt}^* and $\epsilon_{it}^\#$ in $\mathcal{G}_{J,T}(\theta)$ have respectively zero means given the set of product characteristics $(\mathbf{X}, \boldsymbol{\xi}(\theta_d^0, \mathbf{s}^0, P^0))$, we have $\mathbb{E}[\mathcal{G}_{J,T}(\theta, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] = \mathbb{E}[\mathcal{G}_{J,T}(\theta)]$. Thus condition (ii) follows from assumption B2. The condition (iii)(a) can be shown as follows.

$$\begin{aligned} & \sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} \frac{\|\mathcal{G}_J(\theta) - \mathbb{E}[\mathcal{G}_J(\theta)] - \mathcal{G}_J(\theta^0)\|}{J^{-\frac{1}{2}} + \|\mathcal{G}_J(\theta)\| + \|\mathbb{E}[\mathcal{G}_J(\theta)]\|} \\ & \leq \sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \|\mathcal{G}_J(\theta) - \mathbb{E}[\mathcal{G}_J(\theta)] - \mathcal{G}_J(\theta^0)\| \\ & \leq \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \|\mathcal{G}_J^d(\theta_d) - \mathbb{E}[\mathcal{G}_J^d(\theta_d)] - \mathcal{G}_J^d(\theta_d^0)\| + \sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \|\mathcal{G}_J^c(\theta) - \mathbb{E}[\mathcal{G}_J^c(\theta)] - \mathcal{G}_J^c(\theta^0)\| \\ & = o_p(1) + o_p(1) = o_p(1) \end{aligned}$$

where $\sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \|\mathcal{G}_J^d(\theta_d) - \mathbb{E}[\mathcal{G}_J^d(\theta_d)] - \mathcal{G}_J^d(\theta_d^0)\| = o_p(1)$ and $\sup_{\|\theta - \theta^0\| \leq \delta_{J,T}} J^{\frac{1}{2}} \|\mathcal{G}_J^c(\theta) - \mathbb{E}[\mathcal{G}_J^c(\theta)] - \mathcal{G}_J^c(\theta^0)\| = o_p(1)$ come respectively from B3(a) and (b). For condition (iii)(b), we have

$$\begin{aligned} & \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} \frac{\|\mathcal{G}_{J,T}^a(\theta_d) - \mathbb{E}[\mathcal{G}_{J,T}^a(\theta_d)] - \mathcal{G}_{J,T}^a(\theta_d^0)\|}{T^{-\frac{1}{2}} + \|\mathcal{G}_{J,T}^a(\theta_d)\| + \|\mathbb{E}[\mathcal{G}_{J,T}^a(\theta_d)]\|} \\ & \leq \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} T^{\frac{1}{2}} \|\mathcal{G}_{J,T}^a(\theta_d) - \mathbb{E}[\mathcal{G}_{J,T}^a(\theta_d)] - \mathcal{G}_{J,T}^a(\theta_d^0)\| \\ & = \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} T^{\frac{1}{2}} \left\| \mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\theta_d^0)\} + \boldsymbol{\eta}^N - \boldsymbol{\eta}^0 \right. \\ & \quad \left. - \mathbb{E}[\mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] + \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\theta_d^0)\}] - \mathbb{E}[\boldsymbol{\eta}^N - \boldsymbol{\eta}^0] \right. \\ & \quad \left. - \mathbf{G}_{J,T}^a(\theta_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) + \frac{1}{T} \sum_{t=1}^T \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\Upsilon}_t^0 \mathbf{H}_0^{-1} \{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\theta_d^0)\} - \boldsymbol{\eta}^N - \boldsymbol{\eta}^0 \right\| \\ & = \sup_{\|\theta_d - \theta_d^0\| \leq \delta_{J,T}} T^{\frac{1}{2}} \left\| \mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) - \mathbb{E}[\mathbf{G}_{J,T}^a(\theta_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0)] - \mathbf{G}_{J,T}^a(\theta_d^0, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) \right\| \\ & = o_p(1) \end{aligned}$$

from assumption B3(c). Assumption B1 guarantees condition (v). Let us show (iv). Given the expression in (30), or

$$\begin{aligned} \begin{pmatrix} J^{\frac{1}{2}} \mathbf{G}_J(\boldsymbol{\theta}) \\ T^{\frac{1}{2}} \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d) \end{pmatrix} &= \begin{pmatrix} J^{\frac{1}{2}} \mathbf{G}_J(\boldsymbol{\theta}, \mathbf{s}^0, P^0) \\ T^{\frac{1}{2}} \mathbf{G}_{J,T}^a(\boldsymbol{\theta}_d, \mathbf{s}^0, P^0, \boldsymbol{\eta}^0) \end{pmatrix} + \sum_{i=1}^n \mathbf{Y}_{JT_i} + \sum_{r=1}^R \mathbf{Y}_{JT_r} + \sum_{i'=1}^N \begin{pmatrix} \mathbf{0} \\ T^{\frac{1}{2}} \boldsymbol{\epsilon}_{i'}^{\#}/N \end{pmatrix} \\ &\equiv \mathbf{T}_{J,T,1} + \mathbf{T}_{J,T,2} + \mathbf{T}_{J,T,3} + \mathbf{T}_{J,T,4}, \end{aligned}$$

we need to show each of $\mathbf{T}_{J,T,1}$, $\mathbf{T}_{J,T,2}$, $\mathbf{T}_{J,T,3}$ and $\mathbf{T}_{J,T,4}$ converges to the multivariate normal. Notice that since these four terms are conditionally independent given $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$ and thus mutually uncorrelated, the Cramér-Wold device will ensure that the sum of them also converges to the multivariate normal.

The first term $\mathbf{T}_{J,T,1}$: Given the set of product characteristics $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$, the demand and supply components $(\mathbf{z}_j^d \boldsymbol{\xi}_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \mathbf{z}_j^c \boldsymbol{\omega}_j(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0))$ in $\mathbf{T}_{J,T,1}$ are independent across j while the components of the additional moments $(\boldsymbol{\eta}^0 - \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0))$ are independent across t . Therefore when we apply the Lyapunov's central limit theorem to $\mathbf{b}' \mathbf{T}_{J,T,1}$, we have to take into account two sampling processes indexed by j and t simultaneously. Write $\boldsymbol{\psi}_t^0 \equiv \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. The Lyapunov condition is satisfied for this term because

$$\begin{aligned} &\lim_{J,T \rightarrow \infty} \mathbb{V}[\mathbf{b}' \mathbf{T}_{J,T,1}]^{-(2+\delta)/2} \sum_{(j,t)=(1,1)}^{(J,T)} \mathbb{E} \left[\left\| \mathbf{b}' \begin{pmatrix} J^{-1/2} \mathbf{z}_j^d \boldsymbol{\xi}_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) \\ J^{-1/2} \mathbf{z}_j^c \boldsymbol{\omega}_j(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0) \\ T^{-1/2} (\boldsymbol{\eta}^0 - \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t^0) \end{pmatrix} \right\|^{2+\delta} \right] \\ &\leq \lim_{J,T \rightarrow \infty} \{\mathbf{b}' \mathbb{V}[\mathbf{T}_{J,T,1}] \mathbf{b}\}^{-(2+\delta)/2} \|\mathbf{b}'\|^{2+\delta} \sum_{(j,t)=(1,1)}^{(J,T)} \mathbb{E} \left[\left\| \begin{pmatrix} J^{-1/2} \mathbf{z}_j^d \boldsymbol{\xi}_j(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0) \\ J^{-1/2} \mathbf{z}_j^c \boldsymbol{\omega}_j(\boldsymbol{\theta}^0, \mathbf{s}^0, P^0) \\ T^{-1/2} (\boldsymbol{\eta}^0 - \boldsymbol{\nu}_t^{obs} \otimes \boldsymbol{\psi}_t^0) \end{pmatrix} \right\|^{2+\delta} \right] \\ &= \{\mathbf{b}' \boldsymbol{\Phi}_1 \mathbf{b}\}^{-(2+\delta)/2} \|\mathbf{b}'\|^{2+\delta} \cdot 0 = 0 \end{aligned}$$

for some $\delta > 0$ by assumptions B4(a) and B4(e). Thus we obtain $\{\mathbb{V}[\mathbf{b}' \mathbf{T}_{J,T,1}]\}^{-1/2} \mathbf{b}' \mathbf{T}_{J,T,1} \xrightarrow{w} N(0, 1)$ which is equivalent to $\mathbf{T}_{J,T,1} \xrightarrow{w} N(\mathbf{0}, \boldsymbol{\Phi}_1)$.

The second term $\mathbf{T}_{J,T,2}$: Abbreviate $\mathbf{Y}_{J,T,i}^0 \equiv \mathbf{Y}_{J,T,i}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Given $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0), \{\boldsymbol{\nu}_t^{obs}\}_{t=1}^T)$, $\mathbf{Y}_{J,T,i}^0$ has zero mean and conditionally independent across i . The Lyapunov condition for this term is

$$\begin{aligned} &\lim_{n \rightarrow \infty} \{\mathbb{V}[\mathbf{b}' \mathbf{T}_{J,T,2}]\}^{-(2+\delta)/2} \sum_{i=1}^n \mathbb{E}[|\mathbf{b}' \mathbf{Y}_{J,T,i}^0|^{2+\delta}] \\ &\leq \lim_{n \rightarrow \infty} \{\mathbb{V}[\mathbf{b}' \mathbf{T}_{J,T,2}]\}^{-(2+\delta)/2} \|\mathbf{b}'\|^{2+\delta} \sum_{i=1}^n \mathbb{E}[|\mathbf{Y}_{J,T,i}^0|^{2+\delta}] = \{\mathbf{b}' \boldsymbol{\Phi}_2 \mathbf{b}\}^{-(2+\delta)/2} \|\mathbf{b}'\|^{2+\delta} \cdot 0 \\ &= 0 \end{aligned}$$

for some $\delta > 0$ using assumptions B4(b) and B4(f). Thus we obtain $\{\mathbb{V}[\mathbf{b}' \mathbf{T}_{J,T,2}]\}^{-1/2} \mathbf{b}' \mathbf{T}_{J,T,2} \xrightarrow{w} N(0, 1)$ which is equivalent to $\mathbf{T}_{J,T,2} \xrightarrow{w} N(\mathbf{0}, \boldsymbol{\Phi}_2)$.

The third term $\mathbf{T}_{J,T,3}$ requires us to increase the number R of simulation draws. Using B4(c) and B4(g), we obtain $\mathbf{T}_{J,T,3} \xrightarrow{w} N(\mathbf{0}, \boldsymbol{\Phi}_3)$ from the argument similar to the argument for the second term.

The fourth term $\mathbf{T}_{J,T,4}$ has only the components for the additional moments. Given $(\mathbf{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \mathbf{s}^0, P^0))$, $\boldsymbol{\epsilon}_{i'}^{\#}$ has zero mean and conditionally independent across i' . Using B4(d) and B4(h), applying the Lyapunov central limit theorem to these components lead to $\mathbf{T}_{J,T,4} \xrightarrow{w} N(\mathbf{0}, \text{diag}(\mathbf{0}, \boldsymbol{\Phi}_4^a))$.

Since all of the four terms in (30) converge to the normal, their sum also converges to the normal where the asymptotic variance-covariance matrix is $\Phi_1 + \Phi_2 + \Phi_3 + \text{diag}(\mathbf{0}, \Phi_4^a)$.

References

- [1] Berry, S. T., “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25 (1994), 242–262.
- [2] Berry, S., J. Levinsohn, and A. Pakes, “Automobile Prices in Market Equilibrium,” *Econometrica*, 63 (1995), 841–890.
- [3] Berry, S., J. Levinsohn, and A. Pakes, “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market,” *Journal of Political Economy*, 112 (2004), 69–105.
- [4] Berry, S., O. Linton, and A. Pakes, “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems,” *Review of Economic Studies*, 71 (2004), 613–654.
- [5] *Consumer Expenditures and Income*, The US Bureau of Labor Statistics, April, 2007.
- [6] Imbens, G. W., and T. Lancaster, “Combining Micro and Macro Data in Microeconomic Models,” *Review of Economic Studies*, 61 (1994), 655–680.
- [7] Mann, H. B., and A. Wald, “On Stochastic Limit and Order Relationships,” *Annals of Mathematical Statistics*, 15 (1943), 780–799.
- [8] Nevo, A., “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, 69 (2001), 307–342.
- [9] Pakes, A. and D. Pollard., “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57 (1989), 1027–1057.
- [10] Petrin, A., “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, 110 (2002), 705–729.
- [11] Sudhir, K., “Competitive Pricing Behavior in the Auto Market: A Structural Analysis,” *Marketing Science*, 20 (2001), 42–60.