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

| | |
|---|---|
| 著者 | MYOJO Satoshi, KANAZAWA Yuichiro |
| year | 2007-10 |
| シリーズ | Department of Social Systems and Management Discussion Paper Series ~ no. 1185 |
| URL | http://hdl.handle.net/2241/100198 |

Department of Social Systems and Management

Discussion Paper Series

**No. 1185**

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

by

Satoshi MYOJO and Yuichiro KANAZAWA

October 2007

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

**Abstract**

In this paper, we first give asymptotic theorems for the framework proposed by Berry, Levinsohn, and Pakes (1995) to estimate the system of demand and supply models. We then generalize the idea given by Petrin (2002), which extends the framework by adding new moment conditions when demographically-categorized purchasing pattern data are available. We also gives the asymptotic theorems to this GMM estimator and show that the use of the additional moment conditions allows us to estimate of the demand side parameters more precisely. Finally we run Monte Carlo experiments to evaluate these asymptotic theorems and show that the additional summary information on the consumer's choice contributes the precision of the estimate.

# 1 Introduction

Recent studies, extending the framework proposed by Berry, Levinsohn, and Pakes (1995) (hereafter, BLP (1995)), have been trying 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 2000 and 2001) uses individual's income, age and a dummy variable indicating the individual is a child or not in the utility function. The background behind this is that public sources of information such as CPS and IPUMS are widely available. Those sources can 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.

Some recent studies went further and try combining those demographics with the information on consumer's choice under the "extended" BLP frameworks. For instance, Petrin (2002), referring to Imbens and Lancaster (1994), 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 quoted from CEX automobile supplement in the GMM estimation. Petrin (2002)'s framework presupposes the market information on the population average, which is readily accessible through public sources.

Berry, Levinsohn, and Pakes (2004) (hereafter, BLP (2004)), on the other hand, uses detailed consumer-level CAMIP data provided by General Motors, which include not only individuals' choices but also the choices they would have made had the first choice products not been available to them. In their new framework, the model-predicted covariances between the first- and second-choice vehicle characteristics and household attributes are put close to those calculated from CAMIP data as additional moment conditions in the GMM estimation. Although the method proposed by BLP (2004) 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 admitted in the paper: the CAMIP data "are generally not available to researchers outside of the company" (page 79, line 30).

### Asymptotic Properties of the Estimator in the Previous Studies

The moment conditions used in BLP (1995) are orthogonal conditions of the unobserved product quality $\xi_j$ and the unobserved cost shifter $\omega_j$ with the corresponding instrumental variables $\boldsymbol{z}_j^d$ and $\boldsymbol{z}_j^c$. The moments are obtained by averaging $\xi_j \boldsymbol{z}_j^d$ and $\omega_j \boldsymbol{z}_j^c$ over products. As the number of products $J$ grows large, BLP (1995) claimed that the GMM estimator is consistent and asymptotically normal (CAN).

In BLP (1995), $\xi_j$ are not obtained analytically, but numerically obtained as a solution of $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \boldsymbol{s}^n$. The market shares $\sigma_j$ are approximated by the simulated values with random $R$ draws of consumers. This generates the simulation error in the evaluation of the $\xi_j$ and the $\omega_j$. Furthermore, the sampling error produced by the use of the observed market shares $\boldsymbol{s}^n$, which are typically calculated from random $n$ draws of consumers and thus not equivalent to the underlying true market shares $\boldsymbol{s}^0$, also enters the $\xi_j$ and the $\omega_j$. As a result, what we can actually evaluate for the sample moments include the three distinct randomness: stochastic nature of the product characteristics; randomness generated in the simulation process; and randomness generated in the sampling process.

In BLP (1995), the authors were aware that the number $R$ of simulation draws and the size $n$ of consumer sample must grow at rates faster than the number $J$ of products to establish CAN properties of the GMM estimator. They also acknowledged that, even then, the asymptotic variance-covariance matrix of the resulting estimator consists of three distinct components in responses to these three randomnesses. In the paper, they reported that estimating the random coefficient logit model for demand model would require $n$ and $R$ to grow on the order of $J^3$, and that the precise proofs for the asymptotic theorem of the GMM estimator were still in progress.

In Petrin (2002), the additional moments are the set of functions of the expected value of consumer' demographics given specific product characteristics consumers chose (e.g., expected family size of households that purchased minivans). The evaluation of these new moments are also affected by the aforementioned simulation and sampling errors. This is because he evaluates the conditional expectations of consumer demographics assuming that product characteristics $(\boldsymbol{X}, \boldsymbol{\xi})$ are given, and the $\boldsymbol{\xi}$ includes the simulation and the sampling errors for the reasons elaborated at the beginning of this section.

In addition, the extra market information themselves possibly contain another type of sampling error. This is because the extra market information is typically estimate for the population average demographics obtained from the sample of consumers (e.g., CEX sample) separate from the one from which the observed market share $\boldsymbol{s}^n$ is calculated. This error may also affect on the evaluation of the new moments. In summary each of the three errors (the simulation error, the sampling error in the observed market shares, and the sampling error in the extra information) as well as the stochastic natures of the product characteristics and the consumer demographics are likely to affect the new moment conditions. The estimator proposed by Petrin appears to assume that we are able to control the impacts from these errors. Unfortunately, Petrin (2002) did not provide any asymptotic theorems for the estimator.

Berry, Linton, and Pakes (2004) presents the asymptotic theorem for the random coefficient logit models of demand estimated by the demand side moment conditions and showed the rates of $R$ and $n$ at which they are able to establish CAN properties of the GMM estimator relative to $J$. However, the asymptotic theorem for the GMM estimator with the simultaneous use of the demand and the supply side moment conditions are yet to be known, although they claimed that "it is straightforward to add the pricing equation to the analysis" given in what follows (page 618, line 11).

BLP (2004) claimed that if the number of consumers sampled in the CAMIP data grow faster enough when the number of products grows large, the estimator with their new framework is also consistent and asymptotically normal. In the study, the authors take into account the simulation errors and the CAMIP data's sampling error in the calculation of the asymptotic variance of the estimator. They justifiably neglected the sampling error in the observed market share since the precise market share data are readily available in the U.S. automobile market. To objectively and precisely estimate the U.S. consumers' automobile preferences using unbiased publicly-available data, we thought it best to use the framework considering both the demand and supply side with additional demographics information. BLP (2004), as good as they may be, fell short in this regard because they only consider the demand side and they use the CAMIP data generally not available outside of the GM. We therefore choose to proceed following Petrin (2002)'s footsteps.

In this paper, we provide general conditions under which the extension of the GMM estimator originally proposed by Petrin (2002) has CAN properties. The assumptions we make use of for the demand side specification and the notations of the proof generally follow the asymptotic theorems given in Berry, Linton, and Pakes (2004), but we considerably extend their theorem in three directions: first we clearly state that the asymptotics we set forth is not conditioned on the product characteristics, which we will see is stochastic; Second, we incorporate the supply side as well as the demand side; Third we include additional demographics moment conditions.

Except BLP (1995) and BLP (2004), studies in marketing and industrial organization appeared to ignore the effects of the errors generated by the simulation and the sampling processes and thus did not adjust the variance-covariance matrix of the estimator when employing BLP framework (See in Table 1). As for the simulation process, this is probably due to a computa-

Table 1: The Consideration of Errors in the Past Studies

| | Demand Side Moments | | Supply Side Moments | | Additional Moments | | |
|---|---|---|---|---|---|---|---|
| | Simulation Error | Sampling Error | Simulation Error | Sampling Error | Simulation Error | Sampling Error | Extra Information Error |
| BLP (1995) | ○ | ○, but negligible | ○ | ○ | — | — | — |
| Sudhir (2000) | × | × | × | × | — | — | — |
| Nevo (2001) | × | × | × | × | — | — | — |
| Petrin (2002) | × | × | × | × | × | × | × |
| BLP (2004) | ○ | ○, but negligible | — | — | ○ | ○, but negligible | ○ |

The symbol ○ (×) indicates the error was (not) took into account in the evaluation of the moment.
"—" means that the study did not use the corresponding moment conditions.

tional burden incurred to evaluate the simulation error. To numerically isolate the magnitude of the simulation error, for instance, researchers have to repeat the estimation algorithm with many independent sets of $R$ simulation draws of consumers with the observed market share fixed.

## 2 Background on the BLP (1995)'s Framework

### 2.1 Demand Side Model

The discrete choice differentiated product demand systems formulates that the utility of consumer $i$ for product $j$ is a function of parameters, $\boldsymbol{\theta}_d$, observed product characteristics, $\boldsymbol{x}_j$, unobserved (by the econometricians) product characteristics, $\xi_j$, and random consumer tastes, $\nu_{ij}$. Given the product characteristics $(\boldsymbol{x}_j, \xi_j)$ for the all $(J)$ products marketed, the consumer either chooses to buy one of the products or not to buy any product, in which case we say the consumer chooses the "outside" good. Each consumer makes the choice that maximizes his/her utility. Different consumers may make different choices because of their tastes, and their tastes follow the distribution denoted by $P^0$.

Although the most product characteristics are not correlated with the unobserved product characteristics $\xi_j \in \Re$, $j = 1, \ldots, J$, some of them (e.g., price) are likely to be correlated with the $\xi_j$.[1] We denote the vector of observed product characteristics by $\boldsymbol{x}_j = (\boldsymbol{x}'_{1j}, \boldsymbol{x}'_{2j})'$ where $\boldsymbol{x}_{1j} \in \Re^{K_1}$ are the ones that are not correlated with the $\xi_j$ in the sense that

$$\mathrm{E}_{\xi|\mathrm{x}_1}[\xi_j|\boldsymbol{x}_{1j}] = 0 \qquad \text{and} \qquad \sup_{1\le j\le J} \mathrm{E}_{\xi|\mathrm{x}_1}[\xi_j^2|\boldsymbol{x}_{1j}] < \infty \tag{1}$$

with probability one. Product characteristics in the $\boldsymbol{x}_{2j} \in \Re^{K_2}$ are correlated with the $\xi_j$. The set of observed product characteristics for all the products is denoted by $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J)'$.

In this framework, we assume the set of exogenous product characteristics $(\boldsymbol{x}_{1j}, \xi_j), j = 1, \ldots, J$ are random sample of product characteristics of size $J$ from the underlying population of product characteristics. Thus, $(\boldsymbol{x}_{1j}, \xi_j)$ are assumed to be independent across $j$, while $\boldsymbol{x}_{2j}$ are in general not independent across $j$ since they are endogenously determined in the market as functions of product characteristics of the other products as well as its own product.

The demand model determines the purchase probability of a consumer as a function of his/her attributes and the product characteristics in the market. A distributional assumption on the consumers' unobservable heterogeneity is made to obtain expected purchase probability conditional on product characteristics and consumer attributes. The conditional purchase probability $\sigma_{ij}$ of product $j$ is a map from consumer $i$'s attributes $\boldsymbol{\nu}_i \in \Re^v$, a demand side parameter vector $\boldsymbol{\theta}_d \in \Theta_d$, and the set of characteristics of all products $(\boldsymbol{X}, \boldsymbol{\xi})$, and is thus denoted as $\sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_i; \boldsymbol{\theta}_d)$. BLP (1995)'s framework generates the vector of market shares, $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, by aggregating over the individual choice probability with the distribution $P$ of the consumer attributes $\boldsymbol{\nu}_i$ as

$$\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \int \sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_i; \boldsymbol{\theta}_d) dP(\boldsymbol{\nu}_i) \tag{2}$$

where $P$ is typically the empirical distribution of the attributes 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 $\boldsymbol{X}$ and $\boldsymbol{\xi}$. If we evaluate equation (2) at $(\boldsymbol{\theta}_d^0, P^0)$, where $\boldsymbol{\theta}_d^0$ is the true

[1] The unobserved product characteristics $\xi_j$ are product characteristics difficult to measure or observe by researchers. They typically include consumers' perception on style, brand equity, effect of promotional activity, and service at point-of-sale.

value of the parameters, it will give the "conditionally true" market shares $\boldsymbol{s}^0$ given the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi})$ in the population, i.e.,

$$\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d^0, P^0) \equiv \boldsymbol{s}^0. \tag{3}$$

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

$$\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \boldsymbol{0} \tag{4}$$

for every $(\boldsymbol{X}, \boldsymbol{\theta}_d, \boldsymbol{s}, P) \in \mathcal{X} \times \Theta_d \times \mathcal{S}_J \times \mathcal{P}$, where $\mathcal{X}$ is a space for the product characteristics $\boldsymbol{X}$, and $\mathcal{P}$ is a family of probability measures. If we solve the identity in (3) with respect to $\boldsymbol{\xi}$ under the conditions that guarantee the uniqueness of the $\boldsymbol{\xi}$ in (4), we are able to retrieve the original $\xi_j$ which we assume are independent across $j$. However, if we solve (4) at any $(\boldsymbol{\theta}_d, \boldsymbol{s}, P) \neq (\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$, the resulting $\xi_j(\boldsymbol{X}, \boldsymbol{\theta}_d, \boldsymbol{s}, P)$ are not equivalent to the true value of $\xi_j$. For this $\xi_j(\boldsymbol{X}, \boldsymbol{\theta}_d, \boldsymbol{s}, P)$, the independence assumption is violated because the two factors for $\xi_j$—the market share $s_j$ and the endogenous product characteristics $\boldsymbol{x}_{2j}$ 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.

### 2.2 Supply Side Model

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 $(m = 1, \ldots, F)$ are modelled as maximizers of profit from the combination of products they are producing. Specifically, supplier $m$ maximizes the following profit function.

$$\mathrm{PR}_m = \sum_{j \in \mathcal{J}_m} (p_j - c_j) M_s \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \quad m = 1, \ldots, F, \tag{5}$$

where $\mathcal{J}_m$ denotes the set of products provided by supplier $m$, and $p_j$ and $c_j$ are respectively price and marginal cost of product $j$, and $M_s$ denotes the potential market size. By assuming the Bertrand-Nash pricing for supplier's strategy, the first order condition in terms of $p_j$ is given as

$$\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) + \sum_{l \in \mathcal{J}_m} (p_l - c_l) \partial \sigma_l(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial p_j = 0 \quad \text{for } j \in \mathcal{J}_m. \tag{6}$$

This equation can be expressed in matrix form

$$\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) + \boldsymbol{\Delta}(\boldsymbol{p} - \boldsymbol{c}) = \boldsymbol{0} \tag{7}$$

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

$$\Delta_{jk} = \begin{cases} \partial \sigma_k(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P) / \partial p_j, & \text{if the products } j \text{ and } k \text{ are} \\ & \text{produced by the same firm;} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Solving (7) with respect to $\boldsymbol{c}$ gives

$$\boldsymbol{c} = \boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \tag{9}$$

where

$$\boldsymbol{m}_g \equiv -\boldsymbol{\Delta}^{-1}\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P) \tag{10}$$

represents the vector of the profit margin for all the products on the market. We suppress $\boldsymbol{X}$ in the expression of $\boldsymbol{m}_g$ for notational simplicity.

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

$$g(c_j) = \mathbf{w}_j'\boldsymbol{\theta}_c + \omega_j \tag{11}$$

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 the exogenous ones $\mathbf{w}_{1j} \in \Re^{L_1}$ as well as endogenous ones $\mathbf{w}_{2j} \in \Re^{L_2}$, and thus we write $\mathbf{w}_j = (\mathbf{w}_{1j}', \mathbf{w}_{2j}')'$ and $\boldsymbol{W} = (\mathbf{w}_1,\ldots,\mathbf{w}_J)'$. The exogenous cost shifters include not only the cost variables determined outside the market under consideration (e.g. crude oil price), but also the product design characteristics that suppliers can not 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. The unobserved cost shifters $\omega_j$ are assumed to be uncorrelated with the exogenous cost shifters $\mathbf{w}_{1j}$, and then satisfy the condition that

$$\mathrm{E}_{\omega|\mathrm{w}_1}[\omega_j|\mathbf{w}_{1j}] = 0, \quad \text{and} \quad \sup_{1\le j\le J} \mathrm{E}_{\omega|\mathrm{w}_1}[\omega_j^2|\mathbf{w}_{1j}] < \infty \tag{12}$$

with probability one.

As in the formulation of $(\boldsymbol{x}_{1j},\xi_j), j = 1,\ldots,J$, on the demand side, we assume the set of exogenous cost shifters $(\mathbf{w}_{1j},\omega_j), j = 1,\ldots,J$ are random sample of cost shifters of size $J$ 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 with respect to $j$ as they are determined in the market as functions of cost shifters of other products.

Substituting (9) for (11) and evaluating $\xi_j$ at $\xi_j(\boldsymbol{X},\boldsymbol{\theta}_d,\boldsymbol{s},P), j = 1,\ldots,J$, gives the reduced form of the unobserved cost shifters $\omega_j$.

$$\omega_j(\boldsymbol{\theta},\boldsymbol{s},P) = g(p_j - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{X},\boldsymbol{\theta}_d,\boldsymbol{s},P),\boldsymbol{\theta}_d,P)) - \mathbf{w}_j'\boldsymbol{\theta}_c \tag{13}$$

where the parameter vector $\boldsymbol{\theta}$ 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 $\xi_j$, but by those of the other products on the market, these $\omega_j$ are in general dependent across $j$ when $(\boldsymbol{\theta},\boldsymbol{s},P) \neq (\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)$. However, when (13) is evaluated at $(\boldsymbol{\theta},\boldsymbol{s},P) = (\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)$, we are able to recover the original $\omega_j, j = 1,\ldots,J$, which are independent across $j$. Define $\boldsymbol{g}(\boldsymbol{x}) \equiv (g(x_1),\ldots,g(x_J))$ and rewrite (13) in vector form

$$\boldsymbol{\omega}(\boldsymbol{\theta},\boldsymbol{s},P) = \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{X},\boldsymbol{\theta}_d,\boldsymbol{s},P),\boldsymbol{\theta}_d,P)) - \boldsymbol{W}\boldsymbol{\theta}_c. \tag{14}$$

### 2.3 GMM Estimation

Zero moment restrictions between unobserved characteristics $(\xi_j,\omega_j)$ and exogenous instrumental variables $(\boldsymbol{z}_j^d,\boldsymbol{z}_j^c)$ will be imposed to estimate $\boldsymbol{\theta}$ by the generalized method of moments (henceforth, GMM).

Let us define the $J \times M_1$ demand side instrument matrix $\boldsymbol{Z}_d = (\boldsymbol{z}_1^d,\ldots,\boldsymbol{z}_J^d)'$ whose components $\boldsymbol{z}_j^d$ can be written as $\boldsymbol{z}_j^d(\boldsymbol{x}_{11},\ldots,\boldsymbol{x}_{1J}) \in \Re^{M_1}$, where $\boldsymbol{z}_j^d(\cdot) : \Re^{K_1\times J} \to \Re^{M_1}$ for $j = 1,\ldots,J$. It should be noted that the demand side instruments $\boldsymbol{z}_j^d$ for product $j$ are assumed to be a function of the exogenous characteristics not only of its own, but of the other products $(\boldsymbol{x}_{11},\ldots,\boldsymbol{x}_{1J})$

in the market. This is because the instruments, by definition, must correlate with the product characteristics $\boldsymbol{x}_{2j}$, and this endogenous variables $\boldsymbol{x}_{2j}$ (e.g. price) are determined by both its own and its competitors' product characteristics as we mentioned above.

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

Considering the stochastic nature of product characteristics $\boldsymbol{X}_1$ as well as of $\boldsymbol{\xi}$, we set forth the demand side restriction as

$$\mathrm{E}_{\mathrm{x}_1,\xi}\left[\boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\right] = \boldsymbol{0} \tag{15}$$

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

$$\mathrm{E}_{\mathrm{w}_1,\omega}\left[\boldsymbol{z}_j^c \omega(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)\right] = \boldsymbol{0} \tag{16}$$

at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. Hereafter, we suppress the dependence on $\boldsymbol{X}$ and $\boldsymbol{W}$ in the expression of $\xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ and $\omega_j(\boldsymbol{\theta}, \boldsymbol{s}, P)$ respectively for notational simplicity. We suppose that the number of restrictions $(M_1 + M_2)$ is equal to or greater than the number $K$ of parameters in $\boldsymbol{\theta}$.

Now let us form the average of $\boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ and $\boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ as

$$\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) \equiv J^{-1} \sum_{j=1}^{J} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \tag{17}$$

$$\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \equiv J^{-1} \sum_{j=1}^{J} \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0). \tag{18}$$

The GMM estimator for $\boldsymbol{\theta}^0$ minimizes the sum of norms of $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ and $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$, that is, it minimizes the norm of

$$\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) = \begin{pmatrix} \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) \\ \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \end{pmatrix}. \tag{19}$$

To derive the asymptotic properties of this estimator, we have to make assumption for how $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ behaves as the number of products $J$ tends to infinity.

We know that the $(\xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0))$ are dependent across $j$ at $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$. Moreover, since $\boldsymbol{z}_j^d$ and $\boldsymbol{z}_j^c$ are respectively functions of the exogenous characteristics $\boldsymbol{X}_1$ and the exogenous cost shifters $\boldsymbol{W}_1$ of all the products, they are also dependent across $j$. This implies that the uniform convergence of the objective function $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$ to $||\,\mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)]||$ over all possible $\boldsymbol{\theta} \in \Theta$ is not guaranteed.[2] As a result, the standard consistency proofs of the GMM estimator that assume uniform convergence of the objective function are not applicable. Instead, we set the condition which bounds $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$ away from zero for all $\boldsymbol{\theta}$ outside of a neighborhood of $\boldsymbol{\theta}^0$ as Berry, Linton and Pakes (2004) did. This condition enables us to use Theorem 3.1 in Pakes and Pollard (1989) to derive the consistency.

If we can further assume that $J^{\frac{1}{2}}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)]]$ converges to $J^{\frac{1}{2}}[\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) - \mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)]]$ in probability as the stochastic $\boldsymbol{\theta}$ converges in probability to $\boldsymbol{\theta}^0$, that is, the process $J^{\frac{1}{2}}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)]]$ is stochastically equicontinuous at $\boldsymbol{\theta}^0$, and that $J^{\frac{1}{2}}\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$ converges weakly to the normal distribution, the GMM estimator for $\boldsymbol{\theta}^0$ can be shown to be asymptotically normal by Theorem 3.3 in Pakes and Pollard (1989).

[2] The expectation symbol $\mathrm{E}[\cdot]$ here means that taking expectation over $(\boldsymbol{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$.

We have two separate problems in the evaluation of $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$. Although $P^0$ is so far assumed to be known, we typically will not be able to calculate $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ analytically and will have to approximate it by a simulator, say $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$, where $P^R$ is the empirical measure of some i.i.d. sample $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_R$ from the underlying distribution $P^0$. Simulated market shares are then given by

$$\begin{aligned} &\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \\ &= \int \sigma_{ij}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_i; \boldsymbol{\theta}_d) dP^R(\boldsymbol{\nu}_i) \equiv \frac{1}{R}\sum_{r=1}^{R} \sigma_{rj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d). \end{aligned} \tag{20}$$

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

$$s_j^n = \frac{1}{n}\sum_{i=1}^{n} 1(C_i = j), \tag{21}$$

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

We substitute $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ given as a solution of $\boldsymbol{s}^n - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \boldsymbol{0}$ for (17) to obtain

$$\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) = J^{-1}\sum_{j=1}^{J} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R). \tag{22}$$

Furthermore, substituting $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) = (\omega_1(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R), \ldots, \omega_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R))'$ obtained from evaluating (13) at $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $P = P^R$ for (18) gives

$$\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) = J^{-1}\sum_{j=1}^{J} \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R). \tag{23}$$

The actual objective function is thus $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||$. Consequently, our estimator of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, satisfies

$$||\boldsymbol{G}_J(\hat{\boldsymbol{\theta}}, \boldsymbol{s}^n, P^R)|| = \inf_{\theta \in \Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||. \tag{24}$$

In the expression of $||\boldsymbol{G}_J(\hat{\boldsymbol{\theta}}, \boldsymbol{s}^n, P^R)||$, there exist three distinct randomness: one generated from the draws of the product characteristics $(\boldsymbol{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$, one generated from the sampling process of consumers for $\boldsymbol{s}^n$, and one generated from the empirical distribution $P^R$. The impact of these randomness on the estimate of $\boldsymbol{\theta}$ will be decided by the relative size of the sample—$J$, $n$, and $R$. Unless $n$ and $R$ are much larger than $J$, the impact from the sampling error and the simulation error may not be negligible. 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 $\boldsymbol{s}^n$ and the true market share $\boldsymbol{s}^0$. Specifically, its component $\epsilon_j^n$ for the product $j$ is

$$\begin{aligned} \epsilon_j^n &\equiv s_j^n - s_j^0 = \frac{1}{n}\sum_{i=1}^{n} 1(C_i = j) - s_j^0 = \frac{1}{n}\sum_{i=1}^{n} \left\{1(C_i = j) - s_j^0\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \epsilon_{ji} \end{aligned} \tag{25}$$

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

Note that from (4), for any $\boldsymbol{\theta}_d \in \Theta_d$, the unique solutions $\boldsymbol{\xi}$ for

$$\boldsymbol{s}^n - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \mathbf{0} \quad \text{and} \quad \boldsymbol{s}^0 - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) = \mathbf{0}$$

are written as $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ respectively. In other words, substituting these $\boldsymbol{\xi}$s back into $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ and $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ retrieves $\boldsymbol{s}^n$ and $\boldsymbol{s}^0$ respectively. Therefore for any $\boldsymbol{\theta}_d \in \Theta_d$

$$\boldsymbol{s}^n = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) \tag{26}$$

and

$$\boldsymbol{s}^0 = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0). \tag{27}$$

If we evaluate (4) with the observed market share $\boldsymbol{s}^n$ and the underlying population $P^0$ of consumers, the resulting $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^0)$ satisfies the equation

$$\boldsymbol{s}^n = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^0), \boldsymbol{\theta}_d, P^0) \tag{28}$$

for all $\boldsymbol{\theta}_d \in \Theta_d$. Furthermore, for all $\boldsymbol{\theta}_d \in \Theta_d$, the $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$ which is obtained by evaluating (4) with the true market share $\boldsymbol{s}^0$ and the empirical population $P^R$ of consumers satisfies

$$\boldsymbol{s}^0 = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R). \tag{29}$$

The simulation process generates the simulation error $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)$, which is for any $\boldsymbol{\theta}_d$ a difference between the simulated market shares in (20) obtained from a sample of $R$ consumers whose distribution follows the empirical distribution $P^R$ and those obtained from the population distribution $P^0$ of all the consumers. That is, the simulation error $\epsilon_j^R$ for product $j$ with sample of $R$ consumers is

$$\epsilon_j^R(\boldsymbol{\theta}_d) \equiv \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$$

for $j = 1, \ldots, J$. From (27), $\epsilon_j^R(\boldsymbol{\theta}_d)$ can be rewritten as

$$\begin{aligned} \epsilon_j^R(\boldsymbol{\theta}_d) &= \frac{1}{R} \sum_{r=1}^{R} \sigma_{rj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) - s_j^0 \\ &= \frac{1}{R} \sum_{r=1}^{R} \left\{ \sigma_{rj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) - s_j^0 \right\} \\ &= \frac{1}{R} \sum_{r=1}^{R} \epsilon_{jr}^*(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d) \end{aligned} \tag{30}$$

where $\epsilon_{jr}^*(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d) = \sigma_{rj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) - \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ are by definition independent across $r$ conditional on $(\boldsymbol{X}, \boldsymbol{\xi})$.

## 2.5 Metrics, Neighborhoods, and Notations

We will work with the product space $\Theta \times \mathcal{S}_J \times \mathcal{P}$. The parameter space $\Theta$ is a compact subset of $\Re^K$ and we use the Euclidean metric on $\Theta$, $\rho_E(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = ||\boldsymbol{\theta} - \boldsymbol{\theta}^*||$. The space for the market share vector $\boldsymbol{s}$ is $J+1$ dimensional unit simplex $\mathcal{S}_J$,

$$\mathcal{S}_J = \left\{ (s_0, \ldots, s_J)' \middle| \quad 0 < s_j < 1 \text{ for } j = 0, \ldots, J, \text{ and } \sum_{j=0}^{J} s_j = 1 \right\}.$$

Since the market share $s_j$ generally shrinks as the number $J$ of the products on the market increases, we need to make sure the speed at which the $s_j$ becoming close to the true share $s_j^0$ ought to be faster than the speed at which $s_j^0$ converges to zero. To ascertain this, we need to use the metric $\rho_{s^0}$ on $\mathcal{S}_J$

$$\rho_{s^0}(\boldsymbol{s}, \boldsymbol{s}^*) = \max_{0 \le j \le J} \left| \frac{s_j - s_j^*}{s_j^0} \right|.$$

The $\mathcal{P}$ is the set of probability measures of consumer's attributes. The $L_\infty$ 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 $\Re^v$, where $v$ is the dimension of the consumer attributes in the purchasing probability. This metric will be used to measure the distance between the empirical distribution $P^R$ and the underlying distribution $P^0$ of consumer's attributes.

Since the dimension of the unobserved product characteristics $\boldsymbol{\xi}$ increases, element by element convergence of $\boldsymbol{\xi}$ to $\boldsymbol{\xi}^*$ does not automatically guarantee that $||\boldsymbol{\xi} - \boldsymbol{\xi}^*|| = o_p(1)$. In the proof, all we need is the convergence of the unobserved product characteristics $\boldsymbol{\xi}$ as 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 of course allow 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 $\delta$ neighborhoods for $\boldsymbol{\theta}^0, P^0$, and $\boldsymbol{s}^0$ respectively as $\mathcal{N}_{\theta^0}(\delta) = \{\boldsymbol{\theta} : \rho_E(\boldsymbol{\theta}, \boldsymbol{\theta}^0) \le \delta\}$, $\mathcal{N}_{P^0}(\delta) = \{P : \rho_P(P, P^0) \le \delta\}$, and $\mathcal{N}_{s^0}(\delta) = \{\boldsymbol{s} : \rho_s(\boldsymbol{s}, \boldsymbol{s}^0) \le \delta\}$. Also for each $\boldsymbol{\theta}$, the $\delta$ neighborhood of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ is defined by $\mathcal{N}_{\xi^0}(\boldsymbol{\theta}; \delta) = \{\boldsymbol{\xi} : \rho_\xi(\boldsymbol{\xi}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)) \le \delta\}$.

The notation we use for the Euclidean norm of any $m \times n$ matrix $\boldsymbol{A}$ is $||\boldsymbol{A}|| = \{\text{tr}(\boldsymbol{A}'\boldsymbol{A})\}^{1/2}$. 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, the symbols should be interpreted element by element. If $\boldsymbol{x}$ is a $k \times 1$ vector, diag$[\boldsymbol{x}]$ denotes a $k \times k$ diagonal matrix with the element of $\boldsymbol{x}$ along its principle diagonal.

# 3 Asymptotic Theory for BLP (1995)

## 3.1 Consistency

In this section, we derive the asymptotic theorems for the BLP framework. Our proofs are different from the one in Berry, Linton, and Pakes (2004) in two ways. First, in Berry, Linton, and Pakes (2004), the asymptotic theorems appear to be established under the condition that $(\boldsymbol{X}, \boldsymbol{\xi})$ is given while the dimension $J$ of the product characteristics grows infinitely. Our proofs for the theorems do not condition on $(\boldsymbol{X}, \boldsymbol{\xi})$. Second, we derive the theorem not only for the demand side model but for the system of demand and supply models. We first describe assumptions needed to obtain the consistency of the estimator.

In Assumption A1(a), we assume that the observed market share $s_j^n$ for product $j$ is the Bernoulli random variables averaged over the $n$ sampled consumers ($i = 1, \ldots, n$). Assumption A1(b) guarantees that the simulation error $\epsilon_{jr}^*$ defined in (30) relative to the number $R$ of the simulation draws is of the same order as the sampling error $\epsilon_{ji}$ relative to the number $n$ of the sample. These are used to control the magnitudes of the respective errors. Note that in A1(a), $\boldsymbol{s}^n$ and $\boldsymbol{s}^0$ are the result of consumer behavior, and the consumers are assumed to be able to observe the true "unobserved" product characteristics, $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$. As a result, we can condition on $\boldsymbol{X}$ and on $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$, but not on a general $\boldsymbol{\xi}$ when evaluating the moments of the difference $\boldsymbol{s}^n - \boldsymbol{s}^0$. On the other hand, in A1(b), $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ and $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$, both of which are model-calculated shares, are just the device researchers use and they are not able to observe the unobserved product characteristics, true or otherwise. As a result, we need to treat $\boldsymbol{\xi}$ as unobserved and unknown, and we need to condition on the unobserved and unknown $\boldsymbol{\xi}$ along with on the $\boldsymbol{X}$.

Assumption A2 is regularity condition for the share function. In A2(a), we first assume that the model-calculated market share $\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ for product $j$ will not abruptly change as the unobserved product quality $\xi_k$ for product $k$ changes. Moreover the $\boldsymbol{H}$ in (36) being invertible means one can quantify the change in unobserved product quality $\partial\xi_j$ for product $j (j = 1, \ldots, J)$ associated with the change in the model-calculated market share $\partial\sigma_k$ for product $k (k = 1, \ldots, J)$. Assumption A2(b) stipulates how the model-calculated market share $\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ 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. The set of assumptions A2(a) and (b) is a sufficient condition for the existence of a unique solution $\boldsymbol{\xi}$ to (4) for every $(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ (See appendix in Berry (1994) for detail).

It looks as if we need a similar setup for the supply side unobserved cost shifter $\omega_j$ relative to the model-calculated market share $\sigma_k$. This is not so, however, because as clearly seen in (13), the $\omega_j(\boldsymbol{\theta}, \boldsymbol{s}, P)$ can be obtained as a function of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ aside from the observed $(p_j, \mathbf{w}_j)$ and the parameters $(\boldsymbol{\theta}_d, \boldsymbol{\theta}_c)$ once we decide to choose which $(\boldsymbol{s}, P)$ to evaluate, enabling the characteristics of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ to transmit to $\omega_j(\boldsymbol{\theta}, \boldsymbol{s}, P)$. Therefore what we need is the fact that there exists a profit margin $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)$ in (10) that is at least locally smooth with respect to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ along with smoothness in $g(\cdot)$. Assumption A2(c) guarantees the existence of $\boldsymbol{\Delta}^{-1}$, which in turn guarantees the existence of $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)$ in (10). We replace local smoothness of $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)$ relative to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ with the assumption A7. We will come back to this when explaining 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.

In the situation we are considering here, the number $J$ of the products in the market increases. This means that the "conditionally" true market shares $\boldsymbol{s}^0$ and also the theoretical market shares $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ generally approach to zero as $J$ grows large. Assumptions A3(a),(b) guarantee that $\boldsymbol{s}^n$ and $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ converge to $\boldsymbol{s}^0$ and $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ faster respectively than the speed at which $\boldsymbol{s}^0$ and $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ converge to zero.

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

Assumption A5 is a condition that bounds $||\boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$ away from $||\boldsymbol{G}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)||$ (which converges to zero in probability) over $\boldsymbol{\theta}$ outside of a neighborhood of $\boldsymbol{\theta}^0$. This condition corresponds to condition (iii) in Theorem 3.1 of Pakes and Pollard (1989).

For all $\boldsymbol{\theta}_d$, the value of $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ that satisfies the equation $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) = \boldsymbol{s}^0$ is assumed unique. Since the sum of the market shares including that of the outside good–$s_0^0$–is fixed to be one, this $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ also satisfies

$$\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)/\sigma_0(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) = \boldsymbol{s}^0/s_0^0.$$

Define a function $\boldsymbol{\tau}_J(\cdot) : \Re^J \to \Re^J$ such that $\boldsymbol{\tau}_J(\boldsymbol{s}) = (\log(s_1/s_0), \ldots, \log(s_J/s_0))$. Then, from (27), the relation is equivalent to saying that

$$\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) \quad = \quad \boldsymbol{\tau}_J(\boldsymbol{s}^0) = \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0))$$

at $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ for all $\boldsymbol{\theta}_d$. Assumption A6 guarantees that any $\boldsymbol{\xi}$ outside the $\delta$ neighborhood of the $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ cannot make $\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0))$ close to $\boldsymbol{\tau}_J(\boldsymbol{s}^0)$ 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 allow for the fact that the dimension of the model-calculated market share $\boldsymbol{\sigma}$ increases. The functional treatment $\boldsymbol{\tau}_J$ is due to making this assumption easier to verify for logit-like demand models.

In assumption A7, we assume the profit margins $J^{-\frac{1}{2}}\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)$ have stochastically equicontinuity-like characteristics in $(\boldsymbol{\xi}, P)$ at $(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), P^0)$ for any $\boldsymbol{\theta}_d \in \Theta_d$. As we see in the proof, we show that $\Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) \notin \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d, \delta)] \to 0$ and $\Pr[P^R \notin \mathcal{N}_{P^0}(\delta)] \to 0$ 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 $\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ and $\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ is close uniformly in $\boldsymbol{\theta}_d \in \Theta_d$. We should note that assumption A7 is not stochastic equicontinuity as defined because the dimension of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ grows large, though $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ converges to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ in probability in averaged Euclidean metric.

One more comment on the behavior of the dimension increasing $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$. 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 $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$ at $J = 500$ must be equal to the all 100 elements of $\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$ 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 $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ at $J = 500$ are equal to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ at $J = 100$.

**Assumption A1** (a) *Given the set of product characteristics* $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$*, the difference* $\boldsymbol{s}^n - \boldsymbol{s}^0$ *between the observed market share* $\boldsymbol{s}^n$ *and the "conditionally" true market share* $\boldsymbol{s}^0$ *have conditional mean*

$$\begin{aligned} & \mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{\epsilon}^n | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \\ &= \mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{s}^n - \boldsymbol{s}^0 | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] = \boldsymbol{0} \end{aligned} \tag{31}$$

*with the conditional variance-covariance matrix*

$$\begin{aligned} \boldsymbol{V}_2 &= \mathrm{E}_{\epsilon|\mathrm{x},\xi}[(\boldsymbol{s}^n - \boldsymbol{s}^0)(\boldsymbol{s}^n - \boldsymbol{s}^0)' | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \\ &= \frac{1}{n}\left(\mathrm{diag}[\boldsymbol{s}^0] - \boldsymbol{s}^0\boldsymbol{s}^{0\prime}\right). \end{aligned} \tag{32}$$

(b) *For each* $\boldsymbol{\theta}_d$*, given the set of product characteristics* $(\boldsymbol{X}, \boldsymbol{\xi})$*, the difference* $\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)$ *have conditional mean*

$$\mathrm{E}_{\epsilon^*|\mathrm{x},\xi}[\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) | \boldsymbol{X}, \boldsymbol{\xi}] = \boldsymbol{0} \tag{33}$$

*with the conditional variance-covariance matrix*

$$\begin{aligned} \boldsymbol{V}_3 &= \mathrm{E}_{\epsilon^*|\mathrm{x},\xi}\Big[\left\{\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)\right\} \\ &\quad \times \left\{\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)\right\}' \Big| \boldsymbol{X}, \boldsymbol{\xi}\Big] \end{aligned} \tag{34}$$

*whose order of magnitude relative to* $R$ *is the same as that of* $\boldsymbol{V}_2$ *relative to* $n$ *or,*

$$R \cdot O(\boldsymbol{V}_3) = n \cdot O(\boldsymbol{V}_2). \tag{35}$$

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

$$\boldsymbol{H}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \partial\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\boldsymbol{\xi}' \tag{36}$$

*is invertible for all* $J$*.*
(b) *For every* $(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$*,* $\partial\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial\xi_j > 0$ *for* $j = 1, \ldots, J$*,*

*and* $\partial \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial \xi_k < 0$ *for* $k, j = 1, \ldots, J, k \neq j$.
(c) *For every finite* $J$, *for all* $\boldsymbol{\theta}_d \in \Theta_d$, *and for all* $P$ *in a neighborhood of* $P^0$, $\partial \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial p_k$ *exists for* $j, k = 1, \ldots, J$, *and the matrix* $\boldsymbol{\Delta}$ *whose* $(j, k)$ *element is defined in* (8) *is invertible for all* $J$ *and continuously differentiable both in* $\boldsymbol{\xi}$ *and* $\boldsymbol{\theta}_d$.

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

$$\text{(a)} \quad \rho_{s^0}(\boldsymbol{s}^n, \boldsymbol{s}^0) = \max_{0 \leq j \leq J} \left| \frac{s_j^n - s_j^0}{s_j^0} \right| = o_p(1). \tag{37}$$

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

$$\begin{aligned} \text{(b)} \quad & \rho_{\sigma(X,\xi,\theta_d,P^0)}(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R), \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) \\ = & \max_{0 \leq j \leq J} \left| \frac{\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) - \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}{\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)} \right| = o_p(1). \end{aligned} \tag{38}$$

*for any* $\boldsymbol{\xi}$ *and* $\boldsymbol{\theta}_d \in \Theta$.

**Assumption A4** (a) *The demand side instrumental variables are such that the matrix* $\boldsymbol{Z}_d'\boldsymbol{Z}_d/J$ *is stochastically bounded, i.e., for all* $\epsilon > 0$ *there exists an* $M_\epsilon$ *such that* $\Pr[||\boldsymbol{Z}_d'\boldsymbol{Z}_d/J|| > M_\epsilon] < \epsilon$. *Moreover, we suppose* $||\boldsymbol{Z}_d'\boldsymbol{Z}_d/J||$ *is uniformly integrable in* $J$, *i.e.,*

$$\lim_{\alpha \to \infty} \sup_J \int ||\boldsymbol{Z}_d'\boldsymbol{Z}_d/J|| \{||\boldsymbol{Z}_d'\boldsymbol{Z}_d/J|| > \alpha\} dP_{\mathrm{x}_1}(\boldsymbol{X}_1) = 0$$

*where* $P_{\mathrm{x}_1}(\cdot)$ *is the joint distribution of* $\boldsymbol{X}_1$.

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

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

$$\lim_{J \to \infty} \Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq C(\delta) \right] = 1. \tag{39}$$

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

$$\begin{aligned} \lim_{J \to \infty} \Pr\Bigg[ & \inf_{\theta_d \in \Theta_d} \inf_{\xi \notin \mathcal{N}_{\xi^0}(\theta_d; \delta)} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)) \\ & -\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0))|| > C(\delta) \Bigg] = 1. \end{aligned} \tag{40}$$

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

$$\begin{aligned} \lim_{J \to \infty} \Pr\Bigg[ & \sup_{(\xi, P) \in \mathcal{N}_{\xi^0}(\theta_d; \delta) \times \mathcal{N}_{P^0}(\delta)} J^{-\frac{1}{2}} ||\boldsymbol{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \\ & -\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)|| > \delta \Bigg] = 0. \end{aligned} \tag{41}$$

**Theorem 1 (Consistency of $\hat{\boldsymbol{\theta}}$)** *Suppose that A1–A7 hold for some* $n(J), R(J) \to \infty$. *Then,* $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^0$.

### 3.2 Asymptotic Normality

We next establish the asymptotic normality of $\hat{\boldsymbol{\theta}}$. Throughout we assume that $\hat{\boldsymbol{\theta}}$ is consistent with respect to $\boldsymbol{\theta}^0$, or assumptions A1–A7 to hold. To derive the asymptotic distribution, we first decompose the unobserved quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ into three random terms—the unobserved quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$, the term generated from the sampling error $\boldsymbol{\epsilon}^n$, and the term generated from the simulation error $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)$ and substitute this relationship for $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ in $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$. We decompose the unobserved cost shifter $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ into three terms likewise and substitute this relationship for $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ in $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$.

**Demand Side Derivation**

Write

$$\begin{aligned}\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) &= \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) + \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\right\} \\ &\quad + \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\right\}. \end{aligned} \tag{42}$$

For fixed $\boldsymbol{\theta}_d$, we use Taylor series approximation to the second and the third terms in (42). Specifically, by the mean value theorem

$$\begin{aligned}\mathbf{0} &= \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{s}^n \\ &= \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{s}^n \\ &\quad + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\right\} \\ &= \boldsymbol{s}^0 - \boldsymbol{s}^n + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\right\} \\ &= -\boldsymbol{\epsilon}^n + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\right\}\end{aligned}$$

where $\bar{\boldsymbol{\xi}}$ is $J \times 1$ vector of the values between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$. Notice that we write

$$\frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} = \begin{pmatrix} \frac{\partial \sigma_1}{\partial \xi_1}\Big|_{\bar{\xi}_1} & \cdots & \frac{\partial \sigma_1}{\partial \xi_J}\Big|_{\bar{\xi}_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J}{\partial \xi_1}\Big|_{\bar{\xi}_J} & \cdots & \frac{\partial \sigma_J}{\partial \xi_J}\Big|_{\bar{\xi}_J} \end{pmatrix}.$$

In other words, the matrix $\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)/\partial \boldsymbol{\xi}'$ contains $\bar{\boldsymbol{\xi}}_1, \ldots, \bar{\boldsymbol{\xi}}_J$ in its 1st to the $J$th row, all of which can be distinct. For notational convenience however, we suppress the indices in $\bar{\boldsymbol{\xi}}_j$ and simply write $\bar{\boldsymbol{\xi}}$. From assumption A2(a) the matrix $\boldsymbol{H}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) = \partial \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)/\partial \boldsymbol{\xi}'$ is invertible for each $\boldsymbol{\xi} \in \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d; \epsilon)$ with probability tending to one, we can write

$$\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) = \left\{\frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'}\right\}^{-1} \boldsymbol{\epsilon}^n \tag{43}$$

with probability tending to one. Likewise,

$$\begin{aligned}\mathbf{0} &= \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{s}^0 \\ &= \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{s}^0 \\ &\quad + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \\
&\quad + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{ \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) \right\} \\
&= \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \left\{ \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) \right\}
\end{aligned}
$$

where $\underline{\boldsymbol{\xi}}$ is $J \times 1$ vector of values between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$. By assumption A2(a),

$$
\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) = - \left\{ \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \right\}^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \tag{44}
$$

with probability tending to one. Therefore, by substituting (43) and (44) for (42) and using the notation in (36) we obtain

$$
\begin{aligned}
&\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) \\
&= \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) + \left\{ \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \right\}^{-1} \boldsymbol{\epsilon}^n - \left\{ \frac{\partial \boldsymbol{\sigma}(\boldsymbol{X}, \underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'} \right\}^{-1} \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \\
&= \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) + \boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d).
\end{aligned} \tag{45}
$$

Substituting (45) for $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ in (22) gives

$$
\begin{aligned}
&\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) \\
&= J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) \\
&= J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) + J^{-1}\boldsymbol{Z}_d' \left\{ \boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right\} \\
&= \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) + J^{-1}\boldsymbol{Z}_d' \left\{ \boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \right\}.
\end{aligned} \tag{46}
$$

Now we approximate $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ within the neighborhood of $\boldsymbol{\theta}_d^0$ by the following function $\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)$.

$$
\begin{aligned}
\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d) &= \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) \\
&\quad + J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \left\{ \boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \right\}.
\end{aligned} \tag{47}
$$

**Cost Side Derivation**

Write

$$
\begin{aligned}
\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) &= \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) + \{\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R)\} \\
&\quad + \{\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)\}.
\end{aligned} \tag{48}
$$

Since $g(\cdot)$ is assumed to be continuously differentiable, the $j$-th element of the second term in (48) can be rewritten by the mean value theorem as

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

$$
\begin{aligned}
& +\frac{\partial g(p_j - m_{g_j}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R))}{\partial \boldsymbol{\xi}'}\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\} \\
= \ & -\dot{g}(p_j - m_{g_j}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R))\frac{\partial m_{g_j}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'}\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)\}
\end{aligned} \tag{49}
$$

where $\bar{\bar{\boldsymbol{\xi}}}$ is between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$. By substituting (43) for (49) and using the notation in (36), we obtain

$$
\begin{aligned}
& \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \omega_j(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R) \\
= \ & -\dot{g}(p_j - m_{g_j}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R))\frac{\partial m_{g_j}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)}{\partial \boldsymbol{\xi}'}\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n.
\end{aligned}
$$

In vector form, this can be expressed as

$$
\begin{aligned}
& \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R) \\
= \ & -\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n
\end{aligned} \tag{50}
$$

where

$$
\boldsymbol{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \frac{\partial \boldsymbol{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)}{\partial \boldsymbol{\xi}'} \tag{51}
$$

and

$$
\begin{aligned}
& \boldsymbol{L}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \\
= \ & \begin{pmatrix} \dot{g}(p_1 - m_{g_1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \dot{g}(p_J - m_{g_J}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) \end{pmatrix}.
\end{aligned} \tag{52}
$$

Actually, $J \times J$ matrices $\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)$ and $\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)$ contain $\bar{\bar{\boldsymbol{\xi}}}_1, \ldots, \bar{\bar{\boldsymbol{\xi}}}_J$ in its 1st to the $J$th rows, all of which can be distinct, but we here suppress this fact for notational simplicity. Similarly, we rewrite the third term in (48) by the mean value theorem,

$$
\begin{aligned}
& \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \\
= \ & \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
= \ & \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
& -\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\}
\end{aligned} \tag{53}
$$

where $\underline{\underline{\boldsymbol{\xi}}}$ is between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$. Substituting (44) for (53) gives

$$
\begin{aligned}
& \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \\
= \ & \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
& +\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d).
\end{aligned} \tag{54}
$$

By substituting (50) and (54) for (48), we have

$$
\begin{aligned}
& \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) \\
= \ & \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \\
& +\boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)) \\
& -\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n \\
& +\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d).
\end{aligned} \tag{55}
$$

Thus, the supply side moments $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) = J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ are rewritten by (55) as

$$\begin{aligned}
&\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) \\
&= J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) \\
&= \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) \\
&\quad + J^{-1}\boldsymbol{Z}_c'\left\{\boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{g}(\boldsymbol{p} - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0))\right\} \\
&\quad - J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n \\
&\quad + J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d). \qquad (56)
\end{aligned}$$

We approximate the supply side moments $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ within the neighborhood of $\boldsymbol{\theta}^0$ by the following function $\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})$.

$$\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}) = \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\left\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\} \qquad (57)$$

where $\boldsymbol{H}_0 = \boldsymbol{H}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$, $\boldsymbol{L}_0 = \boldsymbol{L}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$, and $\boldsymbol{M}_0 = \boldsymbol{M}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Let

$$\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d) \\ \boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}) \end{pmatrix}. \qquad (58)$$

The first term in $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ is the sample moment evaluated at $(\boldsymbol{s}, P) = (\boldsymbol{s}^0, P^0)$ and thus contains neither the sampling nor simulation errors, while the second term is an approximation for the difference between $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$. Note that the three components in $\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)$—$\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$, $J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n$, and $J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$—are not mutually independent because they all include the product characteristics $\boldsymbol{X}$ as well as the unobserved product quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$, both of which are random. However they are uncorrelated if evaluated at $\boldsymbol{\theta}_d = \boldsymbol{\theta}_d^0$ as shown below due to (31) and (33) in assumption A1. For the covariance between $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$ and $J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n$, we have

$$\begin{aligned}
&\mathrm{Cov}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n] \\
&= \mathrm{E}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \cdot J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n] \\
&\quad - \mathrm{E}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]\,\mathrm{E}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n] \\
&= \mathrm{E}_{\mathrm{x},\xi}[\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \cdot J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]] \\
&\quad - \mathrm{E}_{\mathrm{x}_1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]\,\mathrm{E}_{\mathrm{x},\xi}[\mathrm{E}_{\epsilon|\mathrm{x},\xi}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]] \\
&= \mathrm{E}_{\mathrm{x},\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \cdot J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{\epsilon}^n|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]] \\
&\quad - \mathrm{E}_{\mathrm{x}_1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]\,\mathrm{E}_{\mathrm{x},\xi}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{\epsilon}^n|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]] \\
&= \mathrm{E}_{\mathrm{x},\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \cdot J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1} \cdot \boldsymbol{0}] \\
&\quad - \mathrm{E}_{\mathrm{x}_1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)]\,\mathrm{E}_{\mathrm{x},\xi}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1} \cdot \boldsymbol{0}] \\
&= \boldsymbol{0}.
\end{aligned}$$

Similarly, we obtain $\mathrm{Cov}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)] = \boldsymbol{0}$. Since $\boldsymbol{\epsilon}^n$ and $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$ are generated by the distinct sampling process given $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$, they are conditionally independent. Thus, for the covariance between $J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n$ and $J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$, we also obtain $\mathrm{Cov}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n, J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)] = \boldsymbol{0}$.

On the supply side, we can similarly show that the three components in $\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}^0)$—$\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$, $J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n$, and $J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$—are mutually uncorrelated by using A1. These facts enable us to calculate the asymptotic variance-covariance matrix of $J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$ as a sum of the three variance-covariance matrices, each derived from the three separate components in $\mathcal{G}_J(\boldsymbol{\theta}^0)$.

We prove that (1) the difference between $J^{\frac{1}{2}}\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ and $J^{\frac{1}{2}}\mathcal{G}_J(\boldsymbol{\theta})$ to be $o_p(1)$ within any shrinking neighborhood of $\boldsymbol{\theta}^0$, and thus the estimator $\bar{\boldsymbol{\theta}}$ which minimizes $||\mathcal{G}_J(\boldsymbol{\theta})||$ has the same asymptotic distribution as $\hat{\boldsymbol{\theta}}$ which minimizes $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||$. Then we prove that (2) $\bar{\boldsymbol{\theta}}$ is asymptotically normally distributed with variance-covariance matrix consisting of the three components corresponding to the term $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$, the term involving $\boldsymbol{\epsilon}^n$ and the term consisting of $\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)$ by applying a version of Theorem 3.3 in Pakes and Pollard (1989).

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

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 (differentiable in $\boldsymbol{\theta}$) for the expectation of $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$. Given assumption B2, B3 implies that $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ can be approximated by $\boldsymbol{\Gamma}_J(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$ near $\boldsymbol{\theta}^0$. Assumptions B4(a)–(c) determine the magnitude of the three components in $J^{\frac{1}{2}}\mathcal{G}_J(\boldsymbol{\theta}^0)$, where each component is shown to follow asymptotically normal, while assumptions B4(d)–(f) are the Lyapunov conditions used in the central limit theorem. Assumption B6 is the regularity condition for the profit margin $\boldsymbol{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ which guarantees its smoothness in terms of $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d$.

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

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

$$\begin{aligned} \mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)] &= \begin{pmatrix} \mathrm{E}_{\mathrm{x}1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \\ \mathrm{E}_{\mathrm{w}1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)] \end{pmatrix} \\ &= \boldsymbol{\Gamma}_J(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + o(||\boldsymbol{\theta} - \boldsymbol{\theta}^0||) \end{aligned} \tag{59}$$

*uniformly in* $J$. *The matrix* $\boldsymbol{\Gamma}_J = (\boldsymbol{\Gamma}_J^{d\prime}, \boldsymbol{\Gamma}_J^{c\prime})' \to \boldsymbol{\Gamma} = (\boldsymbol{\Gamma}^{d\prime}, \boldsymbol{\Gamma}^{c\prime})'$ *as* $J \to \infty$, *where* $\boldsymbol{\Gamma}_J$ *has full column rank.*

**Assumption B3** *For all sequences of positive numbers* $\delta_J$ *such that* $\delta_J \to 0$,

$$\begin{aligned} \text{(a)} \quad \sup_{||\theta_d - \theta_d^0|| \le \delta_J} \Big|\Big| & J^{\frac{1}{2}} \left\{ \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{x}1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \right\} \\ & - J^{\frac{1}{2}} \left\{ \boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{x}1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \right\} \Big|\Big| = o_p(1) \end{aligned} \tag{60}$$

*and*

$$\begin{aligned} \text{(b)} \quad \sup_{||\theta - \theta^0|| \le \delta_J} \Big|\Big| & J^{\frac{1}{2}} \left\{ \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{w}1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)] \right\} \\ & - J^{\frac{1}{2}} \left\{ \boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{w}1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)] \right\} \Big|\Big| = o_p(1). \end{aligned} \tag{61}$$

**Assumption B4** *Let*

$$\begin{aligned}\boldsymbol{Z}_d'\boldsymbol{H}^{-1}(\boldsymbol{\xi},\boldsymbol{\theta},P) &\equiv (\boldsymbol{a}_1^d(\boldsymbol{\xi},\boldsymbol{\theta}_d,P),\ldots,\boldsymbol{a}_J^d(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)),\\ -\boldsymbol{Z}_c'\boldsymbol{L}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\boldsymbol{M}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) &\equiv (\boldsymbol{a}_1^c(\boldsymbol{\xi},\boldsymbol{\theta}_d,P),\ldots,\boldsymbol{a}_J^c(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)).\end{aligned}$$

*Set*

$$\begin{aligned}\boldsymbol{Y}_{Ji}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) &\equiv \frac{1}{nJ^{\frac{1}{2}}}\sum_{j=1}^{J}\begin{pmatrix}\boldsymbol{a}_j^d(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{ji},\\ \boldsymbol{a}_j^c(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{ji}\end{pmatrix},\\ \boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) &\equiv \frac{1}{RJ^{\frac{1}{2}}}\sum_{j=1}^{J}\begin{pmatrix}\boldsymbol{a}_j^d(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{jr}^*(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d)\\ \boldsymbol{a}_j^c(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{jr}^*(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d)\end{pmatrix}.\end{aligned}$$

*Suppose that*

$$\text{(a)}\qquad \lim_{J\to\infty}\mathrm{V}_{\mathrm{x}_1,\xi,\mathrm{w}_1,\omega}\left[\begin{pmatrix}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)/J^{\frac{1}{2}}\\ \boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)/J^{\frac{1}{2}}\end{pmatrix}\right]=\boldsymbol{\Phi}_1, \tag{62}$$

$$\text{(b)}\qquad \lim_{n,J\to\infty} n\,\mathrm{V}_{\epsilon,\mathrm{x},\xi}[\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)]=\boldsymbol{\Phi}_2, \tag{63}$$

$$\text{(c)}\qquad \lim_{R,J\to\infty} R\,\mathrm{V}_{\epsilon^*,\mathrm{x},\xi}[\boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)]=\boldsymbol{\Phi}_3 \tag{64}$$

*for finite positive definite matrices* $\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2$ *and* $\boldsymbol{\Phi}_3$*. Suppose that the following Lyapunov conditions hold.*

$$\text{(d)}\qquad \sum_{j=1}^{J}\mathrm{E}_{\mathrm{x}_1,\xi,\mathrm{w}_1,\omega}\left[\left|\left|\begin{pmatrix}\boldsymbol{z}_j^d\xi_j(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)/J^{\frac{1}{2}}\\ \boldsymbol{z}_j^c\omega_j(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)/J^{\frac{1}{2}}\end{pmatrix}\right|\right|^{2+\delta}\right]=o(1), \tag{65}$$

$$\text{(e)}\qquad n\,\mathrm{E}_{\epsilon,\mathrm{x},\xi}[||\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)||^{2+\delta}]=o(1), \tag{66}$$

$$\text{(f)}\qquad R\,\mathrm{E}_{\epsilon^*,\mathrm{x},\xi}[||\boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)||^{2+\delta}]=o(1) \tag{67}$$

*for some* $\delta>0$*.*

**Assumption B5** *For all sequences of positive numbers* $\delta_J$ *with* $\delta_J\to 0$*, we assume*

$$\text{(a)}\qquad \sup_{||\theta_d-\theta_d^0||\le\delta_J}\ \sup_{(\xi_1,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^J\times\mathcal{N}_{P^0}(\delta_J)}\left|\left|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\left\{\boldsymbol{H}^{-1}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)-\boldsymbol{H}_0^{-1}\right\}\boldsymbol{\epsilon}^n\right|\right| = o_p(1); \tag{68}$$

$$\text{(b)}\qquad \sup_{||\theta_d-\theta_d^0||\le\delta_J}\ \sup_{(\xi_1,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^J\times\mathcal{N}_{P^0}(\delta_J)}\left|\left|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\left\{\boldsymbol{H}^{-1}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) -\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\}\right|\right| = o_p(1); \tag{69}$$

$$\text{(c)}\qquad \sup_{||\theta_d-\theta_d^0||\le\delta_J}\ \sup_{(\xi_1,\xi_2,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^{2J}\times\mathcal{N}_{P^0}(\delta_J)}\left|\left|J^{-\frac{1}{2}}\boldsymbol{Z}_c' \times\left\{\boldsymbol{L}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{M}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2,\boldsymbol{\theta}_d,P)-\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\right\}\boldsymbol{\epsilon}^n\right|\right| = o_p(1); \tag{70}$$

$$\text{(d)}\qquad \sup_{||\theta_d-\theta_d^0||\le\delta_J}\ \sup_{(\xi_1,\xi_2,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^{2J}\times\mathcal{N}_{P^0}(\delta_J)}\left|\left|J^{-\frac{1}{2}}\boldsymbol{Z}_c' \times\left\{\boldsymbol{L}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{M}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2,\boldsymbol{\theta}_d,P)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)-\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\}\right|\right| = o_p(1); \tag{71}$$

$$\text{(e)}\qquad \sup_{||\theta_d-\theta_d^0||\le\delta_J}\ \sup_{P\in\mathcal{N}_{P^0}(\delta_J)}\left|\left|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\{\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P) -\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\}\right|\right| = o_p(1) \tag{72}$$

*where* $\boldsymbol{\xi}_1 = (\boldsymbol{\xi}_{11}, \ldots, \boldsymbol{\xi}_{1J})$ *and* $\boldsymbol{\xi}_2 = (\boldsymbol{\xi}_{21}, \ldots, \boldsymbol{\xi}_{2J})$ *are respectively a set of distinct J vectors, each vector corresponds to each row of* $J \times J$ *matrices* $\boldsymbol{L}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$, $\boldsymbol{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ *and* $\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$.

**Assumption B6** *For every finite* $J$, *for all* $\boldsymbol{\theta}_d \in \Theta_d$, *and for all* $P$ *in a neighborhood of* $P^0$,

$$\boldsymbol{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \partial \boldsymbol{m}_g(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial \boldsymbol{\xi}' \tag{73}$$

*exists and continuous both in* $\boldsymbol{\xi}$ *and* $\boldsymbol{\theta}_d$.

**Theorem 2 (Asymptotic Normality of $\hat{\boldsymbol{\theta}}$)** *Suppose that A1–A7 and B1–B6 hold for some* $n(J), R(J) \to \infty$. *Then, the estimator* $\hat{\boldsymbol{\theta}}$ *that minimizes* $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||$ *is asymptotically normal at the rate of* $J^{\frac{1}{2}}$:

$$J^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \overset{w}{\rightsquigarrow} N[\mathbf{0}, (\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}] \tag{74}$$

*with* $\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \boldsymbol{\Phi}_3$.

# 4 Estimating Demand and Supply Systems with Purchasing Information on the Consumer's Demographics

## 4.1 Additional Moments with Purchasing Information

The framework in BLP(1995) uses the orthogonal conditions between the unobserved product characteristics $(\xi_j, \omega_j)$ and the exogenous instrumental variables $(\boldsymbol{z}_j^d, \boldsymbol{z}_j^c)$ to obtain the GMM estimate of the parameter $\boldsymbol{\theta}$. For some markets, however, market summaries such as averaged demographics of consumers who purchased specific type of products are publicly available, even if their detailed individual-level data such as purchasing history are not. In the U.S. automobile market, for instance, we know the median income of consumers who purchased domestic, European, or Japanese vehicles from publications such as the *Ward's Motor Vehicle Facts & Figures*. In this section, we first generalize the idea given by Petrin (2002), who extends the BLP framework by additional moment conditions constructed from the market summary data to the GMM. We then give the asymptotic theorem to this GMM estimator and uncover the conditions under which the use of the additional moment conditions allows us to estimate of the demand side parameters more precisely.

First we define some words and notations. *Discriminating attributes* is the product characteristic or attribute that enables consumers to discriminate some products from others. When we say consumer $i$ takes a discriminating attribute $q$, this means that consumer chooses a product from a group of products whose characteristic or attribute have discrimating attribute $q$. An automobile attribute "imports" is one of such discriminating attributes. When we say a consumer chooses this attribute, what we mean is that the consumer purchases an imports. Similarly, "minivan" and "costing less than \$10,000" are examples of the discriminating attribute as we define here. We consider a finite number of discriminating attributes ($q = 1, \ldots, N_p$) and denote all the products involved in attribute $q$ as $\mathcal{Q}_q$. By definition, discriminating attributes for outside good is undefined.

We next consider expectation of consumer's demographics conditional on a specific discriminating attribute. Suppose that some information on demographics for consumer $t$ are available. Demographic variables such as age, family size, or, income, is already numerical, but for other demographics such as having children, belonging to certain age group, choice of residential area, can be numerically expressd using indicators. We denote this numerically represented $D$ dimensional demographics as $\boldsymbol{\nu}_t^o = (\nu_{t1}^o, \ldots, \nu_{tD}^o)'$. We assume that the joint distribution of demographics $\boldsymbol{\nu}_t^o$ has a bounded support. The consumer $t$'s observed demographic $\nu_{td}^o, d = 1, \ldots, D$

is averaged over consumers who choose discriminating attribute $q$ in the population to obtain the conditional expectation $\eta^0_{dq} = \mathrm{E}[\nu^o_{td}|C_t \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0)]$. An example of this conditional expectation would be the expected value of income of consumers in the population $P^0$ who purchased imported vehicles.

Since the conditional expectation can be written as

$$
\begin{aligned}
&\mathrm{E}[\nu^o_{td}|C_t \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \\
&= \int \nu^o_{td} \Pr[d\nu^o_{td}|C_t \in \mathcal{Q}_q, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] \\
&= \frac{\int \nu^o_{td} \Pr[C_i \in \mathcal{Q}_q|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \nu^o_{td}] P^0(d\nu^o_{td})}{\Pr[C_i \in \mathcal{Q}_q|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)]} \\
&= \frac{\int \nu^o_{td} \Pr[C_t \in \mathcal{Q}_q|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t] P^0(d\boldsymbol{\nu}_t)}{\Pr[C_t \in \mathcal{Q}_q|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)]} \\
&= \int \nu^o_{td} \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} P^0(d\boldsymbol{\nu}_t), \qquad (75)
\end{aligned}
$$

we can form an identity

$$
\eta^0_{dq} - \int \nu^o_{td} \frac{\sum_{j \in \mathcal{Q}_q} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_q} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)} P^0(d\boldsymbol{\nu}_t) = 0 \qquad (76)
$$

at $\boldsymbol{\theta}_d = \boldsymbol{\theta}^0_d$ for $q = 1, \ldots, N_p, d = 1, \ldots, D$. Although $P^0$ is so far assumed known, we typically will not be able to calculate the second term on the left-hand side of (76) analytically and will have to approximate it by the i.i.d. sample $\boldsymbol{\nu}_t, t = 1, \ldots, T$ from the underlying distribution $P^0$. The sample moments $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$ corresponding to (76) are

$$
\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = \boldsymbol{\eta}^0 - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \qquad (77)
$$

where

$$
\boldsymbol{\eta}^0 = \begin{pmatrix} \eta^0_{11} \\ \vdots \\ \eta^0_{1N_p} \\ \vdots \\ \eta^0_{D1} \\ \vdots \\ \eta^0_{DN_p} \end{pmatrix}, \quad \boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \begin{pmatrix} \frac{\sum_{j \in \mathcal{Q}_1} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_1} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \\ \vdots \\ \frac{\sum_{j \in \mathcal{Q}_{N_p}} \sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu}_t; \boldsymbol{\theta}_d)}{\sum_{j \in \mathcal{Q}_{N_p}} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P)} \end{pmatrix}. \qquad (78)
$$

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. This random sample $\boldsymbol{\nu}_t, t = 1, \ldots, T$ of consumers is taken independent of the sample $\boldsymbol{\nu}_r, r = 1, \ldots, R$ in (20) for calculating the simulated market shares $\boldsymbol{\sigma}_j(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$. Note that these additional moment conditions are conditional on product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0))$, and thus depend on the indices $J$ and $T$.

Suppose that we do not know the conditional expectation of demographics $\eta^0_{dq}$, instead, we have its estimate $\eta^N_{dq}$ from independent sources such as CEX automobile supplement in the case of Petrin (2002). We assume $N$ independent consumer draws with their purchasing histories are used to construct $\boldsymbol{\eta}^N = (\eta_{11}^N, \ldots, \eta_{1N_p}^N, \ldots, \eta_{D1}^N, \ldots, \eta_{DN_p}^N)'$ and define the sampling error $\boldsymbol{\epsilon}^N$ contained in $\boldsymbol{\eta}^N$ as follows.

$$\boldsymbol{\epsilon}^N = \boldsymbol{\eta}^N - \boldsymbol{\eta}^0 = \frac{1}{N}\sum_{i'=1}^{N} \boldsymbol{\epsilon}_{i'}^{\#}. \tag{79}$$

In short, we assume here that $\boldsymbol{\eta}^N$ is the sum of $N$ conditionally independent random variables given the set of product characteristics $(\boldsymbol{X}, \boldsymbol{\xi})$ of all products. Note that quantities $n$ and $N$ are respectively the number of samples taken to calculate the observed market share and the observed demographic average of consumers purchasing product with discriminating attribute. As such they are beyond the control of researchers. On the other hand quantities $R$ and $T$ are respectively the number of samples taken to simulate the model-calculated market share as well as the model-calculated demographic average of consumers purchasing product with discriminating attribute from the population $P^0$ of consumers. They are both chosen by the researchers and these two samples must be independent.

Since we evaluate the unobserved quality $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P)$ at $(\boldsymbol{s}, P) = (\boldsymbol{s}^n, P^R)$ in (77), the sample moments we can calculate are

$$\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) = \boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}_t^o \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) \tag{80}$$

for $\boldsymbol{\theta}_d \in \Theta_d$. As an extension to BLP(1995), we use $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$ to estimate $\boldsymbol{\theta}$, in addition to the two sample moments $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ in (22) and $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ in (23). The objective function we minimize in the GMM estimation is the sum of norm of $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$, $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$, and $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$, that is, the norm of

$$\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) = \begin{pmatrix} \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) \\ \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) \\ \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) \end{pmatrix}. \tag{81}$$

In the following, we derive the CAN properties for the GMM estimator $\breve{\boldsymbol{\theta}}$ which minimizes $||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)||$. Notice that the first two moments $\boldsymbol{G}_J^d$ and $\boldsymbol{G}_J^c$ in $\boldsymbol{G}_{J,T}$ are sample moments averaged over products $j = 1, \ldots, J$, while the third moment $\boldsymbol{G}_{J,T}^a$ is averaged over consumers $t = 1, \ldots, T$. To derive asymptotics for $\breve{\boldsymbol{\theta}}$, we have to increase two distinct sample size indices $J$ and $T$ simultaneously. We assume the sample size $T$ of consumers is always greater than the number of products $J$, and then $T$ grows faster than $J$, that is, $J/T \to 0$ as $J \to \infty$.

## 4.2 Consistency

For any $\delta > 0$, we show that $\lim_{J,T\to\infty} \Pr[||\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}^0|| > \delta] \to 0$. The proof is a straightforward extension to the consistency proof for $\hat{\boldsymbol{\theta}}$ in Theorem 1.

Assumption A8 bounds $||\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||$ away from $||\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||$ over $\boldsymbol{\theta}_d$ outside of a neighborhood of $\boldsymbol{\theta}_d^0$. This condition parallels assumption A5, which bounds $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ away from $\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$.

In assumption A9, we assume an asymptotic property the discriminating attributes $q, q = 1, \ldots, N_p$ must obey. We guarantee non-zero aggregate market shares for products with discriminating attribute $q$ when the number of products $J$ grows large. With this assumption and the following assumption A10(b), the additional moment defined in (77) has finite variance at $\boldsymbol{\theta}_d = \boldsymbol{\theta}_d^0$.

Assumption A10(a) specifies properties for error contained in the additional information $\eta_{dq}^N$. We assume $\eta_{dq}^N$ is unbiased for the true value $\eta_{dq}^0$ and consistent at a rate of $N^{1/2}$ given the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$. Assumption A10(b) guarantees a finiteness for $\eta_{dq}^0$.

Assumption A11 is on the proportion of the probabilities taking discriminating attributes between individual $t$ and population $P$, $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$. We assume that the average absolute distance between $\boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ and $\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ converges to zero in probability within the $\delta$ neighborhood of $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ for any $\boldsymbol{\theta}_d \in \Theta_d$. This assumption will be used to guarantee that we can bring the sample analogue of the additional moments, $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$, close enough to $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^N)$ for any $\boldsymbol{\theta}_d$.

**Assumption A8** *For all $\delta > 0$, there exists $C(\delta)$ such that*

$$\lim_{J,T\to\infty} \Pr\left[\inf_{\theta_d \notin \mathcal{N}_{\theta_d^0}(\delta)} ||\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| \geq C(\delta)\right] = 1. \tag{82}$$

**Assumption A9** *For all discriminating attributes $q = 1, \ldots, N_p$,*

$$\left\{\sum_{j\in\mathcal{Q}_q} \sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\right\}^{-2}$$

*has a finite mean and variance for every $J$.*

**Assumption A10** (a) *For all observed consumer's demographics $d = 1, \ldots, D$ and for all discriminating attributes $q = 1, \ldots, N_p$, the sampling error $\eta_{dq}^N - \eta_{dq}^0$ has zero mean and variance of order $1/N$ conditional on product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$ of all products, i.e.,*

$$\mathrm{E}_{\epsilon^\#|\mathrm{x},\xi}\left[\eta_{dq}^N - \eta_{dq}^0 | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] = 0, \tag{83}$$

$$\mathrm{V}_{\epsilon^\#|\mathrm{x},\xi}\left[\eta_{dq}^N - \eta_{dq}^0 | \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] = O_p(1/N). \tag{84}$$

(b) *For all observed consumer's demographics $d = 1, \ldots, D$ and for all discriminating attributes $q = 1, \ldots, N_p$, $\eta_{dq}^0$ has a finite mean and variance for all $J$, i.e., $\mathrm{E}_{\mathrm{x},\xi}[\eta_{dq}^0] < \infty$ and $\mathrm{V}_{\mathrm{x},\xi}[\eta_{dq}^0] < \infty$.*

**Assumption A11** *For any $\boldsymbol{\theta}_d \in \Theta_d$, and for all $\delta > 0$,*

$$\lim_{J,T\to\infty} \Pr\left[\sup_{(\xi,P)\in\mathcal{N}_{\xi^0}(\theta_d;\delta)\times\mathcal{N}_{P^0}(\delta)} T^{-1/2}||\boldsymbol{\Psi}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)|| > \delta\right] = 0, \tag{85}$$

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

**Theorem 3 (Consistency of $\breve{\boldsymbol{\theta}}$)** *Suppose that A1–A11 hold for some $n(J,T), R(J,T)$, and $N$, all of which grow infinitely as $J$ and $T$ grow infinitely. Then, $\breve{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^0$.*

## 4.3 Asymptotic Normality

To derive the asymptotic normality of $\hat{\boldsymbol{\theta}}$ in Theorem 2, we approximated the demand side moments $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and the supply side moments $\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ respectively by $\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)$

and $\boldsymbol{\mathcal{G}}^c_J(\boldsymbol{\theta})$ within the shrinking neighborhood of $\boldsymbol{\theta}^0$. Similarly, we will use an approximation to the additional moments $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$. Decompose the additional moments $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$ into four terms.

$$\begin{aligned}
&\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) \\
&= \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) + \{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R, \boldsymbol{\eta}^N)\} \\
&\quad + \{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^N)\} \\
&\quad + \{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}. \qquad (86)
\end{aligned}$$

The second term in (86) can be written as

$$\begin{aligned}
&\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R, \boldsymbol{\eta}^N) \\
&= \boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) \\
&\quad - \left\{\boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R)\right\} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R)\} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \Big\{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) \\
&\quad + \frac{\partial \boldsymbol{\psi}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)}{\boldsymbol{\xi}'}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R))\Big\} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)) \qquad (87)
\end{aligned}$$

where $\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \partial \boldsymbol{\psi}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial \boldsymbol{\xi}'$ and $\boldsymbol{\xi}^\dagger = (\boldsymbol{\xi}^\dagger_1, \ldots, \boldsymbol{\xi}^\dagger_J)$ is the set of intermediate vectors between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$. Substituting (43) for (87) gives

$$\begin{aligned}
&\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R, \boldsymbol{\eta}^N) \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n. \qquad (88)
\end{aligned}$$

The third term in (86) is

$$\begin{aligned}
&\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^N) \\
&= \boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) \\
&\quad - \left\{\boldsymbol{\eta}^N - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\right\} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\nu}^o_t \otimes \Big\{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\partial \psi_t(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)}{\boldsymbol{\xi}'}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^R)-\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0))\Big\} \\
= \;& -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \Big\{\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R)-\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0) \\
&+\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^R)-\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0))\Big\} \qquad (89)
\end{aligned}
$$

where $\boldsymbol{\xi}^{\ddagger}=(\boldsymbol{\xi}_1^{\ddagger},\ldots,\boldsymbol{\xi}_J^{\ddagger})$ is the set of intermediate vectors between $\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)$. Substituting (44) for (89) gives

$$
\begin{aligned}
&\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^R,\boldsymbol{\eta}^N)-\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^N) \\
= \;& -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \Big\{\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R)-\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0) \\
&-\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\Big\}. \qquad (90)
\end{aligned}
$$

The fourth term in (86) is

$$
\begin{aligned}
&\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^N)-\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0) \\
= \;& \boldsymbol{\eta}^N-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0) \\
&-\Big\{\boldsymbol{\eta}^0-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\Big\} \\
= \;& \boldsymbol{\eta}^N-\boldsymbol{\eta}^0. \qquad (91)
\end{aligned}
$$

Consequently, by substituting (88), (90) and (91) for (86), we can rewrite the additional moments $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$ as follows.

$$
\begin{aligned}
&\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N) \\
= \;& \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0) \\
&-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \Big\{\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R)-\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0) \\
&+\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^{\dagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^n-\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\Big\} \\
&+\boldsymbol{\eta}^N-\boldsymbol{\eta}^0. \qquad (92)
\end{aligned}
$$

We use the following approximation $\boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d)$ to $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}^d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$.

$$
\begin{aligned}
\boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d) \;=\;& \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\nu}_t^o \otimes \boldsymbol{\Upsilon}_t^0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\} \\
&+\boldsymbol{\eta}^N-\boldsymbol{\eta}^0. \qquad (93)
\end{aligned}
$$

where $\boldsymbol{\Upsilon}_t^0 \equiv \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$.

In order to obtain the asymptotic normality of $\breve{\boldsymbol{\theta}}$, we will take the same path as the proof of Theorem 2, that is, we first show that the sample moments $\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$ in (81) are well approximated by

$$
\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})=\begin{pmatrix}\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)\\ \boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})\\ \boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d)\end{pmatrix} \qquad (94)
$$

within the $\delta_{J,T}$ neighborhood of $\boldsymbol{\theta}^0$ where $\delta_{J,T}$ is converges to 0 as $J, T \to \infty$, and then show that the estimator which minimizes the norm of $\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})$ is asymptotically normal.

Assumption B7 plays the same role on the additional moments $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$ as assumption B5 does on the $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$, or it guarantees that the difference between $\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)$ and $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N)$ is stochastically small enough within the neighborhood of $\boldsymbol{\theta}^0_d$.

Assumption B8 and B9 are used in a same way as assumption B2 and B3. Assumption B8 is just a differentiability condition for the expectation of $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$ at $\boldsymbol{\theta}^0_d$. Given B8, assumption B9 approximates $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$ by $\boldsymbol{\Gamma}^a_{J,T}(\boldsymbol{\theta}_d - \boldsymbol{\theta}^0_d) + \boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$ near $\boldsymbol{\theta}^0_d$.

In assumptions B10(a)–(d), we specify the asymptotic covariance for the four terms in $T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)$, or $T^{\frac{1}{2}}\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$, $T^{-\frac{1}{2}}\sum_{t=1}^T \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}^0_t \boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n$, $T^{-\frac{1}{2}}\sum_{t=1}^T \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}^0_t \boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d)$, and $T^{\frac{1}{2}}(\boldsymbol{\eta}^N - \boldsymbol{\eta}^0)$. These terms are mutually independent conditional on the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0))$, and thus the asymptotic covariance of $T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)$ is the sum of the four covariance matrices. Assumptions B10(e)–(h) are respectively Lyapunov conditions necessary to ensure the four terms converge to the normal distribution.

**Assumption B7** *For all discriminating attributes $q (q = 1, \ldots, N_p)$, and for any $\delta_{J,T}$ such that $\delta_{J,T} \to 0$ as $J, T \to \infty$,*

(a)
$$\sup_{||\theta_d - \theta^0_d|| < \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta^0_d; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \Bigg|\Bigg| T^{-\frac{1}{2}} \sum_{t=1}^T \Big[ \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P)\boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)\boldsymbol{\epsilon}^n \Big] \Bigg|\Bigg| = o_p(1); \tag{95}$$

(b)
$$\sup_{||\theta_d - \theta^0_d|| < \delta_{J,T}} \sup_{(\xi_1, \xi_2, P) \in \{\mathcal{N}_{\xi^0}(\theta^0_d; \delta_{J,T})\}^{2J} \times \mathcal{N}_{P^0}(\delta_{J,T})} \Bigg|\Bigg| T^{-\frac{1}{2}} \sum_{t=1}^T \Big[ \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) - \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d) \Big] \Bigg|\Bigg| = o_p(1); \tag{96}$$

(c)
$$\sup_{||\theta_d - \theta^0_d|| < \delta_{J,T}} T^{\frac{1}{2}} \sum_{j \in \mathcal{Q}_q} \epsilon^R_j(\boldsymbol{\theta}_d) = o_p(1). \tag{97}$$

**Assumption B8** *For all $\boldsymbol{\theta}_d$ in some $\delta > 0$ neighborhood of $\boldsymbol{\theta}^0_d$,*

$$\mathrm{E}[\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)] = \boldsymbol{\Gamma}^a_{J,T}(\boldsymbol{\theta}_d - \boldsymbol{\theta}^0_d) + o(||\boldsymbol{\theta}_d - \boldsymbol{\theta}^0_d||) \tag{98}$$

*uniformly in $J$ and $T$. The Matrix $\boldsymbol{\Gamma}^a_{J,T} \to \boldsymbol{\Gamma}^a$ as $J, T \to \infty$, where $\boldsymbol{\Gamma}^a_{J,T}$ has full column rank.*

**Assumption B9** *For all sequence of positive numbers $\delta_{J,T}$ such that $\delta_{J,T} \to 0$ as $J, T \to \infty$,*

$$\sup_{||\theta_d - \theta^0_d|| \le \delta_{J,T}} \Bigg|\Bigg| T^{\frac{1}{2}}\{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \mathrm{E}[\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)]\} - T^{\frac{1}{2}}\{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \mathrm{E}[\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)]\} \Bigg|\Bigg| = o_p(1). \tag{99}$$

**Assumption B10** *Let*

$$-\sum_{t=1}^T \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) \equiv (\boldsymbol{a}^a_1(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \ldots, \boldsymbol{a}^a_J(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))$$

*and set*

$$\boldsymbol{Y}^a_{J,T,i}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) \equiv \frac{1}{n\sqrt{T}}\sum_{j=1}^{J}\boldsymbol{a}^a_j(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{ji},$$

$$\boldsymbol{Y}^{*a}_{J,T,r}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) \equiv \frac{1}{R\sqrt{T}}\sum_{j=1}^{J}\boldsymbol{a}^a_j(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon^*_{jr}(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d).$$

*Suppose that*

$$\text{(a)} \quad \lim_{J,T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{V}_{\nu,\mathrm{x},\xi}\left[\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)\right]=\boldsymbol{\Phi}^a_1, \tag{100}$$

$$\text{(b)} \quad \lim_{J,T,n\to\infty} n\,\mathrm{V}_{\epsilon,\nu,\mathrm{x},\xi}[\boldsymbol{Y}^a_{J,T,i}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)]=\boldsymbol{\Phi}^a_2, \tag{101}$$

$$\text{(c)} \quad \lim_{J,T,R\to\infty} R\,\mathrm{V}_{\epsilon^*,\nu,\mathrm{x},\xi}[\boldsymbol{Y}^{*a}_{J,T,r}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)]=\boldsymbol{\Phi}^a_3 \tag{102}$$

$$\text{(d)} \quad \lim_{J,T,N\to\infty} N\,\mathrm{V}_{\epsilon^\#,\mathrm{x},\xi}[T^{\frac{1}{2}}N^{-1}\boldsymbol{\epsilon}^\#_{i'}]=\boldsymbol{\Phi}^a_4 \tag{103}$$

*for finite positive definite matrices* $\boldsymbol{\Phi}^a_1, \boldsymbol{\Phi}^a_2$, $\boldsymbol{\Phi}^a_3$ *and* $\boldsymbol{\Phi}^a_4$. *Suppose that for some* $\delta > 0$,

$$\text{(e)} \quad \sum_{t=1}^{T}\mathrm{E}_{\nu,\mathrm{x},\xi}[||\{\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)\}/\sqrt{T}||^{2+\delta}]=o(1), \tag{104}$$

$$\text{(f)} \quad n\,\mathrm{E}_{\epsilon,\nu,\mathrm{x},\xi}[||\boldsymbol{Y}^a_{J,T,i}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)||^{2+\delta}]=o(1), \tag{105}$$

$$\text{(g)} \quad R\,\mathrm{E}_{\epsilon^*,\nu,\mathrm{x},\xi}[||\boldsymbol{Y}^{*a}_{J,T,r}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)||^{2+\delta}]=o(1), \tag{106}$$

$$\text{(h)} \quad N\,\mathrm{E}_{\epsilon^\#,\mathrm{x},\xi}[||T^{\frac{1}{2}}N^{-1}\boldsymbol{\epsilon}^\#_{i'}||^{2+\delta}]=o(1). \tag{107}$$

**Theorem 4 (Asymptotic Normality of $\breve{\boldsymbol{\theta}}$)** *Suppose that A1–A11 and B1–B10 hold for some increasing* $n(J,T), R(J,T), N$, *such that* $T/J\to\infty$ *as* $J\to\infty$ *and* $N\to\infty$. *Then, the estimator* $\breve{\boldsymbol{\theta}}$ *that minimizes* $||\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)||$ *is asymptotically normal at the rate of* $J^{\frac{1}{2}}$*:*

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

*The variance-covariance matrix* $\boldsymbol{V}$ *is written as*

$$\boldsymbol{V}=(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Gamma}^a)^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Gamma}^a)^{-1}$$

*where* $\boldsymbol{\Phi}=\boldsymbol{\Phi}_1+\boldsymbol{\Phi}_2+\boldsymbol{\Phi}_3$.

**Remark 1** The variance reduction of the estimates through the use of additional moments is due to the component $\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Gamma}^a$ in the asymptotic covariance matrix in Theorem 4. Notice also that this asymptotic covariance matrix assumes the ratio of the two size indices, $J/T$, converges to 0 as $J$ goes to infinity. For the finite sample where $T$ does not dominate $J$, the covariance matrix will be

$$\boldsymbol{V}=(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Gamma}^a)^{-1}\left(\boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}+\frac{J}{T}\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Phi}^a\boldsymbol{\Gamma}^a\right)(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}+\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Gamma}^a)^{-1} \tag{108}$$

where $\boldsymbol{\Phi}^a=\boldsymbol{\Phi}^a_1+\boldsymbol{\Phi}^a_2+\boldsymbol{\Phi}^a_3+\boldsymbol{\Phi}^a_4$. The term $(J/T)\boldsymbol{\Gamma}^{a\prime}\boldsymbol{\Phi}^a\boldsymbol{\Gamma}^a$ increases variance of the estimated parameters. Consequently, the use of the additional moments does not necessarily improve the accuracy of the estimates.

# 5 Concrete Examples

In this section, we discuss the conditions that guarantee the assumptions in the previous sections. When the number $J$ of products in the market grows large, the dimension of the market share vector increases. This implies that almost all elements of the market share vector decrease to zero. The rate at which the market share converges to zero and the response of market share to the change of the unobserved product quality, both of which determine the appropriateness of the assumptions, depend on the underlying distribution of the product characteristics and the consumer heterogeneity as well as the structure of competition in the market.

In the following, we consider two primal examples to examine the assumptions. The first is the simple logit model in which we can analytically solve the equation (4) in terms of $\boldsymbol{\xi}$ and thus do not incur the simulation error in the model. Without the simulation error, it is fairly easy to verify the assumptions for the logit model. The second is the random coefficient logit model. As discussed in BLP (1995), this model has useful properties when product characteristics and consumers' taste are multi-dimensionally distributed and then nature of competition among products is complex. Our main concern in the previous section is also in the efficient estimation for the random coefficient logit model. However, the random coefficient logit model has no closed-form solution for (4) and for the inverse of $\boldsymbol{H}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$. Thus, our examination has to rely on its stochastic approximation.

### Logit Model

The utility function of consumer $i$ for product $j$ in one of the simplest logit model is given by

$$u_{ij} = \delta_j + \nu_{ij}, \quad \delta_j = \theta_p p_j + \theta_x x_j + \xi_j \tag{109}$$

where $p_j$ and $x_j$ are respectively the price and the characteristic of product $j$, and $(\theta_p, \theta_x)$ is the set of demand parameters $\boldsymbol{\theta}_d$. The assumption that the consumer heterogeneity $\nu_{ij}$ being extreme-value distributed derives the probability of consumer $i$ choosing product $j$ as

$$\sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}, P) = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)}. \tag{110}$$

If we assume that the distribution of $\delta_j$ has a bounded support, the stochastic magnitude of $\sigma_j$ is $O_p(1/J)$. This implies that when the number of products grows large, the market share for each product, including outside good, decreases to zero at the same rate. Therefore we can reasonably assume the following condition on the rate at which the market share approaches zero when we use the logit model for demand.

**Condition S1**(a) *There exists positive finite constants $\underline{c}$ and $\overline{c}$ such that with probability one*

$$\frac{\underline{c}}{J} \leq s_j^0 \leq \frac{\overline{c}}{J}, \qquad j = 0, 1, \ldots, J. \tag{111}$$

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

Condition S1(a) leads us to $s_j^0 = O_p(1/J)$. In addition, this condition bounds the market share for each product away from zero for any fixed $J$, and then the inverse of the market share is stochastically of order of $J$, i.e., $1/s_j^0 = O_p(J)$. By condition S1(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 stochastic magnitude is of order one, i.e., $1/(1-\sum_{j\in\mathcal{J}_m} s_j^0) = O_p(1)$.

The limiting behavior of the market shares, both observed and model-calculated, are assumed in assumption A3. Assumptions A3(a) and (b) control the way in which $\boldsymbol{s}^n$ and $\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ approach to the true market share $\boldsymbol{s}^0$ 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 $R$ as $J$ grows large as well as on the limiting behavior of the true market share $\boldsymbol{s}^0$. We below derive the growth rates of $n$ and $R$ necessary to ensure A3 when condition S1 is satisfied.

First, we derive the rate for assumption A3(a). For any $\delta > 0$,

$$
\begin{aligned}
&\Pr\left[\rho_{s^0}(\boldsymbol{s}^n, \boldsymbol{s}^0) > \delta\right] \\
&= \Pr\left[\max_{0\le j\le J}\left|\frac{s_j^n - s_j^0}{s_j^0}\right| > \delta\right] \\
&\le \sum_{j=0}^{J}\Pr\left[\left|\frac{s_j^n - s_j^0}{s_j^0}\right| > \delta\right] \\
&= \sum_{j=0}^{J}\Pr\left[\frac{s_j^n - s_j^0}{s_j^0} > \delta\right] + \sum_{j=0}^{J}\Pr\left[\frac{s_j^n - s_j^0}{s_j^0} < -\delta\right] \\
&= \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\Pr\left[\frac{s_j^n - s_j^0}{s_j^0} > \delta \,\middle|\, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad + \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\Pr\left[\frac{s_j^n - s_j^0}{s_j^0} < -\delta \,\middle|\, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\Pr\left[\sum_{i=1}^{n}\epsilon_{ji} > n\delta s_j^0 \,\middle|\, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad + \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\Pr\left[\sum_{i=1}^{n}\epsilon_{ji} < -n\delta s_j^0 \,\middle|\, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right]. \qquad (112)
\end{aligned}
$$

Since $|\epsilon_{ji}| < 1$ and $\epsilon_{ji}$ are independently distributed across $i$ conditional on $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$ with conditional mean zero and conditional variance $s_j^0(1-s_j^0)$ by assumption A1(a), under condition S1(a), we can rewrite the first term in (112) by the Bernstein inequality as

$$
\begin{aligned}
&\sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\Pr\left[\sum_{i=1}^{n}\epsilon_{ji} > n\delta s_j^0 \,\middle|\, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\le \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\exp\left(-\frac{(n\delta s_j^0)^2}{2\,\mathrm{V}_{\epsilon|\mathrm{x},\xi}\left[\sum_{i=1}^{n}\epsilon_{ji}|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] + 2n\delta s_j^0}\right)\right] \\
&= \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\exp\left(-\frac{(n\delta s_j^0)^2}{2ns_j^0(1-s_j^0) + 2n\delta s_j^0}\right)\right] \\
&= \sum_{j=0}^{J}\mathrm{E}_{\mathrm{x},\xi}\left[\exp\left(-\frac{\delta^2}{2(1-s_j^0)/(ns_j^0) + 2\delta/(ns_j^0)}\right)\right] \\
&= J\,\mathrm{E}_{\mathrm{x},\xi}[\exp(-\delta^2 O_p(n/J))]. \qquad (113)
\end{aligned}
$$

The upper bound for the second term on the right hand side of (112) is obtained similarly. If the term $\exp(-\delta^2 O_p(n/J))$ is individually uniformly integrable, the left-hand side of (113) is bounded by $J\exp(-\delta^2 O(n/J))$. By Cauchy's convergence test (ratio test), we have a sufficient condition to ensure $J\exp(-\delta^2 O(n/J))$ to decrease to zero: $J^{1+\epsilon}/n \to 0$ for any $\epsilon > 0$. This guarantees assumption A3(a). Notice that since the logit model incurs no simulation error in the evaluation of $\boldsymbol{\xi}$, we do not need to take account of assumption A3(b) for the case of the logit model.[3]

In assumption A4, we simply assume that the instrumental matrices $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$ are respectively stochastically bounded.

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

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

in probability tending to one as $J \to \infty$. In the following, we examine what it means to have $\partial \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}'$ 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 $\boldsymbol{\theta}_c$, as noted on page 6. This means that the matrix $\partial \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}'$ takes the following form

$$\frac{\partial \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} \partial \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_d' & \boldsymbol{0} \\ \partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_d' & \partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_c' \end{pmatrix}. \tag{114}$$

This matrix is full-column rank if the components $\partial \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_d'$ and $\partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_c'$ are respectively of full-column rank, regardless of the value of $\partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_d'$. Moreover, we know that $\partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_c' = -J^{-1}\boldsymbol{Z}_c'\boldsymbol{W}$ by the definition of the cost side moment in (18) and the assumed linear dependence of $\boldsymbol{\omega}$ on $\boldsymbol{W}$ in (14). By properly choosing the cost side instruments $\boldsymbol{Z}_c$ and cost shifter $\boldsymbol{W}$, we can construct $\partial \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_c'$ to be of full-column rank for all $J$. Therefore we only need to check $\partial \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}_d'$ below. The first order derivative of $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ in terms of $\boldsymbol{\theta}_d$ can be rewritten as

$$\begin{aligned}&\frac{\partial \boldsymbol{G}_J(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)}{\partial \boldsymbol{\theta}_d'} \\ &= J^{-1}\boldsymbol{Z}_d' \frac{\partial \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)}{\partial \boldsymbol{\theta}_d'} \\ &= -J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)}{\partial \boldsymbol{\theta}_d'}\end{aligned} \tag{115}$$

since $\partial \boldsymbol{\sigma}(\cdot)/\partial \boldsymbol{\xi}' \cdot \partial \boldsymbol{\xi}/\partial \boldsymbol{\theta}_d' + \partial \boldsymbol{\sigma}(\cdot)/\partial \boldsymbol{\theta}_d' = \boldsymbol{0}$ from the implicit function theorem.

For the case of the logit model, we have

$$\boldsymbol{H}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P) = \boldsymbol{S} - \boldsymbol{s}\boldsymbol{s}', \quad \text{and} \quad \boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P) = \boldsymbol{S}^{-1} + \boldsymbol{i}\boldsymbol{i}'/s_0, \tag{116}$$

where $\boldsymbol{S} = \text{diag}[\boldsymbol{s}]$ and $\boldsymbol{i} = (1, \ldots, 1)'$. Furthermore,

$$\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)}{\partial \boldsymbol{\theta}_d'} = \begin{pmatrix} s_1(p_1 - \sum p_j s_j) & s_1(x_1 - \sum x_j s_j) \\ \vdots & \vdots \\ s_J(p_J - \sum p_j s_j) & s_J(x_J - \sum x_j s_j) \end{pmatrix}. \tag{117}$$

[3]A sufficient condition for assumption A3(b) could have been shown to be $J^{1+\epsilon}/R \to 0$ under condition S1(a) by the similarly way, but this condition would have to hold uniformly over $\boldsymbol{\theta}_d$.

Substituting (116) and (117) for (115) gives us $\partial \boldsymbol{G}^d_J(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}'_d = -J^{-1}(\sum \boldsymbol{z}^d_j p_j, \sum \boldsymbol{z}^d_j x_j)$. Therefore, unless the price $p_j$ is a linear function of the product characteristics $x_j$, $\partial \boldsymbol{G}^d_J(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)/\partial \boldsymbol{\theta}'_d$ with the logit model will be automatically of full column rank.

Assumption A6 can be verified by the similar way as A5, that is, to see whether the first order derivative of $\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))$ with respect to $\boldsymbol{\xi}$ is of full-rank, whereas the dimension of $\partial \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))/\partial \boldsymbol{\xi}'$ increases as $J$ grows large. In the logit model case, this matrix is of full-rank since $\partial \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))/\partial \boldsymbol{\xi}' = \boldsymbol{I}$.

In assumption A7, we guarantee that the profit margin $\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ shows the same distributional characteristics as $\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ as $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $P^R$ converge to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ and $P^0$ respectively. Using the logit model for demand determines the structure of the profit margin of product $j$ via the response of market share to the price change

$$\frac{\partial \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)}{\partial p_l} = \begin{cases} \theta_p \sigma_j (1-\sigma_j) & (l = j) \\ -\theta_p \sigma_j \sigma_l & (l \neq j) \end{cases} . \tag{118}$$

The profit margin of product $j$ with the logit model is calculated as

$$\begin{aligned} m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P) &= -\{\boldsymbol{\Delta}^{-1}\boldsymbol{\sigma}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\theta}_d, P)\}_j \\ &= -\frac{1}{\theta_p(1-\sum_{l\in\mathcal{J}^j_m} s_l)} \end{aligned} \tag{119}$$

where $\mathcal{J}^j_m$ denotes the set of all products of the firm that produces product $j$, i.e., $\mathcal{J}^j_m = \mathcal{J}_m$ if $j \in \mathcal{J}_m$, and $\theta_p$ in (109) is expected to be negative. The (119) implies that when we employ the logit model for demand, the profit margin is the same across the products one firm produces and is increasing in the firm's aggregate market share. Hence, we obtain the fact that $J/n \to 0$ guarantees assumption A7 under condition S1 as follows.

$$\begin{aligned} &J^{-1}||\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)||^2 \\ &= J^{-1}\sum_{m=1}^F \sum_{j\in\mathcal{J}_m} \left[m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\right]^2 \\ &= J^{-1}\sum_{m=1}^F \sum_{j\in\mathcal{J}_m} \left[-\frac{1}{\theta_p(1-\sum_{l\in\mathcal{J}^j_m} s^n_l)} + \frac{1}{\theta_p(1-\sum_{l\in\mathcal{J}^j_m} s^0_0)}\right]^2 \\ &= J^{-1}\theta_p^{-2}\sum_{m=1}^F J_m \left[\frac{\sum_{l\in\mathcal{J}_m}(s^n_l - s^0_l)}{(1-\sum_{l\in\mathcal{J}_m} s^n_l)(1-\sum_{l\in\mathcal{J}_{\Updownarrow}} s^0_l)}\right]^2 \\ &= J^{-1}\theta_p^{-2}\sum_{m=1}^F J_m \left[\frac{b_m}{1-b_m}\cdot\frac{1}{1-\sum_{l\in\mathcal{J}_m} s^0_l}\right]^2, \end{aligned}$$

where $b_m = \sum_{l\in\mathcal{J}_m}(s^n_l - s^0_l)/(1-\sum_{l\in\mathcal{J}_m} s^0_l)$. We know that $1/(1-\sum_{l\in\mathcal{J}_m} s^0_l) = O_p(1)$ by condition S1, and that $s^n_l - s^0_l = O_p(1/\sqrt{nJ})$ by assumption A1(a). Therefore, $b_m = J_m O_p(1/\sqrt{nJ}) \cdot O_p(1) \leq (J/\bar{c})O_p(1/\sqrt{nJ}) = O_p(\sqrt{J/n})$ by condition S1(b). This gives

$$\begin{aligned} &J^{-1}||\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)||^2 \\ &\leq \theta_p^{-2}\frac{F}{\bar{c}}\left[\frac{O_p(\sqrt{J/n})}{1-O_p(\sqrt{J/n})}\cdot O_p(1)\right]^2 \\ &= O_p(J/n), \end{aligned} \tag{120}$$

assuming the parameter associated with the price is negative and away from zero.

We next examine the asymptotic normality in Theorem 2. In Theorem 2, the variance of the GMM estimator consists of the three components, $\mathbf{\Phi}_1, \mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$, each of which is due to the randomness of the product characteristics, the sampling error, and the simulation error respectively. Assumption B4(a), (b), and (c) bound these variance components as $J$ goes to infinity. In the logit model case, $\mathbf{\Phi}_3 = \mathbf{0}$ because there is no need for simulation, and thus no simulation error. We focus on B4(b) here. Without loss of generality, we assume below that the instrument matrices, $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$, are respectively $J \times 1$ vectors. Then, since the conditional variance of $\epsilon_{ji}$ is given as $s_j^0(1-s_j^0)$ in assumption A1, $\mathbf{\Phi}_2$ generally takes the form of

$$\begin{aligned}\mathbf{\Phi}_2 \equiv& \begin{bmatrix} \Phi_2^{dd} & \Phi_2^{dc} \\ \Phi_2^{dc} & \Phi_2^{cc} \end{bmatrix} \\ =& \lim_{J,n\to\infty} \frac{1}{nJ} \times \\ & \mathrm{E}_{\mathrm{x},\xi} \begin{bmatrix} \sum_j (a_j^d)^2 s_j^0 - (\sum_j a_j^d s_j^0)^2 & \sum_j a_j^d a_j^c s_j^0 - (\sum_j a_j^d s_j^0)(\sum_j a_j^c s_j^0) \\ \sum_j a_j^d a_j^c s_j^0 - (\sum_j a_j^d s_j^0)(\sum_j a_j^c s_j^0) & \sum_j (a_j^c)^2 s_j^0 - (\sum_j a_j^c s_j^0)^2 \end{bmatrix} \end{aligned} \tag{121}$$

where $a_j^d$ and $a_j^c$ are respectively $j$th elements of $\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}$ and $-\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}$. If we simply use $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x}$ as the cost function in (14), the logit model derives

$$a_j^d = \frac{z_j^d}{s_j^0} + \frac{\sum_l z_l^d}{s_0^0}, \quad a_j^c = \frac{\sum_{l\in\mathcal{J}_m^j} z_l^c}{\theta_p^0(1-\sum_{l\in\mathcal{J}_m^j} s_l^0)^2}, \quad j = 1,\dots,J. \tag{122}$$

Let $\alpha(J) = J\bar{z}^d/s_0^0, \beta_j = z_j^d s_0^0/(J\bar{z}^d s_j^0)$ and thus $a_j^d = \alpha(J)(1+\beta_j)$, then

$$\begin{aligned} &\textstyle\sum_j (a_j^d)^2 s_j^0 - (\sum_j a_j^d s_j^0)^2 \\ =& \textstyle\sum_j \alpha(J)^2(1+\beta_j)^2 s_j^0 - (\sum_j \alpha(J)(1+\beta_j)s_j^0)^2 \\ =& \alpha(J)^2 \left[\textstyle\sum_j s_j^0 - (\sum_j s_j^0)^2 + 2(1-\sum_j s_j^0)(\sum_j \beta_j s_j^0) + \sum_j \beta_j^2 s_j^0 - (\sum_j \beta_j s_j^0)^2\right] \\ =& \alpha(J)^2 \left[s_0^0(1-s_0^0) + 2s_0^0 \textstyle\sum_j \beta_j s_j^0 - (\sum_j \beta_j s_j^0)^2 + \sum_j \beta_j^2 s_j^0\right] \\ \le& \alpha(J)^2 \left[s_0^0(1-s_0^0) + 2s_0^0 \max_j|\beta_j| \cdot \textstyle\sum_j s_j^0 + \max_j|\beta_j|^2 \cdot \sum_j s_j^0\right] \\ =& \alpha(J)^2(1-s_0^0)\left[s_0^0 + 2s_0^0\max_j|\beta_j| + \max_j|\beta_j|^2\right]. \end{aligned}$$

Assuming $z_j^d/\bar{z}^d = O_p(1)$, we have $\alpha(J) = O_p(J^2)$ and $\beta_j = O_p(1/J)$ under condition S1 and assumption A4. Then, the $(1,1)$ element of $\mathbf{\Phi}_2$ is

$$\Phi_2^{dd} = \lim_{J,n\to\infty} \frac{1}{nJ} \mathrm{E}_{\mathrm{x},\xi}\left[\textstyle\sum_j (a_j^d)^2 s_j^0 - (\sum_j a_j^d s_j^0)^2\right] = O(J^2/n). \tag{123}$$

By the similar calculation, we obtain $\Phi_2^{cc} = O_p(J/n)$ and $\Phi_2^{dc} = O_p(J^2/n)$. Therefore, we need to increase $n$ at least as fast as $J^2$ in order to bound $\mathbf{\Phi}_2$ finite.

Assumptions B4(d), (e), and (f) are the Lyapunov condition necessary to guarantee that the three terms in $J^{1/2}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$ follows asymptotically normal respectively. We just check assumption B4(e). Hence,

$$\begin{aligned} & n\,\mathrm{E}_{\epsilon,\mathrm{x},\xi}[||\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)||^{2+\delta}] \\ =& (n^{1+\delta}J^{(2+\delta)/2})^{-1}\,\mathrm{E}_{\epsilon,\mathrm{x},\xi}\left[\left\{(\textstyle\sum_j a_j^d\epsilon_{ji})^2 + (\sum_j a_j^c\epsilon_{ji})^2\right\}^{(2+\delta)/2}\right]. \end{aligned}$$

We obtain

$$\begin{aligned} |\textstyle\sum_j a_j^d\epsilon_{ji}| \le& \textstyle\sum_j |a_j^d\epsilon_{ji}| \le \max_{1\le j\le J}|a_j^d| \cdot \sum_j|\epsilon_{ji}| = \max_{1\le j\le J}|a_j^d| \cdot \sum_j |1(C_i=j) - s_j^0| \\ \le& \max_{1\le j\le J}|a_j^d| \cdot \textstyle\sum_j\{1(C_i=j)+s_j^0\} \\ \le& 2\max_{1\le j\le J}|a_j^d|. \end{aligned}$$

Similarly, we have $|\sum_j a_j^c \epsilon_{ji}| \leq 2\max_{1\leq j\leq J}|a_j^c|$. Under condition S1, $a_j^d$ and $a_j^c$ for the logit model given in (122) are respectively $O_p(J^2)$ and $O_p(J)$. Therefore,

$$
\begin{aligned}
& n\,\mathrm{E}_{\epsilon,\mathrm{x},\xi}[||\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)||^{2+\delta}] \\
\leq\ & (n^{1+\delta}J^{(2+\delta)/2})^{-1}\,\mathrm{E}_{\epsilon,\mathrm{x},\xi}\left[\{O_p(J^2)^2+O_p(J)^2\}^{(2+\delta)/2}\right] \\
=\ & \mathrm{E}_{\epsilon,\mathrm{x},\xi}\left[O_p\left(\frac{J^{3+3\delta/2}}{n^{1+\delta}}\right)\right].
\end{aligned} \tag{124}
$$

If we impose that $n$ increases as fast as $J^2$, i.e., $n=O(J^2)$, the Lyapunov condition B4(e) follows for $\delta>2$ by $(3+3\delta/2)-2(1+\delta)=1-\delta/2<0$.[4]

Finally, we examine assumption B5. The equicontinuity-like conditions in B5 guarantee that $\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^n,P^R)$ is well approximated by $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ near the neighborhood of $(\boldsymbol{\theta}^0,\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),P^0)$ and then the first order residual terms in Taylor approximation can be negligible as $J$ goes large. B5(b) and B5(d) are assumptions respectively on the demand and cost side residuals caused by the simulation error, and B5(d) is on the properties of the cost function $\boldsymbol{g}(\cdot)$ and of the profit margin $\boldsymbol{m}_g(\cdot)$ near $P^0$, they are all unnecessary to check in the logit model case. Then the remained to check are B5(a) and B5(c). Since the $j$th element of $\boldsymbol{Z}_d'\boldsymbol{H}^{-1}$ is $z_j^d/s_j+J\bar{z}^d/s_0$ for the logit model, the residual for the demand side moment in B5(a) evaluated at the observed market share $\boldsymbol{s}^n$ can be bounded as follows.

$$
\begin{aligned}
& |J^{-1/2}\boldsymbol{Z}_d'\{\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^0),\boldsymbol{\theta}_d,P^0)-\boldsymbol{H}_0^{-1}\}\boldsymbol{\epsilon}^n| \\
=\ & |J^{-1/2}\textstyle\sum_{j=1}^J\{z_j^d/s_j^n+J\bar{z}^d/s_0^n-z_j^d/s_j^0-J\bar{z}^d/s_0^0\}\epsilon_j^n| \\
=\ & |J^{-1/2}\textstyle\sum_{j=1}^J\{z_j^d(1/s_j^n-1/s_j^0)(s_j^n-s_j^0)+J\bar{z}^d(1/s_0^n-1/s_0^0)(s_j^n-s_j^0)\}| \\
=\ & |J^{-1/2}\textstyle\sum_{j=1}^J\{z_j^d(s_j^n-s_j^0)^2/(s_j^n s_j^0)+J\bar{z}^d(s_j^n-s_j^0)(s_0^n-s_0^0)/(s_0^n s_0^0)\}| \\
\leq\ & J^{-1/2}\textstyle\sum_{j=1}^J|z_j^d b_j^2/(1+b_j)|+J^{1/2}|\bar{z}^d|\cdot|b_0/(1+b_0)|\cdot(1/s_0^0)\cdot\sum_{j=1}^J|s_j^n-s_j^0| \\
\leq\ & J^{-1/2}\max_{1\leq j\leq J}|z_j^d|\cdot\textstyle\sum_{j=1}^J|b_j^2/(1+b_j)|+J^{1/2}|\bar{z}^d|\cdot|b_0/(1+b_0)|\cdot(s_j^0/s_0^0)\cdot\sum_{j=1}^J|b_j|,
\end{aligned} \tag{125}
$$

where $b_j=(s_j^n-s_j^0)/s_j^0$. From condition S1, assumptions A1(a), and A3(a), we have $b_j=O_p(\sqrt{J/n})=o_p(1)$. Thus $b_j^2/(1+b_j)=O_p(J/n)/(1+O_p(1))=O_p(J/n)$. Assuming $\max_j|z_j^d|=O_p(1)$, both of the first and second terms of the right hand side in the above inequality are $O_p(J^{3/2}/n)$. Therefore, we need $n$ to grow faster than $J^{3/2}$. For B5(c), let us abbreviate $s_{\mathcal{J}_m}^n=\sum_{j\in\mathcal{J}_m}s_j^n$ and $s_{\mathcal{J}_m}^0=\sum_{j\in\mathcal{J}_m}s_j^0$ and assume $\bar{z}_{\mathcal{J}_m}^c=J_m^{-1}\sum_{j\in\mathcal{J}_m}z_j^c=O_p(1)$, then

$$
\begin{aligned}
& |J^{-1/2}\boldsymbol{Z}_c'\{\boldsymbol{L}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^0),\boldsymbol{\theta}_d,P^0)\boldsymbol{M}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^0),\boldsymbol{\theta}_d,P^0)\boldsymbol{H}^{-1}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^0),\boldsymbol{\theta}_d,P^0) \\
& \quad -\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\}\boldsymbol{\epsilon}^n| \\
=\ & |J^{-1/2}\textstyle\sum_{j=1}^J\{a_j^c(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^0),\boldsymbol{\theta}_d,P^0)-a_j^c(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\}(s_j^n-s_j^0)| \\
=\ & \left|J^{-1/2}\sum_{m=1}^F\sum_{j\in\mathcal{J}_m}\left\{\frac{\sum_{l\in\mathcal{J}_m^j}z_l^c}{\theta_p(1-\sum_{l\in\mathcal{J}_m^j}s_l^n)^2}-\frac{\sum_{l\in\mathcal{J}_m^j}z_l^c}{\theta_p(1-\sum_{l\in\mathcal{J}_m^j}s_l^0)^2}\right\}(s_j^n-s_j^0)\right| \\
=\ & \left|J^{-1/2}\sum_{m=1}^F\theta_p^{-1}J_m\bar{z}_{\mathcal{J}_m}^c\left\{\frac{1}{(1-s_{\mathcal{J}_m}^n)^2}-\frac{1}{(1-s_{\mathcal{J}_m}^0)^2}\right\}(s_{\mathcal{J}_m}^n-s_{\mathcal{J}_m}^0)\right| \\
=\ & \left|J^{-1/2}\sum_{m=1}^F\theta_p^{-1}J_m\bar{z}_{\mathcal{J}_m}^c\left\{\frac{1}{1-2\frac{s_{\mathcal{J}_m}^n-s_{\mathcal{J}_m}^0}{1-s_{\mathcal{J}_m}^0}+\left(\frac{s_{\mathcal{J}_m}^n-s_{\mathcal{J}_m}^0}{1-s_{\mathcal{J}_m}^0}\right)^2}-1\right\}\frac{s_{\mathcal{J}_m}^n-s_{\mathcal{J}_m}^0}{(1-s_{\mathcal{J}_m}^0)^2}\right|
\end{aligned}
$$

[4]Obviously, if we allow $n$ to grow at the order of $J^3$, this requirement of $\delta>2$ can be relaxed to $\delta>0$ as BLP (1995) claimed.

$$
\begin{aligned}
&= \left| J^{-1/2} \sum_{m=1}^{F} \theta_p^{-1} J_m \bar{z}^c_{\mathcal{J}_m} \left\{ \frac{2\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}} - \left(\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}}\right)^2}{1-2\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}} + \left(\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}}\right)^2} \right\} \frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{(1-s^0_{\mathcal{J}_m})^2} \right| \\
&\leq |\theta_p|^{-1} J^{-1/2} \sum_{m=1}^{F} J_m |\bar{z}^c_{\mathcal{J}_m}| \left| \frac{2\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}} - \left(\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}}\right)^2}{1-2\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}} + \left(\frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{1-s^0_{\mathcal{J}_m}}\right)^2} \right| \cdot \left| \frac{s^n_{\mathcal{J}_m}-s^0_{\mathcal{J}_m}}{(1-s^0_{\mathcal{J}_m})^2} \right| \\
&= |\theta_p|^{-1} J^{-1/2} \cdot F \cdot O(J) \cdot O_p(1) \left( \frac{2 \cdot O_p\left(\sqrt{J/n}\right) - O_p(J/n)}{1 - 2 \cdot O_p\left(\sqrt{J/n}\right) + O_p(J/n)} \right) \cdot O_p\left(\sqrt{J/n}\right) \\
&= O_p(J^{3/2}/n) \qquad (126)
\end{aligned}
$$

where, by condition S1 and assumption A1(a), we use $(1-s^0_{\mathcal{J}_m})^{-1} = O_p(1)$ and $s^n_{\mathcal{J}_m} - s^0_{\mathcal{J}_m} = \sum_{\mathcal{J}_m}(s^n_l - s^0_l) = J_m O_p(1/\sqrt{nJ}) = O_p(\sqrt{J/n})$, and thus $(s^n_{\mathcal{J}_m} - s^0_{\mathcal{J}_m})/(1-s^0_{\mathcal{J}_m}) = O_p(\sqrt{J/n})$.

To summarize, when we use the logit model for demand, the rate of increase for $n$ relative to $J$ required to guarantee the consistency of the GMM estimator is of order of $J^{1+\epsilon}$ by the argument following (113) and (120), while the rate for the asymptotic normality is of order of $J^2$ based on the argument following (123)–(126).

We should note that, to guarantee the CAN property of the estimator in Theorems 1 and 2 for the use of the logit model, we have assumed that the number $J_m$ of the products produced by firm $m$ increases as the number $J$ of products in the market grows. Instead, the CAN property is equally obtained if we fix the number of products a firm produces to be one and increase the number $F$ of firms in the market, i.e, $J_m = 1$ and $F = J \to \infty$. As seen in (119), the logit model cannot have different profit margins across the products produced by the same firm, and accordingly, a number of empirical studies that use the logit model have assumed that each firm produces a product or a composite product in the market. This empirical use of the logit model implicitly assumes that the number of firms in the market grows. Nevertheless the CAN property of the logit estimates can be similarly obtained with the slight modification on the setup of Theorems 1 and 2.

As for Theorems 3 and 4, it should be noted that additional demographically-categorized purchasing information does not lend itself to finer and more accurate estimates for logit model. This is because, for logit model, consumers' demographic information are all summarized in the error term and is integrated out. As a result, individual purchasing probability for a product is the same accross consumers and agree with the market share.

Therefore we defer to the next subsection of the random coefficient logit model on the examination of how fast the number $T$ of consumers drawn to match the observed demographically-categorized purchasing information must increase relative to the number $J$ of products on the market and the number $R$ of consumers used to simulation in order for us to have Theorems 3 and 4. We also see that the number $N$ of the sample size to calculate such purchasing information must increase infinitely relative to $T$.

## Random Coefficient Logit Model

In what follows, we assume a random coefficient logit model with one random coefficient:

$$
u_{ij} = \delta_j + \theta^u_x \nu^x_i x_j + \nu_{ij} \quad \text{with} \quad \delta_j = \theta_p p_j + \theta_x x_j + \xi_j \qquad (127)
$$

where $\nu^x_i$ represents consumer $i$'s random preference on the characteristic $x_j$ relative to the price. The parameter $\theta^u_x$ shows the magnitude of the preference, and when $\theta^u_x = 0$, the model is simple logit model. Provided that $\nu_{ij}$'s are i.i.d. extreme value, the probability $\sigma_{ij}$ of consumer $i$ choosing product $j$ is given by

$$\sigma_{ij}(\boldsymbol{\xi}, \nu_i; \boldsymbol{\theta}_d) = \frac{\exp(\delta_j + \theta_x^u \nu_i^x x_j)}{1 + \sum_{k=1}^J \exp(\delta_k + \theta_x^u \nu_i^x x_k)}. \tag{128}$$

The market share of product $j$ is obtained by integrating (128) in terms of $\nu_i^x$ over the population $P^0$. We simulate it with a random sample of $R$ individuals as

$$\sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}, P^R) \equiv \frac{1}{R}\sum_{r=1}^R \sigma_{rj}(\boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) = \frac{1}{R}\sum_{r=1}^R \frac{\exp(\delta_j + \theta_x^u \nu_r^x x_j)}{1 + \sum_{k=1}^J \exp(\delta_k + \theta_x^u \nu_r^x x_k)} \tag{129}$$

In the following, we put forward Condition S2 on the magnitude of the individual choice probability stronger than Condition S1(a). Although the condition makes individual's behavior restrictive, this treatment allows us to calculate the rate of $n$, $R$, $N$, and $T$ relative to $J$, at which the random coefficient logit model follows our limiting theorems.

**Condition S2** *For all consumer $r$ with the demographics $\boldsymbol{\nu}_r$, and for all possible value of the product characteristics $(\boldsymbol{X}, \boldsymbol{\xi})$, there exists positive finite constants $\underline{c}$ and $\overline{c}$ such that with probability one*

$$\begin{aligned} \frac{\underline{c}}{J} &\leq \inf_{\theta_d \in \Theta_d} \sigma_{rj}(\boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) \\ &\leq \sup_{\theta_d \in \Theta_d} \sigma_{rj}(\boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d) \leq \frac{\overline{c}}{J}, \quad j = 0, 1, \ldots, J. \end{aligned} \tag{130}$$

Obviously, Condition S2 is a sufficient condion of Condition S1(a) 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 Condition S1(a). With Condition S2, the individual choice probability $\sigma_{rj}(\boldsymbol{\xi}, \boldsymbol{\nu}_r; \boldsymbol{\theta}_d)$ and its inverse are respectively $O_p(1/J)$ and $O_p(J)$. We assume that our two sets of simulation draws of individuals $\boldsymbol{\nu}_r, r = 1, \ldots, R$ and of the individuals $\boldsymbol{\nu}_t, t = 1, \ldots, T$ satisfy condition S2.

As stated above, the random coefficient logit model has no closed-form solution to the inverse of $\boldsymbol{H}$. However, under condition S2, we can approximate it by

$$\begin{aligned} &\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R) + \frac{1}{\sigma_0(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)}(1 + O_p(1/J))\, \boldsymbol{i}\boldsymbol{i}', \end{aligned} \tag{131}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \mathrm{diag}(\sigma_1(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \ldots, \sigma_J(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P))$. In the appendix of Berry, Linton, and Pakes (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 (46) 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 for $R$. Furthermore, in the appendix of this paper, we show that the arguments above apply to our supply side specification too. Therefore, for Theorem 2 to hold for the random coefficient logit model, the number $n$ of the sample size for caluculating the observed market share must increase at the rate of $J^2$ and the number $R$ of the simulation draws must increase at the rate of $J^2$ as well.

Applicability of assumptions A5 and A6 in Theorem 1 to the random coefficient logit model would have to be checked via numerical computations on a case-by-case basis because these assumptions require us to examine full-rankness of the matrices that contain the inverse of $\boldsymbol{H}$. Assumption A7, on the other hand, can be verified relatively easily using (131).

Now we turn our attention to cases 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 $\theta_x^u$ in (127) more accurately by using the information on consumers who choose specific sets of attributes in products. Denote the set of products having this attribute by $\mathcal{Q}$. Hereinafter, assume that we have a consistent estimate $\eta^N$, which was constructed from $N$ independent consumer draws from the population $P^0$, separate from the $n$ independent draws from $P^0$ for calculating the observed market share, with the expectation $\eta^0$ of $\nu_i^x$ conditional on the individual choosing a product in $\mathcal{Q}$. We further assume that $\eta^N$ satisfies assumption A10, that is, $\eta^N$ has the conditional expectation, corresponding to (83) but written in the spirit of (75),

$$\eta^0 = \mathrm{E}[\nu_t^x | C_t \in \mathcal{Q}, \boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \tag{132}$$

and the conditional variance of order $O_p(1/N)$ for (84). Given $\eta^N$, we will draw $T$ individuals from the population $P^0$ to construct an additional moment,

$$\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) = \eta^N - \frac{1}{T}\sum_{t=1}^{T} \nu_t^x \psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) \tag{133}$$

where $\psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) = \sum_{j\in\mathcal{Q}} \sigma_{tj}(\boldsymbol{\xi}, \boldsymbol{\nu}_t, \boldsymbol{\theta}_d) / \sum_{j\in\mathcal{Q}} \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$. In the following, we will derive the condition to guarantee that the specification above satisfies the assumptions in Theorems 3 and 4 under Condition S2.

On assumption A8, we require that the $1 \times 3$ matrix $\partial \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)/\partial \boldsymbol{\theta}_d'$ is of full colmun rank. We can rewrite this matrix as

$$\begin{aligned} &\frac{\partial \boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)}{\partial \boldsymbol{\theta}_d'} \\ &= \frac{1}{T}\sum_{t=1}^{T} \nu_t^x \left[ \frac{\partial \psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}{\partial \boldsymbol{\xi}'} \boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0) \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}{\partial \boldsymbol{\theta}_d'} + \frac{\psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0)}{\partial \boldsymbol{\theta}_d'} \right] \Bigg|_{\xi = \xi(\theta_d, s^0, P^0)}. \end{aligned}$$

Here, the component $\boldsymbol{H}^{-1}$ has no closed form expression, while we can approximate it within $O_p(1/J)/\sigma_0$ error by taking $R \to \infty$ in (131). As a result, to verify assumption A8, we would have to have the representative utility $\delta_j$, consumer's random preference $\nu_i^x$, and its associated parameter value $\theta_x^u$ fixed. We will check the singularity of $\partial \boldsymbol{G}_{J,T}^a / \partial \boldsymbol{\theta}_d'$ in our computational example in the next section.

For assumption A9, 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 S2.

To check assumption A11, we decompose

$$\begin{aligned} &T^{-1/2} || \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) || \\ &\leq T^{-1/2} || \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) || \\ &\quad + T^{-1/2} || \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) || \end{aligned} \tag{134}$$

where $\boldsymbol{\Psi} = (\psi_1, \ldots, \psi_T)'$ is a $T \times 1$ matrix. The square of the first term in (134) is bounded by

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

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

where $\boldsymbol{\xi}^*$ is between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$. In the proof of Theorem 1 (equation (A.6)), we have shown that $J^{-1}||\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)||^2 = o_p(1)$. Thus, it remains to show that $||\partial \boldsymbol{\Psi}(\boldsymbol{\xi}^*, \boldsymbol{\theta}_d, P^R)/\partial \boldsymbol{\xi}'||^2 = O_p(T/J)$ to guarantee this whole term to be $o_p(1)$. For the random coefficient logit model, we obtain the $j$th element of $\boldsymbol{\Upsilon}_t$ as

$$\begin{aligned}
\{\boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\}_j &\equiv \frac{\partial \psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)}{\partial \xi_j} \\
&= \frac{\sigma_{tj}(1\{j \in \mathcal{Q}\} - \sum_{k\in\mathcal{Q}} \sigma_{tk})}{\sum_{k\in\mathcal{Q}} \sigma_k} \\
&\quad - \frac{\sum_{k\in\mathcal{Q}} \sigma_{tk}}{\sum_{k\in\mathcal{Q}} \sigma_k} \cdot \frac{1\{j \in \mathcal{Q}\} \int \sigma_{rj} dP - \sum_{k\in\mathcal{Q}} \int \sigma_{rj}\sigma_{rk} dP}{\sum_{k\in\mathcal{Q}} \sigma_k} \qquad (135)
\end{aligned}$$

where $\sigma_{rj} = \sigma_{rj}(\boldsymbol{\xi}, \boldsymbol{\nu}_r, \boldsymbol{\theta}_d)$, $\sigma_{tj} = \sigma_{tj}(\boldsymbol{\xi}, \boldsymbol{\nu}_t, \boldsymbol{\theta}_d)$ and $\sigma_j = \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$. Under Condition S2, both of $\sigma_{rj}$ and $\sigma_j$ are $O_p(1/J)$, while $\sum_{j\in\mathcal{Q}} \sigma_j$ and $1/\sum_{j\in\mathcal{Q}} \sigma_j$ are both $O_p(1)$. Thus, we have $\partial \psi_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)/\partial \xi_j = O_p(1/J)$, and so

$$\begin{aligned}
\left|\left|\frac{\partial \boldsymbol{\Psi}(\boldsymbol{\xi}^*, \boldsymbol{\theta}_d, P)}{\partial \boldsymbol{\xi}'}\right|\right|^2 &= \sum_{j=1}^{J}\sum_{t=1}^{T}\left(\frac{\partial \psi_t(\boldsymbol{\xi}^*, \boldsymbol{\theta}_d, P)}{\partial \xi_j}\right)^2 = J \cdot T \cdot O_p(1/J)^2 \\
&= O_p(T/J).
\end{aligned}$$

The square of the second term of (134) is

$$\begin{aligned}
& T^{-1}||\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)||^2 \\
&= T^{-1}\sum_{t=1}^{T}\{\psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \psi_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\}^2 \\
&= T^{-1}\sum_{t=1}^{T}\left\{\frac{\sum_{j\in\mathcal{Q}} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t, \boldsymbol{\theta}_d)}{\sum_{j\in\mathcal{Q}} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R)} - \frac{\sum_{j\in\mathcal{Q}} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t, \boldsymbol{\theta}_d)}{\sum_{j\in\mathcal{Q}} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)}\right\}^2 \\
&= \left\{\frac{\sum_{j\in\mathcal{Q}}\{\sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) - \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)\}}{\sum_{j\in\mathcal{Q}} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^R) \cdot \sum_{j\in\mathcal{Q}} \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)}\right\}^2 \\
&\qquad \times T^{-1}\sum_{t=1}^{T}\left\{\sum_{j\in\mathcal{Q}} \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t, \boldsymbol{\theta}_d)\right\}^2 \\
&= \left\{\frac{\sum_{j\in\mathcal{Q}} O_p(1/\sqrt{RJ})}{\sum_{j\in\mathcal{Q}} O_p(1/J) \cdot \sum_{j\in\mathcal{Q}} O_p(1/J)}\right\}^2 \times T^{-1}\sum_{t=1}^{T}\left\{\sum_{j\in\mathcal{Q}} O_p(1/J)\right\}^2 \\
&= O_p(J/R)
\end{aligned}$$

under assumption A1(b) and Condition S2. As a result, $R$ is required to grow slightly faster than $J$.

We next move on to assumptions in Theorem 4. For assumption B7(a), it is sufficient to show that two components in the norm of (95) is respectively $o_p(1)$. Write $\sigma_j^R = \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ and $\sigma_j^T = \sigma_j(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^T)$ for notational simplicy, and then we approximate the $j$th element of $T^{-1}\sum_{t=1}^{T} \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)$ by using $\boldsymbol{H}^{-1}$ in (131) and $\partial \psi_t/\partial \xi_j$ in (135) as follows.

$$\left\{T^{-1}\sum_{t=1}^{T} \boldsymbol{\Upsilon}_t(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)\right\}_j$$

$$
\begin{aligned}
&= T^{-1}\sum_{t=1}^{T}\sum_{l=1}^{J}\frac{\partial\psi_t(\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)}{\partial\xi_l}H_{lj}^{-1}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)\\
&= \left[\frac{T^{-1}\sum_{t=1}^{T}(\sum_{l\in\mathcal{Q}}\sigma_{tl})\sigma_{t0}}{\sum_{l\in\mathcal{Q}}\sigma_l^R}-\frac{(\sum_{l\in\mathcal{Q}}\sigma_l^T)R^{-1}\sum_{r=1}^{R}(\sum_{l\in\mathcal{Q}}\sigma_{rl})\sigma_{r0}}{(\sum_{l\in\mathcal{Q}}\sigma_l^R)^2}\right]\cdot\frac{1}{\sigma_0^R}(1+O_p(1/J))\\
&\quad+\left[\frac{\sigma_j^T\cdot 1\{j\in\mathcal{Q}\}-T^{-1}\sum_{t=1}^{T}(\sum_{l\in\mathcal{Q}}\sigma_{tl})\sigma_{tj}}{\sum_{l\in\mathcal{Q}}\sigma_l^R}\right.\\
&\quad\left.-\frac{(\sum_{l\in\mathcal{Q}}\sigma_l^T)\{\sigma_j^R\cdot 1\{j\in\mathcal{Q}\}-R^{-1}\sum_{r=1}^{R}(\sum_{l\in\mathcal{Q}}\sigma_{rl})\sigma_{rj}\}}{(\sum_{l\in\mathcal{Q}}\sigma_l^R)^2}\right]\cdot\frac{1}{\sigma_j^R}.
\end{aligned}
\tag{136}
$$

As for the first component of (95), under Condition S2, we obtain from (136),

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

We can also obtain for the second component, $||T^{-1}\sum_{t=1}^{T}\boldsymbol{\Upsilon}_t^0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n|| = O_p(\sqrt{J/n})$ using (131) and (135) with $(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)=(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$. As a whole, we have

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

$$= O_p\left(\sqrt{T \cdot J/n}\right).$$

Therefore, we have to increase $n$ faster than $TJ$. We notice that the requirement above for assumption B7(a) is stronger than what is required for theorem 2, that is, $n$ grows faster than $J^2$, because we assume the number $T$ of consumers used in evaluating the additional moment is greater than the number $J$ of the products in the market.

As for assumption B7(b), through a quite similar calculation as for assumption B7(a), we can show that the number $R$ of simulation draws is needed to grow faster than $TJ$.

We can easily see that assumption B7(c) requires $R$ grows faster than $TJ$ as follows.

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

In assumption B10(a), we need to keep the variance of $\sum_{i=1}^n \boldsymbol{Y}_{J,T,i}^{a0}$, which is the residual component in the additional moment $T^{1/2}\boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d^0)$ in terms of the sampling error, bounded. Write $\sigma_{tj}^0 = \sigma_{tj}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t, \boldsymbol{\theta}_d^0)$ and $\sigma_j^0 = \sigma_j(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$, then

$$\begin{aligned}
a_j^{a0} &\equiv a_j^a(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \\
&= \{-\textstyle\sum_{t=1}^T \nu_t^x \boldsymbol{\Upsilon}_t^0 \boldsymbol{H}_0^{-1}\}_j \\
&= -\frac{\sum_{t=1}^T \nu_t^x \sigma_{tj}^0(1\{j\in\mathcal{Q}\} - \sum_{l\in\mathcal{Q}}\sigma_{tl}^0)}{\sum_{l\in\mathcal{Q}} s_l^0}\cdot\frac{1}{s_j^0} - \frac{\sum_{t=1}^T \nu_t^x \sigma_{t0}^0 \sum_{l\in\mathcal{Q}}\sigma_{tl}^0}{\sum_{l\in\mathcal{Q}} s_l^0}\cdot\frac{1}{s_0^0}(1+O_p(1/J)) \\
&= \alpha(1+\beta_j+O_p(1/J)) \qquad (137)
\end{aligned}$$

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 $\alpha$ and $\beta$ are respectively $O_p(T)$ and $O_p(1)$ under Condition S2. Using $a_j^{a0}$ calculated above, the expectation of the principal component of $\boldsymbol{Y}_{J,T,i}^{a0}$ with respect to $\epsilon_{ji}$ conditional on $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$ is calculated as follows.

$$\begin{aligned}
&\textstyle\sum_{j=1}^J (a_j^{a0})^2 s_j^0 - (\sum_{j=1}^J a_j^{a0} s_j^0)^2 \\
&= \textstyle\sum_{j=1}^J \alpha^2(1+\beta_j+O_p(1/J))^2 s_j^0 - \{\sum_{j=1}^J \alpha(1+\beta_j+O_p(1/J))s_j^0\}^2 \\
&= \alpha^2\left[s_0^0(1-s_0^0)(1+O_p(1/J))^2 + 2(\textstyle\sum_{j=1}^J \beta_j s_j^0)s_0^0(1+O_p(1/J)) + \sum_{j=1}^J \beta_j^2 s_j^0 - (\sum_{j=1}^J \beta_j s_j^0)^2\right] \\
&\leq \alpha^2\left[s_0^0(1-s_0^0)(1+O_p(1/J))^2 + 2\max_j|\beta_j|\cdot(\textstyle\sum_{j=1}^J s_j^0)s_0^0(1+O_p(1/J)) + \max_j|\beta_j|^2\cdot\sum_{j=1}^J s_j^0\right] \\
&= \alpha^2(1-s^0)\left[s_0^0(1+O_p(1/J))^2 + 2\max_j|\beta_j|s_0^0(1+O_p(1/J)) + \max_j|\beta_j|^2\right].
\end{aligned}$$

Substituting $\alpha = O_p(T)$ and $\beta_j = O_p(1)$, we further obtain

$$\begin{aligned}
&\textstyle\sum_{j=1}^J (a_j^{a0})^2 s_j^0 - (\sum_{j=1}^J a_j^{a0} s_j^0)^2 \\
&= O_p(T)^2(1-O_p(1/J))\left[O_p(1/J)(1+O_p(1/J))^2 + 2O_p(1)O_p(1/J)(1+O_p(1/J)) + O_p(1)^2\right] \\
&= O_p(T^2).
\end{aligned}$$

Therefore the variance of $\sum_{i=1}^n \boldsymbol{Y}_{J,T,i}^{a0}$ is

$$
\begin{aligned}
& \mathrm{V}_{\epsilon,\nu,\mathrm{x},\xi}[\textstyle\sum_{i=1}^n \boldsymbol{Y}_{J,T,i}^{a0}] \\
= \ & \textstyle\sum_{i=1}^n \mathrm{E}_{\epsilon,\nu,\mathrm{x},\xi}[(1/n^2T)(\sum_{j=1}^J a_j^{a0}\epsilon_{ji})^2] \\
= \ & (1/nT)\,\mathrm{E}_{\nu,\mathrm{x},\xi}\Big[\textstyle\sum_{j=1}^J (a_j^{a0})^2\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\epsilon_{ji}^2|\mathrm{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)] \\
& \qquad + \textstyle\sum_{j\neq k} a_j^{a0}a_k^{a0}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\epsilon_{ji}\epsilon_{ki}|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)]\Big] \\
= \ & (1/nT)\,\mathrm{E}_{\nu,\mathrm{x},\xi}\Big[\textstyle\sum_{j=1}^J (a_j^{a0})^2 s_j^0(1-s_j^0) - \sum_{j\neq k} a_j^{a0}a_k^{a0}s_j^0s_k^0\Big] \\
= \ & (1/nT)\,\mathrm{E}_{\nu,\mathrm{x},\xi}\Big[\textstyle\sum_{j=1}^J (a_j^{a0})^2 s_j^0 - (\sum_{j=1}^J a_j^{a0}s_j^0)^2\Big] \\
= \ & (1/nT)\,\mathrm{E}_{\nu,\mathrm{x},\xi}[O_p(T^2)] \\
= \ & \mathrm{E}_{\nu,\mathrm{x},\xi}[O_p(J/n)].
\end{aligned}
$$

To keep this variance bounded, $n$ is needed to grow as fast as $J$.

Similar calculation holds for assumption B10(c) and derives that $R$ is required to grow as fast as $J$.

We assume in A10(a) that the additional information $\eta^N$ is $\sqrt{N}$ consistent with $\eta^0$. In assumption B10(d), we bound the variance of the residual term in the additional moment $T^{1/2}\boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d^0)$ corresponding to the sampling error contained in the additional information. We see

$$
\begin{aligned}
N\,\mathrm{V}_{\epsilon^\#,\mathrm{x},\xi}[T^{1/2}N^{-1}\epsilon_{i'}^\#] &= \mathrm{E}_{\mathrm{x},\xi}[\mathrm{V}_{\epsilon^\#|\mathrm{x},\xi}[T^{1/2}(\eta^N-\eta^0)|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)]] \\
&= \mathrm{E}_{\mathrm{x},\xi}[O_p(T/N)].
\end{aligned}
$$

To hold B10(d), we require that the sample size $N$ for additional information grows as fast as the sample size $T$ of our consumer draws in constructing the additional moment does.

Assumption B10(f) gives the Lyapunov condition the residual term $\sum_{i=1}^n \boldsymbol{Y}_{J,T,i}^{a0}$ in the additional moment follows. Since $a_j^{a0}$ in (137) is $O_p(T)$ under Condition S2, we obtain

$$
\begin{aligned}
& n\,\mathrm{E}_{\epsilon,\nu,\mathrm{x},\xi}[||\boldsymbol{Y}_{J,T,i}^{a0}||^{2+\delta}] \\
= \ & \tfrac{1}{n^{1+\delta}T^{(2+\delta)/2}}\,\mathrm{E}_{\epsilon,\nu,\mathrm{x},\xi}[|\textstyle\sum_{j=1}^J a_j^{a0}\epsilon_{ji}|^{2+\delta}] \\
\leq \ & \tfrac{1}{n^{1+\delta}T^{(2+\delta)/2}}\,\mathrm{E}_{\nu,\mathrm{x},\xi}[2^{2+\delta}\max_j|a_j^{a0}|^{2+\delta}] \\
= \ & \mathrm{E}_{\nu,\mathrm{x},\xi}[O_p(n^{-(1+\delta)}T^{(2+\delta)/2})].
\end{aligned}
$$

Substituting $n = O(T^k)$ and solving $(2+\delta)/2 - k(1+\delta) < 0$ gives $k > 1$ for any $\delta > 0$, which means that $n$ is necessary to grow faster than $T$.

By similar argument for assumption B10(g) and B10(h), we obtain the fact that $R$ and $N$ are required to grow faster than $T$ respectively.

In summary, for the random coefficient logit model, the estimator with the additional moment has consistency in Theorem 3 when $n$ and $R$ grow faster than $J$. The asymptotic normality in Theorem 4, on the other hand, requires that $n$ and $R$ to grow faster than $TJ$. Moreover, $N$ has to grow faster than $T$.

# 6 Computational Results

In this section, we run Monte Carlo experiments to evaluate the theorems derived in the previous sections. By repeatedly estimating a demand and supply system with randomly generated data sets, we verify the asymptotic normality of the GMM estimator. Through experiments, we examine how the sampling and simulation errors in the observed data and the simulated market share affect the results. Furthemore, we show that the use of additional consumer purchasing information well contributes the accuracy of the resulting random coefficient estimate.

The consumer's utility function we specify here is the following random coefficient logit model.

$$u_{ij} = -\alpha p_j + \beta x_j \nu_i^o + \xi_j + \epsilon_{ij} \tag{138}$$

where the unobserved quality $\xi_j$ and the exogenous product characteristics $x_j$ are respectively random draws from $N(0,1)$ and $N(1,1)$. Unless otherwise stated, the random draws in the data set are i.i.d. The price of product $p_j$ is, on the other hand, treated as endogenous and then determined in the market. The $\nu_i^o$ is a consumer's taste for $x_j$ and distributed from $N(0,1)$. The $\epsilon_{ij}$'s are i.i.d. extreme value draws. We set the demand side parameters $\alpha = 1.0$ and $\beta = 1.0$. The market share $\sigma_j$ is calculated by

$$\sigma_j = \int \frac{\exp(-\alpha p_j + \beta x_j \nu_i^o + \xi_j)}{1 + \sum_{l=1}^{J} \exp(-\alpha p_l + \beta x_l \nu_i^o + \xi_l)} P(d\nu_i^o). \tag{139}$$

The true market share $s_j^0$ is obtained by evaluating (139) with the underlying distribution $P^0$ of $\nu_i^o$. We draw 10,000 consumers from N(0,1) as the underlying population.

For the supply side, we assume there exist five oligopolistic suppliers in the market and they produce the same number of products. These suppliers are assumed to have the same cost function

$$c_j = x_j \gamma + \omega_j \tag{140}$$

where the unobserved cost shifter $\omega_j$ is a random draw from $N(0,1)$. For cost side parameter, we set $\gamma = 1.5$. At the Bertrand-Nash equilibrium, the suppliers determine the price of their products to satisfy

$$\boldsymbol{f}(\boldsymbol{p}) = \boldsymbol{c} - \boldsymbol{p} - \boldsymbol{\Delta}^{-1}\boldsymbol{\sigma} = \boldsymbol{0} \tag{141}$$

under the population $P^0$. The $(j,k)$ element of the $J \times J$ gradient matrix $\boldsymbol{\Delta}$ is given by

$$\Delta_{jk} = \begin{cases} \partial\sigma_k/\partial p_j, & \text{if the products } j \text{ and } k \text{ are} \\ & \text{produced by the same firm;} \\ 0, & \text{otherwise.} \end{cases} \tag{142}$$

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 (141), that is, $J$ dimensional nonlinear simultaneous equations. In practice, an iteration algorithm is required to solve (141), and we adopt the Newton-Raphson method.

We first estimate the system of demand and supply given in (139) and (140) by the BLP framework. To estimate the models, we construct the three instrumental variables from $x_j$, one is $x_j$ itself, one is the company average of $x_j$, and one is the average of $x_j$ over other companies. Table 2 gives the result for the mean estimated values across 100 Monte Carlo experiments when $n = \infty$ fixed, i.e., the observed market shares have no sampling error. Each column corresponds to the different number $J$ of products, while each row corresponds to the different number $R$ of consumer draws used in the simulation process. The values in parenthesis show the simulated standard error—the standard error of the estimated parameters across the simulation. In the table, we can observe the simulated standard errors of parameters decrease as $J$ increases. For $J$ fixed, the increasing $R$ also contributes the reduction of the standard errors. The standard error

Table 2: Monte Carlo Results for the BLP Framework, 100 repetitions, $n = \infty$

| # of Consumer Draws ($R$) | $\alpha(1.0)$ # of products ($J$) 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| 10 | 0.974 | 0.953 | 0.952 | 0.934 |
| | (0.266) | (0.174) | (0.138) | (0.134) |
| 50 | 0.974 | 0.990 | 0.989 | 0.971 |
| | (0.166) | (0.110) | (0.079) | (0.060) |
| 100 | 0.982 | 0.997 | 0.989 | 0.979 |
| | (0.156) | (0.123) | (0.058) | (0.045) |
| $10J$ | 0.982 | 0.993 | 0.994 | 0.982 |
| | (0.156) | (0.099) | (0.056) | (0.036) |
| $J^2$ | 0.982 | 0.988 | 0.992 | 0.982 |
| | (0.156) | (0.093) | (0.055) | (0.035) |

| # of Consumer Draws ($R$) | $\beta(1.0)$ # of products ($J$) 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| 10 | 1.303 | 1.385 | 1.223 | 1.177 |
| | (1.207) | (1.172) | (0.909) | (0.760) |
| 50 | 0.957 | 0.983 | 0.958 | 0.936 |
| | (0.702) | (0.539) | (0.406) | (0.306) |
| 100 | 0.909 | 0.981 | 0.912 | 0.935 |
| | (0.749) | (0.692) | (0.363) | (0.274) |
| $10J$ | 0.909 | 0.919 | 0.887 | 0.900 |
| | (0.749) | (0.543) | (0.347) | (0.238) |
| $J^2$ | 0.909 | 0.930 | 0.886 | 0.896 |
| | (0.749) | (0.605) | (0.325) | (0.240) |

| # of Consumer Draws ($R$) | $\gamma(1.5)$ # of products ($J$) 10 | 25 | 50 | 100 |
|---|---|---|---|---|
| 10 | 1.558 | 1.543 | 1.546 | 1.518 |
| | (0.388) | (0.265) | (0.191) | (0.176) |
| 50 | 1.595 | 1.609 | 1.602 | 1.574 |
| | (0.316) | (0.164) | (0.121) | (0.089) |
| 100 | 1.583 | 1.613 | 1.605 | 1.582 |
| | (0.246) | (0.164) | (0.101) | (0.071) |
| $10J$ | 1.583 | 1.614 | 1.610 | 1.586 |
| | (0.246) | (0.158) | (0.097) | (0.073) |
| $J^2$ | 1.583 | 1.610 | 1.608 | 1.587 |
| | (0.246) | (0.156) | (0.098) | (0.073) |

Standard error across repetitions stands in the parenthesis.

for $\beta$ is much larger than those for $\alpha$ and $\gamma$. This is because $\beta$ is the coefficient for the random term depending on the consumer taste $\nu_i^o$ as well as the product characteristics $x_j$ and thus the uncorrelated relationship between the unobserved quality $\xi_j$ and the instrumental variables involves less information on $\beta$. In particular, when the number of simulation draws is small ($R = 10$), the estimated value of $\beta$ is upwardly biased.

Table 3 gives the result for the case where the observed market share $s_j^n$ contains the sampling error. Here, we fixed the number $R = 100$ for the simulation draws of consumer. We construct the observed market share $s_j^n$ from a multinomial sample of size $n$ with the response probability $(s_0^0, \ldots, s_J^0)$. When $n$ is not large enough, there are zero-share products. We remove these products in estimating parameters. In the table, we observe the larger $n$ becomes, the smaller the simulated standard error becomes for any fixed $J$.

We next implement the Monte Carlo simulation for the extended framework with the additional moments. As the additional moment, we suppose to have the information on (1) the expected value of $\nu_i^o$ over consumers who choose products priced higher than the average price; and (2) the expected value of $\nu_i^o$ over consumers who choose products with $x_j$ greater than the average of $x_j$. That is, the additional moments are

$$\eta_1^0 = \mathrm{E}[\nu_i^o | C_i \in \mathcal{Q}\{p_j \geq \bar{p}\}, x, \xi], \tag{143}$$

$$\eta_2^0 = \mathrm{E}[\nu_i^o | C_i \in \mathcal{Q}\{x_j \geq \bar{x}\}, x, \xi] \tag{144}$$

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 with $x$ greater than the average $\bar{x}$.

Table 4 is the result for the case where we know the expected values in (143) and (144) exactly and no sampling error in the additional information ($N = \infty$). To calculate the additional sample moments, we draws $T$ consumers from the population and then calculate the conditional average of $\nu_i^o$ by using their purchasing probabilities. To make the effect of the additional moments clear, we use the true market share $s_j^0$ as the observed market share ($n = \infty$) and fix $R = 100$. The result indicates if the number of consumer draws $T$ is large enough, the additional information considerably reduce the standard error of $\beta$. For the case of $J = 50, T = 1000$, the standard error of $\beta$ with the additional moments decreases to 0.137 in table 4 from 0.363, which is the value without the additional moments in table 2 ($R = 100$ row, $J = 50$ column). On the other hand, if $T$ is small, the standard error of $\beta$ increases rather than decreases by using the additional moments. The standard error of $\beta$ at $T = 50$ and $J = 50$ increase to 0.392 in table 4 from 0.363 in table 2. Moreover, the additional moments have slight influence on the standard errors of $\alpha$ and $\gamma$ in any value of $T$. This is because the additional information is on the consumer's taste $\nu_i^o$ and contains less information on $\alpha$ and $\gamma$.[5]

We next consider the case where the additional information contain the sampling error. Drawing $N$ consumers from the population, we use the following estimate $\eta^N$ instead of $\eta^0$ as the additional information,

$$\eta_1^N = \sum_{i'=1}^{N} \frac{\nu_{i'}^o \cdot 1\{C_{i'} \in \mathcal{Q}\{p_j \geq \bar{p}\}\}}{N_p} \tag{145}$$

$$\eta_2^N = \sum_{i'=1}^{N} \frac{\nu_{i'}^o \cdot 1\{C_{i'} \in \mathcal{Q}\{x_j \geq \bar{x}\}\}}{N_x}. \tag{146}$$

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 and the number of consumers who choose the product with $x$ greater than the average. This estimators are

[5]The first order derivatives of the additional moments in terms of $\alpha$ are almost zero, while that for $\gamma$ is just zero.

Table 3: Monte Carlo Results for the BLP Framework, 100 repetitions, $R = 100$

| # of Consumer Draws ($n$) | $\alpha(1.0)$ # of products ($J$) | | | | # of Consumer Draws ($n$) | $\beta(1.0)$ # of products ($J$) | | | | # of Consumer Draws ($n$) | $\gamma(1.5)$ # of products ($J$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 25 | 50 | 100 | | 10 | 25 | 50 | 100 | | 10 | 25 | 50 | 100 |
| 500 | 0.978 | 0.978 | 0.891 | 0.857 | 500 | 1.004 | 1.206 | 1.029 | 1.209 | 500 | 1.495 | 1.471 | 1.362 | 1.276 |
| | (0.180) | (0.235) | (0.107) | (0.082) | | (0.824) | (1.348) | (0.476) | (0.457) | | (0.274) | (0.189) | (0.178) | (0.134) |
| 1000 | 0.987 | 0.988 | 0.935 | 0.918 | 1000 | 0.972 | 1.108 | 1.000 | 1.115 | 1000 | 1.528 | 1.529 | 1.458 | 1.396 |
| | (0.160) | (0.186) | (0.088) | (0.072) | | (0.829) | (1.066) | (0.505) | (0.398) | | (0.241) | (0.174) | (0.134) | (0.105) |
| 2000 | 0.980 | 0.991 | 0.961 | 0.959 | 2000 | 0.938 | 1.005 | 0.977 | 1.055 | 2000 | 1.536 | 1.554 | 1.520 | 1.484 |
| | (0.164) | (0.134) | (0.078) | (0.058) | | (0.787) | (0.698) | (0.454) | (0.328) | | (0.241) | (0.161) | (0.110) | (0.084) |
| $10J$ | 0.917 | 0.925 | 0.891 | 0.918 | $10J$ | 1.054 | 1.290 | 1.029 | 1.115 | $10J$ | 1.329 | 1.377 | 1.362 | 1.396 |
| | (0.194) | (0.155) | (0.107) | (0.072) | | (0.913) | (1.483) | (0.476) | (0.398) | | (0.365) | (0.228) | (0.178) | (0.105) |
| $J^2$ | 0.917 | 0.974 | 0.963 | 0.984 | $J^2$ | 1.054 | 1.127 | 0.978 | 0.945 | $J^2$ | 1.329 | 1.493 | 1.520 | 1.570 |
| | (0.194) | (0.134) | (0.086) | (0.046) | | (0.913) | (1.206) | (0.557) | (0.267) | | (0.365) | (0.186) | (0.124) | (0.067) |
| $\infty$ | 0.982 | 0.997 | 0.989 | 0.979 | $\infty$ | 0.909 | 0.981 | 0.912 | 0.935 | $\infty$ | 1.583 | 1.613 | 1.605 | 1.582 |
| | (0.156) | (0.123) | (0.058) | (0.045) | | (0.749) | (0.692) | (0.363) | (0.274) | | (0.246) | (0.164) | (0.101) | (0.071) |

Standard error across repetitions stands in the parenthesis.

Table 4: Monte Carlo Results for the Extended BLP framework, 100 repetitions, $n = \infty, N = \infty, R = 100$

| # of Consumer | $\alpha(1.0)$ # of products ($J$) | | | |
|---|---|---|---|---|
| $T$ | 10 | 25 | 50 | 100 |
| 10 | 0.985 | 0.978 | 0.989 | 0.993 |
| | (0.139) | (0.100) | (0.071) | (0.061) |
| 50 | 1.007 | 0.985 | 0.989 | 0.993 |
| | (0.126) | (0.089) | (0.067) | (0.055) |
| 100 | 1.019 | 0.988 | 0.997 | 0.996 |
| | (0.135) | (0.084) | (0.066) | (0.057) |
| 500 | 1.017 | 0.988 | 0.996 | 1.008 |
| | (0.122) | (0.075) | (0.062) | (0.057) |
| 1000 | 1.025 | 0.982 | 0.992 | 1.002 |
| | (0.133) | (0.072) | (0.062) | (0.054) |
| $10J$ | 1.019 | 0.983 | 0.996 | 1.002 |
| | (0.135) | (0.078) | (0.062) | (0.054) |
| $J^2$ | 1.019 | 0.992 | 0.996 | 0.999 |
| | (0.135) | (0.079) | (0.056) | (0.062) |

| # of Consumer | $\beta(1.0)$ # of products ($J$) | | | |
|---|---|---|---|---|
| $T$ | 10 | 25 | 50 | 100 |
| 10 | 0.930 | 1.039 | 0.954 | 0.999 |
| | (0.568) | (0.683) | (0.469) | (0.530) |
| 50 | 0.978 | 0.999 | 0.978 | 0.958 |
| | (0.411) | (0.356) | (0.392) | (0.316) |
| 100 | 0.974 | 0.991 | 0.953 | 0.933 |
| | (0.336) | (0.284) | (0.317) | (0.249) |
| 500 | 0.991 | 0.961 | 0.981 | 0.958 |
| | (0.271) | (0.227) | (0.169) | (0.148) |
| 1000 | 0.989 | 0.929 | 0.956 | 0.967 |
| | (0.234) | (0.134) | (0.137) | (0.134) |
| $10J$ | 0.974 | 0.967 | 0.981 | 0.967 |
| | (0.336) | (0.233) | (0.169) | (0.134) |
| $J^2$ | 0.974 | 0.959 | 0.954 | 0.955 |
| | (0.336) | (0.184) | (0.125) | (0.087) |

| # of Consumer | $\gamma(1.5)$ # of products ($J$) | | | |
|---|---|---|---|---|
| $T$ | 10 | 25 | 50 | 100 |
| 10 | 1.630 | 1.594 | 1.620 | 1.607 |
| | (0.229) | (0.168) | (0.110) | (0.085) |
| 50 | 1.648 | 1.605 | 1.621 | 1.608 |
| | (0.236) | (0.163) | (0.115) | (0.080) |
| 100 | 1.677 | 1.610 | 1.629 | 1.610 |
| | (0.250) | (0.159) | (0.107) | (0.083) |
| 500 | 1.676 | 1.617 | 1.620 | 1.615 |
| | (0.241) | (0.134) | (0.089) | (0.083) |
| 1000 | 1.682 | 1.614 | 1.617 | 1.610 |
| | (0.238) | (0.139) | (0.097) | (0.087) |
| $10J$ | 1.677 | 1.612 | 1.620 | 1.610 |
| | (0.250) | (0.143) | (0.089) | (0.087) |
| $J^2$ | 1.677 | 1.620 | 1.621 | 1.606 |
| | (0.250) | (0.142) | (0.092) | (0.087) |

Standard error across repetitions stands in the parenthesis.

unbiased for $\eta^0$ conditional on $x$ and $\xi$.[6] Table 5 shows the result for this case. In the table, we can observe the standard error of $\beta$ is decreasing in $N$.

Next, we evaluate the asymptotic theorem in the previous sections that gives the asymptotic distribution and the asymptotic variance of the parameter estimates. For $J = 25, R = 2000, n = 2000$ fixed, we implement 1,000 Monte Carlo simulation using the BLP framework, and then we calculate the average and standard error of the estimate across these different simulation data-set. We also obtain the asymptotic variances of the GMM estimates given in (74). For each data-set, we calculate the moment conditions and their derivatives in terms of parameters (the parameters are fixed at true values). By averaging resulting values over data-sets, we obtain the estimate for the expected values $\mathbf{\Gamma}_{J,T}$ and $\mathbf{\Phi}$ respectively. For the extended framework, we implement the same simulation with $J = 25, R = 2000, n = 2000, N = 2000, T = 500$ fixed. The variances of the estimates are obtained using (108). Table 6 shows the result. In the table, the simulated standard errors of estimates are relatively consistent with the asymptotic standard errors.

Finally, we make density trace plots for the estimated parameters from the 1,000 estimates used in table 6. (To make these plots, we use the command in the S-plus package with default options.) The solid lines in Figure 1 and Figure 2 show the densities of the estimated parameters, while the dotted lines show their asymptotic distributions using the true values of parameters and the asymptotic variance in Table 6 as mean and variance. In the figures, the simulated distributions of the estimates for the demand parameters $\alpha$ and $\beta$ look fitting well in the asymptotic distributions, while that for the cost side parameter $\gamma$ does not seem so much. However, the shape of the simulated distribution is relatively close to that for the normal. We consider our asymptotic distribution in the theory is a relatively good approximation for the asymptotic distribution of the parameter estimates.

# Appendix Proofs

### Proof of Theorem 1

The consistency argument is established by showing that

[6] All random variables included in $\eta_1^N$ are $x$, $\xi$, $\nu_{i'}^o$, and $1\{C_{i'} \in \mathcal{Q}\{p_j \geq \bar{p}\}$. Abbreviate $C_{i'}^p \equiv 1\{C_{i'} \in \mathcal{Q}\{p_j \geq \bar{p}\}\}$ and then

$$\begin{aligned}
\mathrm{E}_{\nu_{i'}^o, C_{i'}^p|x,\xi}[\eta_1^N - \eta_1^0|x,\xi] &= \mathrm{E}_{C_{i'}^p|x,\xi}\left[\mathrm{E}_{\nu_{i'}^o|x,\xi,C_{i'}^p}[\eta^N - \eta_1^0|x,\xi,C_{i'}^p]\Big|x,\xi\right] \\
&= \mathrm{E}_{C_{i'}^p|x,\xi}\left[\mathrm{E}_{\nu_{i'}^o|x,\xi,C_{i'}^p}\left[\sum_{i'=1}^{N}\frac{\nu_{i'}^o C_{i'}^p}{N_p} - \eta_1^0\Bigg|x,\xi,C_{i'}^p\right]\Bigg|x,\xi\right] \\
&= \mathrm{E}_{C_{i'}^p|x,\xi}\left[\sum_{i'=1}^{N}\mathrm{E}_{\nu_{i'}^o|x,\xi,C_{i'}^p}[\nu_{i'}^o|x,\xi,C_{i'}^p]\frac{C_{i'}^p}{N_p} - \eta_1^0\Bigg|x,\xi\right] \\
&= \mathrm{E}_{C_{i'}^p|x,\xi}\left[\mathrm{E}_{\nu_{i'}^o|x,\xi,C_{i'}^p}[\nu_{i'}^o|x,\xi,C_{i'}^p]\frac{\sum_{i'=1}^{N} C_{i'}^p}{N_p} - \eta_1^0\Bigg|x,\xi\right] \\
&= \mathrm{E}_{C_{i'}^p|x,\xi}\left[\mathrm{E}_{\nu_{i'}^o|x,\xi,C_{i'}^p}[\nu_{i'}^o|x,\xi,C_{i'}^p] - \eta_1^0\Big|x,\xi\right] \\
&= \mathrm{E}_{C_{i'}^p|x,\xi}[0|x,\xi] \\
&= 0.
\end{aligned}$$

Table 5: Monte Carlo Results for the Extended BLP framework, 100 repetitions, $n = \infty, R = 100, T = 1000$

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

Standard error across repetitions stands in the parenthesis.

Figure 1: Kernel Density Estimate of Parameters, BLP Framework, J=25, n=2000, R=2000

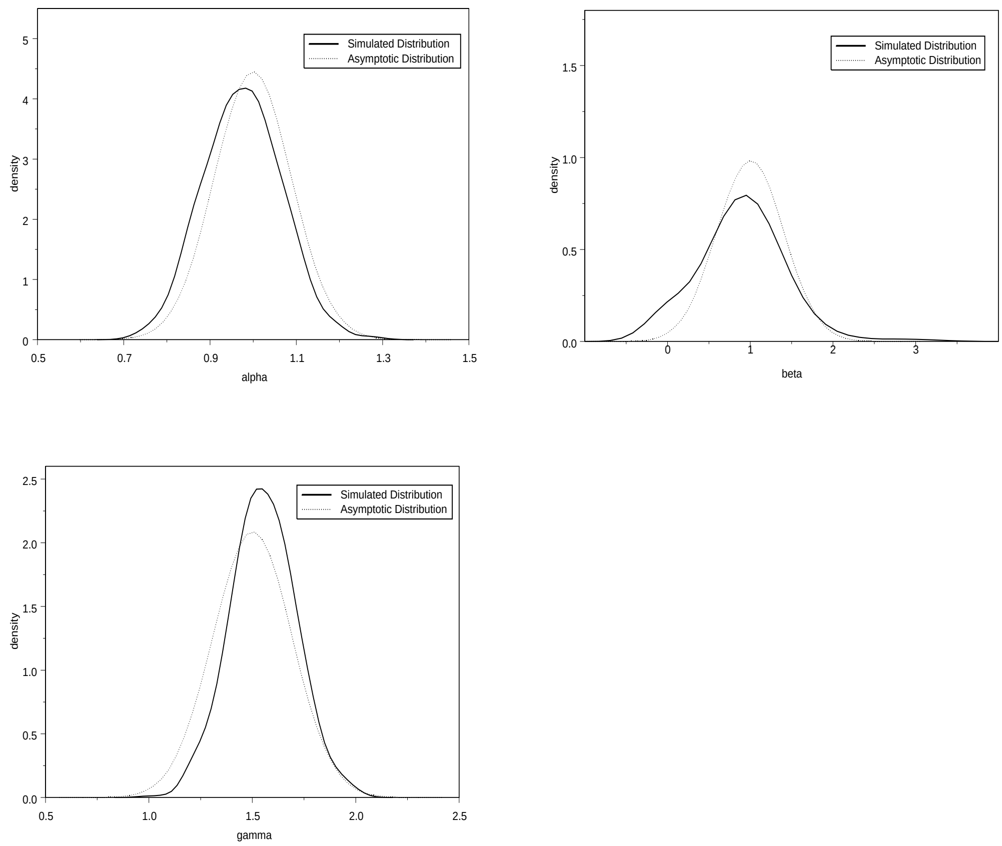

Figure 2: Kernel Density Estimate of Parameters, Additional Moment Framework, J=25, n=2000, R=2000, T=500, N=2000

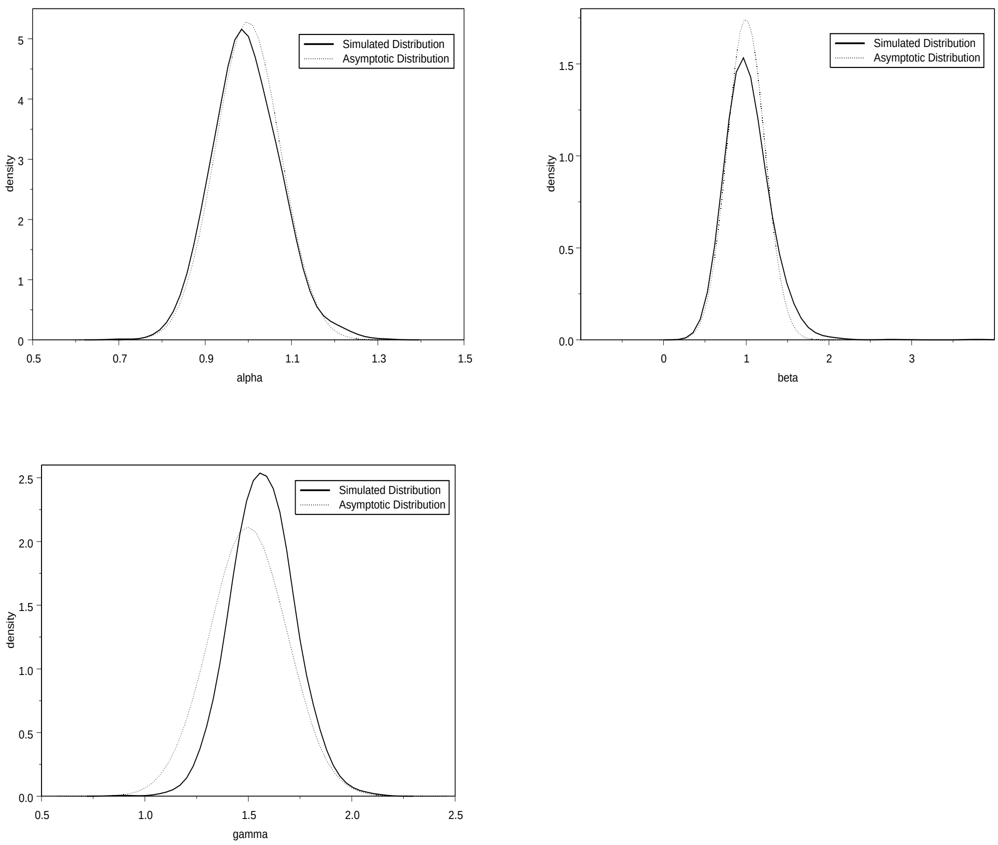

Table 6: Simulated and Estimated Standard Errors ($J = 25, n = 2000, R = 2000, N = 2000, T = 500$)

| | | $\alpha$ | $\beta$ | $\gamma$ |
|---|---|---|---|---|
| BLP 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 |
| Additional Moment Method | 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 |

(1-i) the estimator $\tilde{\boldsymbol{\theta}}$ defined as any sequence that satisfies

$$||\boldsymbol{G}_J(\tilde{\boldsymbol{\theta}}, \boldsymbol{s}^0, P^0)|| = \inf_{\theta \in \Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| + o_p(1) \tag{A.1}$$

is consistent for $\boldsymbol{\theta}^0$, and

(1-ii) $\sup_{\theta\in\Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$ converges to zero in probability.

A consequence of (1-ii) is that $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||$ and $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||$ have a 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 to be consistent for $\boldsymbol{\theta}^0$ from (1-i).

We first 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. Their theorem guarantees that an estimator $\tilde{\boldsymbol{\theta}}$ that satisfies (A.1) is consistent for $\boldsymbol{\theta}^0$ if

(i-a) $\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) = o_p(1)$, and

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

**Proof of (i-a)**

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

$$\begin{aligned}
&\mathrm{E}_{\mathrm{x}_1,\xi}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] \\
&= \mathrm{E}_{\mathrm{x}_1}\left[\mathrm{E}_{\xi|\mathrm{x}_1}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&= \mathrm{E}_{\mathrm{x}_1}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \, \mathrm{E}_{\xi|\mathrm{x}_1}\left[\xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&= 0,
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{V}_{\mathrm{x}_1,\xi}[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \\
&\quad = \mathrm{E}_{\mathrm{x}_1}\left[\mathrm{V}_{\xi|\mathrm{x}_1}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&\qquad + \mathrm{V}_{\mathrm{x}_1}\left[\mathrm{E}_{\xi|\mathrm{x}_1}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&\quad = \mathrm{E}_{\mathrm{x}_1}\left[\mathrm{V}_{\xi|\mathrm{x}_1}\left[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&\quad = \mathrm{E}_{\mathrm{x}_1}\left[J^{-2}\textstyle\sum_{j=1}^J (z_{jm}^d)^2\, \mathrm{V}_{\xi|\mathrm{x}_1}\left[\xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right] \\
&\quad = \mathrm{E}_{\mathrm{x}_1}\left[J^{-2}\textstyle\sum_{j=1}^J (z_{jm}^d)^2\, \mathrm{E}_{\xi|\mathrm{x}_1}\left[\xi_j^2(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right]\right].
\end{aligned}
$$

Since the conditional variance of $\xi_j$ in (1) is bounded by some constant $M > 0$ or $\mathrm{E}_{\xi|\mathrm{x}_1}[\xi_j^2(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|\boldsymbol{X}_1] < M$ with probability one, we have

$$
J^{-2}\textstyle\sum_{j=1}^J (z_{jm}^d)^2\, \mathrm{E}_{\xi|\mathrm{x}_1}\left[\xi_j^2(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\middle| \boldsymbol{X}_1\right] \le (1/J)(\textstyle\sum_{j=1}^J (z_{jm}^d)^2/J)M.
$$

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 $\mathrm{E}_{\mathrm{x}_1}[\sum_{j=1}^J (z_{jm}^d)^2/J] = O(1)$. Hence

$$
\begin{aligned}
&\mathrm{V}_{\mathrm{x}_1,\xi}[J^{-1}\textstyle\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] \\
&\quad = \mathrm{E}_{\mathrm{x}_1}\left[J^{-2}\textstyle\sum_{j=1}^J (z_{jm}^d)^2\, \mathrm{E}_{\xi|\mathrm{x}_1}[\xi_j^2(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)|\boldsymbol{X}_1]\right] \\
&\quad \le \frac{M}{J}\, \mathrm{E}_{\mathrm{x}_1}[\textstyle\sum_{j=1}^J (z_{jm}^d)^2/J] \\
&\quad = \frac{M}{J}\cdot O(1) \to 0 \text{ as } J \to \infty.
\end{aligned}
$$

Bernoulli's weak law of large numbers ensures that the $m$-th element of $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)$ converges to the corresponding element of $\mathrm{E}_{\mathrm{x}_1,\xi}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] = 0$ in probability, i.e.,

$$
\begin{aligned}
\lim_{J\to\infty} \Pr[|\{\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\}_m| > \epsilon] &= \lim_{J\to\infty} \Pr\left[\left|\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)/J\right| > \epsilon\right] \\
&\le \lim_{J\to\infty} \frac{1}{\epsilon^2}\, \mathrm{V}_{\mathrm{x}_1,\xi}\left[\sum_{j=1}^J z_{jm}^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)/J\right] \\
&\le \frac{1}{\epsilon^2}\lim_{J\to\infty}\frac{M}{J}\cdot O(1) \\
&= 0.
\end{aligned}
$$

Thus $||\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)|| = o_p(1)$. Similarly, we can show that the supply side moments $\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$ converge to $\mathrm{E}_{\mathrm{w}_1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)] = 0$ in probability by (12) and A4(b). Hence $||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| = o_p(1)$.

### Proof of (i-b)

Next we show (i-b). For every $(\epsilon, \delta) > (0, 0)$ and any positive function of $\delta$, $C(\delta)$, following relationship holds in general.

$$
\begin{aligned}
&\left\{\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \ge C(\delta)\right\} \\
&\quad \subset \left\{\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| + ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \ge C(\delta)\right\}
\end{aligned}
$$

$$\subset \left\{ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| \geq C(\delta) - \frac{\epsilon}{2} \right\} \cup \left\{ ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \frac{\epsilon}{2} \right\}. \tag{A.2}$$

Taking probability of both side of (A.2) gives

$$\begin{aligned}
&\Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq C(\delta) \right] \\
&\leq \Pr\left[ \left\{ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| \geq C(\delta) - \frac{\epsilon}{2} \right\} \cup \left\{ ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \frac{\epsilon}{2} \right\} \right] \\
&\leq \Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| \geq C(\delta) - \frac{\epsilon}{2} \right] + \Pr\left[ ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \frac{\epsilon}{2} \right].
\end{aligned}$$

We thus obtain

$$\begin{aligned}
&\Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| \geq C(\delta) - \frac{\epsilon}{2} \right] \\
&\geq \Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq C(\delta) \right] \\
&\qquad - \Pr\left[ ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \frac{\epsilon}{2} \right].
\end{aligned} \tag{A.3}$$

Since $||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| = o_p(1)$, for any $\epsilon$ there exists $J_1(\epsilon)$ such that if $J \geq J_1(\epsilon)$, $\Pr[||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \epsilon/2] \leq \epsilon/2$. By assumption A5, for any $(\epsilon, \delta) > (0, 0)$, there exists $C(\delta) > 0$ and $J_2(\epsilon, \delta)$ such that when $J \geq J_2(\epsilon, \delta)$

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

Therefore, from (A.3), for any $(\epsilon, \delta) > (0, 0)$ there exists $C(\delta) > 0$, $J_1(\epsilon)$ and $J_2(\epsilon, \delta)$ such that when $J \geq \max\{J_1(\epsilon), J_2(\epsilon, \delta)\}$,

$$\begin{aligned}
&\Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| \geq C(\delta) - \frac{\epsilon}{2} \right] \\
&\geq \Pr\left[ \inf_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq C(\delta) \right] \\
&\qquad - \Pr\left[ ||\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)|| \geq \frac{\epsilon}{2} \right] \\
&\geq 1 - \frac{\epsilon}{2} - \frac{\epsilon}{2} = 1 - \epsilon.
\end{aligned}$$

Thus we have

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

by setting $C^*(\epsilon, \delta) = C(\delta) - \epsilon/2$. This is equivalent to (i-b), i.e.,

$$\lim_{J \to \infty} \Pr\left[ \sup_{\theta \notin \mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||^{-1} > C^{\#}(\epsilon, \delta) \right] < \epsilon$$

with $C^{\#}(\epsilon, \delta) = 1/C^*(\epsilon, \delta)$.

We next turn to show (1-ii), or

$$\sup_{\theta \in \Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)|| = o_p(1). \tag{A.4}$$

From the definitions of $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ in (17), (18), (22), and (23), we have

$$\begin{aligned}
&\sup_{\theta \in \Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||^2 \\
&\leq \sup_{\theta_d \in \Theta_d} ||J^{-1} \boldsymbol{Z}_d' \{\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\}||^2 \\
&\quad + \sup_{\theta \in \Theta} ||J^{-1} \boldsymbol{Z}_c' \{\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)\}||^2 \\
&\leq J^{-1} ||\boldsymbol{Z}_d' \boldsymbol{Z}_d|| \times \sup_{\theta_d \in \Theta_d} J^{-1} ||\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)||^2 \\
&\quad + J^{-1} ||\boldsymbol{Z}_c' \boldsymbol{Z}_c|| \times \sup_{\theta \in \Theta} J^{-1} ||\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||^2
\end{aligned} \tag{A.5}$$

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

$$\sup_{\theta_d \in \Theta_d} J^{-1} ||\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R) - \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)||^2 = o_p(1), \tag{A.6}$$

and

$$\sup_{\theta \in \Theta} J^{-1} ||\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R) - \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)||^2 = o_p(1). \tag{A.7}$$

In order to show (A.6), we first show that

$$\begin{aligned}
&\sup_{\theta_d \in \Theta_d} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^0)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0))|| \\
&\quad = o_p(1)
\end{aligned} \tag{A.8}$$

and then show that (A.8) implies (A.6) by using assumption A6. The proof for (A.7) is directly derived from (A.6) and assumption A7.

**Proof of (A.6)**

Since for any $\boldsymbol{\theta}_d$, $\boldsymbol{s}^n = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ from (26) and $\boldsymbol{s}^0 = \boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$ from (27), the left-hand side of (A.8) is bounded by

$$\begin{aligned}
&\sup_{\theta_d \in \Theta_d} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^0)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0))|| \\
&= \sup_{\theta_d \in \Theta_d} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^0)) - \boldsymbol{\tau}_J(\boldsymbol{s}^0) \\
&\qquad + \{\boldsymbol{\tau}_J(\boldsymbol{s}^n) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R))\}|| \\
&\leq J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{s}^n) - \boldsymbol{\tau}_J(\boldsymbol{s}^0)|| \\
&\quad + \sup_{\theta_d \in \Theta_d} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^0))|| \\
&\leq J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{s}^n) - \boldsymbol{\tau}_J(\boldsymbol{s}^0)|| \\
&\quad + \sup_{\theta_d \in \Theta_d} \sup_{\xi} J^{-\frac{1}{2}} ||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0))||.
\end{aligned} \tag{A.9}$$

In the following we show that both the two terms in (A.9) are $o_p(1)$ as $J \to \infty$. By the mean value theorem, for some intermediate values $\bar{s}_j = s_j^0 + q_j(s_j^n - s_j^0)$ $(0 \le q_j \le 1), j = 1, \ldots, J$, the square of the first term is

$$
\begin{aligned}
& J^{-1}||\boldsymbol{\tau}_J(\boldsymbol{s}^n) - \boldsymbol{\tau}_J(\boldsymbol{s}^0)||^2 \\
= \;& J^{-1}\sum_{j=1}^{J}[\log(s_j^n/s_0^n) - \log(s_j^0/s_0^0)]^2 \\
= \;& J^{-1}\sum_{j=1}^{J}[\log(s_j^n) - \log(s_j^0) - \{\log(s_0^n) - \log(s_0^0)\}]^2 \\
= \;& J^{-1}\sum_{j=1}^{J}\left[\frac{\partial}{\partial s_j}\log(s_j)\bigg|_{s_j=\bar{s}_j}(s_j^n - s_j^0) - \frac{\partial}{\partial s_0}\log(s_0)\bigg|_{s_0=\bar{s}_0}(s_0^n - s_0^0)\right]^2 \\
= \;& J^{-1}\sum_{j=1}^{J}\left[\frac{s_j^n - s_j^0}{\bar{s}_j} - \frac{s_0^n - s_0^0}{\bar{s}_0}\right]^2 \\
= \;& J^{-1}\sum_{j=1}^{J}\left(\frac{s_j^n - s_j^0}{\bar{s}_j}\right)^2 - 2J^{-1}\sum_{j=1}^{J}\left(\frac{s_j^n - s_j^0}{\bar{s}_j}\right)\left(\frac{s_0^n - s_0^0}{\bar{s}_0}\right) + J^{-1}\sum_{j=1}^{J}\left(\frac{s_0^n - s_0^0}{\bar{s}_0}\right)^2 \\
\le \;& J^{-1}\sum_{j=1}^{J}\left(\frac{s_j^0}{\bar{s}_j}\right)^2\left(\frac{s_j^n - s_j^0}{s_j^0}\right)^2 + 2\left(\frac{s_0^0}{\bar{s}_0}\right)\left|\frac{s_0^n - s_0^0}{s_0^0}\right| J^{-1}\left|\sum_{j=1}^{J}\left(\frac{s_j^0}{\bar{s}_j}\right)\left(\frac{s_j^n - s_j^0}{s_j^0}\right)\right| \\
& + \left(\frac{s_0^0}{\bar{s}_0}\right)^2\left(\frac{s_0^n - s_0^0}{s_0^0}\right)^2 \\
\le \;& \max_{1\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right)^2 \cdot \max_{1\le j\le J}\left(\frac{s_j^n - s_j^0}{s_j^0}\right)^2 + 2\left(\frac{s_0^0}{\bar{s}_0}\right)\left|\frac{s_0^n - s_0^0}{s_0^0}\right| \max_{1\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right)\max_{1\le j\le J}\left|\frac{s_j^n - s_j^0}{s_j^0}\right| \\
& + \left(\frac{s_0^0}{\bar{s}_0}\right)^2\left(\frac{s_0^n - s_0^0}{s_0^0}\right)^2 \\
= \;& \max_{1\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right) \cdot \max_{1\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right) \cdot \max_{1\le j\le J}\left(\frac{s_j^n - s_j^0}{s_j^0}\right) \cdot \max_{1\le j\le J}\left(\frac{s_j^n - s_j^0}{s_j^0}\right) \\
& +2\left(\frac{s_0^0}{\bar{s}_0}\right)\left|\frac{s_0^n - s_0^0}{s_0^0}\right| \max_{1\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right)\max_{1\le j\le J}\left|\frac{s_j^n - s_j^0}{s_j^0}\right| \\
& + \left(\frac{s_0^0}{\bar{s}_0}\right) \cdot \left(\frac{s_0^0}{\bar{s}_0}\right) \cdot \left(\frac{s_0^n - s_0^0}{s_0^0}\right) \cdot \left(\frac{s_0^n - s_0^0}{s_0^0}\right) \\
\le \;& O_p(1) \cdot O_p(1) \cdot o_p(1) \cdot o_p(1) + O_p(1) \cdot o_p(1) \cdot O_p(1) \cdot o_p(1) \\
& + O_p(1) \cdot O_p(1) \cdot o_p(1) \cdot o_p(1) \\
= \;& o_p(1) \qquad\qquad \text{(A.10)}
\end{aligned}
$$

where $o_p(1)$ terms come from A3(a), while $O_p(1)$ terms follow the next equation.

$$
\begin{aligned}
\max_{0\le j\le J}\left(\frac{s_j^0}{\bar{s}_j}\right) &= \max_{0\le j\le J}\left(\frac{s_j^0}{s_j^0 + q_j(s_j^n - s_j^0)}\right) = \max_{0\le j\le J}\left(\frac{1}{1 + q_j(s_j^n - s_j^0)/s_j^0}\right) \\
&= \max_{0\le j\le J}\left(\frac{1}{1 + q_j \cdot o_p(1)}\right) = O_p(1).
\end{aligned}
$$

For the second term of (A.9), by the mean value theorem, we obtain for given $(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d)$,

$$
J^{-1}||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^R)) - \boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d, P^0))||^2
$$

$$
\begin{aligned}
&= J^{-1}\sum_{j=1}^{J}[\log(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)/\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)) \\
&\qquad -\log(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)/\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0))]^2 \\
&= J^{-1}\sum_{j=1}^{J}[\log(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)) - \log(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)) \\
&\qquad -\{\log(\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)) - \log(\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0))\}]^2 \\
&= J^{-1}\sum_{j=1}^{J}\left[\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right. \\
&\qquad\qquad \left.-\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right]^2 \\
&= J^{-1}\sum_{j=1}^{J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)^2 \\
&\quad -2J^{-1}\sum_{j=1}^{J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right) \\
&\qquad\qquad \times\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right) \\
&\quad +J^{-1}\sum_{j=1}^{J}\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right)^2 \\
&\le J^{-1}\sum_{j=1}^{J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)^2\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right)^2 \\
&\quad +2\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right)\left|\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right| \\
&\quad \times J^{-1}\left|\sum_{j=1}^{J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right)\right| \\
&\quad +\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right)^2 \\
&\le \max_{0\le j\le J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)^2\max_{0\le j\le J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right)^2 \\
&\quad +2\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right)\left|\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right| \\
&\quad \times\max_{0\le j\le J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)\max_{0\le j\le J}\left|\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right| \\
&\quad +\left(\frac{\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_0(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_0}\right)^2 \qquad\qquad \text{(A.11)}
\end{aligned}
$$

where $\bar{\sigma}_j, j = 0, \ldots, J$ are values between $\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)$ and $\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)$. We need to show that (A.11) is $o_p(1)$ uniformly over $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d \in \Theta_d$. A straightforward application of A3(b) to the relative difference share terms in (A.11) yields that they are all of order $o_p(1)$ uniformly over $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d \in \Theta_d$. As for the relative share term,

$$\begin{aligned}
&\max_{0\le j\le J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\bar{\sigma}_j}\right)\\
&= \max_{0\le j\le J}\left(\frac{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}{\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)+q_j(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0))}\right)\\
&= \max_{0\le j\le J}\left(\frac{1}{1+q_j(\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R)-\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0))/\sigma_j(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0)}\right)\\
&= \max_{0\le j\le J}\left(\frac{1}{1+q_j\cdot o_p(1)}\right) = O_p(1) \qquad \text{(A.12)}
\end{aligned}$$

where $0 \le q_j \le 1$. Again utilizing A3(b) yields that (A.12) holds uniformly over $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_d \in \Theta_d$. Thus

$$\sup_{\theta_d\in\Theta_d}\sup_{\xi} J^{-\frac{1}{2}}||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^R))-\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi},\boldsymbol{\theta}_d,P^0))||=o_p(1).$$

Hence we obtain (A.8).

By assumption A6, for all $\boldsymbol{\theta}_d \in \Theta_d$, if $J^{-1}||\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R)-\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)||^2 \ge \delta$ for some $\delta > 0$, then there exists $C(\delta)$ such that

$$\inf_{\theta_d\in\Theta_d} J^{-\frac{1}{2}}||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^0))-\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0))|| \ge C(\delta)$$

with probability tending to one as $J \to \infty$. In other words, its contrapositive statement is that whenever

$$\sup_{\theta_d\in\Theta_d} J^{-\frac{1}{2}}||\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^0))-\boldsymbol{\tau}_J(\boldsymbol{\sigma}(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0))|| = o_p(1)$$

holds, A6 implies $\sup_{\theta_d\in\Theta_d} J^{-1}||\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R)-\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)||^2 \le \delta$, or in the presence of A6, (A.8) implies (A.6), i.e., for any $\boldsymbol{\theta}_d \in \Theta_d$ and $\delta > 0$,

$$\Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R) \notin \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d;\delta)] \to 0. \qquad \text{(A.13)}$$

**Proof of (A.7)**

By the Glivenko-Cantelli theorem,

$$\Pr[P^R \notin \mathcal{N}_{P^0}(\delta)] \to 0 \qquad \text{(A.14)}$$

for $\delta > 0$ as $R \to \infty$. From (A.13) and (A.14) as $J, R \to \infty$, for given $\delta > 0$, $\Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R)] \in \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d;\delta), P^R \in \mathcal{N}_{P^0}(\delta)] \to 1$ or

$$\Pr[(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),P^R) \in \mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d;\delta)\times\mathcal{N}_{P^0}(\delta)] \to 1.$$

Thus assumption A7 guarantees that the differences in the profit margin behave uniformly over $\boldsymbol{\theta}_d \in \Theta_d$ as

$$\sup_{\theta_d\in\Theta_d} J^{-\frac{1}{2}}||\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^R)-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)|| = o_p(1). \qquad \text{(A.15)}$$

Since $\dot{g}(\cdot)$ is assumed finite for all realizable values of cost, we can derive (A.7) by using (A.15) in the following inequality with the definition of $\omega_j(\boldsymbol{\theta}, \boldsymbol{s}, P)$ in (13).

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

where $\overline{m_{g_j}}$ are between $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R), \boldsymbol{\theta}_d, P^R)$ and $m_{g_j}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0)$. Notice that $p_j - m_{g_j}$ generally represents the marginal cost. We should also note that the difference between $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)$ includes only the demand side parameters $\boldsymbol{\theta}_d$ because of the linear dependence of $\boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}, P)$ on the supply side parameters $\boldsymbol{\theta}_c$ as seen in (13). □

**Proof of Theorem 2**

To establish Theorem 2, we show that for the approximation $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) = (\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)', \boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})')'$ defined in (47) and (57) to $\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)$,

(2-i) $\sup_{||\theta-\theta^0||\leq\delta_J}\left|\left|J^{\frac{1}{2}}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) - \boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)]\right|\right| \xrightarrow{p} 0$ when $\delta_J \to 0$, and

(2-ii) an estimator that minimizes $||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||$ over $\boldsymbol{\theta} \in \Theta$ would be; (1) asymptotically normal at rate $J^{\frac{1}{2}}$, and (2) have a variance-covariance matrix which is the sum of three mutually uncorrelated terms (one resulting from randomness in the draws on exogenous variables $(\boldsymbol{x}_{1j}, \xi_j, \mathbf{w}_{1j}, \omega_j)$, one from sampling error $\epsilon_j^n$, 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 $||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||$, has the same limiting distribution as our estimator that minimizes $||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)||$. Since the former is easier to analyze, we work with it.

**proof of (2-i)**

We show (2-i) by establishing that for any $\delta_J \to 0$,

$$\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\left|\left|J^{\frac{1}{2}}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d) - \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)]\right|\right| = o_p(1), \quad (A.16)$$

$$\sup_{||\theta-\theta^0||\leq\delta_J}\left|\left|J^{\frac{1}{2}}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}) - \boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R)]\right|\right| = o_p(1). \quad (A.17)$$

We first show (A.16). From (46) and (47), $||J^{\frac{1}{2}}[\mathcal{G}_J^d(\boldsymbol{\theta}_d) - \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)]||$ can be rewritten as

$$\begin{aligned}
&\Big|\Big|J^{\frac{1}{2}}[\mathcal{G}_J^d(\boldsymbol{\theta}_d) - \boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)]\Big|\Big| \\
&= \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\Big[\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\} \\
&\quad -\{\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\}\Big]\Big|\Big| \\
&\leq \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1} - \boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\}\boldsymbol{\epsilon}^n\Big|\Big| \\
&\quad + \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\}\Big|\Big|. \qquad \text{(A.18)}
\end{aligned}$$

We show the two terms in the right-hand side of (A.18) are respectively $o_p(1)$ uniformly in $\boldsymbol{\theta}_d$ within the shrinking neighborhood of $\boldsymbol{\theta}_d^0$. We know that for each $\boldsymbol{\theta}_d$ both $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$ converge to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$ in probability in terms of averaged Euclidean distance as $n$ and $R$ grow. Since $\bar{\boldsymbol{\xi}}$ is intermediate between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^n, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$, it also converges to $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$. Thus, for any sequence $\delta_J \to 0$, $\Pr[(\bar{\boldsymbol{\xi}}_1, \ldots, \bar{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_J)\}^J] \to 0$. Moreover, for any $\delta_J \to 0$, we have $\Pr[\rho_P(P^R, P^0) \geq \delta_J] \to 0$ as $R$ grows by the Glivenko-Cantelli theorem. Therefore, by using assumption B5(a), we have

$$\begin{aligned}
&\Pr\Big[\sup_{||\theta_d - \theta_d^0|| \leq \delta_J} \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1} - \boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\}\boldsymbol{\epsilon}^n\Big|\Big| > c\Big] \\
&\leq \Pr\Big[\sup_{||\theta_d - \theta_d^0|| \leq \delta_J} \sup_{(\xi, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_J)\}^J \times \mathcal{N}_{P^0}(\delta_J)} \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1} \\
&\quad -\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\}\boldsymbol{\epsilon}^n\Big|\Big| > c\Big] \\
&\quad + \Pr[(\bar{\boldsymbol{\xi}}_1, \ldots, \bar{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_J)\}^J] + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_J)] \\
&\to 0. \qquad \text{(A.19)}
\end{aligned}$$

Notice that in the expression of $\boldsymbol{H}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)$, as mentioned before, we have suppressed the fact there exist different $\bar{\boldsymbol{\xi}}$s for different rows in $\boldsymbol{H}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)$. Therefore, in (A.19), we have to evaluate $\boldsymbol{H}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)$ row by row with distinct $\bar{\boldsymbol{\xi}}_j, j = 1, \ldots, J$.

For the intermediate vectors $\underline{\boldsymbol{\xi}}_j, j = 1, \ldots, J$ between $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^R)$ and $\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)$, we have $\Pr[(\underline{\boldsymbol{\xi}}_1, \ldots, \underline{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0; \delta_J)\}^J] \to 0$ for any $\delta_J \to 0$. Thus for the second term in (A.18), by assumption B5(b),

$$\begin{aligned}
&\Pr\Big[\sup_{||\theta_d - \theta_d^0|| \leq \delta_J} \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\}\Big|\Big| > c\Big] \\
&\leq \Pr\Big[\sup_{||\theta_d - \theta_d^0|| \leq \delta_J} \sup_{(\xi, P) \in \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_J)\}^J \times \mathcal{N}_{P^0}(\delta_J)} \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_d'\{\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) \\
&\quad -\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\}\Big|\Big| > c\Big] \\
&\quad + \Pr[(\underline{\boldsymbol{\xi}}_1, \ldots, \underline{\boldsymbol{\xi}}_J) \notin \{\mathcal{N}_{\xi^0}(\theta_d^0; \delta_J)\}^J] + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_J)] \\
&\to 0. \qquad \text{(A.20)}
\end{aligned}$$

We next show (A.17). From (56) and (57), we know that

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

$$
\begin{aligned}
&-\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^n \\
&+\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\Big]\Big|\Big| \\
\leq\ & \Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\Big[\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R))-\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0))\Big]\Big|\Big| \\
&+\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\{\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}-\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\}\boldsymbol{\epsilon}^n\Big|\Big| \\
&+\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\{\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)-\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\}\Big|\Big|.
\end{aligned}
\tag{A.21}
$$

We need to show that the three terms in the right-hand side of (A.21) are respectively $o_p(1)$ within the $\delta_J$ neighborhood of $\boldsymbol{\theta}_d^0$, From assumption B5(e), we know that

$$
\begin{aligned}
\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'[\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R)) \\
-\boldsymbol{g}(\boldsymbol{p}-\boldsymbol{m}_g(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0))]\Big|\Big| = o_p(1).
\end{aligned}
\tag{A.22}
$$

With the argument similar to obtain (A.19), we can derive for the second term on the right-hand side of (A.21) by using B5(c),

$$
\begin{aligned}
&\Pr\Big[\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\{\boldsymbol{L}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\bar{\bar{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R) \\
&\qquad -\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\}\boldsymbol{\epsilon}^n\Big|\Big| > c\Big] \\
\leq\ & \Pr\Big[\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\ \sup_{(\xi_1,\xi_2,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^{2J}\times\mathcal{N}_{P^0}(\delta_J)}\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c' \\
&\times\{\boldsymbol{L}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{M}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2,\boldsymbol{\theta}_d,P)-\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\}\boldsymbol{\epsilon}^n\Big|\Big| > c\Big] \\
&+\Pr[(\bar{\bar{\boldsymbol{\xi}}}_1,\ldots,\bar{\bar{\boldsymbol{\xi}}}_J)\notin\{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0;\delta_J)\}^J]+\Pr[(\bar{\boldsymbol{\xi}}_1,\ldots,\bar{\boldsymbol{\xi}}_J)\notin\{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0;\delta_J)\}^J] \\
&+\Pr[P^R\notin\mathcal{N}_{P^0}(\delta_J)] \\
\to\ & 0.
\end{aligned}
$$

For the third term on the right-hand side of (A.21), we obtain by assumption B5(d)

$$
\begin{aligned}
&\Pr\Big[\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c'\{\boldsymbol{L}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{M}(\underline{\underline{\boldsymbol{\xi}}},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d) \\
&\qquad -\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}\Big|\Big| > c\Big] \\
\leq\ & \Pr\Big[\sup_{||\theta_d-\theta_d^0||\leq\delta_J}\ \sup_{(\xi_1,\xi_2,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_J)\}^{2J}\times\mathcal{N}_{P^0}(\delta_J)}\Big|\Big|J^{-\frac{1}{2}}\boldsymbol{Z}_c' \\
&\times\{\boldsymbol{L}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{M}(\boldsymbol{\xi}_1,\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2,\boldsymbol{\theta}_d,P)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)-\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}\Big|\Big| > c\Big] \\
&+\Pr[(\underline{\underline{\boldsymbol{\xi}}}_1,\ldots,\underline{\underline{\boldsymbol{\xi}}}_J)\notin\{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0;\delta_J)\}^J]+\Pr[(\underline{\boldsymbol{\xi}}_1,\ldots,\underline{\boldsymbol{\xi}}_J)\notin\{\mathcal{N}_{\xi^0}(\boldsymbol{\theta}_d^0;\delta_J)\}^J] \\
&+\Pr[P^R\notin\mathcal{N}_{P^0}(\delta_J)] \\
\to\ & 0.
\end{aligned}
$$

**proof of (2-ii)**

We now turn to show (2-ii). In order to show that the estimator that minimizes the norm of $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ is asymptotically normally distributed we apply a version of Theorem 3.3 in Pakes and Pollard (1989). A difference here is that the expectation of $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ could vary with $J$. This is because the derivative of $(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}, P), \boldsymbol{\omega}(\boldsymbol{\theta}, \boldsymbol{s}, P))$ with respect to $\boldsymbol{\theta}$ and the instrumental variables $(\boldsymbol{Z}_d, \boldsymbol{Z}_c)$ both depend on the number and characteristics of the all products marketed. The version of the theorem we use is:

*Let $\bar{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}^0$, the unique point of $\Theta$ for which $\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)] = 0$. If:*

(i) $||\boldsymbol{\mathcal{G}}_J(\bar{\boldsymbol{\theta}})|| \leq o_p(J^{-\frac{1}{2}}) + \inf_\theta ||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||$;

(ii) $\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]$ *is differentiable at $\boldsymbol{\theta}^0$ with a derivative matrix $\boldsymbol{\Gamma}_J$ of full rank, and $\boldsymbol{\Gamma}_J \to \boldsymbol{\Gamma}$ as $J \to \infty$;*

(iii) *for every sequence $\{\delta_J\}$ of positive numbers that converges to zero,*

$$\sup_{||\theta-\theta^0||\leq\delta_J} \frac{||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) - \mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})] - \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)||}{J^{-\frac{1}{2}} + ||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})|| + ||\,\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]||} = o_p(1);$$

(iv) $J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0) \overset{w}{\rightsquigarrow} N(\boldsymbol{0}, \boldsymbol{\Phi})$;

(v) $\boldsymbol{\theta}^0$ *is an interior point of* $\Theta$*;*

*then*

$$J^{\frac{1}{2}}(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \overset{w}{\rightsquigarrow} N(\boldsymbol{0}, (\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma})^{-1}).$$

The set of assumptions, $\mathrm{E}_{\xi|\mathrm{x}_1}[\xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)|\boldsymbol{x}_{1j}] = 0$ given in (1), $\mathrm{E}_{\omega|\mathrm{w}_1}[\omega_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)|\mathbf{w}_{1j}] = 0$ given in (12), $\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{\epsilon}^n|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] = \boldsymbol{0}$ given in A1(a), and $\mathrm{E}_{\epsilon^*|\mathrm{x},\xi}[\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)] = \boldsymbol{0}$ for each $\boldsymbol{\theta}_d$ given in assumption A1(b) ensures that the unconditional expectation $\mathrm{E}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d^0)] = \boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}^0)] = \boldsymbol{0}$. Noting the fact that under $(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$, $\mathrm{E}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] = \boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)] = \boldsymbol{0}$.

$$\begin{aligned}
&\mathrm{E}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d^0)] \\
&= \mathrm{E}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) + J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n - J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right] \\
&= \mathrm{E}_{\mathrm{x}_1,\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] + \mathrm{E}_{\epsilon,\mathrm{x}_1,\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n\right] - \mathrm{E}_{\epsilon^*,\mathrm{x}_1,\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right] \\
&= \mathrm{E}_{\mathrm{x}_1,\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] + \mathrm{E}_{\epsilon,\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n\right] - \mathrm{E}_{\epsilon^*,\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right] \\
&= \mathrm{E}_{\mathrm{x}_1}\left[\mathrm{E}_{\xi|\mathrm{x}_1}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\Big|\,\boldsymbol{X}_1\right]\right] \\
&\quad + \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{E}_{\epsilon|\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n\Big|\,\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad - \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{E}_{\epsilon^*|\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\Big|\,\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \mathrm{E}_{\mathrm{x}_1}\left[J^{-1}\boldsymbol{Z}_d'\,\mathrm{E}_{\xi|\mathrm{x}_1}\left[\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\Big|\,\boldsymbol{X}_1\right]\right] \\
&\quad + \mathrm{E}_{\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}\left[\boldsymbol{\epsilon}^n\Big|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad - \mathrm{E}_{\mathrm{x},\xi}\left[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon^*|\mathrm{x},\xi}\left[\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\Big|\,\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \boldsymbol{0}, \\
&\mathrm{E}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}^0)] \\
&= \mathrm{E}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) - J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}] \\
&= \mathrm{E}_{\mathrm{w}_1,\omega}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)] - \mathrm{E}_{\epsilon,\mathrm{x}_1,\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n] \\
&\quad + \mathrm{E}_{\epsilon^*,\mathrm{x}_1,\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)] \\
&= \mathrm{E}_{\mathrm{w}_1,\omega}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)] - \mathrm{E}_{\epsilon,\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n]
\end{aligned}$$

$$
\begin{aligned}
& + \mathrm{E}_{\epsilon^*,\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)] \\
= \; & \mathrm{E}_{\mathrm{w}_1}[\mathrm{E}_{\omega|\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{\omega}(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)|\boldsymbol{W}_1]] \\
& - \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[\mathrm{E}_{\epsilon|\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{W}_1]] \\
& + \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[\mathrm{E}_{\epsilon^*|\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{W}_1]] \\
= \; & \mathrm{E}_{\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\,\mathrm{E}_{\omega|\mathrm{w}_1}[\boldsymbol{\omega}(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)|\boldsymbol{W}_1]] \\
& - \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi,\mathrm{w}_1}[\boldsymbol{\epsilon}^n|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{W}_1]] \\
& + \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon^*|\mathrm{x},\xi,\mathrm{w}_1}[\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{W}_1]] \\
= \; & \mathrm{E}_{\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\,\mathrm{E}_{\omega|\mathrm{w}_1}[\boldsymbol{\omega}(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)|\boldsymbol{W}_1]] \\
& - \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon|\mathrm{x},\xi}[\boldsymbol{\epsilon}^n|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)]] \\
& + \mathrm{E}_{\mathrm{x},\xi,\mathrm{w}_1}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\,\mathrm{E}_{\epsilon^*|\mathrm{x},\xi}[\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)|\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0)]] \\
= \; & \mathbf{0}.
\end{aligned}
$$

We confirm that under the assumptions we give in the theorem each of the conditions (i)–(v) is satisfied. Any estimator that minimizes $||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||$ satisfies (i). Since $\mathrm{E}[J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)] = \mathbf{0}$, we have from (47)

$$
\mathrm{E}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)] = \mathrm{E}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)].
$$

Similarly, since $\mathrm{E}[J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}] = \mathbf{0}$, we obtain from (57)

$$
\mathrm{E}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})] = \mathrm{E}[\boldsymbol{G}_J^c(\boldsymbol{\theta},\boldsymbol{s}^0,P^0)].
$$

Thus

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\theta}'}\,\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})] &= \frac{\partial}{\partial\boldsymbol{\theta}'}\,\mathrm{E}[\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^0,P^0)] = \left(\left\{\frac{\partial\,\mathrm{E}[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d)]}{\partial\boldsymbol{\theta}'}\right\}',\left\{\frac{\partial\,\mathrm{E}[\boldsymbol{G}_J^c(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}'}\right\}'\right)' \\
&= ({\boldsymbol{\Gamma}_J^d}',{\boldsymbol{\Gamma}_J^c}')' \qquad \text{(A.23)}
\end{aligned}
$$

by assumption B2 and condition (ii) is satisfied. We can show (iii) as follows.

$$
\begin{aligned}
& \sup_{||\theta-\theta^0||\le\delta_J}\frac{||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})-\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]-\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)||}{J^{-\frac{1}{2}}+||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||+||\,\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]||} \\
\le & \sup_{||\theta-\theta^0||\le\delta_J} J^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})-\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]-\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)|| \\
\le & \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)-\mathrm{E}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)]-\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d^0)|| \\
& + \sup_{||\theta-\theta^0||\le\delta_J} J^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})-\mathrm{E}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})]-\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}^0)|| \\
= & \; o_p(1)+o_p(1) \\
= & \; o_p(1)
\end{aligned}
$$

where the first $o_p(1)$ term comes from

$$
\begin{aligned}
& \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)-\mathrm{E}[\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d)]-\boldsymbol{\mathcal{G}}_J^d(\boldsymbol{\theta}_d^0)|| \\
& = \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}\left|\left|\boldsymbol{G}_J^d(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)+J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\left\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\}\right.\right. \\
& \quad - \mathrm{E}_{\epsilon,\epsilon^*,\mathrm{x}_1,\xi}\left[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0)+J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\left\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&-\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) - J^{-1}\boldsymbol{Z}_d'\boldsymbol{H}_0^{-1}\left\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\}\Big|\Big| \\
= & \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}\Big|\Big|\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{x}_1,\xi}\left[\boldsymbol{G}_J^d(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0)\right] - \boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\Big|\Big| \\
= & o_p(1)
\end{aligned}
$$

by assumption B3(a), and the second $o_p(1)$ term comes from

$$
\begin{aligned}
&\sup_{||\theta-\theta^0||\le\delta_J} J^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}) - \mathrm{E}[\boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta})] - \boldsymbol{\mathcal{G}}_J^c(\boldsymbol{\theta}^0)|| \\
\le & \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}\Big|\Big|\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\} \\
&- \mathrm{E}_{\epsilon,\epsilon^*,\mathrm{x}_1,\xi,\mathrm{w}_1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}] \\
&-\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) + J^{-1}\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\}\Big|\Big| \\
= & \sup_{||\theta_d-\theta_d^0||\le\delta_J} J^{\frac{1}{2}}\Big|\Big|\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \mathrm{E}_{\mathrm{w}_1,\omega}[\boldsymbol{G}_J^c(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0)] - \boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)\Big|\Big| \\
= & o_p(1)
\end{aligned}
$$

by assumption B3(b). Assumption B1 ensures condition (v). Let us turn to show (iv). We set

$$
(\boldsymbol{a}_1^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \ldots, \boldsymbol{a}_J^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) \equiv \boldsymbol{Z}_d'\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \tag{A.24}
$$

$$
(\boldsymbol{a}_1^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P), \ldots, \boldsymbol{a}_J^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)) \equiv -\boldsymbol{Z}_c'\boldsymbol{L}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\boldsymbol{M}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\boldsymbol{H}^{-1}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P). \tag{A.25}
$$

Decompose $J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$ into the tree terms:

$$
\begin{aligned}
&J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0) \\
= & J^{\frac{1}{2}}\boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) + J^{-\frac{1}{2}}\begin{pmatrix} \boldsymbol{Z}_d'\boldsymbol{H}_0^{-1} \\ -\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\end{pmatrix}\left\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\} \\
= & \sum_{j=1}^{J}\begin{pmatrix} J^{-\frac{1}{2}}\boldsymbol{z}_j^d\xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ J^{-\frac{1}{2}}\boldsymbol{z}_j^c\omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)\end{pmatrix} + J^{-\frac{1}{2}}\begin{pmatrix} \boldsymbol{Z}_d'\boldsymbol{H}_0^{-1} \\ -\boldsymbol{Z}_c'\boldsymbol{L}_0\boldsymbol{M}_0\boldsymbol{H}_0^{-1}\end{pmatrix}\left\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0)\right\} \\
= & \sum_{j=1}^{J}\begin{pmatrix} J^{-\frac{1}{2}}\boldsymbol{z}_j^d\xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ J^{-\frac{1}{2}}\boldsymbol{z}_j^c\omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)\end{pmatrix} + \sum_{i=1}^{n}\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0) \\
&-\sum_{r=1}^{R}\boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)
\end{aligned} \tag{A.26}
$$

where

$$
\begin{aligned}
\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) &= \frac{1}{nJ^{\frac{1}{2}}}\sum_{j=1}^{J}\begin{pmatrix} \boldsymbol{a}_j^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\epsilon_{ji} \\ \boldsymbol{a}_j^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\epsilon_{ji}\end{pmatrix}, \\
\boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P) &= \frac{1}{RJ^{\frac{1}{2}}}\sum_{j=1}^{J}\begin{pmatrix} \boldsymbol{a}_j^d(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\epsilon_{jr}^*(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d) \\ \boldsymbol{a}_j^c(\boldsymbol{\xi}, \boldsymbol{\theta}_d, P)\epsilon_{jr}^*(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\theta}_d)\end{pmatrix}.
\end{aligned}
$$

Note that the first term on the right-hand side of (A.26) is random because of the product characteristics $(\boldsymbol{X}_1, \boldsymbol{\xi})$ and the cost shifter $(\boldsymbol{W}_1, \boldsymbol{\omega})$. However, at $(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)$, these $\xi_j$'s and $\omega_j$'s are independent as stated in page 5. This forces us to condition only on $(\boldsymbol{X}_1, \boldsymbol{W}_1)$ to make the each component on the term independent. On the other hand, the second term on the right-hand side of (A.26), originating from the sampling error in calculating the observed market share, is dependent on $(\boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{W}, \boldsymbol{\omega})$. Similarly for the third term corresponding to the simulation error in calculating the market share. We show that

$$\left\{\mathrm{V}\left[\boldsymbol{b}' J^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)\right]\right\}^{-1/2} \boldsymbol{b}' J^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0) \tag{A.27}$$

is asymptotically normal with mean zero and variance one for any real constant vector $\boldsymbol{b}$ such that $\boldsymbol{b}'\boldsymbol{b} = 1$. Then the Cramér-Wold device says that $J^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$ converges to multivariate normal. Since the three terms in (A.26), denoted $\boldsymbol{T}_{J1}, \boldsymbol{T}_{J2}, \boldsymbol{T}_{J3}$, have mean zero and are mutually uncorrelated, it is sufficient to show that each of $\{\mathrm{V}[\boldsymbol{b}'\boldsymbol{T}_{Jl}]\}^{-1/2}\boldsymbol{b}'\boldsymbol{T}_{Jl}, l = 1, 2, 3$ is asymptotically normal.[7] Notice that each element of $\boldsymbol{T}_{Jl}$ is the sum of non-independent, but conditionally independent random sequence. Thus we have to use a version of central limit theorem which is applicable to conditionally independent random sequences. In appendix **??**, we derive the version for Lyapunov central limit theorem.

The first term $\boldsymbol{b}'\boldsymbol{T}_{J1}$:

Given $(\boldsymbol{X}_1, \boldsymbol{W}_1)$, $(\boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0))$ are conditionally independent across $j$. Set

$$\zeta_i = \left\{\mathrm{V}\left[\boldsymbol{b}' J^{-\frac{1}{2}} \sum_{j=1}^{J} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right]\right\}^{-1/2} \boldsymbol{b}' J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}$$

and $Z = (\boldsymbol{X}_1, \boldsymbol{W}_1)$ for the central limit theorem in the appendix. Then, by assumption B4(a) and B4(d), we can show that the Lyapunov condition is satisfied for the first term as follows.

$$\begin{aligned}
&\lim_{J\to\infty} \sum_{j=1}^{J} \mathrm{E}\left[\left|\left\{\mathrm{V}\left[\boldsymbol{b}' J^{-\frac{1}{2}} \sum_{j=1}^{J} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right]\right\}^{-1/2} \boldsymbol{b}' J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right|^{2+\delta}\right] \\
&= \lim_{J\to\infty} \sum_{j=1}^{J} \left\{\mathrm{V}\left[\boldsymbol{b}' J^{-\frac{1}{2}} \sum_{j=1}^{J} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right]\right\}^{-(2+\delta)/2} \mathrm{E}\left[\left|\boldsymbol{b}' J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right|^{2+\delta}\right] \\
&\leq \lim_{J\to\infty} \sum_{j=1}^{J} \left\{\boldsymbol{b}' \mathrm{V}\left[J^{-\frac{1}{2}} \sum_{j=1}^{J} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right] \boldsymbol{b}\right\}^{-(2+\delta)/2} \\
&\qquad \times ||\boldsymbol{b}'||^{2+\delta} \mathrm{E}\left[\left|\left|J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right|\right|^{2+\delta}\right] \\
&= \{\boldsymbol{b}'\boldsymbol{\Phi}_1\boldsymbol{b}\}^{-(2+\delta)/2} ||\boldsymbol{b}'||^{2+\delta} \lim_{J\to\infty} \sum_{j=1}^{J} \mathrm{E}\left[\left|\left|J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right|\right|^{2+\delta}\right] \\
&= 0
\end{aligned}$$

for some $\delta > 0$. Thus we have

$$\begin{aligned}
&\left\{\mathrm{V}\left[\boldsymbol{b}' J^{-\frac{1}{2}} \sum_{j=1}^{J} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix}\right]\right\}^{-1/2} \sum_{j=1}^{J} \boldsymbol{b}' J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix} \\
&\overset{w}{\rightsquigarrow} N(0, 1)
\end{aligned}$$

which is equivalent to saying that

$$\boldsymbol{b}' \sum_{j=1}^{J} J^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{z}_j^d \xi_j(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) \\ \boldsymbol{z}_j^c \omega_j(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) \end{pmatrix} \overset{w}{\rightsquigarrow} N(0, \boldsymbol{b}'\boldsymbol{\Phi}_1\boldsymbol{b}). \tag{A.28}$$

[7] These three terms are not mutually independent due to inclusion of the common random variables $\boldsymbol{X}$ and $\boldsymbol{\xi}$.

The second term $\boldsymbol{b}'\boldsymbol{T}_{J2} = \boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$:
Abbreviate $\boldsymbol{Y}_{Ji}^0 = \boldsymbol{Y}_{Ji}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Given $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{W}, \boldsymbol{\omega}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0))$, $\boldsymbol{Y}_{Ji}^0$ are conditionally independent across $i$.
Set $\zeta_i = \{V[\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}^0]\}^{-\frac{1}{2}}\boldsymbol{b}'\boldsymbol{Y}_{Ji}^0$ and $Z = (\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{W}, \boldsymbol{\omega}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0))$ for the central limit theorem in the appendix. The Lyapunov condition for this term is

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}\left[\left|\left\{\mathrm{V}\left[\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}^0\right]\right\}^{-\frac{1}{2}}\boldsymbol{b}'\boldsymbol{Y}_{Ji}^0\right|^{2+\delta}\right] \\
&= \lim_{n\to\infty}\left\{\mathrm{V}\left[\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}^0\right]\right\}^{-(2+\delta)/2}\sum_{i=1}^{n}\mathrm{E}[|\boldsymbol{b}'\boldsymbol{Y}_{Ji}^0|^{2+\delta}] \\
&\leq \lim_{n\to\infty}\left\{\boldsymbol{b}'\,\mathrm{V}\left[\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}^0\right]\boldsymbol{b}\right\}^{-(2+\delta)/2}||\boldsymbol{b}'||^3\sum_{i=1}^{n}\mathrm{E}[||\boldsymbol{Y}_{Ji}^0||^{2+\delta}] \\
&= \{\boldsymbol{b}'\boldsymbol{\Phi}_2\boldsymbol{b}\}^{-(2+\delta)/2}||\boldsymbol{b}'||^{2+\delta}\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}[||\boldsymbol{Y}_{Ji}^0||^{2+\delta}] \\
&= 0
\end{aligned}$$

by assumption B4(b) and B4(e). Thus we have

$$\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{Ji}^0 \overset{w}{\rightsquigarrow} N(0, \boldsymbol{b}'\boldsymbol{\Phi}_2\boldsymbol{b}). \tag{A.29}$$

The third term $\boldsymbol{b}'\boldsymbol{T}_{J3} = \boldsymbol{b}'\sum_{r=1}^{R}\boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$:
The argument is completely same as that for the second term.
Abbreviate $\boldsymbol{Y}_{Jr}^{*0} = \boldsymbol{Y}_{Jr}^*(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)$. Then, by using the central limit theorem with B4(c) and B4(f), we have

$$\boldsymbol{b}'\sum_{r=1}^{n}\boldsymbol{Y}_{Jr}^{*0} \overset{w}{\rightsquigarrow} N(0, \boldsymbol{b}'\boldsymbol{\Phi}_3\boldsymbol{b}). \tag{A.30}$$

Since the three terms in $\boldsymbol{b}'J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$ converges respectively to normal each of them are uncorrelated, so is $\boldsymbol{b}'J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)$.

$$\boldsymbol{b}'J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0) \overset{w}{\rightsquigarrow} N(0, \boldsymbol{b}'\boldsymbol{\Phi}\boldsymbol{b}) \tag{A.31}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \boldsymbol{\Phi}_3$. This completes the proof for the theorem 2. □

**Proof of Theorem 3**

We will show that

(1-i)' the estimator $\tilde{\boldsymbol{\theta}}$ defined as any sequence that satisfies

$$||\boldsymbol{G}_{J,T}(\tilde{\boldsymbol{\theta}}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| = \inf_{\theta\in\Theta}||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| + o_p(1)$$

is consistent for $\boldsymbol{\theta}^0$, and

(1-ii)' $\sup_{\theta\in\Theta}||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^n, P^R, \boldsymbol{\eta}^N) - \boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| = o_p(1)$.

To show (1-i)', Theorem 3.1 of Pakes and Pollard (1989) requires

(i-a)' $\boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)$, and

(i-b)' $\sup_{\theta\notin\mathcal{N}_{\theta^0}(\delta)}||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||^{-1} = O_p(1)$ for each $\delta > 0$.

**Proof of (i-a)'**

Since we have shown that $\boldsymbol{G}_J^d(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0) = o_p(1)$ and $\boldsymbol{G}_J^c(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0) = o_p(1)$, the remaining is to show that $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)$. We apply Bernoulli's weak law of large number to each row of $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)$. We denote the element of $\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}, P, \boldsymbol{\eta})$ corresponding to consumer's demographic $d$ and discriminating attribute $q$ as $\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d, \boldsymbol{s}, P, \boldsymbol{\eta})\}_{d,q}, d = 1, \ldots, D, q = 1, \ldots, N_p$. By the definition of $\eta_{dq}^0$ given in (76), the expectation and the variance of $\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}$ are respectively

$$
\begin{aligned}
&\mathrm{E}[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{E}_{\nu|\mathrm{x},\xi}\left[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\eta_{dq}^0 - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\nu|\mathrm{x},\xi}\left[\nu_{td}^o\frac{\sum_{j\in\mathcal{Q}_q}\sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0)}{\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\eta_{dq}^0 - \frac{1}{T}\sum_{t=1}^{T}\eta_{dq}^0\right] \\
&= 0, \\
&\mathrm{V}[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{V}_{\nu|\mathrm{x},\xi}\left[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad + \mathrm{V}_{\mathrm{x},\xi}\left[\mathrm{E}_{\nu|\mathrm{x},\xi}\left[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{V}_{\nu|\mathrm{x},\xi}\left[\{\boldsymbol{G}_{J,T}^a(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{V}_{\nu|\mathrm{x},\xi}\left[\eta_{d,q}^0 - \frac{1}{T}\sum_{t=1}^{T}\nu_{td}^o\frac{\sum_{j\in\mathcal{Q}_q}\sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0)}{\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{V}_{\nu|\mathrm{x},\xi}\left[\nu_{td}^o\frac{\sum_{j\in\mathcal{Q}_q}\sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0)}{\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)}\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&= \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[\mathrm{E}_{\nu|\mathrm{x},\xi}\left[\left\{\nu_{td}^o\frac{\sum_{j\in\mathcal{Q}_q}\sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0)}{\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)}\right\}^2\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right] - \left(\eta_{dq}^0\right)^2\right] \\
&= \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[1\Big/\left\{\textstyle\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\right\}^2\right. \\
&\qquad \left.\times\,\mathrm{E}_{\nu|\mathrm{x},\xi}\left[\left\{\nu_{td}^o\textstyle\sum_{j\in\mathcal{Q}_q}\sigma_{tj}(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\nu}_t; \boldsymbol{\theta}_d^0)\right\}^2\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)\right]\right] \\
&\quad -\frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[\left(\eta_{dq}^0\right)^2\right] \\
&= \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[1\Big/\left\{\textstyle\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\right\}^2\right. \\
&\qquad \left.\times\,\mathrm{E}_{\nu|\mathrm{x},\xi}\left[(\nu_{td}^o)^2\middle|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_0^d, \boldsymbol{s}^0, P^0)\right]\right] - \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[\left(\eta_{dq}^0\right)^2\right]
\end{aligned}
$$

Since the distributional support of consumer's demographic is assumed bounded, its second moment is finite, i.e., $\mathrm{E}_{\nu|\mathrm{x},\xi}[(\nu_{td}^o)^2|\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0)] = \mathrm{E}_\nu[(\nu_{td}^o)^2] \leq M$ for some constant $M < \infty$. Assumption A9 guarantees that

$$
\mathrm{E}_{\mathrm{x},\xi}\left[1\Big/\left\{\textstyle\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\right\}^2\right] = O(1).
$$

Moreover, we have $\mathrm{E}_{\mathrm{x},\xi}\left[\left(\eta^0_{dq}\right)^2\right] = O(1)$ from assumption A10(b). Thus the variance of $\{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}$ is

$$\begin{aligned}
&\mathrm{V}[\{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)\}_{d,q}] \\
&\leq \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[M\Big/\left\{\textstyle\sum_{j\in\mathcal{Q}_q}\sigma_j(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}^0_d, P^0)\right\}^2\right] - \frac{1}{T}\mathrm{E}_{\mathrm{x},\xi}\left[\left(\eta^0_{dq}\right)^2\right] \\
&\leq O_p(1/T) + O_p(1/T) = o_p(1).
\end{aligned}$$

Thus Bernoulli's weak law of large number ensures that $\{\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)\}_{d,q}$ as $T \to \infty$ (and hence $J \to \infty$).

**Proof of (i-b)'**

From argument similar to deriving (A.3), for any $(\epsilon, \delta) > (0, 0)$ and $C(\delta)$, the relationship

$$\begin{aligned}
&\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| \geq C(\delta) - \epsilon/2\right] \\
&\geq \Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| \geq C(\delta)\right] \\
&\qquad - \Pr\left[||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| \geq \epsilon/2\right] \qquad (A.32)
\end{aligned}$$

holds in general. Since $\boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) = o_p(1)$, for any $\epsilon > 0$, there exist $J_1(\epsilon)$ and $T_1(\epsilon)$ such that when $J > J_1$ and $T > T_1$

$$\Pr\left[||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)|| \geq \epsilon/2\right] \leq \epsilon/2. \qquad (A.33)$$

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

$$\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)||^2 < C_2(\delta)\right] < \frac{\epsilon}{4}.$$

From assumption A8, for the $(\epsilon, \delta)$, there exists $C_3(\delta)$, $J_3(\epsilon, \delta)$ and $T_3(\epsilon, \delta)$ such that when $J > J_3$ and $T > T_3$

$$\Pr\left[\inf_{\theta_d\notin\mathcal{N}_{\theta^0_d}(\delta)} ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||^2 < C_3(\delta)\right] < \frac{\epsilon}{4}.$$

Thus when $J > \min(J_2, J_3)$ and $T > T_3$

$$\begin{aligned}
&\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||^2 < C_2(\delta) + C_3(\delta)\right] \\
&= \Pr\Big[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} \Big\{||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)||^2 \\
&\qquad + ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||^2\Big\} < C_2(\delta) + C_3(\delta)\Big] \\
&\leq \Pr\Big[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0) - \boldsymbol{G}_J(\boldsymbol{\theta}^0, \boldsymbol{s}^0, P^0)||^2 \\
&\qquad + \inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)||^2 < C_2(\delta) + C_3(\delta)\Big]
\end{aligned}$$

$$
\begin{aligned}
&\leq \Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^0,P^0)-\boldsymbol{G}_J(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0)||^2 < C_2(\delta)\right] \\
&\quad + \Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)-\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)||^2 < C_3(\delta)\right] \\
&\leq \frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{\epsilon}{2}.
\end{aligned}
$$

By setting $C(\delta)=\{C_2(\delta)+C_3(\delta)\}^{\frac{1}{2}}$, we have

$$
\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)-\boldsymbol{G}_{J,T}(\boldsymbol{\theta}^0,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)|| \geq C(\delta)\right] \geq 1-\frac{\epsilon}{2}. \tag{A.34}
$$

By substituting (A.33) and (A.34) for (A.32), when $J>\max(J_1,J_2,J_3)$ and $T>\max(T_1,T_2,T_3)$,

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

Then we have

$$
\limsup_{J,T}\Pr\left[\inf_{\theta\notin\mathcal{N}_{\theta^0}(\delta)} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)|| > C^*(\epsilon,\delta)\right] \geq 1-\epsilon \tag{A.35}
$$

for $C^*(\epsilon,\delta)=C(\delta)-\epsilon/2$ and hence (i-b)' is shown.

**Proof of (1-ii)'**

We show

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

From (1-ii) in the proof of Theorem 1, we know that the first term of the right-hand side in the following inequality converges to zero in probability as $J$ goes to infinity.

$$
\begin{aligned}
&\sup_{\theta\in\Theta} ||\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)-\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)|| \\
&\quad \leq \sup_{\theta\in\Theta} ||\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^n,P^R)-\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^0,P^0)|| \\
&\quad\quad + \sup_{\theta_d\in\Theta_d} ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)-\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)||.
\end{aligned} \tag{A.36}
$$

In the following, we see the second term in (A.36) to be $o_p(1)$.

$$
\begin{aligned}
&\sup_{\theta_d\in\Theta_d} ||\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)-\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)|| \\
&= \sup_{\theta_d\in\Theta_d} ||\boldsymbol{\eta}^N - T^{-1}\textstyle\sum_{t=1}^T \boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^R),\boldsymbol{\theta}_d,P^R) \\
&\quad\quad -\{\boldsymbol{\eta}^0 - T^{-1}\textstyle\sum_{t=1}^T \boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0),\boldsymbol{\theta}_d,P^0)\}|| \\
&\leq ||\boldsymbol{\eta}^N-\boldsymbol{\eta}^0|| \\
&\quad + \sup_{\theta_d\in\Theta_d}\left|\left|T^{-1}\textstyle\sum_{t=1}^T \boldsymbol{\nu}^o_t\otimes\{\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^R)-\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\}\right|\right| \\
&= ||\boldsymbol{\eta}^N-\boldsymbol{\eta}^0||
\end{aligned}
$$

$$
\begin{aligned}
&+\sup_{\theta_d\in\Theta_d} T^{-1}||(\boldsymbol{\nu}^o)'\{\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^R)-\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\}|| \\
\leq\ & ||\boldsymbol{\eta}^N-\boldsymbol{\eta}^0|| \\
&+T^{-1/2}||\boldsymbol{\nu}^0||\cdot\sup_{\theta_d\in\Theta_d} T^{-1/2}||\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R),\boldsymbol{\theta}_d,P^R)-\boldsymbol{\Psi}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)||, \\
=\ & O_p(N^{-1/2})+O_p(1)\cdot o_p(1)=o_p(1)
\end{aligned}
$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)=(\boldsymbol{\psi}_1(\boldsymbol{\xi},\boldsymbol{\theta}_d,P),\ldots,\boldsymbol{\psi}_T(\boldsymbol{\xi},\boldsymbol{\theta}_d,P))'$ and $\boldsymbol{\nu}^o=(\boldsymbol{\nu}_1^o,\ldots,\boldsymbol{\nu}_T^o)'$. In the last equality above, $||\boldsymbol{\eta}^N-\boldsymbol{\eta}^0||=O_p(N^{-1/2})$ comes from A10(a), and $T^{-1/2}||\boldsymbol{\nu}^o||=O_p(1)$ is because the observed consumer demographics $\boldsymbol{\nu}_t^o$ are assumed bounded. The $o_p(1)$ term follows the next inequaility with assumption A11:

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

where $\Pr[\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R)\notin\mathcal{N}_{\xi^0(\theta_d;\delta)}]\to 0$ and $\Pr[P^R\notin\mathcal{N}_{P^0(\delta)}]\to 0$. □

**Proof of Theorem 4**

In the proof of Theorem 2, we shown that the difference between $\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^n,P^R)$ and $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ is $o_p(J^{-\frac{1}{2}})$ near $\boldsymbol{\theta}^0$, or $\sup_{||\theta-\theta^0||\leq\delta_J} J^{\frac{1}{2}}||\boldsymbol{G}_J(\boldsymbol{\theta},\boldsymbol{s}^n,P^R)-\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})||=o_p(1)$. We show below that $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$ in (92) and $\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)$ in (93) is $o_p(T^{-\frac{1}{2}})$ within the $\delta_{J,T}$ neighborhood of $\boldsymbol{\theta}^0$. This makes the difference between $\boldsymbol{G}_{J,T}(\boldsymbol{\theta},\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$ in (81) and $\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta}_d)$ in (94) is stochastically small enough near $\boldsymbol{\theta}^0$.

For the element of $\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^n,P^R,\boldsymbol{\eta}^N)$ corresponding to consumer demographics $d$ and discriminating attribute $q$, we have

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

$$-\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\Big]\Bigg| \tag{A.37}$$

where $\boldsymbol{\Upsilon}_{tq\cdot}$ is the $q$th row vector of $\boldsymbol{\Upsilon}_t$. Thus, it is sufficient to show that the three terms in the right-hand side of (A.37) are respectively $o_p(1)$ or,

$$\sup_{||\theta_d-\theta_d^0||<\delta_{J,T}}\left|T^{-1/2}\sum_{t=1}^{T}\nu_{td}^o\{\psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^R)-\psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d,P^0)\}\right| = o_p(1), \tag{A.38}$$

$$\sup_{||\theta_d-\theta_d^0||<\delta_{J,T}}\Bigg|T^{-1/2}\sum_{t=1}^{T}\nu_{td}^o\Big[\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n - \boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}^{\dagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^n\Big]\Bigg| = o_p(1), \tag{A.39}$$

$$\sup_{||\theta_d-\theta_d^0||<\delta_{J,T}}\Bigg|T^{-1/2}\sum_{t=1}^{T}\nu_{td}^o\Big[\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d^0) - \boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}^{\ddagger},\boldsymbol{\theta}_d,P^R)\boldsymbol{H}^{-1}(\underline{\boldsymbol{\xi}},\boldsymbol{\theta}_d,P^R)\boldsymbol{\epsilon}^R(\boldsymbol{\theta}_d)\Big]\Bigg| = o_p(1). \tag{A.40}$$

We can obtain (A.38) as follows.

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

$$
\begin{aligned}
&= \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| T^{-1}\sum_{t=1}^{T} \nu_{td}^o \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \\
&\quad \times \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| \frac{\sum_{j\in\mathcal{Q}_q} T^{1/2}\epsilon_j^R(\boldsymbol{\theta}_d)}{\sum_{j\in\mathcal{Q}_q} s_j^0 + \sum_{j\in\mathcal{Q}_q}\epsilon_j^R(\boldsymbol{\theta}_d)} \right| \\
&= \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| T^{-1}\sum_{t=1}^{T} \nu_{td}^o \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \\
&\quad \times \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| \frac{\left(\sum_{j\in\mathcal{Q}_q} s_j^0\right)^{-1} \sum_{j\in\mathcal{Q}_q} T^{1/2}\epsilon_j^R(\boldsymbol{\theta}_d)}{1+\left(\sum_{j\in\mathcal{Q}_q} s_j^0\right)^{-1} T^{-1/2}\sum_{j\in\mathcal{Q}_q} T^{1/2}\epsilon_j^R(\boldsymbol{\theta}_d)} \right| \\
&= \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| T^{-1}\sum_{t=1}^{T} \nu_{td}^o \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \\
&\quad \times \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| \frac{O_p(1)\cdot o_p(1)}{1+O_p(1)\cdot T^{-1/2} o_p(1)} \right| \\
&= \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| T^{-1}\sum_{t=1}^{T} \nu_{td}^o \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \cdot o_p(1) \\
&= o_p(1) \qquad\qquad \text{(A.41)}
\end{aligned}
$$

where we use assumption A9 for $(\sum_{j\in\mathcal{Q}_q} s_j^0)^{-1} = O_p(1)$ and assumption B7(c) for $\sum_{j\in\mathcal{Q}_q} T^{1/2}\epsilon_j^R(\boldsymbol{\theta}_d) = o_p(1)$. For the last equality in (A.41), we use the law of large number as follows.

$$
\sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \left| T^{-1}\sum_{t=1}^{T} \nu_{td}^o \psi_{tq}(\boldsymbol{\xi}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d, P^0) \right| \xrightarrow{p} |\eta_{dq}^0| = |O_p(1)|
$$

where $\eta_{dq}^0 = O_p(1)$ follows from assumption A10(b). For (A.39), we have

$$
\begin{aligned}
&\Pr\Bigg[ \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \Bigg| T^{-1/2}\sum_{t=1}^{T} \nu_{td}^o \Big[ \boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n \\
&\qquad -\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n \Big] \Bigg| > c \Bigg] \\
\leq\ &\Pr\Bigg[ \max_t |\nu_{td}^o| \cdot \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \Bigg| T^{-1/2}\sum_{t=1}^{T} \Big[ \boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n \\
&\qquad -\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}^\dagger, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n \Big] \Bigg| > c \Bigg] \\
\leq\ &\Pr\Bigg[ \max_t |\nu_{td}^o| \cdot \sup_{||\theta_d-\theta_d^0||<\delta_{J,T}} \ \sup_{(\xi_1,\xi_2,P)\in\{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_{J,T})\}^{2J}\times\mathcal{N}_{P^0}(\delta_{J,T})} \Bigg| T^{-1/2} \\
&\quad \times \sum_{t=1}^{T} \Big[ \boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0), \boldsymbol{\theta}_d^0, P^0)\boldsymbol{H}_0^{-1}\boldsymbol{\epsilon}^n \\
&\qquad -\boldsymbol{\Upsilon}_{tq\cdot}(\boldsymbol{\xi}_1, \boldsymbol{\theta}_d, P^R)\boldsymbol{H}^{-1}(\boldsymbol{\xi}_2, \boldsymbol{\theta}_d, P^R)\boldsymbol{\epsilon}^n \Big] \Bigg| > c \Bigg] \\
&\quad + \Pr[\boldsymbol{\xi}^\dagger \notin \{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_{J,T})\}^J] + \Pr[\bar{\boldsymbol{\xi}} \notin \{\mathcal{N}_{\xi^0}(\theta_d^0;\delta_{J,T})\}^J] \\
&\quad + \Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \\
=\ &o(1)
\end{aligned}
$$

where we use assumption that $\max_t |\nu^o_{td}| < M$(constant), B7(a) and the facts $\Pr[\boldsymbol{\xi}^\dagger \notin \{\mathcal{N}_{\xi^0}(\theta^0_d; \delta_{J,T})\}^J] \to 0$, $\Pr[\bar{\boldsymbol{\xi}} \notin \{\mathcal{N}_{\xi^0}(\theta^0_d; \delta_{J,T})\}^J] \to 0$, $\Pr[P^R \notin \mathcal{N}_{P^0}(\delta_{J,T})] \to 0$. We can also obtain (A.40) by similar argument as for (A.39) by using assumption B7(b).

What we next show is the asymptotic normality of the estimator $\check{\boldsymbol{\theta}}$ that minimizes the norm of $\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})$ in (94). To do this, we use a version of Theorem 3.3 in Pakes and Pollard (1989) described in appendix **??**, which gives asymptotic normality to the estimator indexed by two distinct indices. From the theorem, if we can show the following five conditions,

**(i)'** $||\boldsymbol{\mathcal{G}}_{J,T}(\check{\boldsymbol{\theta}})|| = o_p(J^{-\frac{1}{2}}) + o_p(T^{-\frac{1}{2}}) + \inf_\theta ||\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})||$;

**(ii)'** $\mathrm{E}[\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})]$ is differentiable at $\boldsymbol{\theta}^0$ with a derivative matrix $\boldsymbol{\Gamma}_{J,T} = (\boldsymbol{\Gamma}'_J, \boldsymbol{\Gamma}^a_{J,T})'$ of full rank where $\boldsymbol{\Gamma}_{J,T}$ converges to $(\boldsymbol{\Gamma}', \boldsymbol{\Gamma}^a)'$ as $J, T \to \infty$;

**(iii)'** for every sequence $\{\delta_{J,T}\}$ of positive numbers that converges to zero as $J, T$ goes to infinity,

$$\text{(a)} \qquad \sup_{||\theta-\theta^0|| \le \delta_{J,T}} \frac{||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}) - \mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})] - \boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)||}{J^{-\frac{1}{2}} + ||\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})|| + ||\,\mathrm{E}[\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})]||} = o_p(1);$$

$$\text{(b)} \qquad \sup_{||\theta_d-\theta_d^0|| \le \delta_{J,T}} \frac{||\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d) - \mathrm{E}[\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)] - \boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)||}{T^{-\frac{1}{2}} + ||\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)|| + ||\,\mathrm{E}[\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)]||} = o_p(1);$$

**(iv)'**

$$\begin{pmatrix} J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0) \\ T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d) \end{pmatrix} \overset{w}{\rightsquigarrow} N\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}^a \end{pmatrix}\right);$$

**(v)'** $\boldsymbol{\theta}^0$ is an interior point of $\Theta$,

**(vi)'** The size index $T$ grows faster than $J$ ($T/J \to \infty$ as $J \to \infty$),

then, we have $\check{\boldsymbol{\theta}} \overset{w}{\rightsquigarrow} N(\mathbf{0}, \boldsymbol{V})$ where

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

We are considering the situation where the number $T$ of consumer draws used to evaluate the additional moments is larger and grows faster than the number $J$ of products, and thus (vi)' is satisfied. Our estimator $\check{\boldsymbol{\theta}}$ satisfies (i)'. Since the three random variables $\epsilon_{ji}$, $\epsilon^*_{jr}$ and $\boldsymbol{\epsilon}^{\#}_{i'}$ in $\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})$ have respectively zero means given the set of product characteristics $(\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0))$, we have $\mathrm{E}[\boldsymbol{G}_{J,T}(\boldsymbol{\theta}, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)] = \mathrm{E}[\boldsymbol{\mathcal{G}}_{J,T}(\boldsymbol{\theta})]$. Thus condition (ii)' follows from assumptions B2 and B8. We shown condition (iii)'(a) in the proof of Theorem 2. For condition (iii)'(b), we have

$$\begin{aligned}
&\sup_{||\theta_d-\theta_d^0|| \le \delta_{J,T}} \frac{||\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d) - \mathrm{E}[\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)] - \boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)||}{T^{-\frac{1}{2}} + ||\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)|| + ||\,\mathrm{E}[\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)]||} \\
\le\ & \sup_{||\theta_d-\theta_d^0|| \le \delta_{J,T}} T^{\frac{1}{2}}||\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d) - \mathrm{E}[\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)] - \boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)|| \\
=\ & \sup_{||\theta_d-\theta_d^0|| \le \delta_{J,T}} T^{\frac{1}{2}}\Bigg|\Bigg|\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) - \frac{1}{T}\sum_{t=1}^T \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}^0_t \boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d)\} + \boldsymbol{\eta}^N - \boldsymbol{\eta}^0 \\
&- \mathrm{E}[\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0)] + \frac{1}{T}\sum_{t=1}^T \mathrm{E}\left[\boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}^0_t \boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d)\}\right] - \mathrm{E}[\boldsymbol{\eta}^N - \boldsymbol{\eta}^0] \\
&- \boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d, \boldsymbol{s}^0, P^0, \boldsymbol{\eta}^0) + \frac{1}{T}\sum_{t=1}^T \boldsymbol{\nu}^o_t \otimes \boldsymbol{\Upsilon}^0_t \boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n - \boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d)\} - \boldsymbol{\eta}^N - \boldsymbol{\eta}^0\Bigg|\Bigg|
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{||\theta_d-\theta_d^0||\le\delta_{J,T}} T^{\frac{1}{2}}\left\|\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)-\mathrm{E}[\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)]-\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0)\right\| \\
&= o_p(1)
\end{aligned}
$$

from assumption B9. Assumption B1 guarantees condition (v)'. Let us show (iv)'. The additional moments $\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)$ includes two random draws of consumer $\boldsymbol{\nu}^o_t, t=1,\ldots,T$ and $\boldsymbol{\epsilon}^{\#}_{i'}, i'=1,\ldots,N$, which are not included in $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$. Thus $\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}_d)$ and $\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta})$ are conditionally independent, conditional on the set of product characteristics $(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0))$, and then uncorrelated each other. Since we also know that $J^{\frac{1}{2}}\boldsymbol{\mathcal{G}}_J(\boldsymbol{\theta}^0)\overset{w}{\rightsquigarrow}N(\boldsymbol{0},\boldsymbol{\Phi})$ as $J\to\infty$, what we have to show is $T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)\overset{w}{\rightsquigarrow}N(\boldsymbol{0},\boldsymbol{\Phi}^a)$ as $J,T\to\infty$. Set

$$
(\boldsymbol{a}^a_1(\boldsymbol{\xi},\boldsymbol{\theta}_d,P),\ldots,\boldsymbol{a}^a_1(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)) \equiv -\sum_{t=1}^{T}\boldsymbol{\nu}^o_t\otimes\boldsymbol{\Upsilon}_t(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\boldsymbol{H}^{-1}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P). \tag{A.42}
$$

Decompose $T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)$ into the four terms:

$$
\begin{aligned}
&T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d) \\
&= T^{\frac{1}{2}}\boldsymbol{G}^a_{J,T}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0,\boldsymbol{\eta}^0) - T^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{\nu}^o_t\otimes\boldsymbol{\Upsilon}^0_t\boldsymbol{H}_0^{-1}\{\boldsymbol{\epsilon}^n-\boldsymbol{\epsilon}^R(\boldsymbol{\theta}^0_d)\} \\
&\quad +T^{\frac{1}{2}}(\boldsymbol{\eta}^N-\boldsymbol{\eta}^0) \\
&= \sum_{t=1}^{T}T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)+\sum_{i=1}^{n}\boldsymbol{Y}^a_{J,T,i}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0) \\
&\quad -\sum_{r=1}^{R}\boldsymbol{Y}^{*a}_{J,T,r}(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)+\sum_{i'=1}^{N}T^{\frac{1}{2}}N^{-1}\boldsymbol{\epsilon}^{\#}_{i'}
\end{aligned} \tag{A.43}
$$

where

$$
\begin{aligned}
\boldsymbol{Y}^a_{J,T,i}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) &= \frac{1}{nT^{\frac{1}{2}}}\sum_{j=1}^{J}\boldsymbol{a}^a_j(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon_{ji}, \\
\boldsymbol{Y}^{*a}_{J,T,r}(\boldsymbol{\xi},\boldsymbol{\theta}_d,P) &= \frac{1}{RT^{\frac{1}{2}}}\sum_{j=1}^{J}\boldsymbol{a}^a_j(\boldsymbol{\xi},\boldsymbol{\theta}_d,P)\epsilon^*_{jr}.
\end{aligned}
$$

Since the four terms of $T^{\frac{1}{2}}\boldsymbol{\mathcal{G}}^a_{J,T}(\boldsymbol{\theta}^0_d)$ in (A.43) are conditionally independent given $(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0))$ and thus mutually uncorrelated, we will show that each of them, denoted by $\boldsymbol{T}^a_{J,T,1}$, $\boldsymbol{T}^a_{J,T,2}$, $\boldsymbol{T}^a_{J,T,3}$ and $\boldsymbol{T}^a_{J,T,4}$, are respectively asymptotically multivariate normal by using the Cramér-Wold device. We show that for any constant vector $\boldsymbol{b}$ such that $\boldsymbol{b}'\boldsymbol{b}=1$, $\{\mathrm{V}[\boldsymbol{b}'\boldsymbol{T}^a_{J,T,l}]\}^{-1/2}\boldsymbol{b}'\boldsymbol{T}^a_{J,T,l}$ for $l=1,2,3,4$ is respectively asymptotically standard normal.

The first term $\boldsymbol{b}'\boldsymbol{T}^a_{J,T,1}=\boldsymbol{b}'\sum_{t=1}^{T}T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)$:

Given $(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0))$, $\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0))$ have zero mean and are conditionally independent across $t$. Write $\boldsymbol{\psi}^0_t\equiv\boldsymbol{\psi}_t(\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}^0_d,P^0)$ and set $\zeta_i=\{\mathrm{V}[\boldsymbol{b}'T^{-\frac{1}{2}}\sum_{t=1}^{T}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}^0_t)]\}^{-1/2}\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}^0_t)$ and $Z=(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}^0_d,\boldsymbol{s}^0,P^0)))$ in the central limit theorem given in appendix **??**. Then, the Lyapunov condition for this term is

$$
\lim_{T\to\infty}\sum_{t=1}^{T}\mathrm{E}\left[\left|\left\{\mathrm{V}\left[\boldsymbol{b}'T^{-\frac{1}{2}}\sum_{t=1}^{T}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}^0_t)\right]\right\}^{-1/2}\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}^o_t\otimes\boldsymbol{\psi}^0_t)\right|^{2+\delta}\right]
$$

$$
\begin{aligned}
&= \lim_{T\to\infty}\left\{\mathrm{V}\left[\boldsymbol{b}'T^{-\frac{1}{2}}\sum_{t=1}^{T}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right]\right\}^{-(2+\delta)/2}\sum_{t=1}^{T}\mathrm{E}\left[\left|\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right|^{2+\delta}\right]\\
&\le \lim_{T\to\infty}\left\{\boldsymbol{b}'\,\mathrm{V}\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right]\boldsymbol{b}\right\}^{-(2+\delta)/2}\sum_{t=1}^{T}||\boldsymbol{b}'||^{2+\delta}\,\mathrm{E}\left[\left|\left|T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right|\right|^{2+\delta}\right]\\
&= \{\boldsymbol{b}'\boldsymbol{\Phi}_1^a\boldsymbol{b}\}^{-(2+\delta)/2}\,||\boldsymbol{b}'||^{2+\delta}\lim_{T\to\infty}\sum_{t=1}^{T}\mathrm{E}\left[\left|\left|T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right|\right|^{2+\delta}\right]\\
&= 0
\end{aligned}
$$

for some $\delta > 0$ by assumption B10(a) and B10(e). Thus we obtain

$$
\left\{\mathrm{V}\left[\boldsymbol{b}'T^{-\frac{1}{2}}\sum_{t=1}^{T}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\right]\right\}^{-1/2}\sum_{t=1}^{T}\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\overset{w}{\rightsquigarrow}N(0,1),
$$

which is equivalent to

$$
\sum_{t=1}^{T}\boldsymbol{b}'T^{-\frac{1}{2}}(\boldsymbol{\eta}^0-\boldsymbol{\nu}_t^o\otimes\boldsymbol{\psi}_t^0)\overset{w}{\rightsquigarrow}N(0,\boldsymbol{b}'\boldsymbol{\Phi}_1^a\boldsymbol{b}). \tag{A.44}
$$

The second term $\boldsymbol{b}'\boldsymbol{T}_{J,T,2}^a=\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{J,T,i}^a(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$:
Abbreviate $\boldsymbol{Y}_{J,T,i}^{a0}\equiv\boldsymbol{Y}_{J,T,i}^a(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$. Given $(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\{\boldsymbol{\nu}_t^o\}_{t=1}^T)$, $\boldsymbol{Y}_{J,T,i}^{a0}$ have zero mean and conditionally independent across $i$. Suppose $\zeta_i=\{\mathrm{V}[\boldsymbol{b}'\sum_{i=1}^n\boldsymbol{Y}_{J,T,i}^{a0}]\}^{-1/2}\boldsymbol{b}'\boldsymbol{Y}_{J,T,i}^{a0}$ and $Z=(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0))$ in the central limit theorem in appendix **??**. Then the Lyapunov condition for this term is

$$
\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}\left[\left|\left\{\mathrm{V}\left[\boldsymbol{b}'\sum_{i=1}^{n}\boldsymbol{Y}_{J,T,i}^{a0}\right]\right\}^{-1/2}\boldsymbol{b}'\boldsymbol{Y}_{J,T,i}^{a0}\right|^{2+\delta}\right]\\
&= \lim_{n\to\infty}\left\{\boldsymbol{b}'\,\mathrm{V}\left[\sum_{i=1}^{n}\boldsymbol{Y}_{J,T,i}^{a0}\right]\boldsymbol{b}\right\}^{-(2+\delta)/2}\sum_{i=1}^{n}\mathrm{E}\left[\left|\boldsymbol{b}'\boldsymbol{Y}_{J,T,i}^{a0}\right|^{2+\delta}\right]\\
&\le \{\boldsymbol{b}'\boldsymbol{\Phi}_2^a\boldsymbol{b}\}^{-(2+\delta)/2}||\boldsymbol{b}'||^{2+\delta}\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}\left[\left|\left|\boldsymbol{Y}_{J,T,i}^{a0}\right|\right|^{2+\delta}\right]\\
&= 0
\end{aligned}
$$

by assumption B10(b) and B10(f). Thus we obtain

$$
\sum_{i=1}^{n}\boldsymbol{b}'\boldsymbol{Y}_{J,T,i}^{a0}\overset{w}{\rightsquigarrow}N(0,\boldsymbol{b}'\boldsymbol{\Phi}_2^a\boldsymbol{b}). \tag{A.45}
$$

The third term $\boldsymbol{b}'\boldsymbol{T}_{J,T,3}^a=\boldsymbol{b}'\sum_{r=1}^{R}\boldsymbol{Y}_{J,T,r}^{a*}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$:
For this term, we can obtain the asymptotic normality from a similar argument as for the second term. Abbreviate $\boldsymbol{Y}_{J,T,r}^{a*0}\equiv\boldsymbol{Y}_{J,T,r}^{a*}(\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0),\boldsymbol{\theta}_d^0,P^0)$. By using assumption B10(c) and B10(g), we obtain

$$
\sum_{r=1}^{R}\boldsymbol{b}'\boldsymbol{Y}_{J,T,r}^{a*0}\overset{w}{\rightsquigarrow}N(0,\boldsymbol{b}'\boldsymbol{\Phi}_3^a\boldsymbol{b}). \tag{A.46}
$$

The forth term $\boldsymbol{b}'\sum_{i'=1}^{N}T^{\frac{1}{2}}N^{-1}\boldsymbol{\epsilon}_{i'}^{\#}$:
Given $(\boldsymbol{X},\boldsymbol{\xi}(\boldsymbol{\theta}_d^0,\boldsymbol{s}^0,P^0))$, $\boldsymbol{\epsilon}_{i'}^{\#}$ have zero mean and conditionally independent across $i'$. Suppose

$\zeta_i = \left\{ \mathrm{V}\left[ \boldsymbol{b}' \sum_{i'=1}^{N} T^{1/2} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right] \right\}^{-1/2} \boldsymbol{b}' T^{1/2} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#}$ and $Z = (\boldsymbol{X}, \boldsymbol{\xi}(\boldsymbol{\theta}_d^0, \boldsymbol{s}^0, P^0))$ in the central limit theorem in appendix **??**. The Lyapunov condition for this term is

$$
\begin{aligned}
& \lim_{N\to\infty} \sum_{i'=1}^{N} \mathrm{E}\left[ \left| \left\{ \mathrm{V}\left[ \boldsymbol{b}' \sum_{i'=1}^{N} T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right] \right\}^{-1/2} \boldsymbol{b}' T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right|^{2+\delta} \right] \\
= & \lim_{N\to\infty} \left\{ \mathrm{V}\left[ \boldsymbol{b}' \sum_{i'=1}^{N} T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right] \right\}^{-(2+\delta)/2} \sum_{i'=1}^{N} \mathrm{E}\left[ \left| \boldsymbol{b}' T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right|^{2+\delta} \right] \\
\leq & \lim_{N\to\infty} \left\{ \mathrm{V}\left[ \boldsymbol{b}' \sum_{i'=1}^{N} T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right] \right\}^{-(2+\delta)/2} ||\boldsymbol{b}'||^{2+\delta} \sum_{i'=1}^{N} \mathrm{E}\left[ \left|\left| T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right|\right|^{2+\delta} \right] \\
= & \{ \boldsymbol{b}' \boldsymbol{\Phi}_4^a \boldsymbol{b} \}^{-(2+\delta)/2} ||\boldsymbol{b}'||^{2+\delta} \lim_{N\to\infty} \sum_{i'=1}^{N} \mathrm{E}\left[ \left|\left| T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \right|\right|^{2+\delta} \right] \\
= & \ 0
\end{aligned}
$$

by assumption B10(d) and B10(h). Thus we obtain

$$
\sum_{i'=1}^{N} \boldsymbol{b}' T^{\frac{1}{2}} N^{-1} \boldsymbol{\epsilon}_{i'}^{\#} \overset{w}{\leadsto} N(0, \boldsymbol{b}' \boldsymbol{\Phi}_4^a \boldsymbol{b}). \tag{A.47}
$$

The four terms in $\boldsymbol{b}' T^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d^0)$ respectively converge to the normal. Accordingly, $\boldsymbol{b}' T^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d^0)$ converges to the normal. Then the Cramér-Wold device leads us to obtain

$$
T^{\frac{1}{2}} \boldsymbol{\mathcal{G}}_{J,T}^a(\boldsymbol{\theta}_d^0) \overset{w}{\leadsto} N(\boldsymbol{0}, \boldsymbol{\Phi}^a) \tag{A.48}
$$

where $\boldsymbol{\Phi}^a = \boldsymbol{\Phi}_1^a + \boldsymbol{\Phi}_2^a + \boldsymbol{\Phi}_3^a + \boldsymbol{\Phi}_4^a$. Therefore condition (iv)' is satisfied and thus this ends the proof of Theorem 4.□