Semi-Nonparametric Estimation of Random Coefficients Logit Model for Aggregate Demand

Zhentong Lu*   Xiaoxia Shi†   Jing Tao‡

Abstract

In this paper, we propose a two-step semi-nonparametric estimator for the widely used random coefficients logit demand model. The approach applies to the same setup as Berry, Levinsohn, and Pakes (1995, BLP)-type of models with many products, but has the advantage of not requiring computing demand inversion. In particular, the first step of our approach estimates the fixed coefficients via a computationally very easy linear sieve generalized method of moments (GMM). The second step uncovers the distribution of the random coefficient via a sieve minimum distance or GMM procedure. We show identification and derive the asymptotic properties of the estimator in a large market environment. Monte Carlo simulations and empirical illustrations support the theoretical results and demonstrate the usefulness of our estimator in practice.

Keywords: Demand Estimation, Differentiated Products, Random Coefficients Logit, Semi-Nonparametric Estimation

JEL: C01, C14, L10, L62

*Financial Stability Department, Bank of Canada. Email: zlu@bankofcanada.ca
†Department of Economics, University of Wisconsin-Madison. Email: xshi@ssc.wisc.edu
‡Corresponding author. Department of Economics, University of Washington. Email: jingtao@uw.edu

1 Introduction

Demand estimation for differentiated products plays a central role in modern empirical industrial organization. The groundbreaking works of Berry (1994) and Berry et al. (1995) (henceforth, BLP) provide an important framework for analyzing aggregate demand by jointly modeling consumer preference heterogeneity and addressing price endogeneity. In this framework, consumer preference heterogeneity is represented by random coefficients, and price endogeneity is explicitly modeled by the dependence of price on a market/product level demand shock. By inverting the demand system and imposing a mean independence assumption on the demand shock with instrumental variables, a nested fixed point generalized method of moment (GMM) estimator can be employed to estimate the model. The framework has been used extensively to estimate demand in various markets/industries, which in turn provides bases for analyzing market outcomes and policy issues, see Berry and Haile (2014) for a synthesis of the empirical literature applying the BLP framework.

Nevertheless, estimating a flexible model within the BLP framework is still a challenging task for many empirical applications. First, the standard estimation procedure, nested fixed point GMM, is computationally intensive and can be numerically unstable (see the discussions in Knittel and Metaxoglou (2014)). Although there has been important progress on this issue, e.g., Dubé et al. (2012), Lee and Seo (2015), and Conlon and Gortmaker (2020), among others, computational complexity is a hurdle for many applied researchers. Furthermore, largely because of this difficulty, researchers have to impose strong parametric assumptions on the distribution of random coefficients, e.g., normal distribution (almost exclusively used in practice) to reduce the number of parameters and thus to simplify the estimation problem.

The main contribution of this paper is a two-step semi-nonparametric estimator for the random coefficients logit model that is computationally easy to implement, and as a result, makes more flexible parametric and even nonparametric specifications of random coefficients feasible. The approach applies to BLP-type models with many products but has the advantage of not requiring costly numerical demand inversion.

In the first step, we transform the original demand system into a partial linear model, where the linear part captures the utility contribution of the product/market characteristics with fixed (non-random) coefficients, and the nonparametric part captures that of those with random coefficients. Approximating the nonparametric part with a linear sieve, one can easily estimate the partial linear model with 2SLS or linear GMM. This step formalizes an approach for researchers to quickly estimate the fixed (non-random) coefficients on product/market characteristics *without* even specifying the distribution of the random coefficients. In fact, it is a common practice among empirical researchers to add polynomial terms of the product characteristics in a logit regression. Our theory confirms that this is a useful way to control the effects of the random coefficients.[1] [2]

In the second step, we substitute the estimated mean utility from the first step back into the original demand system and estimate the distribution of random coefficients by minimum distance (MD) or GMM. The random coefficient distribution can be parametric or non-parametric. When it is non-parametric, we propose suitable sieve approximations designed for distribution functions. Since our approach avoids costly numerical demand inversion, allowing a flexible sieve approximation for the random coefficient distribution is much more computationally tractable than in the standard approach. We believe this is valuable in many applications. For example, the shape of the random coefficient distribution itself may be of central interest when we want to understand the distributional impacts of product characteristics on demand or to analyze the welfare implications of certain event or policy. Moreover, the shape of the random coefficients distribution may have important implications on the substitution patterns among products such as cross-product price elasticities. In some cases, normal random coefficients may have undesirable implications for the substitution

[1]Bordley (2013) provides a justification for adding quadratic terms under the assumptions that both the random coefficients and the product characteristics with random coefficients are normally distributed and that the inclusive value converges to a deterministic limit. We do not impose either assumption.

[2]Recently, Salanié and Wolak (2019) propose a fully linear model to approximate the BLP model; their estimator can be interpreted as a second-order truncation of ours.

pattern in large markets. A more flexible random coefficient distribution may then be necessary to generate realistic substitution patterns.[3]

The key for our approach to work is the logit preference shock and the large number of products (J) framework. The logit shock is needed to write the demand model as a partial linear equation in the first step. It rules out interesting models like probit, ordered logit, or pure characteristic models. But since the mixed multinomial logit model is a workhorse model in empirical IO, our specification does cover a large class of models used in empirical work.[4] The large J framework is necessary to treat the non-linear part of the partial linear equation as a functional parameter. The large J asymptotic theory is less straightforward than the large T (number of markets) one, simply because products in the same markets interact with each other and as a result, it is difficult to characterize the dependence among them. We adopt the triangular array framework of Berry et al. (2004), which relies on the conditional independence of the unobservable product characteristics for deriving the asymptotic properties of the estimator.[5] Simulations show that our estimator works well for J's as small as 5. In demand estimation data sets, it is common to have J bigger than 5.[6]

Our estimator draws on the nonparametric instrumental variable literature (e.g. Ai and Chen (2003), Chen and Pouzo (2015), Newey and Powell (2003), Hall and Horowitz (2005), and Chen and Christensen (2018)). We modify the standard asymptotic theory to handle two special features of our setting. The first is that products in oligopoly markets are dependent by design and the dependence may not resemble time series or standard spatial dependence. As mentioned above, we deal with this using the triangular array framework of Berry et al.

[3]For example, in Supplemental Appendix S3, we show that all the cross-product elasticities go to zero at the rate $1/J$ as the number of products J goes to infinity if random coefficients are normally distributed, but that needs not to be the case with a thicker-tailed distribution for the random coefficients. A model with all cross-product elasticities drifting to zero at the same rate provides a poor approximation to many of the large markets studied in industrial organization; see Ackerberg and Rysman (2005) for related discussions.

[4]Indeed, multinomial probit models can get computationally prohibitive fast as J increases, which partially explains the popularity of logit-based models. In supplemental Appendix S4.4, we investigate the performance of our estimator when the logit shock is misspecified.

[5]It may be possible to generalize this to allow appropriate weak dependence between the unobserved product quality ξ_{jt}'s, possibly using techniques in Chiang et al. (2021). We leave this for future work.

[6]For other papers on the asymptotic theory for BLP models, see Freyberger (2015), Armstrong (2016) and Moon et al. (2018).

(2004). The second special feature is that the true value of the functional parameter in the partial linear equation, although constant across products, is *random*. As unconventional as a random true value may seem, we show that it can be handled by similar arguments as those in the seminal works of Ai and Chen (2003) and Chen and Pouzo (2015) after strengthening the identification condition to a uniform one (over a deterministic functional space). We verify the uniform identification condition using the spectrum decomposition approach of Hall and Horowitz (2005) in Supplemental Appendix S2.1.

In the growing J environment, the identification of the random coefficients logit model has not been studied in the literature. We provide a simple argument for the nonparametric point identification of the distribution of random coefficients. The argument is inspired by that in Fox et al. (2011, 2012) and Fox et al. (2016). However, our approach is very different from Fox et al. (2011, 2012) and Fox et al. (2016): While they require an appropriate supply-side model in order to handle price endogeneity, we address the endogeneity problem in the first step using instrumental variables obviating the need to specify a supply-side model. Our identification result also contributes to the growing literature on semi/non-parametric identification of the aggregate demand model including Berry and Haile (2014), Dunker et al. (2017), Compiani (2018), Reynaert and Verboven (2014), Gandhi and Houde (2016). None of these papers study the large J setting that we do. More generally, the result also contributes to the literature on non-parametric identification of discrete choice models including Fox and Gandhi (2016) and Lewbel (2000).

We conduct Monte Carlo simulations to examine the finite sample performance of our estimator and to compare it with the standard parametric BLP estimator. We find that our semi-nonparametric estimator achieves similar performance to the parametric version when the parametric assumption on the distribution of random coefficients is correct, which suggests that giving up the parametric assumption does not result in much efficiency loss. More importantly, the semi-nonparametric estimator outperforms the parametric version when the parametric assumption is incorrect, and thus has the virtue of being robust to misspecification. The findings are robust across different choices of sieve spaces and criterion functions. Finally, we apply our approach to estimating demand using data from Berry et al. (1995)'s application as well as the Chinese new car market and obtain meaningful results.

The rest of the paper is organized as follows. Section 2 lays out the basic setup. Section 3 describes our semi-nonparametric estimator. Section 4 develops the asymptotic theory. Sections 5 and 6 report the results of Monte Carlo experiments and the first empirical application. And Section 7 concludes. Additional identification results, technical proofs, and the second empirical application are in the Supplemental Appendix.

2 Random Coefficients Logit Demand Model

2.1 Setup

We consider the standard BLP framework for aggregate demand. Each market $t = 1, ..., T$ consists of a cross-section of differentiated products, labeled by $j = 0, 1, ..., J_t$, and a population of ex-ante identical consumers. The product labeled by 0 is the outside option; each inside product $j \geq 1$ in market t is characterized by a K-dimensional vector of observable characteristics X_{jt} (typically including price) and an unobserved characteristic $\xi_{jt} \in \mathbb{R}$. To simplify notation, we suppress the subscript t in J_t in the following discussions but our method applies to the case with varying J_t without a problem.

Consumer preference is represented by a standard linear random utility model as in Berry (1994), i.e., the utility that consumer i derives from choosing product j in market t is

$$u_{ijt} = \delta_{jt} + X'_{2,jt} v_i + \varepsilon_{ijt}, \tag{1}$$

where the mean utility takes the form of a linear index

$$\delta_{jt} = X'_{1,jt} \beta^0 + \xi_{jt}. \tag{2}$$

Here, we have partitioned X_{jt} into $(X_{1,jt}, X_{2,jt}) \in \mathcal{X}_1 \times \mathcal{X}_2 \subseteq \mathbb{R}^{d_{x_1}} \times \mathbb{R}^{d_{x_2}}$, $d_{x_1} + d_{x_2} = K$, to distinguish product characteristics without and with random coefficients; the vector β^0 represents the fixed coefficients, the vector v_i comprises random coefficients that jointly follow an unknown distribution (CDF) $F^0(\cdot)$ (with PDF $f^0(\cdot)$) and ε_{ijt} is a Type I extreme value distributed preference shock.[7] As a convention, we normalize the location of the model by setting both the mean utility δ_{0t} and characteristics of the outside option $X_{2,0t}$ to zeroes, that is to make $u_{i0t} = \varepsilon_{i0t}$.

Each consumer i chooses a product in market t that maximizes her/his utility. The aggregation of individual choices yields the aggregate choice probability (i.e., market share) for each product $j = 0, 1, ..., J$, i.e.,

$$\sigma_j\left(\delta_t, X_{2,t}; f^0\right) = \int \frac{\exp\left(\delta_{jt} + X'_{2,jt}v\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + X'_{2,kt}v\right)} f^0(v)dv, \tag{3}$$

where $\delta_t := (\delta_{1t}, ..., \delta_{Jt})$ and $X_{2,t} := \left(X'_{2,1t}, \ldots, X'_{2,Jt}\right)'$ is a $J \times d_{x_2}$ matrix.

The primary empirical objective is to estimate the parameters of interest $\theta^0 = (\beta^0, f^0(\cdot))$ based on the demand system

$$s_{jt} = \sigma_j\left(\delta_t, X_{2,t}; f^0\right), \; j = 1, ..., J, \; \forall t, \tag{4}$$

where $s_t := (s_{1t}, ..., s_{Jt})$ (with $s_{0t} = 1 - \sum_{j=1}^{J} s_{jt} > 0$) are observed market shares.

2.2 The Standard BLP Estimator

The standard BLP estimator is constructed based on two building blocks:

1. Assume that $f^0(\cdot)$ has a parametric form $f(\cdot|\lambda^0)$ known up to λ^0. For any $\theta^0 = (\beta^0, \lambda^0)$ and market t, invert the demand system (based on the invertibility results in Berry

[7] We could extend the model to allow for market-specific distributions of random coefficients, i.e., $F_t^0(\cdot)$, with more cumbersome notation. The identification of $F_t^0(\cdot)$ would then require sufficient variation of $X_{2,jt}$ within each market t.

(1994) and Berry et al. (2013))

$$s_{jt} = \sigma_j \left(\delta_t, X_{2,t}; f(\cdot|\lambda^0)\right),\ j = 1, ..., J$$

to obtain

$$\delta_{jt} = \sigma_j^{-1} \left(s_t, X_{2,t}; f(\cdot|\lambda^0)\right),\ j = 1, ..., J,$$

where $\sigma_j^{-1}(\cdot, X_{2,t}; f(\cdot|\lambda^0))$ is the inverse demand function.

2. Find a set of instrumental variables Z_{jt} such that

$$E\left[\xi_{jt}\middle|\, Z_{jt}\right] = 0,\ j = 1, ..., J,\ \forall t. \tag{5}$$

Condition (5), together with (2), in turn, implies the following moment condition

$$E\left[\sigma_j^{-1}\left(s_t, X_{2,t}; f(\cdot|\lambda^0)\right) - X'_{1,jt}\beta^0 \middle|\, Z_{jt}\right] = 0,\ j = 1, ..., J,\ \forall t. \tag{6}$$

Thus the standard BLP GMM estimator with a parametrically specified $f(\cdot\,|\lambda)$ can be defined as

$$(\hat{\beta}, \hat{\lambda}) = \arg\min_{(\beta,\lambda)} \left\| \frac{1}{JT} \sum_{j,t} \left\{ \left[\sigma_j^{-1}\left(s_t, X_{2,t}; f(\cdot\,|\lambda)\right) - X'_{1,jt}\beta\right] Z_{jt} \right\} \right\|_W, \tag{7}$$

where $\|g\|_W^2 = g'Wg$ with a weight matrix W.

In practice, the main hurdle of implementing the BLP estimator comes from the fact that the inversion $\left\{\sigma_j^{-1}(\cdot, X_{2,t}; f(\cdot\,|\lambda)) : j = 1, ..., J\right\}$, which can only be solved numerically (via e.g. the BLP contraction mapping) for each trial of the parameter λ, is nested in the nonlinear optimization problem. Getting a stable solution to the nonlinear optimization problem requires the inversion to be solved repeatedly with high precision which is computationally expensive; see Knittel and Metaxoglou (2014) for more detailed discussion. In the next section, we introduce our semi-nonparametric estimator which offers a new way to address these challenges.

3 Our Semi-Nonparametric Estimator

3.1 A Transformation to A Partially Linear Model

A key observation that leads to our estimation strategy is the following separability property of the random coefficients logit model,

$$\begin{aligned} s_{jt} &= \int \frac{\exp\left(\delta_{jt} + X_{2,jt}'v\right)}{1+\sum_{k=1}^{J}\exp\left(\delta_{kt} + X_{2,kt}'v\right)} f^0(v)dv \\ &= \exp\left(\delta_{jt}\right) \cdot \int \frac{\exp\left(X_{2,jt}'v\right)}{1+\sum_{k=1}^{J}\exp\left(\delta_{kt} + X_{2,kt}'v\right)} f^0(v)dv. \end{aligned} \tag{8}$$

Next, we divide both sides of (8) by the outside share and take logarithm to obtain

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = X_{1,jt}'\beta^0 + \log\left[\frac{\int \frac{\exp(X_{2,jt}'v)}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X_{2,kt}'v)} f^0(v)dv}{\int \frac{1}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X_{2,kt}'v)} f^0(v)dv}\right] + \xi_{jt}. \tag{9}$$

Now observe that $\log\left[\frac{\int \frac{\exp(X_{2,jt}'v)}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X_{2,kt}'v)} f^0(v)dv}{\int \frac{1}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X_{2,kt}'v)} f^0(v)dv}\right]$ varies across j only via $X_{2,jt}$. It depends on the other products' characteristics δ_{kt} and $X_{2,kt}$ only through the so-called inclusive value (see McFadden (1974))): $S_{J,t} = S_{J,t}(v) := \log\left(1 + \sum_{k=1}^{J}\exp(\delta_{kt} + X_{2,kt}'v)\right)$ which does not vary across j.[8] Define

$$\psi^0(x_{2,jt}; S_{J,t}) = \log\left[\frac{\int \frac{\exp(x_{2,jt}'v)}{\exp(S_{J,t})} f^0(v)dv}{\int \frac{1}{\exp(S_{J,t})} f^0(v)dv}\right].$$

[8] This feature resembles that of the market-specific "price function" proposed by Bajari and Benkard (2005), in which a product's price could be written as a function of its *own* characteristics. The price function also depends on all the primitives, such as consumer preferences and product characteristics, in a given market, through terms that do not vary across products.

Then we have,

$$\log(s_{jt}/s_{0t}) = X'_{1,jt}\beta^0 + \psi^0(X_{2,jt}; S_{J,t}) + \xi_{jt}. \tag{10}$$

We treat $\psi^0(\cdot, S_{J,t})$ as a functional parameter, and estimate it along with β^0 using partially linear instrumental variable (PLIV) methods as we detail in later sections. Before going into the estimation, we would like to remark on the random parameter $\psi^0(\cdot, S_{J,t})$, which arguably is an unconventional feature of the partially linear model (10).

First, treating $\psi^0(\cdot, S_{J,t})$ as a functional parameter is the key to the simple linear structure of the first step of the estimation procedure that we describe in the next subsection. The linearity is not preserved if we instead treat $S_{J,t}(\cdot)$, or $\delta_{kt} : k = 1, \ldots, J$ as the parameter(s). Also importantly, $\psi^0(\cdot, S_{J,t})$ subsumes all the unknown $\delta_{kt} : k = 1, \ldots, J$ which would be incidental parameters and would cause estimation to break down if we insisted on estimating them individually. Although it is unconventional for a parameter to have a random true value (at least in the frequentist framework that we adopt), we show in Section 4 below that standard sieve estimation theory still goes through with some modifications.[9]

Second, the identification of $\psi^0(\cdot, S_{J,t})$ is based on the integral equation

$$\begin{aligned}
&\int\int (x_1'\beta^0 + \psi^0(x_2; S_{J,t})) f_{X_{1,jt},X_{2,jt}|Z_{jt}}(x_1, x_2|Z_{jt}) dx_1 dx_2 \\
&= \int\int (x_1'\beta + \psi(x_2; S_{J,t})) f_{X_{1,jt},X_{2,jt}|Z_{jt}}(x_1, x_2|Z_{jt}) dx_1 dx_2,
\end{aligned}$$

where the left-hand side is identified from the sample joint distribution of $\log(s_{jt}/s_{0t})$ and Z_{jt}. This differs from the integral equation that standard PLIV models solves in that the left-hand side is not $E[\log(s_{jt}/s_{0t})|Z_{jt}]$. In fact, in the large J environment, $E[\log(s_{jt}/s_{0t})|Z_{jt}]$ may not be identified from the sample distribution of $\log(s_{jt}/s_{0t})$ and Z_{jt} because $\log(s_{jt}/s_{0t})$ may be strongly dependent across j.

The inclusive value $S_{J,t}$ is useful to conceptualize our random functional parameter

[9] To illustrate the usefulness of the idea of a random parameter in the simplest setting possible, we include an example of a mean estimation in Supplemental Appendix S7.

$\psi^0(\cdot, S_{J,t})$, but will serve no purpose beyond that. Thus, to simplify notation, we will from now on write

$$\psi^0_{J,t}(\cdot) = \psi^0(\cdot, S_{J,t})$$

and treat $\psi^0_{J,t}(\cdot)$ as a function-valued market fixed effect to be estimated.

3.2 A Two-Step Estimator

After writing out the partially linear form, we estimate the structural parameter of interest (β^0, f^0) in two steps. In the first step, we estimate β^0, as well as the "reduced form" functions $\psi^0_{J,t}(\cdot)$'s; and in the second step, we estimate f^0. If the mean and standard deviation of the random coefficients are of primary interest, their estimators can be easily deduced from the estimators of f^0.

To begin, we define some notation. First, we approximate the space of the realizations of $\psi^0_{J,t}(\cdot)$, denoted as Ψ (Ψ will be rigorously defined in Section 4) by a sieve space $\Psi_{k_J,t}$, where k_J is the dimension of the sieve space. In particular, we use the linear sieve $\psi_{k_J,t}(X_{2,jt}) := \sum_{\ell=1}^{k_J} \vartheta_{\ell,t}\psi_\ell(X_{2,jt})$, where $(\vartheta_{1,t}, \ldots, \vartheta_{k_J,t})$ are unknown sieve coefficients to be estimated, and $(\psi_1(\cdot), \ldots, \psi_{k_J}(\cdot))$ are user-specified basis functions. The commonly used basis functions include the polynomial series, the Fourier series, splines and so on. Let $\beta \in \mathcal{B} \subseteq \mathbb{R}^{d_{x_1}}$, $\theta_{k_J,t} = (\beta, \psi_{k_J,t}(\cdot))$ and $\theta_{k_J} = (\beta, \psi_{k_J,1}, \ldots, \psi_{k_J,T}) \in \Theta_{k_J}$, where $\Theta_{k_J} := \mathcal{B} \times \Psi_{k_J,1} \times \cdots \times \Psi_{k_J,T}$ is the sieve space for the parameter space $\Theta := \mathcal{B} \times \Psi^T$. Also, note that the sieve approximation should satisfy the restriction that $\psi_{k_J,t}(0) = 0$ because $\psi^0_{J,t}(0) = 0$ by definition.

Second, with the conditional mean restriction in (5), we have the freedom to choose the basis functions of Z_{jt} as instruments. Suppose that we use a vector of functions $I^{\varsigma_J}(Z_{jt}) := (I_1(Z_{jt})', \ldots, I_{\varsigma_J}(Z_{jt})')'$. Then we can define the sample moments as

$$\bar{g}_t(\theta_{k_J,t}) := \frac{1}{J}\sum_{j=1}^{J} g_{jt}(\theta_{k_J,t}), \tag{11}$$

$$g_{jt}(\theta_{k_J,t}) := \left\{\log\left(s_{jt}/s_{0t}\right) - X'_{1,jt}\beta - \psi_{k_J,t}(X_{2,jt})\right\} \cdot I^{\varsigma_J}(Z_{jt}).$$

Note that the function $\psi_{k_J,t}(\cdot)$ and its associated sieve coefficients $(\vartheta_{1,t},\ldots,\vartheta_{k_J,t})$, are free to vary across markets (t).

Step 1

In the first step, we obtain the estimator of $\theta_J^0 := \left(\beta^0, \psi_{J,1}^0(\cdot), ..., \psi_{J,T}^0(\cdot)\right)$ as the minimizer of the following GMM criterion function in the space Θ_{k_J} :

$$\widehat{\theta}_J := \left(\hat{\beta}_J, \widehat{\psi}_{k_J,1}, \ldots, \widehat{\psi}_{k_J,T}\right) = \arg\min_{\theta_{k_J}\in\Theta_{k_J}} \hat{\mathcal{L}}(\theta_{k_J}), \tag{12}$$

where

$$\hat{\mathcal{L}}_J(\theta_{k_J}) := \sum_{t=1}^{T} \bar{g}_t(\theta_{k_J,t})'\hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1}\bar{g}_t\left(\theta_{k_J,t}\right),$$

$\hat{\Omega}_t(\tilde{\theta}_{J,t}) := \frac{1}{J}\sum_{j=1}^{J} g_{jt}(\tilde{\theta}_{J,t})g_{jt}\left(\tilde{\theta}_{J,t}\right)'$, $t = 1,\ldots,T$, and $\tilde{\theta}_t = (\tilde{\beta}_J, \tilde{\psi}_{k_J,t})$ is a preliminary GMM estimator with positive definite weighting matrices $W_1,\ldots,W_T$, that is,

$$\tilde{\theta}_J := (\tilde{\beta}_J, \tilde{\psi}_{k_J,1}\ldots,\tilde{\psi}_{k_J,T}) = \arg\min_{\theta_{k_J}\in\Theta_{k_J}} \sum_{t=1}^{T} \bar{g}_t(\theta_{k_J,t})'W_t\bar{g}_t(\theta_{k_J,t}). \tag{13}$$

A particularly convenient weight matrix to use when obtaining the preliminary GMM estimator is $W_t^{2sls} = \left[J^{-1}\sum_{j=1}^{J}\left(I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})'\right)\right]^{-1}$ because this leads to the two-stage least square estimator of regressing $\log(s_{jt}/s_{0t})$ on $X_{1,jt}$, and $(\psi_1(X_{2,jt}), ..., \psi_{k_J}(X_{2,jt}))$ using the interactions of $I^{\varsigma_J}(Z_{jt})$ and the market dummies as instrumental variables. Again, note that this first step estimation effectively deals with a linear model and thus can be easily implemented in programs like STATA.

Note that after the first stage estimation, we can obtain the demand shock estimates $\hat{\xi}_{jt} = \log\left(\frac{s_{jt}}{s_{0t}}\right) - X_{1,jt}'\hat{\beta}_J - \hat{\psi}_{k_J,t}(X_{2,jt})$ and hence the mean utility estimates $\hat{\delta}_{jt} = X_{1,jt}'\hat{\beta}_J + \hat{\xi}_{jt}$, which will be used in the second stage estimation. Finally, we define $\widehat{\theta}_{J,t} = (\hat{\beta}_J, \hat{\psi}_{k_J,t})$ for later use.

Step 2

In the second step, we estimate $f^0(\cdot)$ nonparametrically. We first approximate $f(\cdot)$ by a sieve $f_{M_J}(\cdot)$ in a sieve space $\mathcal{F}_{M_J}$, and hence use the following approximation:

$$\int \frac{\exp(X'_{2,jt}v)}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X'_{2,kt}v)}f(v)dv \simeq \int \frac{\exp(X'_{2,jt}v)}{1+\sum_{k=1}^{J}\exp(\delta_{kt}+X'_{2,kt}v)}f_{M_J}(v)dv.$$

Then we can define a MD estimator as

$$\hat{f}_{MD} := \arg\min_{f_{M_J}\in\mathcal{F}_{M_J}} \sum_{t=1}^{T}\sum_{j=1}^{J}\left(\hat{\psi}_{k_J t}(X_{2,jt}) - \log\left[\frac{\int \frac{\exp(X'_{2,jt}v)}{1+\sum_{k=1}^{J}\exp(\hat{\delta}_{kt}+X'_{2,kt}v)}f_{M_J}(v)dv}{\int \frac{1}{1+\sum_{k=1}^{J}\exp(\hat{\delta}_{kt}+X'_{2,kt}v)}f_{M_J}(v)dv}\right]\right)^2, \quad (14)$$

where $\hat{\delta}_{jt}$ is obtained from Step 1. Alternatively, we could estimate f^0 by minimizing a GMM criterion. The GMM criterion has two variants in our case:

$$\hat{f}_{GMM1} := \arg\min_{f_{M_J}\in\mathcal{F}_{M_J}} \sum_{t=1}^{T} \bar{g}_t(\hat{\beta}_J, f_{M_J})'\hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1}\bar{g}_t(\hat{\beta}_J, f_{M_J}), \quad (15)$$

$$\hat{f}_{GMM2} := \arg\min_{f_{M_J}\in\mathcal{F}_{M_J}} \min_{\beta} \sum_{t=1}^{T} \bar{g}_t(\beta, f_{M_J})'\hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1}\bar{g}_t(\beta, f_{M_J}), \quad (16)$$

where

$$\bar{g}_t(\beta, f_{M_J}) = \frac{1}{J}\sum_{j=1}^{J}\left\{\log\left(\frac{s_{jt}}{s_{0t}}\right) - X'_{1,jt}\beta - \log\left[\frac{\int \frac{\exp(X'_{2,jt}v)}{1+\sum_{k=1}^{J}\exp(\hat{\delta}_{kt}+X'_{2,kt}v)}f_{M_J}(v)dv}{\int \frac{1}{1+\sum_{k=1}^{J}\exp(\hat{\delta}_{kt}+X'_{2,kt}v)}f_{M_J}(v)dv}\right]\right\}\cdot I^{\varsigma_J}(Z_{jt}).$$

The only difference between the two GMM estimators of f^0 is that (15) takes the first step estimate $\hat{\beta}_J$ as given while (16) treats β as a nuisance parameter. GMM2 is somewhat more difficult to compute, but we find in Monte Carlos that it yields better performing estimates for the mean and standard deviation of the random coefficient distribution.

We would also like to point out that the second step estimation of $f^0(\cdot)$ can be parametric as well: we just need to replace the sieve approximation $f_{M_J}(\cdot)$ with a certain parametrization

$f(\cdot|\lambda)$ (as in (7)) when implementing our estimator (14). In the Monte Carlo section, we examine this estimation strategy and find that it achieves a very similar performance as the standard BLP parametric estimator when $f(\cdot|\lambda)$ is correctly specified.

3.3 Remarks on the Large J Asymptotic Framework

Before we move on to the formal asymptotic justification for our model in the next section, a few remarks regarding the asymptotic framework are in order.

As in Berry et al. (2004), we study the limiting behavior of our estimator as the number of products $J \to \infty$, while keeping T fixed. In our case, the reasons for considering the large J environment are twofold. From the applied perspective, large markets are common for many empirical scenarios, e.g., national auto market (Berry et al. (1995)), PC market (Goeree (2008), Bajari and Benkard (2005)), housing market (Bayer et al. (2007)), online marketplace (Quan and Williams (2018)), and scanner data with products defined at UPC level. From the theoretical perspective, the key to the linear structure of the first step of our estimation procedure is to use the variation within each market for identification as we treat $\psi^0_{J,t}(\cdot)$ as a parameter.

Despite being empirically relevant, the large J asymptotic framework is controversial in the literature. The main fear is that as $J \to \infty$, the model could approach a limiting model which has no meaningful cross-product substitution and hence fails to capture the oligopoly competition that is central to industrial organization questions.[10] However, our asymptotic framework does not require the model to converge to a limiting model. In particular, the inclusive value does not need to converge to a deterministic limit. Moreover, even when there is a limiting model, cross-product substitutions need not vanish as we demonstrate in an example in Appendix S3. Lastly, even if a limiting model exists in which cross-product substitutions vanish, this limiting feature is not used in our identification arguments. The asymptotic framework is only used as a mathematical tool to justify finite J approximations

[10] Armstrong (2016) quantifies one aspect of this concern by showing that the markup approaches a constant as $J \to \infty$ under a specific set of assumptions.

to certain sample statistics. Our Monte Carlo results suggest that even a small J like 5 or 10 is large enough for the estimator to perform well, while $J \geq 5$ is very common in practice and is far from large enough to make cross-product substitution disappear.

4 Asymptotic Theory

In this section, we prove the consistency and asymptotic normality of $\hat{\beta}_J$ and the consistency of the random coefficients distribution estimator. To begin, we introduce some notation as follows.

For column vector a, let $\|a\|$ be the Euclidean norm. For matrix A, let $||A\| = (tr(A'A))^{1/2}$. For $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, for any vector $a = (a_1, \ldots, a_d)'$ of d integers the differential operator is defined as

$$\bigtriangledown^a = \frac{\partial^{a.}}{\partial x_1^{a_1} \ldots \partial x_d^{a_d}},$$

where $a. = \sum_{i=1}^d a_i$. For a constant $\delta_0 \in \mathbb{R}$ and an integer m_0, define the weighted Sobolev L_2 norm on $\mathcal{C}^{m_0}(\mathcal{X})$, the space of m_0-times differentiable functions mapping $\mathcal{X}$ into $\mathbb{R}$, as

$$\|g\|_{m_0,2,\delta_0} := \left(\sum_{0 \leq a. \leq m_0} \int_{\mathcal{X}} |\bigtriangledown^{a.} g(x)|^2 (1 + x'x)^{\delta_0} dx \right)^{1/2},$$

and let the corresponding weighted Sobolev space of $\psi_t(\cdot)$ associated with these norms be

$$\mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2) := \{\psi \in \mathcal{C}^{m_0}(\mathcal{X}_2) \ s.t. \ \|\psi\|_{m_0,2,\delta_0} < \infty\} \tag{17}$$

for $\mathcal{X}_2 \subseteq \mathbb{R}^{d_{x2}}$. Let

$$\Psi = \{\psi \in \mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2) : \|\psi\|_{m_0,2,\delta_0} \leq B\} \tag{18}$$

for a finite constant $B > 0$. Let the square of the "strong norm" on the space of the parameter $\theta := (\beta, \psi_1, \ldots, \psi_T) \in \mathcal{B} \times \Psi^T$ be

$$\|\theta\|_s^2 = T\|\beta\|^2 + \sum_{t=1}^{T} \|\psi_t\|_{0,2,\delta_0}^2 = T\|\beta\|^2 + \sum_{t=1}^{T} \int_{\mathcal{X}_2} \psi_t(x_2)^2 (1 + x_2' x_2)^{\delta_0} dx_2. \tag{19}$$

4.1 First-Step Estimation: Consistency and Asymptotic Normality

In this subsection, we prove the consistency and asymptotic normality of the estimator defined in (12) as $J \to \infty$ and T stays fixed. We treat $\{(X_{jt}, Z_{jt}, \xi_{jt})\}_{j=1}^J$ as a triangular array as the number of products J increases to infinity, which is natural because the marginal distribution of $(X_{jt},\ Z_{jt},\ \xi_{jt})$ may change and the dependence across j may also change as J increases due to equilibrium firm response to changing market structure as the number of products increases. We introduce the following assumption on the triangular array.

Assumption 1. (i) *The unobserved product characteristics ξ_{jt} are conditionally independent across j given $\{Z_{jt}\}_{j=1}^J$ and satisfy $E[\xi_{jt}|\{Z_{jt}\}_{j=1}^J] = 0$ a.s. for each j and t;*

(ii) *the random variables $\{X_{jt}, Z_{jt}, \xi_{jt}\}_{j=1}^J$ are independent across t.*

Remark. Part (i) is identical to the assumptions imposed on the unobserved characteristic ξ as in Berry et al. (2004). Part (ii) assumes independence across market t but does not require X_{jt} and Z_{jt} to be independent across products. The independence across t is not essential but allows us to write down a relatively simple formula for the asymptotic variance of $\hat{\beta}_J$ below.

We also impose the following assumption that regulates the parameter space for the first-stage estimation of β, as well as the moments of relevant variables. Let $\mathcal{Z}_{jt}$ be the support of Z_{jt}. Let $\mathcal{Z}_t$ be the support of $\{Z_{jt}\}_{j=1}^J$.

Assumption 2. (i)*For all t and J, $\psi_{J,t}^0(\cdot) \in \Psi$ almost surely;*

(ii) *$\mathcal{B}$ is compact with $\beta^0 \in int(\mathcal{B})$;*

(iii) *for m_0 in (18):* (a) *$\mathcal{X}_2$ has the uniform cone property defined in Section 4.4 of Adams (1975); and* (b) *for $\mathcal{X}_2$ bounded, $\delta_0 = 0$ and $m_0 > d_{x_2}/2$, while for $\mathcal{X}_2$ unbounded, $m_0 > d_{x_2}/2$*

and $\delta_0 < -1/2$*;*

(iv) *For some constants* $C, c > 0$*,* $\sup_{j,t} E\left[\xi_{jt}^4\right] < C$*,* $\sup_{j,t}\sup_{\mathbf{z}\in\mathcal{Z}_t} E[\xi_{jt}^2|\{Z_{jt}\}_{j=1}^J = \mathbf{z}] < C$*, and* $\inf_{j,t}\inf_{\mathbf{z}\in\mathcal{Z}_t} E[\xi_{jt}^2|\{Z_{jt}\}_{j=1}^J = \mathbf{z}] > c$*;*

(v) $\sup_{j,t}\sup_{z\in\mathcal{Z}_{jt}} E\left[\|X_{1,jt}\|^2|\, Z_{jt} = z\right] < \infty$*; and*

(vi) $\sup_{j,t}\sup_{x_2\in\mathcal{X}_2} |f_{X_{2,jt}}(x_2)/(1 + x_2' x_2)^{\delta_0}| < \infty$*.*

Remark. Part (i) is a standard assumption in semi/non-parametric estimation, except that we need the modifier "almost surely" because our functional parameter is random. Part (ii) and (iii) are needed to show that the parameter space $\Theta = \mathcal{B} \times \Psi^T$ is $\|\cdot\|_s$-compact. When $\mathcal{X}_2$ is bounded, Theorem 1 in Freyberger and Masten (2017) implies that $\mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2)$ is compactly embedded in the space $\mathscr{W}_{0,\delta_0}(\mathcal{X}_2)$, and thus Ψ is relatively $\|\cdot\|_{0,2,\delta_0}$−compact. Moreover, Theorem 2 of Freyberger and Masten (2017) implies that $\mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2)$ is $\|\cdot\|_{0,2,\delta_0}$−closed. Thus, Ψ is $\|\cdot\|_{0,2,\delta_0}$−compact and consequently, Θ is compact under $\|\cdot\|_s$ because $\mathcal{B}$ is compact. When $\mathcal{X}_2$ is unbounded, Theorem 3 of Freyberger and Masten (2017) implies that $\mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2)$ is compactly embedded in the space $\mathscr{W}_{0,\delta_0}(\mathcal{X}_2)$. Moreover, by Lemma 3 of Freyberger and Masten (2017), which reflects Lemma A.1 in Santos (2012), $\mathscr{W}_{m_0,\delta_0}$ is $\|\cdot\|_{0,2,\delta_0}$−closed, and therefore also compact under the norm $\|\cdot\|_{0,2,\delta_0}$, so Θ is compact under $\|\cdot\|_s$. The bound on δ_0 in the unbounded case makes sure that the constant function belongs to Ψ. Part (iv) is a standard finite moment condition on the error term. Part (v) imposes mild restrictions on the moments of the covariates. This part together with parts (i), (ii), and (iv) implies that $\log(s_j/s_0)$ has a bounded second moment.[11] And Part (vi) gives a relationship between the density of x_2 and the weight $(1 + x_2' x_2)^{\delta_0}$. This is a weak assumption and allows rather general tail behavior for X_2.

By allowing the instrumental function vector $I^{\varsigma_J}(Z_{jt})$ to grow in dimension as $J \to \infty$, we are adopting the approach of Donald et al. (2003) to use a growing number of unconditional

[11] Note that this implies that the inside shares and the outside shares go to zero at the same rate as J increases, which Berry et al. (2004) impose directly as an assumption. Our Monte Carlo design satisfies this assumption, as illustrated in the histograms in Figure 3 in Section S6 in the Supplemental Appendix. In the empirical application reported in Section 6, $|\log(s_{jt}/s_{0t})|$ has a maximum value of 14 with a standard deviation of 1.38, and in the empirical application reported in Supplemental Appendix S5, it has a maximum value of 17.5 with a standard deviation of 3.19.

moments to approximate the conditional moments. The following assumption regulates how fast the dimension of the sieve space may grow. Let $eig_{\min}$ and $eig_{\max}$ denote the minimum and the maximum eigenvalues respectively.

Assumption 3. (i) *For each ς_J (implicitly dependent on t), there is a constant ζ_z and matrix S_z such that for $\tilde{I}^{\varsigma_J}(z) = S_z I^{\varsigma_J}(z)$, $\sup_{z\in\mathcal{Z}_{jt}} \|\tilde{I}^{\varsigma_J}(z)\| \lesssim \zeta_z$ for all j, t, $eig_{\max}(\tilde{I}^{\varsigma_J}(z)\tilde{I}^{\varsigma_J}(z)') \leq \bar{\lambda}$, and $eig_{\min}(\tilde{I}^{\varsigma_J}(z)\tilde{I}^{\varsigma_J}(z)') > 0$ for some constant $\bar{\lambda}$;*

(ii) *for some $c > 0$, $\zeta_z^{(2+c)/c}\sqrt{\log \varsigma_J / J} = o(1)$, $\zeta_z \lesssim \sqrt{\varsigma_J}$; and $\varsigma_J/\sqrt{J} \to 0$.*

Remark. Assumption 3 contains mild restrictions on the basis functions, which is commonly imposed for series estimators. Under Assumption 3, it is without loss of generality to assume that $E\left[I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})'\right]$ is a $\varsigma_J \times \varsigma_J$ identity matrix, which we do for the rest of the discussion.

We define the following notation to facilitate our discussion:

$$\rho_{jt}(\theta_t|\theta_t^*) := \xi_{jt} + X_{1,jt}'(\beta^* - \beta) + (\psi_t^*(X_{2,jt}) - \psi_t(X_{2,jt})) \tag{20}$$

for any θ_t, $\theta_t^* \in \Theta_t := \mathcal{B} \times \Psi$. Let $\theta_J^0 = \left(\beta^0, \psi_{J,1}^0, \ldots, \psi_{J,T}^0\right)$ and let $\theta_{J,t}^0 = (\beta^0, \psi_{J,t}^0)$.

The standard consistency proof requires the uniform convergence of the criterion function $\hat{\mathcal{L}}_J(\theta)$ defined in (12), which is cumbersome to verify due to the estimated weight matrix $\hat{\Omega}_t(\tilde{\theta}_{J,t})$. Instead, to show the consistency of $\widehat{\theta}_J = \arg\min_{\theta_{k_J}\in\Theta_{k_J}} \hat{\mathcal{L}}_J(\theta_{k_J})$, we use a similar argument as Lemma A1 of Newey and Powell (2003), which only requires us to verify the uniform convergence of a simpler criterion function

$$\tilde{\mathcal{L}}_J(\theta) := \sum_{t=1}^{T} \bar{g}_t(\theta_t)' W_t^{2sls} \bar{g}_t(\theta_t), \tag{21}$$

to its population counterpart $\bar{\mathcal{L}}_J(\theta|\theta_J^0)$, where for any (deterministic) pair $\theta, \theta^* \in \Theta$,

$$\bar{\mathcal{L}}_J(\theta|\theta^*) = \frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} E\left[\left(E\left[\rho_{jt}(\theta_t|\theta_t^*)|Z_{jt}\right]\right)^2\right]. \tag{22}$$

We then impose the following conditions which we verify in Supplemental Appendix S2.

Assumption 4. (i) *For the sieve approximation of* $\theta^0_{J,t}$, $\theta^0_{J,k_J,t} := \arg\min_{\theta\in\Theta_{k_J,t}=\mathcal{B}\times\Psi_{k_J}} \|\theta - \theta^0_{J,t}\|_s$, *we have* $\|\bar{g}_t(\theta^0_{J,k_J,t})\| = o_p(1)$;

(ii) *for some generic positive constant* C, $1/C \le eig_{\min}(\hat{\Omega}_t(\tilde{\theta}_{J,t})) \le eig_{\max}(\hat{\Omega}_t(\tilde{\theta}_{J,t})) \le C$ *and* $1/C \le eig_{\min}\left(W_t^{2sls}\right) \le eig_{\max}\left(W_t^{2sls}\right) \le C$ *w.p.a.1. for each* $t = 1, \ldots, T$.

Remark. Part (i) defines the sieve approximation $\theta^0_{J,k_J,t}$, of $\theta^0_{J,t}$ and posits that the approximation error vanishes as $J \to \infty$ so that the sample moment evaluated at $\theta^0_{J,k_J,t}$ converges to zero, as if it is evaluated at $\theta^0_{J,t}$. Part (ii) requires the eigenvalues of the estimated variance-covariance matrix $\hat{\Omega}_t(\tilde{\theta}_{J,t})$ with preliminary estimator $\tilde{\theta}_{J,t}$ and W_t^{2sls} to be bounded. Both parts are verified in Supplemental Appendix S2.2.

Assumption 5. (i) *For any* $\varepsilon > 0$, *we have* $\inf_{J=1,\ldots,\infty} \inf_{\theta,\theta^*\in\Theta:\|\theta-\theta^*\|_s>\varepsilon} \bar{\mathcal{L}}_J(\theta|\theta^*) > 0$;

(ii) $\sup_{\theta_{k_J}\in\Theta_{k_J}} |\tilde{\mathcal{L}}_J(\theta_{k_J}) - \bar{\mathcal{L}}_J(\theta_{k_J}|\theta^0_J)| \xrightarrow{p} 0$ as $J \to \infty$.

Remark. Part (i) is a global identification condition. In this assumption lies our first key modification to the standard sieve estimation theory to accommodate the random parameter $\psi^0_{Jt}(\cdot)$: instead of focusing on the identification at one deterministic point in the parameter space Θ, we require the identification to hold uniformly over the entire Θ. This accommodates the different values that $\psi^0_{Jt}(\cdot)$ may take in different states of the world. Part (i) is verified in Section S2.1 in the Supplemental Appendix. Part (ii) is a uniform convergence condition on the sieve space. Note that the "population" criterion function $\bar{\mathcal{L}}_J(\cdot|\theta^0_J)$ is a random function due to the randomness in θ^0_J, which is our second key modification to accommodate the random parameter $\psi^0_{Jt}(\cdot)$. Part (ii) is verified in Section S2.2 in the Supplemental Appendix.

Theorem 1 (Consistency)**.** *Suppose that Assumptions* 1-5 *hold and that the preliminary estimator* $\tilde{\theta}_{J,t}$ *satisfies* $\|\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \frac{1}{J}\sum_{j=1}^J \xi^2_{jt} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})'\| = o_p(1)$ *for all* t. *Then for the estimator* $\widehat{\theta}_J$ *defined in* (12) , *we have*

$$\|\widehat{\theta}_J - \theta_J^0\|_s = o_p(1).$$

A well-known fact in the semi-parametric literature is that consistency of the non-parametric part alone is not enough to establish the asymptotic normality of $\hat{\beta}_J$, which typically needs that $\widehat{\theta}_J$ converges to θ_J^0 at a rate faster than $J^{-1/4}$ under a weaker norm. To obtain the convergence rate of $\widehat{\theta}_J$ and derive the limiting distribution of $\hat{\beta}_J$, we introduce the following notation and some high-level conditions:

Let $\bar{\mathcal{W}}$ be the closure of the linear span of Ψ. Note that by Assumption 2(i) and $0 \in \Psi$, the closure of the linear span of $\Psi - \psi_{J,t}^0 := \left\{\psi - \psi_{J,t}^0 : \psi \in \Psi\right\}$ is equal to $\bar{\mathcal{W}}$ almost surely. Let

$$\Sigma_{j,J,o} = E[\xi_{jt}^2 | \{Z_{jt}\}_{j=1}^J], \tag{23}$$

and $\omega_{J,t}^\dagger = \left(\omega_{J,t,1}^\dagger, \ldots, \omega_{J,t,\ell}^\dagger, \ldots, \omega_{J,t,d_{x_1}}^\dagger\right)$ be the solution to

$$\min_{\omega_{J,t,\ell} \in \bar{\mathcal{W}}} \frac{1}{J} \sum_{j=1}^J E\left[E\left[X_{1\ell,jt} - \omega_{J,t,\ell}\left(X_{2,jt}\right) | Z_{jt}\right]^2 / \Sigma_{j,J,o}\right]$$

for $\ell = 1, \ldots, d_{x_1}$. Let $D_{j,J,t}(Z_{jt}) = E\left[X_{1,jt} - \omega_{J,t}^\dagger(X_{2,jt}) | Z_{jt}\right]$,

$$V_{J,t}^{-1} = \frac{1}{J} \sum_{j=1}^J E\left[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})' / \Sigma_{j,J,o}\right], \tag{24}$$

$V_{J,\beta} = (\sum_{t=1}^T V_{J,t}^{-1})^{-1}$, and $I_{d_{x_1}}$ be a $d_{x_1} \times d_{x_1}$ identity matrix. Let $\hat{\omega}_{J,t}^\dagger = (\hat{\omega}_{J,t,1}^\dagger, \ldots, \hat{\omega}_{J,t,d_{x_1}}^\dagger)$ be an estimator of $\omega_{J,t}^\dagger$ such that

$$\hat{\omega}_{J,t}^\dagger := \min_{\omega_{J,t} \in \Psi_{k_J}} \left(\frac{1}{J} \sum_{j=1}^J \left(X_{1,jt} - \omega_{J,t}(X_{2,jt})\right) I^{\varsigma_J}(Z_{jt})\right)' \hat{\Omega}_{J,t}(\tilde{\theta}_{J,t})^{-1} \left(\frac{1}{J} \sum_{j=1}^J \left(X_{1,jt} - \omega_{J,t}(X_{2,jt})\right) I^{\varsigma_J}(Z_{jt})\right),$$

and $\hat{V}_{J,\beta} = (\sum_{t=1}^{T} \hat{V}_{J,t}^{-1})^{-1}$ with

$$\hat{V}_{J,t}^{-1} = \left(\frac{1}{J} \sum_{j=1}^{J} \left(X_{1,jt} - \hat{\omega}_{J,t}^{\dagger}(X_{2,jt}) \right) I^{\varsigma_J}(Z_{jt}) \right)' \hat{\Omega}_{J,t}(\widehat{\theta}_t)^{-1} \left(\frac{1}{J} \sum_{j=1}^{J} \left(X_{1,jt} - \hat{\omega}_{J,t}^{\dagger}(X_{2,jt}) \right) I^{\varsigma_J}(Z_{jt}) \right),$$

where $\hat{\Omega}_{J,t}(\widehat{\theta}_t) = \frac{1}{J} \sum_{j=1}^{J} g_{jt}(\widehat{\theta}_t) g_{jt}(\widehat{\theta}_t)'$ for $t = 1, \ldots, T$.

Assumption 6. (i) For any sequence $\lambda_J \in \mathbb{R}^{d_{x_1}}$ such that $\sup_J \|\lambda_J\| < \infty$,we have

$$\sqrt{J} \lambda_J' \left(\hat{\beta}_J - \beta^0 \right) = -\lambda_J' V_{J,\beta} \sum_{t=1}^{T} \left\{ \left(\frac{1}{\sqrt{J}} \sum_{j=1}^{J} D_{j,J,t}(Z_{jt}) \Sigma_{j,J,o}^{-1} \xi_{jt} \right) \right\} + o_p(1);$$

(ii) $\hat{V}_{J,\beta} - V_{J,\beta} \xrightarrow{p} 0$.

Remark. Part (i) assumes a local linearization of the estimator for β^0 and Part (ii) assumes that $\hat{V}_{J,\beta}$ is a consistent estimator of $V_{J,\beta}$. We verify Assumption 6 in Supplemental Appendix S2.2.

Theorem 2. *Suppose that Assumptions 1-6 hold. Suppose that for each t, we have both $eig_{\min}(\frac{1}{J} \sum_{j=1}^{J} E[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})'])^{-1} = O(1)$ and $\sup_{j,J,t} \sup_{z \in \mathcal{Z}_{jt}} \|D_{j,J,t}(z)\| = O(1)$. Then*

$$\sqrt{J} \hat{V}_{J,\beta}^{-1/2} \left(\hat{\beta}_J - \beta^0 \right) \xrightarrow{d} N\left(0, I_{d_{x_1}}\right),$$

where $I_{d_{x_1}}$ is a $d_{x_1} \times d_{x_1}$ identity matrix.

Remark. Theorem 2 provides the asymptotic distribution of $\hat{\beta}_J$ when the number of products J goes to infinity. One can use the above result to construct a $100\tau\%$ valid confidence region for β^0 by $\left[\hat{\beta}_J \pm c_{1-\tau} \sqrt{\hat{V}_{J,\beta}/J} \right]$, where $c_{1-\tau}$ is the $100\tau\%$ critical value from a standard normal distribution.

4.2 Second-Stage Estimation: Consistency

In this section, we show that the MD estimator and the GMM2 estimator proposed in (14) are consistent estimators of f^0. The GMM1 estimator proposed therein is similar and thus omitted. We make heavy use of the norm introduced in Section 4.1. We first introduce the definition of the density space as follows. Let the parameter space $\mathcal{F}$ of f^0 be the space of probability density functions such that $\|f\|_{\alpha_0,2,\mu_0} < B_0$ for some integer $\alpha_0 > d_{x_2}/2$, some bound B_0, and some $\mu_0 \in \mathbb{R}$, and let $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_{M_J} \cdots \subseteq \mathcal{F}$ be a sequence of sieve approximations of $\mathcal{F}$. Define a sup-norm on $\mathcal{F}$ as $\|f\|_\infty = \max_{0\leq a_\cdot \leq \alpha} \sup_{v\in\mathcal{V}} |\nabla^{a_\cdot} f(v)| (1 + v'v)^{\mu/2}$, for a $\mu < \mu_0$.

Let $\varphi^0_{J,t}(v) = \frac{1}{J}\left(1 + \sum_{j=1}^J \exp\left(\delta_{jt} + X'_{2,jt}v\right)\right)$ and let $g_{J,t}(x_2, v) = \frac{\exp(x'_2 v)}{\varphi^0_{J,t}(v)}$. Similarly, let $\hat{\varphi}_{J,t}(v) = \frac{1}{J}\left(1 + \sum_{j=1}^J \exp\left(\hat{\delta}_{jt} + X'_{2,jt}v\right)\right)$ and $\hat{g}_{J,t}(x_2, v) = \frac{\exp(x'_2 v)}{\hat{\varphi}_{J,t}(v)}$. Let

$$G_{J,t}(x_2; f) = \int g_{J,t}(x_2, v) f(v) dv \text{ and } \widehat{G}_{J,t}(x_2; f) = \int \hat{g}_{J,t}(x_2, v) f(v) dv. \tag{25}$$

Define the following population criterion function:

$$\mathcal{Q}_J(f) = J^{-1} \sum_{j=1}^J \sum_{t=1}^T E_X \left[\left\{\psi^0_{J,t}(X_{2,jt}) - \log\left(G_{J,t}(X_{2,jt}; f)/G_{J,t}(\mathbf{0}; f)\right)\right\}^2\right], \tag{26}$$

where $E_X h_J(X_{2,jt}) = \int h_J(x_2) dF_{X_{2,jt}}(x_2)$ for any possibly random function h_J. We make the following assumption to ensure the point identification of f^0.

Assumption 7. (i) *The support of $X_{2,jt}$, denoted $\mathcal{X}_{2,t}$, contains a bounded open $\mathbb{R}^{d_{x_2}}$-ball for some t; denote this open ball by $\mathcal{B}_0$.*

(ii) $\sup_{x_2\in\mathcal{B}_0} \int \exp(x'_2 v) f^0(v) dv < \infty$.

Remark. Part (i) rules out discrete variables in $X_{2,jt}$. This is expected because it is not possible to nonparametrically identify the distribution of a continuous random coefficient based on discrete variation of the covariate. Part (ii) of the assumption requires that the random coefficient distribution has an exponential or subexponential tail. This is satisfied by Gaussian distributions, for example.

Lemma 1. *Suppose that Assumptions 1 and 7 hold. Then*

(a) *for any distribution* $f \in \mathcal{F}$ *such that* $f \neq f^0$*, we have that for some* $x_2 \in \mathcal{X}_{2,t}$ *for the* t*'s such that* $\mathcal{X}_{2,t}$ *contains an open* $\mathbb{R}^{d_{x_2}}$*-ball,*

$$\log \int \frac{\exp(x_2' v)}{\varphi^0_{J,t}(v)} f(v)dv - \log \int \frac{1}{\varphi^0_{J,t}(v)} f(v)dv \neq \psi^0_{J,t}(x_2) \text{ almost surely, and}$$

(b) $\mathcal{Q}_J(f) > \mathcal{Q}_J(f^0)$ *for all* $f \in \mathcal{F}$ *such that* $f \neq f^0$*.*

Remark. Lemma 1 provides a simple argument for the nonparametric point identification of the random coefficients distribution in the growing J environment.

Define the intermediate and the sample criterion functions respectively as:

$$\tilde{\mathcal{Q}}_J(f) = \frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ \psi^0_{J,t}(X_{2,jt}) - \log\left(G_{J,t}(X_{2,jt}; f)/G_{J,t}(\mathbf{0}; f)\right) \right\}^2, \text{ and}$$
$$\hat{\mathcal{Q}}_J(f) = \frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ \hat{\psi}_{k_J,t}(X_{2,jt}) - \log(\widehat{G}_{J,t}(X_{2,jt}; f)/\widehat{G}_{J,t}(\mathbf{0}; f)) \right\}^2. \tag{27}$$

We introduce the following condition that is required for the consistency of the second-step nonparametric estimator.

Assumption 8. (i) For any $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that with probability approaching one, $\inf_{f \in \mathcal{F}: \|f - f^0\|_\infty > \varepsilon} \{\mathcal{Q}_J(f) - \mathcal{Q}_J(f^0)\} > \delta_\varepsilon$;

(ii) $\frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} \left[\hat{\psi}_{k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt}) \right]^2 \xrightarrow{p} 0$ *and* $\sup_{f \in \mathcal{F}_{M_J}} \left| \tilde{\mathcal{Q}}_J(f) - \mathcal{Q}_J(f) \right| \xrightarrow{p} 0$*;*

(iii) $\sup_{f \in \mathcal{F}_{M_J}} \mathcal{Q}_J(f) = O_p(1)$*;*

(iv) *for the sieve approximation of* f^0*:* $f^0_{M_J} = \arg\min_{f \in \mathcal{F}_{M_J}} \|f - f^0\|_\infty$*, we have that* $\|f^0_{M_J} - f^0\|_\infty \to 0$ *and* $\left| \mathcal{Q}_J(f^0_{M_J}) - \mathcal{Q}_J(f^0) \right| \xrightarrow{p} 0$ *as* $M_J \to \infty$ *with* $J \to \infty$*.*

Remark. Part (i) is a uniform identification condition, a weaker version of which is verified in Lemma 1. We need this uniform version because we are in a triangular array asymptotic framework. In this framework, the population criterion function drifts with J. Part (ii) can be verified using Theorem 1 combined with a uniform law of large number under suitable dependence assumption on $\{X_{2,jt}\}$ across j in each market t. This part is verified in Section

S2.3 in the Supplemental Appendix under the independence assumption, which can be a reasonable assumption when the firms determine their $X_{2,jt}$ based on independent private information. Common shocks can be accommodated if the expectations with respect to $X_{2,jt}$ are understood as conditional expectations given the common shocks. Part (iii) requires the population criterion function to be uniformly bounded on the sieve space; this is a weak assumption given that the space of f^0 is already assumed to be compact. Finally, Part (iv) assumes the convergence of sieve approximation $f^0_{M_J}$ to the true parameter f^0.

Theorem 3. *Suppose that Assumptions in Theorem 1 and Assumption 8 hold. Then for $\hat{f}_J = \hat{f}_{MD}$ defined in (14) and $\hat{f}_J = \hat{f}_{GMM2}$ defined in (16), we have*

$$\|\hat{f}_J - f^0\|_\infty \xrightarrow{p} 0. \tag{28}$$

Remark. Theorem 3 shows the consistency of the nonparametric estimators $\hat{f}_{MD}$ and $\hat{f}_{GMM2}$. Compared to Fox et al. (2016), we account for the estimation effect of ξ_{jt} from the first stage. We provide the proof of Theorem 3 in Appendix S1.4. The consistency of $\hat{f}_{GMM1}$ defined in (15) can be derived similarly to that of $\hat{f}_{GMM2}$.

After obtaining an estimator for the distribution of the random coefficients, other parameters of interest may also be recovered such as the CDF function, the substitution pattern among products such as cross-product elasticities, and many others. It is worth mentioning that the GMM estimator in Berry et al. (2004) could achieve semi-parametric efficiency when the model is correctly specified and the optimal weighting matrix is used in a limited information sense[12]. In contrast, our two-step estimator provides a computationally attractive alternative that could sacrifice efficiency. We will investigate the extent of the efficiency loss in the Monte Carlo simulations.

[12] See the discussion on pp. 633 of Berry et al. (2004).

5 Monte Carlo Simulations

In this section, we perform a series of Monte Carlo simulations to examine the performance of our semi-nonparametric estimator and compare it with several alternative estimation strategies. We first focus on the case of a single random coefficient in Section 5.1-5.3 and then extend to the setting with multiple random coefficients in Section 5.5. In all simulation exercises, the number of Monte Carlo repetitions is 1000.

5.1 Data Generating Process

We simulate T markets, each of which has J products. A product $j \in \{0, 1, ..., J\}$ in market $t \in \{1, ..., T\}$ is associated with an exogenous characteristic $X_{jt} \sim N(0, 1)$, an unobserved characteristic $\xi_{jt} \sim N(0, .3^2)$ and an endogenous price.[13] For simplicity, we assume that price equals marginal cost:

$$P_{jt} = mc_{jt} = 0.5X_{jt} + W_{jt} + \xi_{jt} + \zeta_{jt},$$

where mc_{jt} is the marginal cost of product j in market t that is a linear function of X_{jt}, ξ_{jt}, an exogenous observable cost shifter $W_{jt} \sim N(0, 1)$ and cost shock $\zeta_{jt} \sim N(0, .1^2)$. Note that price is endogenous in the sense that it depends on the demand shock ξ_{jt}. Also, we have introduced W_{jt} to make it available as an IV for price in the demand estimation, which provides a convenient way to handle the price endogeneity issue.

The demand specification follows Section 2 closely with a random coefficient on price, so the market share of product j in market t is

$$\sigma_j\left(\delta_t, P_t; F^0\right) = \int \frac{\exp\left(\delta_{jt} + \upsilon P_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + \upsilon P_{kt}\right)} dF^0\left(\upsilon\right),$$

[13] Unreported Monte Carlo simulation shows that increasing the variance of ξ_{jt} (and ζ_{jt} below) causes both the BLP and the SN estimators to have larger variances, but the comparison between estimators is qualitatively unchanged.

where $\delta_{jt} = \alpha^0 + \beta^0 X_{jt} + \xi_{jt}$ and $P_t = (P_{1t}, ..., P_{Jt})$. Given this formula, the market share data are simulated as

$$s_{jt} = \frac{1}{R} \sum_{i=1}^{R} \frac{\exp\left(\delta_{jt} + \upsilon_i P_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + \upsilon_i P_{kt}\right)},$$

where $R = 10,000$ is the number of random draws υ_i's from $F^0(\cdot)$. We consider two designs with distinct F^0's, one is the commonly used normal distribution (symmetric and uni-modal), i.e., $N(-2, .5^2)$ and the other one is an asymmetric, bi-modal mixed normal distribution, i.e., $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$.

Thus, a simulated data set can be written as $\{(s_{jt}, X_{jt}, P_{jt}, W_{jt}) : j = 1, ..., J; t = 1, ..., T\}$, on which we implement our proposed estimator as well as alternative estimation strategies.

5.2 Implementation Details

The implementation of our semi-nonparametric estimator follows closely the description in Section 3.2. In the first step, we estimate $\theta_J = (\alpha, \beta, \psi_{k_J,1}(\cdot), ..., \psi_{k_J,T}(\cdot))$ using the two-step GMM procedure defined by (12) with the 2SLS estimator used as the initial estimator. Here, the market-specific sieve approximation $\psi_{k_J,t}(P_{jt})$ (for any market t) is specified as k_J-order power series, and $I^{\varsigma_J}(W_{jt})$ is defined as a cubic spline (with ς_J knots).[14] We also tried alternative combinations, say power series for both $\psi_{k_J,t}(P_{jt})$ and $I^{\varsigma_J}(W_{jt})$, and they yield virtually identical results (not shown here).

In the second step, we implement the MD and GMM estimators defined by (14), (15), and (16), coupled with three alternative sieve approximations to f^0 (or F^0):

1. The first sieve approximates the inverse CDF and directly generates random draws from $F^0_{M_J}$: As suggested by Fosgerau and Mabit (2013), we draw $u \sim U[0,1]$ and use a power series (with a set of coefficients to be estimated) to transform u to a new random variable υ_{M_J}, i.e.,

$$\upsilon_{M_J} = \sum_{k=1}^{M_J} b_k u^k.$$

[14] We imposed the restriction $\psi_{k_J,t}(0) = 0$ by setting the constant term in the power series to zero.

Note that υ_{M_J} is a draw from an approximate distribution $F^0_{M_J}$ because the above polynomial can be regarded as a sieve approximation to the inverse of F^0.

2. The second sieve approximates F^0 by a discrete distribution that is characterized by a set of grid points $\mathcal{G}_J$ and probability weight on each point $\upsilon_m \in \mathcal{G}_J$ modeled by a logit formula

$$\frac{\exp\left[\sum_{k=1}^{M_J} b_k \upsilon_m^k\right]}{\sum_{\upsilon_l \in \mathcal{G}_J} \exp\left[\sum_{k=1}^{M_J} b_k \upsilon_l^k\right]}.$$

This approximation is proposed in Train (2016) and is computationally attractive thanks to the smoothness of the logit formula.

3. The third sieve approximation adopts Fosgerau and Bierlaire (2007)'s approach. First of all, we can rewrite F^0 as

$$F^0(\upsilon) = Q(\Phi(\upsilon; \mu, \sigma)),$$

where Q is an unknown CDF function from $[0,1]$ to $[0,1]$ and the normal CDF $\Phi(\cdot; \mu, \sigma)$ is chosen as a base distribution. Then we can differentiate the above expression to obtain

$$f^0(\upsilon) = q(\Phi(\upsilon; \mu, \sigma))\,\phi(\upsilon; \mu, \sigma),$$

where $f^0(\cdot)$ and $q(\cdot)$ are density functions. Next, following Bierens (2008), the unknown density function $q(\cdot)$ is approximated by

$$q_k(x) = \frac{\left[1 + \sum_{i=1}^{k} b_i L_i(x)\right]^2}{1 + \sum_{i=1}^{k} b_i^2},$$

where L_i's are transformed Legendre polynomials and b_i's are the sieve coefficients. Finally, with a few manipulations, the market share of product j in market t can be written as

$$\tilde{\sigma}_j\left(\delta_t, P_t; F_{M_J}\right) = \int_0^1 \frac{\exp\left(\delta_{jt} + \Phi^{-1}\left(z;\mu,\sigma\right) P_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + \Phi^{-1}\left(z;\mu,\sigma\right) P_{kt}\right)} q_{M_J}\left(z\right) dz.$$

And this integral is approximated by simulation with $z \sim U\left[0,1\right]$.

For comparison purposes, we also implement a parametric version of our second step estimation as well as the standard BLP estimator (7), and both assume that F^0 is normal. For the BLP estimator, we use the same choice of instrumental variables as our semi-nonparametric estimator (12); the demand inversion is computed using the standard BLP contraction mapping in the estimation procedure.

5.3 Baseline Results

We shall present the simulation results with a series of tables. In these tables, we label our semi-nonparametric estimator as "SN" and the standard parametric BLP estimator as "BLP". Also, the three alternative sieve approximations to F^0 described in the previous subsection are labeled as "I", "II", and "III", respectively; the parametric version of our second step estimation is labeled as "Para". Also, we consider the minimum distance (MD) estimator defined in (14) and two alternative GMM implementations for the second step: "GMM1" refers to the estimator defined in (15); "GMM2" refers to the estimator defined in (16). Hence, a particular specification, say our semi-nonparametric estimator with sieve approximation "I" and criterion "GMM1", is denoted by "SN-I-GMM1". Also, to save space, we only show second-step results from our preferred specification "SN-III-GMM2" in the main text and delegate the full set of results to Appendix S4.

To examine the performance of our estimator with varying sample sizes, we consider different J's and T's. In the case of the BLP estimator and the SN-Para estimator, we approximate the integral by simulating from the (assumed) normal random coefficient distribution.

The number of simulation draws $\tilde{R}$ is related to J and T in the way described in Table 1[15] – the table that also describes our choices of the number of sieve terms in both steps. In the case of SN-I and SN-III estimators, we simulate the integral according to the description in Section 5.2 also using $\tilde{R}$ in Table 1. In the case of SN-II, we approximate the integral according to the description in Section 5.2 using 100 grid points.

Table 1: Tuning Parameter Choices Across Specifications

	$T = 10$			20			40		
J	25	50	100	25	50	100	25	50	100
$\tilde{R} = \frac{J^2 T}{40}$	156	625	2,500	313	1250	5,000	625	2,500	10,000
k_J and ς_J	3	4	5	3	4	5	3	4	5
M_J	3	4	5	4	5	6	5	6	7

5.3.1 Design I: F^0 is Normal

In the first set of experiments, we let the true distribution of random coefficient F^0 be normal (i.e., $N(-2, .5)$), which is the most commonly used distribution in the empirical application.

Table 2 shows the Monte Carlo results for the estimation of the fixed coefficients in the model, i.e, the coefficient on the exogenous characteristic X and the constant term. The BLP estimator shown in this table is for the benchmark case with a correctly specified F^0, i.e., a normal distribution with mean and variance as parameters to be estimated.

From the table, we can see that the RtMSE of our semi-nonparametric estimator is rather close to (though slightly larger than) the benchmark BLP estimator, which means that we do not lose much efficiency (in terms of estimating the fixed coefficients) by relaxing parametric assumptions on F. And we would like to emphasize again that the SN estimators of α and β are obtained from estimating a linear model, which is very easy to compute.

In Table 3, we report the average (across repetitions) GMM standard errors (labeled "Ave. S.E.") of our SN estimator and compare them with the actual standard deviations (labeled "True S.D."). Overall we can see that the Ave. S.E. is smaller than its corresponding

[15] This choice of $\tilde{R}$ follows the theoretical results of Berry et al. (2004).

Table 2: Monte Carlo Results: Fixed Coefficients

			$T=10$			20			40		
Parameter	Estimator	J	25	50	100	25	50	100	25	50	100
β	SN	RtMSE	.0249	.0169	.0116	.0176	.0119	.0080	.0125	.0085	.0059
		Bias	-.0026	4.08E-4	.0013	-.0020	5.34E-4	.0018	-.0014	.0012	.0014
	BLP	RtMSE	.0218	.0153	.0104	.0174	.0116	.0076	.0144	.0090	.0061
		Bias	-.0106	-.0062	-.0037	-.0105	-.0064	-.0034	-.0107	-.0061	-.0036
α	SN	RtMSE	.0318	.0285	.0234	.0222	.0215	.0166	.0164	.0155	.0120
		Bias	-9.55E-4	-9.70E-5	-2.27E-4	3.95E-4	-7.23E-4	-.0011	3.99E-5	-6.96E-4	2.05E-5
	BLP	RtMSE	.0228	.0172	.0131	.0165	.0118	.0090	.0117	.0088	.0061
		Bias	-.0024	-.0014	-.0048	-7.67E-4	-.0018	-.0026	-.0022	-.0017	-.0018

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, F^0 is $N(-2,.5)$.
2. The distribution of random coefficient for the BLP estimator is correctly specified.

Table 3: Monte Carlo Results: Inference on Fixed Coefficients

		$T=10$			20			40		
Parameter	J	25	50	100	25	50	100	25	50	100
β	True S.D.	.0248	.0169	.0115	.0175	.0119	.0078	.0125	.0084	.0057
	Ave. S.E.	.0191	.0152	.0116	.0140	.0111	.0084	.0100	.0080	.0061
α	True S.D.	.0318	.0285	.0234	.0222	.0215	.0166	.0164	.0155	.0120
	Ave. S.E.	.0254	.0267	.0222	.0181	.0192	.0158	.0128	.0136	.0113

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, F^0 is $N(-2,.5)$.
2. The distribution of random coefficient for the BLP estimator is correctly specified.

True S.D., but the discrepancy gets smaller as the sample size gets large. This pattern in effect resembles the findings in the sieve literature (Chen et al. (2014)). Hence, empirical researchers should be cautious about the potential downward bias of the GMM standard errors when the sample size is not large enough.

Next, we summarize the estimation results of random coefficients in Tables 4 and 5. The first thing to note is that the performance of the BLP estimator and the SN-Para estimator are quite close, which is encouraging since both estimators use the same functional form assumption for the random coefficients distribution. The SN estimators with sieve approximations of F have somewhat larger RtMSEs than BLP and the SN estimator with (correctly specified) parametric F. This is also expected since the sieve estimators do not use the parametric assumption. Finally, from the comparisons of different SN estimators (with nonparametric F) in Table 15 in Appendix S4 we find that the combination "SN-III-GMM2" have an overall better performance (in terms of RtMSEs and biases) than others.[16]

[16]In Appendix S4.5, we document the average computational time (across repetitions) of BLP and SN estimators in Table 22, which illustrates the computational advantage of avoiding demand inversion.

Table 4: Monte Carlo Results: Mean of Random Coefficient

		$T = 10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-III-GMM2	RtMSE	.0379	.0240	.0186	.0255	.0189	.0153	.0204	.0144	.0143
	Bias	-.0035	-.0033	-.0069	.0015	-.0061	-.0087	-.0039	-.0067	-.0103
SN-Para-GMM2	RtMSE	.0318	.0177	.0105	.0263	.0127	.0075	.0214	.0098	.0054
	Bias	.0267	.0126	.0036	.0234	.0068	2.35E-5	.0187	.0068	-1.73E-4
BLP	RtMSE	.0330	.0192	.0118	.0277	.0150	.0092	.0249	.0136	.0080
	Bias	.0280	.0147	.0072	.0249	.0118	.0068	.0232	.0122	.0066

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $N(-2, .5)$.
2. The distribution of the random coefficient for the BLP estimator and SN-Para estimators is correctly specified.

Table 5: Monte Carlo Results: Standard Deviation of Random Coefficient

		$T = 10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-III-GMM2	RtMSE	.0379	.0240	.0186	.0255	.0189	.0153	.0204	.0144	.0143
	Bias	-.0035	-.0033	-.0069	.0015	-.0061	-.0087	-.0039	-.0067	-.0103
SN-Para-GMM2	RtMSE	.0318	.0177	.0105	.0263	.0127	.0075	.0214	.0098	.0054
	Bias	.0267	.0126	.0036	.0234	.0068	2.35E-5	.0187	.0068	-1.73E-4
BLP	RtMSE	.0330	.0192	.0118	.0277	.0150	.0092	.0249	.0136	.0080
	Bias	.0280	.0147	.0072	.0249	.0118	.0068	.0232	.0122	.0066

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $N(-2, .5)$.
2. The distribution of random coefficient for the BLP estimator and SN-Para estimators is correctly specified.

5.3.2 Design II: F^0 is Mixed Normal

To further investigate the performance of our estimator, we deviate from the normal case and let the true distribution of random coefficient F^0 be an asymmetric mixed normal $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$. Note that the BLP and SN-Para estimators, which impose normality, are misspecified in this case.

The results for the fixed coefficients, shown in Tables 13 and 14 in Appendix S4 to conserve space, are very similar to those in Tables 2 and 3. This suggests that the BLP estimator of fixed coefficients is not sensitive to the type of misspecification considered here. Our SN estimators work very similarly to the BLP estimator. Moreover, the first step standard errors of the SN estimator for the fixed coefficients seem to be fairly accurate, as shown in Table 14.

Next, we present the estimation results for the random coefficient in Tables 6 and 7. For both mean and standard deviation, we can see that the SN works better than the BLP estimator and SN-Para. In addition, since F^0 is asymmetric, it is useful to examine the skewness of the estimated random coefficient, the results of which are reported in Table 17 in Appendix S4. From the table, it is clear that the biases of our SN estimators decrease quickly as the sample size gets large, while the parametric estimators have a fixed bias because of the incorrect normality (symmetric) assumption. Finally, "SN-III-GMM2" again has an overall better performance than alternative implementations, so we shall use it for the empirical application.

Table 6: Monte Carlo Results: Mean of Random Coefficient

		$T=10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-III-GMM2	RtMSE	.0416	.0305	.0245	.0301	.0246	.0200	.0231	.0202	.0176
	Bias	-.0094	-.0117	-.0113	-.0133	-.0135	-.0119	-.0109	-.0138	-.0125
SN-Para-GMM2	RtMSE	.0448	.0579	.0632	.0494	.0618	.0636	.0500	.0609	.0641
	Bias	-.0403	-.0562	-.0623	-.0474	-.0610	-.0632	-.0490	-.0605	-.0639
BLP	RtMSE	.0419	.0547	.0597	.0466	.0585	.0597	.0473	.0574	.0602
	Bias	-.0383	-.0533	-.0590	-.0450	-.0578	-.0594	-.0465	-.0571	-.0601

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$.
2. The distribution of random coefficient for the BLP and SN-Para estimators is mis-specified.

Table 7: Monte Carlo Results: S.D. of Random Coefficient

		$T=10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-III-GMM2	RtMSE	.0527	.0437	.0275	.0446	.0304	.0192	.0316	.0219	.0156
	Bias	-.0195	-.0200	-.0075	-.0165	-.0123	-.0029	-.0126	-.0088	-8.88E-4
SN-Para-GMM2	RtMSE	.0400	.0381	.0377	.0426	.0374	.0356	.0372	.0337	.0367
	Bias	-.0301	-.0321	-.0342	-.0389	-.0343	-.0337	-.0350	-.0319	-.0358
BLP	RtMSE	.0385	.0430	.0439	.0466	.0451	.0435	.0408	.0419	.0449
	Bias	-.0324	-.0399	-.0422	-.0440	-.0436	-.0425	-.0394	-.0410	-.0445

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$.
2. The distribution of random coefficient for the BLP and SN-Para estimators is mis-specified.

Besides the mean and standard deviation of the RC, we also examine the whole estimated distribution. In particular, we report the mean and median integrated squared errors (ISEs) of the estimated distribution in Table 8. We can see that, in most cases for the mean and in all cases for the median, the SN-III estimator of the RC distribution has smaller ISEs than the SN-Para and the BLP estimators which mis-specify the random coefficient distribution.

In addition, we plot the estimated PDFs of BLP and SN-III-GMM2 for the case with $T = 40$ and $J = 100$. The PDFs are shown in Figure 1; the solid ones are point-wise averages

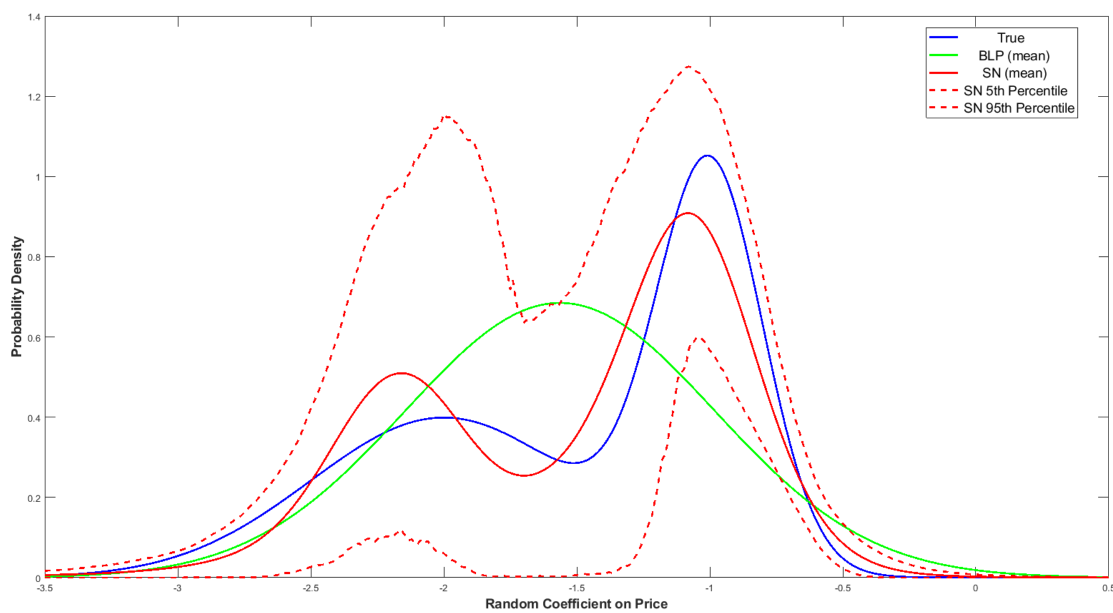
Table 8: Monte Carlo Results: Mean- and Median-Integrated Sequared Errors

		$T = 10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-III-GMM2	Mean-ISE	.0986	.1150	.1215	.1065	.1198	.0983	.1304	.0964	.0934
	Median-ISE	.0969	.0807	.0785	.0811	.0769	.0655	.0817	.0646	.0623
SN-Para-GMM2	Mean-ISE	.1136	.1181	.1199	.1164	.1197	.1203	.1161	.1193	.1204
	Median-ISE	.1133	.1180	.1197	.1163	.1195	.1202	.1159	.1192	.1202
BLP	Mean-ISE	.1137	.1183	.1200	.1165	.1199	.1203	.1162	.1194	.1203
	Median-ISE	.1134	.1182	.1198	.1164	.1197	.1202	.1160	.1194	.1202

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $.5 \times N\left(-1, .2^2\right) + .5 \times N\left(-2, .5^2\right)$.
2. The distribution of random coefficient for the BLP and SN-Para estimators is mis-specified.
3. The ISEs are calculated as $\int \left[\hat{f}_J\left(v\right) - f^0\left(v\right)\right]^2 f^0\left(v\right) dv$ (weighted by the true density) for each repetition. The reported numbers are the means and medians of the ISEs across 1000 repetitions.

across repetitions and the two dashed ones are the 5th and 95th point-wise percentiles of the SN estimator. Looking at the mean, we see that SN estimator can approximate the true bimodal distribution better than the BLP estimator with normal RC. Although the variance of the SN estimator seems rather large, the 5th and 95th confidence band can still rule out the misspecified BLP estimator for some regions in the support of the distribution.[17]

Figure 1: Estimated PDFs of Random Coefficient



[17]The 5% and 95% percentile band of the BLP estimator is very narrow around the mean so is omitted to avoid clutter.

5.4 Results with Small J

Previous results show that our approach works well with a moderate J, e.g., 25; so a natural question is whether it can handle small J's, say less than 10. In Table 9, we show some simulation results for $J = 5, 10$,[18] using the same design as that in 5.3.1. We can see that overall our estimator still performs very well in these cases, though the asymptotic theory requires "large" J. Compared to BLP estimator, the only notable difference is that SN estimator has larger RtMSE and bias for the SD of random coefficient, which is expected and similar to the previous moderate and large J cases.

Table 9: Monte Carlo Results: Small J

			$T = 10$		20		40	
Parameter	Estimator	J	5	10	5	10	5	10
β	SN	RtMSE	.0701	.0532	.0565	.0381	.0508	.0274
		Bias	-.0400	-.0199	-.0400	-.0180	-.0426	-.0143
	BLP	RtMSE	.0960	.0416	.0584	.0358	.0551	.0302
		Bias	-.0393	-.0259	-.0377	-.0270	-.0414	-.0256
α	SN	RtMSE	.0635	.0585	.0424	.0418	.0310	.0301
		Bias	3.45E-4	.0024	5.45E-4	.0010	-1.46E-4	.0010
	BLP	RtMSE	.1052	.0368	.0724	.0269	.0498	.0193
		Bias	-.0021	6.11E-4	.0030	-3.70E-5	.0050	-.0011
Mean of RC	SN-III-GMM2	RtMSE	.1069	.0609	.0954	.0515	.0925	.0422
		Bias	.0849	.0382	.0839	.0379	.0855	.0337
	SN-Para-GMM2	RtMSE	.0915	.0581	.0849	.0562	.0855	.0525
		Bias	.0830	.0519	.0807	.0526	.0828	.0507
	BLP	RtMSE	.0921	.0583	.0846	.0569	.0854	.0530
		Bias	.0812	.0521	.0799	.0533	.0824	.0513
SD of RC	SN-III-GMM2	RtMSE	.0751	.0585	.0529	.0528	.0490	.0403
		Bias	-.0219	-.0153	-.0093	-.0185	-.0121	-.0124
	SN-Para-GMM2	RtMSE	.0488	.0353	.0328	.0247	.0233	.0172
		Bias	-.0014	.0025	7.69E-4	.0026	.0010	.0043
	BLP	RtMSE	.0738	.0356	.0360	.0253	.0259	.0168
		Bias	-.0117	8.19E-4	-.0018	7.20E-4	-4.45E-4	.0024

Note: 1. The design and specification in this set of simulations are the same as those in "Design I: F^0 is Normal" in the Monte Carlo section.
2. For $J = 10$, we set k_J and ς_J to be 3. For $J = 5$, we let k_J and ς_J be 2 so that there are sufficient number of observations to estimate unknown parameters.
3. We setM_J to be 3, 4, 5 for $T = 10$, 20, 40, respectively.
4. For $J = 5$, the first stage estimates of SN estimator are obtained via 2SLS instead of 2-step GMM because the former is numerically more stable in the small J cases.

[18] Our approach cannot work with too small J's, say $J = 3$, because we need a sufficient number of observations to estimate the market-specific sieve coefficients (usually 3 or more of these).

5.5 Extension to Multiple Random Coefficients

In this subsection, we extend the above simulation exercise to the cases with multiple random coefficients. Here we consider a design with three independent random coefficients. A case with two correlated random coefficients is presented in Supplemental Appendix S4.3

We add two exogenous characteristics, $X_{2,jt}$ and $X_{3,jt}$, each of which is associated with new random coefficients, to the data generating process in Section 5.1. And we assume the three random coefficients are independent[19] both in the data generating process and in the estimation.

Given that the random coefficients are independent, the market share equation becomes

$$\sigma_j\left(\delta_t, P_t, X_{2,t}, X_{3,t}; F\right) = \int \frac{\exp\left(\delta_{jt} + \sum_{\ell=1}^{3} \upsilon_\ell X_{\ell,jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + \sum_{\ell=1}^{3} \upsilon_\ell X_{\ell,kt}\right)} dF_1^0\left(\upsilon_1\right) F_2^0\left(\upsilon_2\right) F_3^0\left(\upsilon_3\right),$$

where $X_{1,jt} \equiv P_{jt}$ is generated as in Section 5.1, $X_{2,jt}$ and $X_{3,jt}$ are both drawn from a standard normal distribution, and $F_i^0\left(\upsilon_i\right)$ $(i = 1, 2, 3)$ is normal with means μ_i and standard deviations σ_i. Thus the market share is simulated as

$$s_{jt} = \frac{1}{R} \sum_{i=1}^{R} \frac{\exp\left(\delta_{jt} + \sum_{\ell=1}^{3} \upsilon_{\ell,i} X_{\ell,jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt} + \sum_{\ell=1}^{3} \upsilon_{\ell,i} X_{\ell,kt}\right)}, \tag{29}$$

where $\{\upsilon_{i,r} : r = 1, ..., R\}$ is drawn from $F_i^0\left(\upsilon_i\right)$ for $i = 1, 2, 3$.

The implementation of our SN estimator is adjusted to accommodate the three independent random coefficients. For the first stage estimation defined by (12), we use power series to approximate the three dimensional functions $\{\psi_{k_J,t}(P_{jt}, X_{2,jt}, X_{3,jt}) : t = 1, ..., T\}$ and $I^{\varsigma_J}(W_{jt}, X_{2,jt}, X_{3,jt})$.[20] For the second stage estimation, we focus on the preferred specification "SN-III-GMM2" and apply the third sieve approximation described in Section 5.2 to each

[19] This is a commonly imposed assumption in empirical applications, see among others, Berry et al. (1995); Nevo (2001); Petrin (2002).

[20] We use third order polynomials to approximate these functions.

$F_i^0(\upsilon_i)$ for $i = 1, 2, 3$. So the market share of product j in market t can be written as

$$\begin{aligned}&\tilde{\sigma}_j(\delta_t, P_t; F_{M_J})\\&= \int_0^1\int_0^1\int_0^1 \frac{\exp\left(\delta_{jt} + \sum_{\ell=1}^3 \Phi^{-1}(z_\ell; \tilde{\mu}_\ell, \tilde{\sigma}_\ell) X_{\ell,jt}\right)}{1 + \sum_{k=1}^J \exp\left(\delta_{kt} + \sum_{\ell=1}^3 \Phi^{-1}(z_\ell; \tilde{\mu}_\ell, \tilde{\sigma}_\ell) X_{\ell,kt}\right)} \prod_{i=1}^3 q_{i,M_J}(z_i)\, dz,\end{aligned}$$

where $q_{i,M_J}(z_i)$'s ($i = 1, 2, 3$) have the same functional form but different sieve coefficients, and each z_i follows $U[0, 1]$. For comparison, we also include the "SN-Para-GMM2" with a parametric F^0 in the second stage estimation.

The results for the fixed coefficients are very similar to the previous cases so we defer them to the Appendix (see Table 18) to save space. Here in Table 10, we show the estimation results for the mean and standard deviations of the three random coefficients. Although we have only tried a small subset of the extensive list of specifications examined in the previous cases (with a single random coefficient), the results are still informative to give us a sense of how our SN estimator works in the case of multiple random coefficients.

From the results, we can see that the RtMSEs and biases of our SN estimator for random coefficients are larger than those of the correctly specified parametric estimators, including BLP and SN-Para. And the difference is larger than the previous cases with a single random coefficient. This is not surprising given that there are three functions approximated by sieves and the sample sizes are relatively small. Also, the SN estimator performs better for the standard deviations (σ_i's) than the means (μ_i's) of the random coefficients in terms of both RtMSE and bias, which suggests potential rooms for improvement on the estimation of the means. Finally, the SN-Para estimator performs well and is rather similar to the BLP estimator, which reminds us that our estimation strategy with a parametric F in the second step is a valuable complement to the full-blown SN estimator with a nonparametric F^0.

Table 10: Monte Carlo Results: Independent Random Coefficients

			$T = 10$		20					$T = 10$		20	
Parameter	Estimator		$J = 50$	100	50	100	Parameter	Estimator		$J = 50$	100	50	100
μ_1	BLP	RtMSE	.0345	.0183	.0335	.0167	σ_1	BLP	RtMSE	.0183	.0162	.0128	.0120
		Bias	.0323	.0154	.0323	.0154			Bias	.0014	.0032	.0024	.0038
	SN-Para-GMM2	RtMSE	.0349	.0167	.0357	.0146		SN-Para-GMM2	RtMSE	.0414	.0174	.0435	.0144
		Bias	.0291	.0115	.0309	.0119			Bias	-.0098	.0055	-.0170	.0052
	SN-III-GMM2	RtMSE	.1919	.1522	.1787	.1205		SN-III-GMM2	RtMSE	.0496	.0331	.0444	.0233
		Bias	.1293	.0937	.1188	.0708			Bias	-.0081	.0030	-.0162	-.0034
μ_2	BLP	RtMSE	.0127	.0094	.0091	.0064	σ_2	BLP	RtMSE	.0208	.0163	.0143	.0111
		Bias	-.0008	.0002	.0001	.0002			Bias	-.0001	.0017	.0009	.0023
	SN-Para-GMM2	RtMSE	.0155	.0100	.0127	.0070		SN-Para-GMM2	RtMSE	.0330	.0169	.0321	.0125
		Bias	.0011	.0014	.0014	.0011			Bias	-.0082	.0028	-.0122	.0028
	SN-III-GMM2	RtMSE	.0916	.0764	.0846	.0602		SN-III-GMM2	RtMSE	.0457	.0281	.0404	.0218
		Bias	-.0607	-.0491	-.0554	-.0383			Bias	-.0211	-.0071	-.0223	-.0062
μ_3	BLP	RtMSE	.0128	.0091	.0089	.0064	σ_3	BLP	RtMSE	.0203	.0161	.0145	.0114
		Bias	.0003	-.0001	-.0002	.0003			Bias	-.0001	.0008	.0001	.0023
	SN-Para-GMM2	RtMSE	.0156	.0097	.0126	.0072		SN-Para-GMM2	RtMSE	.0323	.0166	.0325	.0126
		Bias	.0024	.0011	.0008	.0014			Bias	-.0080	.0019	-.0128	.0027
	SN-III-GMM2	RtMSE	.0877	.0748	.0821	.0582		SN-III-GMM2	RtMSE	.0402	.0272	.0366	.0221
		Bias	-.0553	-.0469	-.0517	-.0350			Bias	-.0084	-.0022	-.0133	.0007

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, $\mu_1 = -2$, $\mu_2 = \mu_3 = 1$, $\sigma_1 = \sigma_2 = \sigma_3 = .5$.
2. The distribution of random coefficients for the BLP and SN-Para estimators is correctly specified.

6 Empirical Illustrations

To further illustrate our approach, we consider two empirical applications. The first one revisits the original BLP's application to the US auto market, and the second one applies the proposed estimator to the Chinese auto market. In the first application, we find that the random coefficient on price has a bimodal distribution and that the BLP estimates, which assumes a normal random coefficient distribution, bias toward zero. In the second application, we find that the BLP and the SN give nearly identical results. The first application is presented next, while the second in Appendix S5 to conserve space.

The BLP data records the price, quantity, as well as product characteristics of car models on the US market for each year from 1971 to 1990. As in BLP, a market is naturally defined by a year: there are 20 markets and on average, a market has about 110 models (see the last two rows in Table 11).[21]

We consider a specification with one random coefficient on price and fixed coefficients on other product characteristics. The first two columns in Table 11, labeled "BLP", show results obtained using the standard BLP estimators with simple logit and random coefficient logit specifications, respectively. Also, we implement our SN estimator using the SN-III-GMM2

[21]See Berry et al. (1995) for detailed descriptions of the data.

implementation, which is the preferred choice from the simulation results in the Monte Carlo section. In addition, we use the standard BLP IVs to handle the price endogeneity problem for all the estimators; the F-statistic for these IVs from a price regression, reported in the table, suggests that these IVs are not weak.

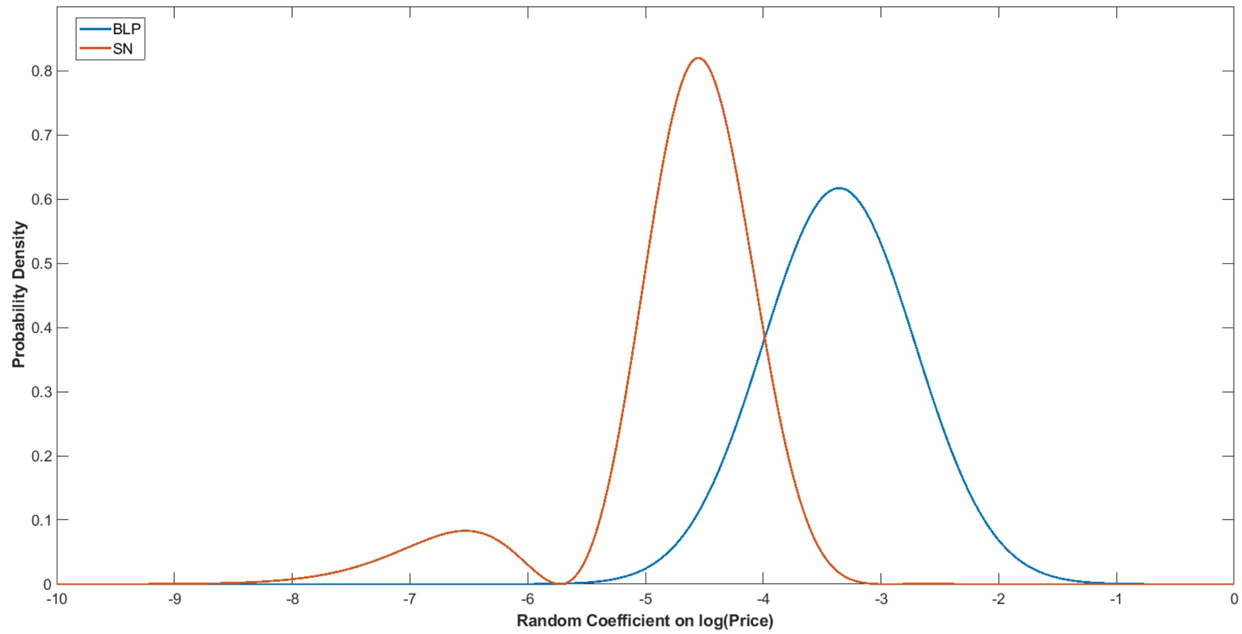
From the results, we can see that our SN estimator yields a larger mean (in absolute value) and standard deviation of price coefficient than the BLP estimator, which translates into an overall more elastic demand. Also, we plot the density functions of the price coefficient based on BLP and SN estimates in Figure 2. Besides the obvious difference in the location of the two distributions, the SN estimator exhibits a bimodal shape (with some very elastic consumers) that indicates a deviation from normality.

Table 11: BLP Auto Data Revisited

	BLP		SN
Fixed Coefficient	Logit	RC-Logit	
HP/Weight (log)	.78 (.15)	.76 (.12)	1.37 (.25)
Size (log)	3.31 (.21)	3.53 (.21)	3.82 (.52)
Dollar per Miles (log)	−.22 (.11)	−.33 (.11)	−.42 (.36)
A/C	.69 (.12)	.62 (.08)	1.01 (.27)
Power Steering	.19 (.08)	.22 (.07)	.55 (.12)
Automatic	.33 (.07)	.30 (.07)	.41 (.10)
FWD	.19 (.06)	.21 (.06)	.18 (.10)
Constant	−3.56 (.25)	−2.43 (.46)	−8.39 (7.57)
RC on Price (log)			
Mean	−2.60 (.17)	−3.36 (.29)	-4.76
Std. Dev.	-	.65 (.12)	.81
Ave. No. of Prod. per Mkt.		110.85	
No. of Mkt.		20	
F-statistic for BLP IV		62.96	

We further examine the implications of the estimation results on price elasticities, which is a key output of demand estimation. To compute the price elasticities, we need not only the parameter estimates $\hat{\theta}$ but also the residuals $\hat{\xi}$'s. We obtain $\hat{\xi}$'s by inverting the demand

Figure 2: Estimated PDF of Random Coefficient



system (4) evaluated at $\hat{\theta}$ via the standard BLP contraction mapping algorithm.[22]

Table 12 reports a sample of price elasticities. Regarding to the own-price elasticities, the SN estimator implies a much more elastic demand than the BLP estimator. Moreover, the pattern of cross-price elasticities is different: compared to the BLP estimator, the SN estimator implies that cars are more substitutable to others, especially for economy cars like Accord. Hence, the rather different patterns of elasticities implied by BLP and SN estimators highlight the importance of the shape of the random coefficients distribution in determining the substitution patterns among products.

[22]The $\hat{\xi}$'s obtained in this way are less noisy than the residuals from the first stage estimation because the former approach takes the full structure of the demand model into account while the latter treats the function $\psi^0_{J,t}(\cdot)$ as a reduced-form, nuisance function. Note that calculating the price elasticity requires solving the contraction mapping only once, and thus it is feasible to obtain a highly precise solution.

Table 12: Price Elasticities for Selected Models in 1990 Market

		735i	Century	Seville	Escort	Taurus	Accord	LS400	323	Maxima	Sentra
BMW	BLP	-2.1163	.0109	.0057	.0183	.0294	.0381	.0076	.0013	.0120	.0031
735i	SN	-3.9038	.0122	.0056	.0245	.0334	.0438	.0075	.0019	.0125	.0042
Buick	BLP	.0010	-2.6209	.0029	.0157	.0208	.0273	.0037	.0012	.0075	.0027
Century	SN	.0011	-4.1714	.0033	.0209	.0259	.0341	.0042	.0016	.0089	.0036
Cadillac	BLP	.0018	.0099	-2.2843	.0177	.0269	.0349	.0062	.0013	.0105	.0030
Seville	SN	.0018	.00113	-4.0174	.0236	.0310	.0408	.0060	.0018	.0112	.0040
Ford	BLP	.0007	.0061	.0020	-2.8271	.0169	.0223	.0026	.0011	.0057	.0024
Escort	SN	.0009	.0081	.0027	-4.2623	.0225	.0298	.0034	.0015	.0076	.0033
Ford	BLP	.0010	.0075	.0028	.0155	-2.6256	.0269	.0036	.0012	.0073	.0027
Taurus	SN	.0011	.0093	.0033	.0208	-4.1632	.0338	.0042	.0016	.0088	.0036
Honda	BLP	.0010	.0074	.0028	.0154	.0202	-2.6344	.0035	.0012	.0072	.0026
Accord	SN	.0011	.0092	.0032	.0206	.0254	-4.1621	.0041	.0016	.0087	.0035
Lexus	BLP	.0019	.0102	.0050	.0179	.0276	.0359	-2.2337	.0013	.0109	.0031
LS400	SN	.0019	.0116	.0048	.0239	.0318	.0417	-3.9879	.0018	.0116	.0041
Mazda	BLP	.0006	.0058	.0019	.0134	.0161	.0213	.0024	-2.8807	.0054	.0023
323	SN	.0008	.0079	.0026	.0186	.0219	.0290	.0033	-4.3006	.0073	.0032
Nissan	BLP	.0013	.0084	.0035	.0165	.0229	.0300	.0045	.0012	-2.5059	.0028
Maxima	SN	.0013	.0100	.0037	.0219	.0276	.0364	.0047	.0017	-4.1199	.0037
Nissan	BLP	.0007	.0061	.0020	.0138	.0169	.0223	.0026	.0011	.0057	-2.8387
Sentra	SN	.0009	.0081	.0027	.0190	.0225	.0298	.0034	.0015	.0076	-4.2781

7 Concluding Remarks

In this paper, we propose a semi-nonparametric approach to estimating the widely used BLP model. The approach is easy to implement and allows for the nonparametric specification of random coefficients. We establish the asymptotic theory of the proposed estimator and show the semi-nonparametric identification of the random coefficients logit demand (BLP) model under the large J framework. Results from simulation studies and empirical applications demonstrate the usefulness of our estimator.

Acknowlegements

We thank the editor, Elie Tamer, the associate editor, and annoymous referees for their constructive suggestions that greatly improve the quality of this work. We are thankful to Patrick Bajari, Steven Berry, Xiaohong Chen, Yanqin Fan, Ying Fan, Amit Gandhi, Philip Haile, Aviv Nevo, Bernard Salanie, Andres Santos and many other individuals for insightful discussions and suggestions; We would also like to thank the participants at the Recent Advances in Econometrics (Toulouse School of Economics), IAAE 2016 Annual Conference, 2017 California Econometric Conference, 2017 SETA Conference, 2018 North American Winter Meeting, 2018 CEME conference, as well as seminar participants at Wisconsin-Madison, Washington-Seattle, Yale, Xiamen U, TAMU, Renmin U, Academia Sinica for their many helpful comments and questions. Finally, we thank Wei-min Hu for sharing the data on the Chinese auto market.

References

Ackerberg, D. A. and M. Rysman (2005): "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects," *The RAND Journal of Economics*, 36, 771–788.

Adams, R. (1975): *Sobolev spaces*, Academic press.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Armstrong, T. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980.

Bajari, P. and C. L. Benkard (2005): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113, 1239–1276.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy.*

Berry, S. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 841–890.

Berry, S., O. Linton, and A. Pakes (2004): "Limit theorems for estimating the parameters of differentiated product demand systems," *Review of Economic Studies*, 71, 613–654.

Berry, S. T. and P. A. Haile (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

Bierens, H. J. (2008): "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results," *Econometric Theory*.

Bordley, R. (2013): "Discrete Choice with Large Choice Sets," *Economic Letters*, 118, 13–15.

Chen, X. and T. M. Christensen (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9, 39–84.

Chen, X., Z. Liao, and Y. Sun (2014): "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics*, 178, 639–658.

Chen, X. and D. Pouzo (2015): "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models," *Econometrica*, 83, 1013–1079.

Chiang, H. D., K. Kato, and Y. Sasaki (2021): "Inference for high-dimensional exchangeable arrays," *Journal of the American Statistical Association*, 1–11.

Compiani, G. (2018): "Nonparametric Demand Estimation in Differentiated Products Markets," *Available at SSRN 3134152*.

Conlon, C. and J. Gortmaker (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51, 1108–1161.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55–93.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80, 2231–2267.

Dunker, F., S. Hoderlain, and H. Kaido (2017): "Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models," *Working Paper*.

Fosgerau, M. and M. Bierlaire (2007): "A practical test for the choice of mixing distribution in discrete choice models," *Transportation Research Part B: Methodological*, 41, 784 – 794.

Fosgerau, M. and S. L. Mabit (2013): "Easy and flexible mixture distributions," *Economics Letters*, 120, 206 – 210.

Fox, J. T. and A. Gandhi (2016): "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47, 118–139.

Fox, J. T., K. il Kim, and C. Yang (2016): "A simple nonparametric approach to estimating the distribution of random coefficients in structural models," *Journal of Econometrics*, 195, 236–254.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011): "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2, 381–418.

——— (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212.

Freyberger, J. (2015): "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, 181, 162–181.

Freyberger, J. and M. A. Masten (2017): "A Practical Guide to Compact Infinite Dimensional Parameter Spaces," *working paper.*

Gandhi, A. and J.-F. Houde (2016): "Measuring Substitution Patterns in Differentiated Products Industries," *Working Paper*.

Goeree, M. S. (2008): "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica*, 76, 1017–1074.

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *The Annals of Statistics*, 33, 2904–2929.

Knittel, C. R. and K. Metaxoglou (2014): "Estimation of random-coefficient demand models: two empiricists' perspective," *Review of Economics and Statistics*, 96, 34–59.

Lee, J. and K. Seo (2015): "A computationally fast estimator for random coefficients logit demand models using aggregate data," *The RAND Journal of Economics*, 46, 86–102.

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

McFadden, D. (1974): "The measurement of urban travel demand," *Journal of Public Economics*, 3, 303 – 328.

Moon, H. R., M. Shum, and M. Weidner (2018): "Estimation of random coefficients logit demand models with interactive fixed effects," *Journal of Econometrics.*

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Petrin, A. (2002): "Quantifying the Benefit of New Products: The Case of Minivan," *Journal of Political Economy*, 110, 705–729.

Quan, T. W. and K. R. Williams (2018): "Product Variety, Across Market Demand Heterogeneity, and the Value of Online Retail," *The RAND Journal of Economics*, 49, 877–913.

Reynaert, M. and F. Verboven (2014): "Improving the performance of random coefficients demand models: the role of optimal instruments," *Journal of Econometrics*, 179, 83–98.

Salanié, B. and F. A. Wolak (2019): "Fast," robust", and approximately correct: estimating mixed demand systems," *working paper*.

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80, 213–275.

Train, K. (2016): "Mixed Logit with a Flexible Mixing Distribution," *Working Paper*.

Online Supplemental Appendix for the "Semi-Nonparametric Estimation of Random Coefficients Logit Model for Aggregate Demand"

Zhentong Lu, Xiaoxia Shi, and Jing Tao

In this supplemental appendix, we include additional proofs and results for "Semi-Nonparametric Estimation of Random Coefficients Logit Model for Aggregate Demand" by Zhentong Lu, Xiaoxia Shi, and Jing Tao (LST for short).

Section S1 proves the theorems and lemmas in LST.

Section S2 gives sufficient conditions for and verifies Assumptions 4-6 and Assumption 8 (ii) in LST.

Section S3 reports some theoretical results on cross-product elasticity in a random coefficients logit model when the number of products grows to infinity.

Section S4 reports additional Monte Carlo simulation results.

Section S5 contains the second empirical application.

Section S6 plots the distribution of $\log(s_{jt}/s_{0t})$ at various J values under the Monte Carlo design of Section 5.

Section S7 presents a simple location estimation example to illustrate the idea of a random true parameter value.

S1 Proof of Theorems and Lemmas in LST

In this section, we prove the theorems and lemmas in LST. The proofs are arranged in the order that the corresponding results appear in LST.

S1.1 Proof of Theorem 1

The proof of this theorem uses Lemma S1 below. The lemma is proved at the end of this subsection.

Lemma S1. *Let* $(\Theta, \|\cdot\|)$ *be a compact metric space and* Θ_{k_J} *be a sieve space of* Θ *such that* $\Theta_{k_1} \subseteq \Theta_{k_2} \subseteq \ldots \subseteq \Theta$. *Let* $\widehat{\theta}_J = \arg\min_{\theta \in \Theta_{k_J}} \widehat{\mathcal{R}}_J(\theta)$ *for a sample criterion function* $\widehat{\mathcal{R}}_J : \Theta \to \mathbb{R}$. *Suppose that*

(a) *there is an auxiliary criterion function* $\widetilde{R}_J(\theta)$ *such that* $\sup_{\theta_{k_J} \in \Theta_{k_J}} |\widetilde{R}_J(\theta_{k_J}) - \mathcal{R}_J(\theta_{k_J})| \xrightarrow{p} 0$ *for a population criterion function* $\mathcal{R}_J(\theta)$ *which may still be random;*

(b) *there exists a sequence of random variables* $\theta_J^0 \in \Theta$ *such that* $\mathcal{R}_J(\theta_J^0) = 0$ *for all* J.

(c) *for any* $\varepsilon > 0$ *there exists* $\delta_\varepsilon > 0$ *such that* $\inf_{\theta \in \Theta : \|\theta - \theta_J^0\| > \varepsilon} (\mathcal{R}_J(\theta) - \mathcal{R}_J(\theta_J^0)) > \delta_\varepsilon$ *with probability approaching one, and*

(d) $\left| \widetilde{R}_J(\widehat{\theta}_J) \right| \xrightarrow{p} 0$.

Then we have $\|\widehat{\theta}_J - \theta_J^0\| \xrightarrow{p} 0$ *as* $J \to \infty$.

Proof of Theorem 1. We prove the theorem by verifying the conditions in Lemma S1. Mapping the notation of this theorem to that of Lemma S1, we note that the population criterion function is $\mathcal{R}_J(\theta) = \bar{\mathcal{L}}_J(\theta|\theta_J^0)$ and the auxiliary criterion function is $\widetilde{R}_J(\theta) = \tilde{\mathcal{L}}_J(\theta)$. Condition (a) in Lemma S1 is guaranteed by Assumption 5(ii). Condition (b) holds because $\bar{\mathcal{L}}(\theta_J^0|\theta_J^0) = 0$ by definition. Condition (c) is guaranteed by Assumption 5(i). It is only left to verify Condition (d) of Lemma S1.

Note that for the two-stage GMM estimator $\widehat{\theta}_J$, for some generic positive constants C', C'', and C''', we have

$$\tilde{\mathcal{L}}_J(\widehat{\theta}_J) = \sum_{t=1}^{T} \bar{g}_t(\widehat{\theta}_{J,t})' W_t^{2SLS} \bar{g}_t(\widehat{\theta}_{J,t}) \leq C' \sum_{t=1}^{T} \|\bar{g}_t(\widehat{\theta}_{J,t})\|^2 \leq C'' \sum_{t=1}^{T} \bar{g}_t(\widehat{\theta}_{J,t})' \hat{\Omega}_t \left(\tilde{\theta}_{J,t}\right)^{-1} \bar{g}_{J,t}\left(\widehat{\theta}_t\right)$$

$$\leq C'' \sum_{t=1}^{T} \bar{g}_t \left(\theta^0_{J,k_J,t}\right)' \hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1} \bar{g}_t(\theta^0_{J,k_J,t}) \leq C''' \sum_{t=1}^{T} \|\bar{g}_t(\theta^0_{J,k_J,t})\|^2 = o_p(1), \tag{S1}$$

where the first two inequalities follow by Assumption 4(ii), the third inequality follows by the definition of $\widehat{\theta}_J$, the last inequality follows by Assumption 4(ii) again, and the last equality follows by Assumption 4(i). This verifies Condition (d) of Lemma S1. □

Next we prove the asymptotic normality result for the parametric estimator $\hat{\beta}_J$.

S1.2 Proof of Theorem 2

Proof of Theorem 2. First recall that

$$V_{J,\beta}^{-1} = \sum_{t=1}^{T} V_{J,t}^{-1} = \sum_{t=1}^{T} \frac{1}{J} \sum_{j=1}^{J} E\left[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})'/\Sigma_{j,J,o}\right],$$

and that T is fixed as we take $J \to \infty$. By Assumption 2 (iv) and the conditions in the theorem, we have

$$\inf_{J,t} eig_{\min}(V_{J,t}) \geq \left(\inf_{J} \inf_{\mathbf{z}\in\mathcal{Z}_t} E[\xi_{jt}^2 | \{Z_{jt}\}_{j=1}^{J} = \mathbf{z}]\right) \left(\sup_{J,t} eig_{\max}\left(\frac{1}{J}\sum_{j=1}^{J} E\left[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})'\right]\right)\right)^{-1}$$
$$> 0. \tag{S2}$$

Thus,

$$\inf_{J} eig_{\min}(V_{J,\beta}) \geq T^{-1} \inf_{J,t} eig_{\min}(V_{J,t}) > 0 \tag{S3}$$

which implies that for any $\lambda_J \in \mathbb{R}^{d_{x1}}$ with $\sup_J \|\lambda_J\| < \infty$,

$$\sup_J \lambda_J' V_{J,\beta}^{-1} \lambda_J \leq \sup_J T^2 \lambda_J' \lambda_J \left(\inf_{J,t} eig_{\min}(V_{J,t}) \right)^{-1} < \infty.$$

Then Assumption 6 (i) applies and gives us

$$\begin{aligned} &\sqrt{J}\lambda_J' V_{J,\beta}^{-1/2} \left(\hat{\beta}_J - \beta^0 \right) \\ &= - \lambda_J' V_{J,\beta}^{1/2} \sum_{t=1}^{T} \left\{ \left(\frac{1}{\sqrt{J}} \sum_{j=1}^{J} D_{j,J,t}(Z_{jt})' \Sigma_{j,J,o}^{-1} \xi_{jt} \right) \right\} + o_p(1). \end{aligned} \tag{S4}$$

For each market t let $\mathcal{C}_{j,J,t,}$ be the σ-field generated by $\{\{\xi_{j't}\}_{j'=1}^{j}, \{Z_{jt}\}_{j=1}^{J}\}$ and let

$$\mathbb{A}_{j,J,t} = \frac{\lambda_J' V_{J,\beta}^{1/2} D_{j,J,t}(Z_{jt})}{\Sigma_{j,J,o} \sqrt{\lambda_J' V_{J,\beta}^{1/2} V_{J,t}^{-1} V_{J,\beta}^{1/2} \lambda_J}}. \tag{S5}$$

Then for each t,$\{\{(\mathbb{A}_{j,J,t}\xi_{jt}, \mathcal{C}_{j,J,t})\}_{j=1}^{J}\}_{J=1}^{\infty}$ is a martingale difference array by Assumption 1(i). We first show that

$$\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \mathbb{A}_{j,J,t}\xi_{jt} \xrightarrow{d} N(0,1). \tag{S6}$$

To show this, we follow Theorem 1.3 in Alj et al. (2014), which requires us to verify that (i) $E\left[|\mathbb{A}_{j,J,t}\xi_{jt}|^{2+c}\right] < B < \infty$ for some constants $B > 0$ and $c > 0$ for $j = 1, \ldots, J$ and for all J, and (ii) $\frac{1}{J}\sum_{j=1}^{J} E\left[|\mathbb{A}_{j,J,t}\xi_{jt}|^2 | \mathcal{C}_{j-1,J,t}\right] \xrightarrow{p} 1$.

To verify (i), first note that $\sup_{j,J,t} \sup_{z \in \mathcal{Z}_{jt}} \|D_{J,t}(z)\| < \infty$ by the condition in the theorem. Let $\underline{\sigma}^2 = \inf_{\mathbf{z} \in \mathcal{Z}_t, j=1,\ldots,J; J \geq 1, t \geq 1} E[\xi_{jt}^2 | \{Z_{jt}\} = \mathbf{z}]$. Then $\underline{\sigma} > 0$ with probability one by Assumption 2(iv). Let

$$C = \sup_{j,J,t} \sup_{z \in \mathcal{Z}_{jt}} \|D_{j,J,t}(z)\| / \underline{\sigma}^2.$$

Then for each J, t, we have with probability one,

$$\sup_{j,J,t} \|\mathbb{A}_{j,J,t}\| \leq C \times \frac{\sqrt{\lambda_J' V_{J,\beta} \lambda_J}}{\sqrt{\lambda_J' V_{J,\beta}^{1/2} V_{J,t}^{-1} V_{J,\beta}^{1/2} \lambda_J}} \leq C \times \sqrt{eig_{\max}(V_{J,t})}. \tag{S7}$$

By Assumption 2 (iv) and the condition in the theorem, we have

$$\begin{aligned}
&\sup_{J,t} eig_{\max}(V_{J,t}) \\
&= \sup_{J,t} \left(eig_{\min} \left(\frac{1}{J} \sum_{j=1}^{J} E\left[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})' / \Sigma_{j,J,o} \right] \right) \right)^{-1} \\
&\leq \sup_{j,J,t} \sup_{\mathbf{z} \in \mathcal{Z}_t} E[\xi_{jt}^2 | \{Z_{j't}\}_{j'=1}^{J} = \mathbf{z}] \left(\inf_{J,t} eig_{\min} \left(\frac{1}{J} \sum_{j=1}^{J} E\left[D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})' \right] \right) \right)^{-1} \\
&< \infty. \qquad \text{(S8)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_{J,t} E[|\mathbb{A}_{j,J,t}\xi_{jt}|^4] &= \sup_{j,J,t} E[E[\xi_{jt}^4 | \{Z_{jt}\}_{j=1}^{J}] \|\mathbb{A}_{j,J,t}\|^4] \\
&\leq C^4 (eig_{\max}(V_{J,t}))^2 \sup_{J,t} E[E[\xi_{jt}^4 | \{Z_{jt}\}_{j=1}^{J}]] \\
&= C^4 (eig_{\max}(V_{J,t}))^2 \sup_{J,t} E[\xi_{jt}^4] \\
&< \infty,
\end{aligned}$$

where the two equalities hold by the law of iterated expectations, the first inequality holds by (S7), the second inequality holds by (S8) and Assumption 2(iv). This verifies condition (i) above by setting $c = 2$.

To verify (ii), note that

$$\frac{1}{J} \sum_{j=1}^{J} E\left[\mathbb{A}_{jJ,t}^2 \xi_{jt}^2 | \mathcal{C}_{j-1,J,t} \right] = \frac{1}{J} \sum_{j=1}^{J} \mathbb{A}_{j,J,t}^2 E[\xi_{jt}^2 | \mathcal{C}_{j-1,J,t}] = \frac{1}{J} \sum_{j=1}^{J} \mathbb{A}_{j,J,t}^2 E[\xi_{jt}^2 | \{Z_{jt}\}_{j=1}^{J}]$$

$$= \frac{1}{J}\sum_{j=1}^{J} \frac{\lambda_J' V_{J,\beta}^{1/2} D_{j,J,t}(Z_{jt}) D_{j,J,t}(Z_{jt})' V_{J,\beta}^{1/2}\lambda_J}{\Sigma_{j,J,o}^2 \lambda_J' V_{J,\beta}^{1/2} V_{J,t}^{-1} V_{J,\beta}^{1/2}\lambda_J}\Sigma_{j,J,o} = 1,$$

where the first equality holds because $\mathbb{A}_{j,J,t}$ is measurable with respect to $\{Z_{jt}\}_{j=1}^J$ (since $\Sigma_{j,J,o}$ is a function of $\{Z_{jt}\}_{j=1}^J$, the second equality holds by Assumption 1(i), the third equality holds by the definitions of $\mathbb{A}_{jt}$ and $\Sigma_{j,J,o}$, and the last equality holds by the definition of $V_{J,t}$. Thus, (ii) is verified. Therefore, (S6) holds.

Next note that

$$\begin{aligned}\frac{\sqrt{J}\lambda_J' V_{J,\beta}^{-1/2}(\hat{\beta}_J - \beta^0)}{\sqrt{\lambda_J'\lambda_J}} &= \frac{-\lambda_J' V_{J,\beta}^{1/2}\sum_{t=1}^T\left\{\left(\frac{1}{\sqrt{J}}\sum_{j=1}^J D_{j,J,t}(Z_{jt})'\Sigma_{j,J,o}^{-1}\xi_{jt}\right)\right\}}{\sqrt{\lambda_J'\lambda_J}} + o_p(1)\\ &= \frac{-\sum_{t=1}^T\left\{\sqrt{\lambda_J' V_{J,\beta}^{1/2} V_{J,t}^{-1} V_{J,\beta}^{1/2}\lambda_J}\left(\frac{1}{\sqrt{J}}\sum_{j=1}^J \mathbb{A}_{J,t}(Z_{jt})\xi_{jt}\right)\right\}}{\sqrt{\lambda_J'\lambda_J}} + o_p(1),\end{aligned}$$

where the first equality follows from Assumption 6(i).

For each t, since $V_{J,t}$ is bounded and $\|\lambda_J\| = 1$, for any subsequence of $(\{V_{J,t}\}, \lambda_J)$, there exists a further subsequence that converges to some $(V_{\infty,t}, \lambda_\infty)$. For a given converging subsequence, let $V_{\infty,\beta} = (\sum_{t=1}^T V_{\infty,t}^{-1})^{-1}$. Then along this converging subsequence, $\frac{\sqrt{J}\lambda_J' V_{J,\beta}^{-/2}(\hat{\beta}_J - \beta^0)}{\sqrt{\lambda_J'\lambda_J}}$ converges in distribution to

$$\frac{-\sum_{t=1}^T\left\{\sqrt{\lambda_\infty' V_{\infty,\beta}^{1/2} V_{\infty,t}^{-1} V_{\infty,\beta}^{1/2}\lambda_\infty}\Phi_t\right\}}{\sqrt{\lambda_\infty'\lambda_\infty}} \sim N(0,1), \tag{S9}$$

where $(\Phi_1, \Phi_2, \ldots, \Phi_T)' \sim N(0, I_T)$, where the independence across t follows from Assumption 1(ii). That is, we have shown, any subsequence of the sequence $\frac{\sqrt{J}\lambda_J' V_{J,\beta}^{-1/2}(\hat{\beta}_J - \beta^0)}{\sqrt{\lambda_J'\lambda_J}}$ has a subsequence that converges in distribution to $N(0,1)$. This implies that

$$\sqrt{J}\lambda_J' V_{J,\beta}^{-1/2}(\hat{\beta}_J - \beta^0)/\sqrt{\lambda_J'\lambda_J} \xrightarrow{d} N(0,1).$$

Then the Cramer-Wold device implies that $\sqrt{J}V_{J,\beta}^{-1/2}(\hat{\beta}_J - \beta^0) \xrightarrow{d} N(0, I_{d_{x_1}})$. This and

Assumption 6(ii) as well as (S3) together proves the theorem. □

S1.3 Proof of Lemma 1

Proof of Lemma 1. Part (a): By Assumption 7(i), there exists t_0 such that the support of X_{2,jt_0} contains the bounded open $\mathbb{R}^{d_{x_2}}$-ball $\mathcal{B}_0$. Suppose that the center of $\mathcal{B}_0$ is x_{20}.

Observe that $M_{F,t_0}(y) := \frac{\int \frac{\exp(y'\upsilon)\exp(x'_{20}\upsilon)}{\varphi^0_{J,t_0}(\upsilon)}dF(\upsilon)}{\int \frac{1}{\varphi^0_{J,t_0}(\upsilon)}dF(\upsilon)}$ is the moment generating function of the density function

$$\tilde{f}(\upsilon) := \frac{\exp(x'_{20}\upsilon)f(\upsilon)}{\varphi^0_{J,t_0}(\upsilon)\int \frac{1}{\varphi^0_{J,t_0}(u)}dF(u)},$$

and $M_{F^0,t_0}(y)$ exists for all values of y in an open ball around the origin by Assumption 7(ii). Then by Theorem 2.3.11(b) of Casella and Berger (2001), if $M_{F,t_0}(x_2 - x_{20}) = M_{F^0,t_0}(x_2 - x_{20})$ for all x_2 in an open ball around x_{20}, we have

$$\frac{\exp(x'_{20}\upsilon)f(\upsilon)}{\int \frac{1}{\varphi^0_{J,t}(u)}dF(u)} = \frac{\exp(x'_{20}\upsilon)f^0(\upsilon)}{\int \frac{1}{\varphi^0_{J,t}(u)}dF^0(u)} \quad \text{for all } \upsilon. \tag{S10}$$

Now suppose that $f \neq f^0$, we show that (S10) cannot hold. Suppose the contrary. Since $f \neq f^0$, there exists υ_0 such that $f(\upsilon_0) \neq f^0(\upsilon_0)$. Without loss of generality, suppose that $f(\upsilon_0) < f^0(\upsilon_0)$. Let $\lambda = f(\upsilon_0)/f^0(\upsilon_0) < 1$. Equation (S10) implies that,

$$\frac{\int \frac{1}{\varphi^0_{J,t}(u)}dF(u)}{\int \frac{1}{\varphi^0_{J,t}(u)}dF^0(u)} = \lambda < 1. \tag{S11}$$

Applying (S10) again, we have $f(\upsilon) = \lambda f^0(\upsilon)$ for all υ. But this contradicts the fact that both f and f^0 are density functions and thus must both integrate up to 1. Therefore, for $f \neq f^0$, (S10) cannot hold, which in turn implies that it cannot be true that $M_{F,t_0}(x_2 - x_{20}) = M_{F^0,t_0}(x_2 - x_{20})$ for all $x_2 \in \mathcal{B}_0$. Therefore, there exists $x_2 \in \mathcal{B}_0$ such that

$$\log M_{F,t_0}(x_2 - x_{20}) := \log \int \frac{\exp(x'_2\upsilon)}{\varphi^0_{J,t_0}(\upsilon)}dF(\upsilon) - \log \int \frac{1}{\varphi^0_{J,t_0}(\upsilon)}dF(\upsilon)$$

$$
\begin{aligned}
&\neq \psi^0_{J,t}(x_2) \\
&:= \log \int \frac{\exp(x_2' \upsilon)}{\varphi^0_{J,t_0}(\upsilon)} dF^0(\upsilon) - \log \int \frac{1}{\varphi^0_{J,t_0}(\upsilon)} dF^0(\upsilon) \\
&:= \log M_{F^0,t_0}(x_2 - x_{20}). \qquad \text{(S12)}
\end{aligned}
$$

Therefore,

$$
\log \int \frac{\exp(x_2' \upsilon)}{\varphi^0_{J,t_0}(\upsilon)} dF(\upsilon) - \log \int \frac{1}{\varphi^0_{J,t_0}(\upsilon)} dF(\upsilon) \neq \psi^0_{J,t}(x_2) \text{ almost surely,} \qquad \text{(S13)}
$$

proving Part (a).

Part (b): By Part (a), $f \neq f^0$ implies that for some t and some $x_2 \in \mathcal{X}_{2,t}$,

$$
\log \int \frac{\exp(x_2' \upsilon)}{\varphi^0_{J,t}(\upsilon)} dF(\upsilon) - \log \int \frac{1}{\varphi^0_{J,t}(\upsilon)} dF(\upsilon) \neq \psi^0_{J,t}(x_2) \text{ almost surely .}
$$

Due to the continuity of the functions on both sides in x_2, the inequality holds for x_2 in a subset of $\mathcal{X}_{2,t}$ with positive $X_{2,jt}$-measure. Thus,

$$
\begin{aligned}
Q_J(f) = \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} E_X \left(\left[\log \int \frac{\exp(X_{2,jt}' \upsilon)}{\varphi^0_{J,t}(\upsilon)} f(\upsilon) d\upsilon - \log \int \frac{1}{\varphi^0_{J,t}(\upsilon)} f(\upsilon) d\upsilon - \psi^0_{J,t}(X_{2,jt}) \right]^2 \right) \\
> 0
\end{aligned}
$$

almost surely. ☐

S1.4 Proof of Theorem 3

Proof of Theorem 3. We first show the consistency of $\hat{f}_J = \hat{f}_{MD}$ defined in (14) as follows.

We obtain the consistency of $\hat{f}_J$ by verifying the conditions in Lemma S1. Specifically, we show that

(i) $\sup_{f \in \mathcal{F}_{M_J}} \left| \hat{\mathcal{Q}}_J(f) - \mathcal{Q}_J(f) \right| \xrightarrow{p} 0$,

(ii) $\mathcal{Q}_J(f^0) = 0$,

(iii) for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $\inf_{f \in \mathcal{F}: \|f - f^0\|_\infty > \varepsilon} \left(\mathcal{Q}_J(f) - \mathcal{Q}_J(f^0)\right) > \delta_\varepsilon$ with probability approaching one, and

(iv) $\left|\hat{\mathcal{Q}}_J(\hat{f}_J)\right| \xrightarrow{p} 0$.

Condition (ii) holds by definition. Condition (iii) is guaranteed by Assumption 8(i). To verify Condition (i), first note that the triangular inequality implies that

$$\left|\hat{\mathcal{Q}}_J(f) - \mathcal{Q}_J(f)\right| \leq \left|\hat{\mathcal{Q}}_J(f) - \tilde{\mathcal{Q}}_J(f)\right| + \left|\tilde{\mathcal{Q}}_J(f) - \mathcal{Q}_J(f)\right|. \tag{S14}$$

Assumption 8(ii) implies that the second term is $o_p(1)$ uniformly over $f \in \mathcal{F}_{M_J}$. Thus, it is left to show

$$\sup_{f \in \mathcal{F}_{M_J}} |\hat{\mathcal{Q}}_J(f) - \tilde{\mathcal{Q}}_J(f)| \xrightarrow{p} 0. \tag{S15}$$

Note that

$$\left|\hat{\mathcal{Q}}_J(f) - \tilde{\mathcal{Q}}_J(f)\right| \leq I_J(f) + 2\sqrt{\tilde{Q}_J(f)}\sqrt{I_J(f)}, \tag{S16}$$

where

$$I_J(f) := \frac{1}{J}\sum_{j=1}^{T}\sum_{j=1}^{J}\left\{\hat{\psi}_{k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt}) + \log\left[\frac{G_{J,t}(X_{2,jt}; f)}{G_{J,t}(0; f)}\right] - \log\left[\frac{\widehat{G}_{J,t}(X_{2,jt}; f)}{\widehat{G}_{J,t}(0; f)}\right]\right\}^2.$$

For $I_J(f)$, we have

$$\begin{aligned} I_J(f) \leq & \frac{2}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left[\hat{\psi}_{k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})\right]^2 + \frac{4}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\log G_{J,t}(X_{2,jt}; f) - \log \widehat{G}_{J,t}(X_{2,jt}; f)\right)^2 \\ & + \frac{4}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\log G_{J,t}(0; f) - \log \widehat{G}_{J,t}(0; f)\right)^2 \end{aligned} \tag{S17}$$

where we use the inequality that $(a+b)^2 \leq 2a^2 + 2b^2$. Moreover,

$$\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left[\hat{\psi}_{k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})\right]^2 = o_p(1) \tag{S18}$$

by Assumption 8(ii). In addition, by Taylor expansion and the fact that $\hat{\delta}_{jt} - \delta_{jt} = \psi^0_{J,t}(X_{2,jt}) - \hat{\psi}_{k_J,t}(X_{2,jt})$, we have, uniformly over $f \in \mathcal{F}_{M_J}$,

$$\begin{aligned}&\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\log G_{J,t}(X_{2,jt}; f) - \log \widehat{G}_{J,t}(X_{2,jt}; f)\right)^2\\ =&o_p(1) + O(\|\hat{\psi}_{k_J,t} - \psi^0_{J,t}\|^2_{0,2,\delta_0}) = o_p(1),\end{aligned} \tag{S19}$$

where the last equality follows from Theorem 1. Similarly, uniformly over $f \in \mathcal{F}_{M_J}$,

$$\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\log G_{J,t}(\mathbf{0}; f) - \log \widehat{G}_{J,t}(\mathbf{0}; f)\right)^2 = \quad o_p(1) + O(\|\hat{\psi}_{k_J,t} - \psi^0_{J,t}\|^2_{0,2,\delta_0}) = o_p(1). \tag{S20}$$

Equations (S17)-(S20) together imply that $\sup_{f\in\mathcal{F}_{M_J}} I_J(f) = o_p(1)$. This and (S16), combined with Assumptions 8(ii)-(iii), yield (S15) and in turn verify Condition (i).

For Condition (iv), note that $Q_J(f^0) = 0$. This and Assumption 8(iv) together imply that $Q_J(f^0_{M_J}) = o_p(1)$. Subsequently, Condition (i) implies that $\hat{Q}_J(f^0_{M_J}) = o_p(1)$. By the definition of $\hat{f}_J$, $\hat{Q}_J(\hat{f}_J) \leq \hat{Q}_J(f^0_{M_J})$, we have $\hat{Q}_J(\hat{f}_J) = o_p(1)$ verifying Condition (iv). The conclusion of the theorem follows from Lemma S1.

Let

$$\begin{aligned}\mathcal{Q}_{J,GMM2}(\beta, f) &= \frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} E\left[\left(E\left[\left\{\log\left(\frac{s_{jt}}{s_{0t}}\right) - X'_{1,jt}\beta - \log\left(\frac{G_{J,t}(X_{2,jt}; f)}{G_{J,t}(\mathbf{0}; f)}\right)\right\}|Z_{jt}\right]\right)^2\right],\\ \tilde{Q}_{J,GMM2}(\beta, f) &= \sum_{t=1}^{T}\bar{\bar{g}}_t(\beta, f_{M_J})' W^{2sls}_t \bar{\bar{g}}_t(\beta, f_{M_J}),\end{aligned}$$

where

$$\bar{\hat{g}}_t(\beta, f_{M_J}) = \frac{1}{J}\sum_{j=1}^{J}\left\{\log\left(\frac{s_{jt}}{s_{0t}}\right) - X_{1,jt}'\beta - \log\left(\frac{\hat{G}_{J,t}\left(X_{2,jt}; f_{M_J}\right)}{\hat{G}_{J,t}\left(\mathbf{0}; f_{M_J}\right)}\right)\right\}.$$

Let $\alpha = (\beta, f) \in \mathcal{B} \times \mathcal{F}$, $\alpha_{M_J}^0 = (\beta, f_{M_J}^0) \in \mathcal{B} \times \mathcal{F}_{M_J}$ and $\alpha^0 = (\beta^0, f^0)$. In this proof, for $\alpha \in \mathcal{B} \times \mathcal{F}$, we write

$$\|\alpha\|_s^2 = \|\beta\|^2 + \|f\|_\infty^2.$$

Mapping the notation of this corollary to that of Lemma S1, we note that $\mathcal{R}_J(\alpha) = \mathcal{Q}_{J,GMM2}(\alpha)$ and $\tilde{R}_J(\alpha) = \tilde{Q}_{J,GMM2}(\alpha)$. Specifically, we need to show that

(i) $\sup_{\alpha\in\mathcal{B}\times\mathcal{F}}\left|\tilde{Q}_{J,GMM2}(\alpha) - \mathcal{Q}_{J,GMM2}(\alpha)\right| \xrightarrow{p} 0$,

(ii) $\mathcal{Q}_{J,GMM2}(\alpha^0) = \mathcal{Q}_{J,GMM2}(\beta^0, f^0) = 0$,

(iii) for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that

$$\inf_{\alpha\in\mathcal{B}\times\mathcal{F}:\|\alpha-\alpha^0\|_s>\varepsilon}\left(\mathcal{Q}_{J,GMM2}(\alpha) - \mathcal{Q}_{J,GMM2}(\alpha^0)\right) > \delta_\varepsilon$$

with probability approaching one, and

(iv) $\left|\tilde{\mathcal{Q}}_J(\hat{\alpha}_{GMM2})\right| \xrightarrow{p} 0$.

Condition (ii) holds by definition. Condition (iii) is guaranteed by Assumption 1 and Assumption 7 following the argument in Lemma 1 To verify Condition (i), the proof of Theorem 3 shows that uniformly over $f \in \mathcal{F}$,

$$\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left\{\log\left(\frac{\hat{G}_{J,t}(X_{2,jt}; f)}{\hat{G}_{J,t}(\mathbf{0}; f)}\right) - \log\left(\frac{G_{J,t}(X_{2,jt}; f)}{G_{J,t}(\mathbf{0}; f)}\right)\right\}^2 \xrightarrow{p} 0.$$

Then using a similar argument in Step 2 of Proposition 1, we can verify Condition (i). For Condition (iv), note that $\sum_{t=1}^{T}\bar{\hat{g}}_t(\beta^0, f_{M_J}^0) = o_p(1)$ by Assumption 8 (iv), which implies that $\hat{\mathcal{Q}}_J(\alpha_{M_J}^0) = o_p(1)$. Thus, following the proof of Theorem 1, we have

$$\hat{\mathcal{Q}}_J\left(\hat{\alpha}_{GMM2}\right) \leq \hat{\mathcal{Q}}_J(\alpha_{M_J}^0) = o_p(1)$$

and Condition (iv) is verified. The conclusion follows from Lemma S1. □

S1.5 Proof of Lemma S1

Proof of Lemma S1. Define an arbitrary $\varepsilon > 0$. Consider the following derivation:

$$\begin{aligned}
&\Pr(\|\widehat{\theta}_J - \theta_J^0\| > \varepsilon) \\
&\leq \Pr(\mathcal{R}_J(\widehat{\theta}_J) - \mathcal{R}_J(\theta_J^0) > \delta_\varepsilon) + \Pr(\|\widehat{\theta}_J - \theta_J^0\| > \varepsilon, \mathcal{R}_J(\widehat{\theta}_J) - \mathcal{R}_J(\theta_J^0) \leq \delta_\varepsilon) \\
&= \Pr(\mathcal{R}_J(\widehat{\theta}_J) - \widetilde{\mathcal{R}}_J(\widehat{\theta}_J) + \widetilde{\mathcal{R}}_J(\widehat{\theta}_J) > \delta_\varepsilon) + o(1) \\
&\leq \Pr(|\mathcal{R}_J(\widehat{\theta}_J) - \widetilde{\mathcal{R}}_J(\widehat{\theta}_J)| > \delta_\varepsilon/2) + \Pr(|\widetilde{\mathcal{R}}_J(\widehat{\theta}_J)| > \delta_\varepsilon/2) + o(1) \\
&\to 0,
\end{aligned} \tag{S21}$$

where the first inequality holds by basic set operation, the equality holds by conditions (b) and (c), the second inequality holds by $P(A \cup B) \leq P(A) + P(B)$, and the convergence holds by conditions (a) and (d). □

S2 Verification of Assumptions in LST

S2.1 Verification of Assumption 5(i)

For notational simplicity, in this subsection, we assume that the joint distribution of $(X_{1,jt}', X_{2,jt}', Z_{jt})'$ does not vary across j. It can still vary across t and can change as J changes.

The uniform completeness/rank conditions in Assumption 5(i) are adapted from Hall and Horowitz (2005),[23] following whom we define the operator kernel:

$$\tau_{J,t}(x, w) = \int \frac{f_{XZ,J,t}(x, z) f_{XZ,J,t}(w, z)(1 + x_2'x_2)^{-\delta_0/2}(1 + w_2'w_2)^{-\delta_0/2}}{f_{Z,J,t}(z) f_{X_1|X_2,J,t}(x_1|x_2)^{1/2} f_{X_1|X_2,J,t}(w_1|w_2)^{1/2}} dz,$$

where x stands for the concatenation of the d_{x_1}-vector x_1 and the d_{x_2}-vector x_2, and w has the

[23]We thank Andres Santos for suggesting this approach.

same structure, $f_{XZ,J,t}(\cdot,\cdot)$ stands for the joint density function of X_{jt} and Z_{jt}, $f_{X_1|X_2,J,t}(\cdot)$ stands for the conditional density function of $X_{1,jt}$ given $X_{2,jt}$, and $f_{Z,J,t}(\cdot)$ stands for the marginal density function of Z_{jt}. These densities are allowed to vary with J. We assume below that $\int\int |\tau_{J,t}(x,w)|^2 dxdw < \infty$. Then the operator $\mathcal{T}_{J,t}g(x) = \int \tau_{J,t}(x,w)g(w)dw$ is a self-adjoint Hilbert-Schmidt integral operator. Then following Chapter 5.3 of Horowitz (2009), we note that there exist eigenvalues $\{\lambda_{\ell,J,t} \geq 0\}_{\ell=1}^{\infty}$ and eigen functions $\{\phi_{\ell,J,t} : \mathbb{R}^{d_{x_1}+d_{x_2}} \to \mathbb{R}\}_{\ell=1}^{\infty}$ such that $\int \phi_{\ell,J,t}^2(x)dx = 1$ and $\int \phi_{\ell,J,t}(x)\phi_{\tilde{\ell},J,t}(x)dx = 0$ for all $\ell \neq \tilde{\ell}$, and

$$\tau_{J,t}(x,w) = \sum_{\ell=1}^{\infty} \lambda_{\ell,J,t}\phi_{\ell,J,t}(x)\phi_{\ell,J,t}(w), \tag{S22}$$

and for any function $g : \mathbb{R}^{d_{x_1}+d_{x_2}} \to \mathbb{R}$ such that $\int g^2(x)(1+x_2'x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2)dx < \infty$,

$$g(x)f_{X_1|X_2,J,t}(x_1|x_2)^{1/2}(1+x_2'x_2)^{\delta_0/2} = \sum_{\ell=1}^{\infty} \langle g, \phi_{\ell,J,t}\rangle \phi_{\ell,J,t}(x), \tag{S23}$$

where $\langle g, \phi_{\ell,J,t}\rangle = \int g(x)\phi_{\ell,J,t}(x)f_{X_1|X_2,J,t}(x_1|x_2)^{1/2}(1+x_2'x_2)^{\delta_0/2}dx$.

Assumption 9 (Uniform Completeness)**.** (i) $\frac{1}{J}\sum_{j=1}^{J} E[\|X_{1,jt}\|^2|X_{2,jt}] < \infty$ *almost surely for all* t, J, *and for all* t,J, *and* $x_2 \in \mathcal{X}_2$, *we have*

$$eig_{\min}\left(E\left[\left(X_{1,jt} - E(X_{1,jt}|X_{2,jt}=x_2)\right)\left(X_{1,jt} - E(X_{1,jt}|X_{2,jt}=x_2)\right)'\middle| X_{2,jt}=x_2\right]\right) > c_0,$$

and $\int_{\mathcal{X}_2} \|E(X_{1,jt}|X_{2,jt}=x_2)\|^2(1+x_2'x_2)^{\delta_0}dx_2 < c_0^{-1}$ *for a constant* $c_0 > 0$ *not dependent on* t, J *or* x_2.

(ii) $\int\int |\tau_{J,t}(x,w)|^2dxdw < \infty$ *for all* J, t.

(iii) $\lambda_{\ell,J,t} \geq c_1\ell^{-c_2}$ *for all* ℓ, t,J *and constant* $c_1, c_2 > 0$ *that do not depend on* t *or* J.

(iv) *For any* $(\beta,\psi) \in \mathcal{B}\times\Psi$, *and* $g(x) = \beta'x_1 + \psi(x_2)$, *we have* $|\langle g, \phi_{\ell,J,t}\rangle| \leq c_3\ell^{-c_4}$ *for all* ℓ,t,J *and constants* $0 \leq c_3 < \infty$ *and* $c_4 > 1/2$ *that do not depend on* ℓ, t, *or* J.

Remark. Part (i) is the full-rank condition for identifying β^0. Part (ii) makes sure that $\tau_{J,t}(x,w)$ is the kernel of a self-adjoint Hilbert-Schmidt integral operator which is compact and admits the spectral decomposition shown in (S22). Parts (iii) and (iv) together imply a uniform version of the completeness condition needed for identifying $X_1\beta^0 + \psi^0_{J,t}(X_2)$ using instrument Z. These two conditions are similar to Condition A.3 in Hall and Horowitz (2005).

Lemma S2. *Suppose that Assumptions 1 and 9 hold. Then Assumption* 5(i) *holds.*

Proof. First note that for arbitrary $\theta, \theta^* \in \Theta$,

$$\begin{aligned}
\bar{\mathcal{L}}_J(\theta|\theta^*) &= \frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} E\left[\left(E\left[\rho_{jt}(\theta_t|\theta_t^*)|Z_{jt}\right]\right)^2\right] \\
&= \frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} E\left[\left(E\left[\xi_{jt} + X'_{1,jt}(\beta^* - \beta) + (\psi_t^*(X_{2,jt}) - \psi_t(X_{2,jt}))|Z_{jt}\right]\right)^2\right] \\
&= \sum_{t=1}^{T} E\left[\left(E\left[X'_{1,jt}(\beta^* - \beta) + (\psi_t^*(X_{2,jt}) - \psi_t(X_{2,jt}))|Z_{jt}\right]\right)^2\right],
\end{aligned} \tag{S24}$$

where the first two qualities hold by the definition and the third equality holds by Assumption 1(i). Let $g_t^*(x) = x_1'\beta^* + \psi_t^*(x_2)$ and $g_t(x) = x_1'\beta + \psi_t(x_2)$. Then, by Assumptions 2(i), 2(ii) and 2(iii)(b), and $\frac{1}{J}\sum_{j=1}^{J} E[\|X_{1,jt}\|^2|X_{2,jt}] < \infty$ (Assumption 9(i)), we have that for all J, t,

$$\begin{aligned}
&\int g_t^2(x)(1 + x_2'x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2)dx < \infty \\
&\int g_t^*(x)^2(1 + x_2'x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2)dx < \infty \quad a.s.
\end{aligned} \tag{S25}$$

Thus, they admit the following spectrum decomposition:

$$\begin{aligned}
g_t(x) &= \sum_{\ell=1}^{\infty} \langle g_t, \phi_{\ell,J,t}\rangle\, \phi_{\ell,J,t}(x)(1 + x_2'x_2)^{-\delta_0/2} f_{X_1|X_2,J,t}(x_1|x_2)^{-1/2} \\
g_t^*(x) &= \sum_{\ell=1}^{\infty} \langle g_t^*, \phi_{\ell,J,t}\rangle\, \phi_{\ell,J,t}(x)(1 + x_2'x_2)^{-\delta_0/2} f_{X_1|X_2,J,t}(x_1|x_2)^{-1/2},
\end{aligned} \tag{S26}$$

where $\phi_{\ell,J,t}(x) : \ell = 1, \ldots, \infty$ are the eigen functions of the operator

$$\mathcal{T}_{J,t} g(x) = \int \tau_{J,t}(x,w) g(w) dw,$$

where $\tau_{J,t}(x,w)$ is defined in (S22). Let $b_{\ell,J,t} = \langle g_t, \phi_{\ell,J,t} \rangle - \langle g_t^*, \phi_{\ell,J,t} \rangle$. Then

$$\begin{aligned}
&\bar{\mathcal{L}}_J(\theta|\theta^*) \\
&= \frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} E_Z \left[\left(E_X \left[\sum_{\ell=1}^{\infty} \frac{b_{\ell,J,t} \phi_{\ell,J,t}(X_{jt})(1 + X_{2,jt}' X_{2,jt})^{-\delta_0/2}}{f_{X_1|X_2,J,t}(X_{1,jt}|X_{2,jt})^{1/2}} \middle| Z_{jt} \right] \right)^2 \right] \\
&= \frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} E_Z \left[\left(\sum_{\ell=1}^{\infty} b_{\ell,J,t} \int \frac{\phi_{\ell,J,t}(x)(1 + x_2' x_2)^{-\delta_0/2} f_{XZ,J,t}(x, Z_{jt})}{f_{X_1|X_2,J,t}(x_1|x_2)^{1/2} f_{Z,J,t}(Z_{jt})} dx \right)^2 \right] \\
&= \frac{1}{J} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ \sum_{\ell,\tilde{\ell}}^{\infty} b_{\ell,J,t} b_{\tilde{\ell},J,t} \int \int \phi_{\ell,J,t}(x) \phi_{\tilde{\ell},J,t}(w) \tau_{J,t}(x,w) dx dw \right\} \\
&= \sum_{t=1}^{T} \left\{ \sum_{\ell=1}^{\infty} \lambda_{\ell,J,t} b_{\ell,J,t}^2 \right\}, \qquad \text{(S27)}
\end{aligned}$$

where $\lambda_{\ell,J,t} : \ell = 1, \ldots, \infty$ are the eigenvalues of the integral operator $\mathcal{T}_{J,t}$, and the last equality holds by (S22).

Using (S26), we can derive

$$\begin{aligned}
&\int (g_t(x) - g_t^*(x))^2 (1 + x_2' x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2) dx \\
&= \int \left(\sum_{\ell=1}^{\infty} b_{\ell,J,t} \phi_{\ell,J,t}(x) \right)^2 dx \\
&= \sum_{\ell,\tilde{\ell}} b_{\ell,J,t} b_{\tilde{\ell},J,t} \int \phi_{\ell,J,t}(x) \phi_{\tilde{\ell},J,t}(x) dx \\
&= \sum_{\ell=1}^{\infty} b_{\ell,J,t}^2. \qquad \text{(S28)}
\end{aligned}$$

Assumption 9(iv) implies that

$$|b_{\ell,J,t}| < 2c_3\ell^{-c_4} \text{ for all } J,t.$$

Since $c_4 > 1/2$, the infinite sum $\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2$ converges uniformly over J,t. Thus, for any positive number e_1, there exists an $\overline{\ell}_{e_1}$ (not dependent on J,t) large enough so that

$$\sum_{\ell=1}^{\overline{\ell}_{e_1}} b_{\ell,J,t}^2 > \sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 - e_1/T, \tag{S29}$$

where T is finite in our setting. Equation (S29) implies that, for an arbitrary $e_1 > 0$, we have

$$\bar{\mathcal{L}}_J(\theta|\theta^*) \geq \sum_{t=1}^{T}\sum_{\ell=1}^{\overline{\ell}_{e_1}} \lambda_{\ell,J,t} b_{\ell,J,t}^2 \geq \sum_{t=1}^{T}\sum_{\ell=1}^{\overline{\ell}_{e_1}} c_1 \overline{\ell}_{e_1}^{-c_2} b_{\ell,J,t}^2 \geq c_1 \overline{\ell}_{e_1}^{-c_2} \left(\sum_{t=1}^{T}\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 - e_1 \right). \tag{S30}$$

where the first inequality holds by (S27) and the fact that the eigenvalues $\lambda_{\ell,J,t}$ are non-negative and the second inequality holds by Assumption 9(iii).

It is left to construct a lower bound for $\sum_{t=1}^{T}\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2$ using $\|\theta - \theta^*\|_s^2$. Note that

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 \\
&= \int (g_t(x) - g_t^*(x))^2 (1 + x_2' x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2) dx \\
&= \int (x_1'(\beta - \beta^*) + \psi_t(x_2) - \psi_t^*(x_2))^2 (1 + x_2' x_2)^{\delta_0} f_{X_1|X_2,J,t}(x_1|x_2) dx \\
&= \int E[(X_{1,jt}'(\beta - \beta^*) + \psi_t(x_2) - \psi_t^*(x_2))^2 | X_{2,jt} = x_2](1 + x_2' x_2)^{\delta_0} dx_2 \\
&= (\beta - \beta^*)' \left(\int A_{J,t}(x_2)(1 + x_2' x_2)^{\delta_0} dx_2 \right) (\beta - \beta^*) \\
&\quad + \int (E(X_{1,jt}|X_{2,jt} = x_2)'(\beta - \beta^*) + \psi_t(x_2) - \psi_t^*(x_2))^2 (1 + x_2' x_2)^{\delta_0} dx_2 \qquad \text{(S31)} \\
&= (\beta - \beta^*)' \left(\int A_{J,t}(x_2)(1 + x_2' x_2)^{\delta_0} dx_2 \right) (\beta - \beta^*) + \|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2
\end{aligned}$$

$$+ \int (E(X_{1,jt}|X_{2,jt} = x_2)'(\beta - \beta^*))^2 |X_{2,jt} = x_2](1 + x_2'x_2)^{\delta_0} dx_2$$

$$+ 2(\beta - \beta^*)' \int (E(X_{1,jt}|X_{2,jt} = x_2)(\psi_t(x_2) - \psi_t^*(x_2))(1 + x_2'x_2)^{\delta_0} dx_2$$

$$\geq \|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2$$

$$+ 2(\beta - \beta^*)' \int (E(X_{1,jt}|X_{2,jt} = x_2)(\psi_t(x_2) - \psi_t^*(x_2))(1 + x_2'x_2)^{\delta_0} dx_2$$

$$\geq \|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2$$

$$- 2\|\beta - \beta^*\| \sqrt{\int ||E(X_{1,jt}|X_{2,jt} = x_2)\|^2 (1 + x_2'x_2)^{\delta_0} dx_2} \|\psi_t - \psi_t^*\|_{0,2,\delta_0}, \qquad \text{(S32)}$$

where $A_{J,t}(x_2) = E[(X_{1,jt} - E(X_{1,jt}|X_{2,jt} = x_2))(X_{1,jt} - E(X_{1,jt}|X_{2,jt} = x_2))'|X_{2,jt} = x_2]$. Also note that by Assumption 9(i), we have

$$(\beta - \beta^*)' A_{J,t}(x_2)(\beta - \beta^*) > c_0 \|\beta - \beta^*\|^2, \text{ and} \qquad \text{(S33)}$$

$$\int \|E(X_{1,jt}|X_{2,jt} = x_2)\|^2 (1 + x_2'x_2)^{\delta_0} dx_2 < c_0^{-1}. \qquad \text{(S34)}$$

Consider a constant $a \in (0, 1)$ such that $a + 2c_0^{-1}\sqrt{a} < 1$. Such a constant exists by the mean-value theorem. We consider two cases. In the first case, $\|\beta - \beta^*\|^2 \geq a\|\theta_t - \theta_t^*\|_s^2 := a(\|\beta - \beta^*\|^2 + \|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2)$. Then by (S31) and (S33), we have

$$\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 \geq ac_0 \int (1 + x_2'x_2)^{\delta_0} dx_2 \|\theta_t - \theta_t^*\|_s^2. \qquad \text{(S35)}$$

In the second case, $\|\beta - \beta^*\|^2 < a\|\theta_t - \theta_t^*\|_s^2$. Thus, $\|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2 > (1 - a)\|\theta_t - \theta_t^*\|_s^2$. Then, by (S32) and (S34), we have

$$\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 \geq \|\psi_t - \psi_t^*\|_{0,2,\delta_0}^2 - 2\sqrt{a}\|\theta_t - \theta_t^*\|_s^2 c_0^{-1} > (1 - a - 2c_0^{-1}\sqrt{a})\|\theta_t - \theta_t^*\|_s^2. \qquad \text{(S36)}$$

Therefore, in both cases,

$$\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 \geq \min\{ac_0 \int (1+x_2'x_2)^{\delta_0} dx_2, (1-a-2c_0^{-1}\sqrt{a})\}\|\theta_t - \theta_t^*\|_s^2. \tag{S37}$$

This implies that

$$\sum_{t=1}^{T}\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 \geq \min\{ac_0 \int (1+x_2'x_2)^{\delta_0} dx_2, (1-a-2c_0^{-1}\sqrt{a})\}\|\theta - \theta^*\|_s^2. \tag{S38}$$

Now consider an arbitrarily $\varepsilon > 0$, let $e_1 \in (0, \varepsilon^2 \min\{ac_0 \int (1+x_2'x_2)^{\delta_0} dx_2, (1-a-2c_0^{-1}\sqrt{a})\})$. And let $\delta_\varepsilon = c_1 \overline{\ell}_{e_1}^{-c_2} \left(\min\{ac_0 \int (1+x_2'x_2)^{\delta_0} dx_2, (1-a-2c_0^{-1}\sqrt{a})\}\varepsilon^2 - e_1\right)$. Then $\delta_\varepsilon > 0$ and

$$\bar{\mathcal{L}}_J(\theta|\theta^*) \geq c_1 \overline{\ell}_{e_1}^{-c_2} \left(\sum_{t=1}^{T}\sum_{\ell=1}^{\infty} b_{\ell,J,t}^2 - e_1 \right) \geq \delta_\varepsilon, \tag{S39}$$

where the first inequality holds by (S30) and the second inequality holds by (S38). Since θ and θ^* are arbitrary and δ_ε does not depend on them or J, this verifies Assumption 5(i). □

S2.2 Verification of Assumptions 4-6

Assumption 5(i) has been verified in the previous section. To verify the rest of the conditions in Assumptions 4-6, we introduce the following notation and assumptions. For θ, $\theta^* \in \Theta$, let

$$\|\theta - \theta^*\|_{\mathsf{w}}^2 := \sum_{t=1}^{T} \|\theta_t - \theta_t^*\|_{\mathsf{w}}^2 := \frac{1}{J}\sum_{j=1}^{J}\sum_{t=1}^{T} E\left[\Sigma_{j,J,o}^{-1} E\left[\rho_{jt}(\theta_t|\theta_t^*)|\, Z_{jt}\right]^2\right]. \tag{S40}$$

Note that the definition of $\|\cdot\|_{\mathsf{w}}$ is analogous to the weak norm defined in Equation (14) of Ai and Chen (2003). Similarly, for each t and ψ_t, $\psi_t^* \in \Psi$, define

$$\|\psi_t - \psi_t^*\|_{\mathsf{w}}^2 = \frac{1}{J}\sum_{j=1}^{J}\sum_{t=1}^{T} E\left[\Sigma_{j,J,o}^{-1} E\left[\psi_t(X_{2,jt}) - \psi_t^*(X_{2,jt})|\, Z_{jt}\right]^2\right]. \tag{S41}$$

Let $\psi^0_{J,k_J,t} = \arg\min_{\psi_t \in \Psi_{k_J}} \|\psi_t - \psi^0_{J,t}\|^2_{\mathsf{w}}$.

Let $\bar{\mathcal{V}} = \mathbb{R}^{d_{x_1}} \times \prod_{t=1}^{T} \bar{\mathcal{W}}$. Let $\left(\bar{\mathcal{V}}, \langle \cdot, \cdot \rangle_{\mathsf{w}}\right)$ be a Hilbert space equipped with the following inner product

$$\left\langle \nu^1, \nu^2 \right\rangle_{\mathsf{w}} = \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} E\left[\Sigma^{-1}_{j,J,o} E[X'_{1,jt}\nu^1_\beta + \nu^1_{\psi,t}(X_{2,jt})|Z_{jt}] E[X'_{1,jt}\nu^2_\beta + \nu^2_{\psi,t}(X_{2,jt})|Z_{jt}]\right] \tag{S42}$$

for any $\nu^1 := (\nu^1_\beta, \nu^1_{\psi,1}, \ldots, \nu^1_{\psi,T}), \nu^2 := (\nu^2_\beta, \nu^2_{\psi,1}, \ldots, \nu^2_{\psi,T}) \in \bar{\mathcal{V}}$. For any linear functional $f(\theta) = \lambda'_J \beta$ for a constant vector $\lambda_J \in \mathbb{R}^{d_{x_1}}$ and any $\beta \in \mathcal{B}$, by the Riesz representation theorem, there exists a $\nu^\dagger_J = (\nu^\dagger_{J,\beta}, \nu^\dagger_{J,\psi,1}, \ldots, \nu^\dagger_{J,\psi,T}) \in \bar{\mathcal{V}}$ such that

$$\begin{aligned} f(\theta) - f(\theta^0_J) &= \lambda'_J(\beta - \beta^0) = \left\langle \theta - \theta^0_J, \nu^\dagger_J \right\rangle_{\mathsf{w}} \qquad \text{(S43)} \\ &= \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} E\left[\Sigma^{-1}_{j,J,o} E\left[\rho_{jt}(\theta_t|\theta^0_{J,t})|Z_{jt}\right] E\left[X'_{1,jt}\nu^\dagger_{J,\beta} + \nu^\dagger_{J,\psi,t}(X_{2,jt})|Z_{jt}\right]\right]. \end{aligned}$$

Similar to Ai and Chen (2003) (pp. 1809), we can derive that $\nu^\dagger_{J,\beta} = V_{J,\beta}\lambda_J$, $\nu^\dagger_{J,\psi,t} = -\omega^\dagger_{J,t}\nu^\dagger_{J,\beta}$, where $V_{J,\beta}$ and $\omega^\dagger_{J,t}$ are defined above Assumption 6. Let $\nu^\dagger_{J,t} = (\nu^\dagger_{J,\beta}, \nu^\dagger_{J,\psi,t})$.

Next we define a sieve approximation for $\nu^\dagger_J$. For each ℓ, let the ℓ-th element of $\omega^\dagger_{J,k_J,t}$ be the solution to

$$\min_{\omega_{t,\ell} \in \bar{\mathcal{W}}_{k_J}} \frac{1}{J} \sum_{j=1}^{J} E\left[E\left[X_{1\ell,jt} - \omega_{t,\ell}\left(X_{2,jt}\right)|Z_{jt}\right]^2 / \Sigma_{j,J,o}\right],$$

where $\bar{\mathcal{W}}_{k_J}$ is the closure of the linear span of $\Psi_{k_J} - \psi^0_{J,k_J,t}$, which is the same as the closure of the linear span of Ψ_{k_J} since $\psi^0_{J,k_J,t} \in \Psi_{k_J}$ by definition. Let $\nu^\dagger_{J,k_J,t} = (\nu^\dagger_{J,\beta}, \nu^\dagger_{\psi,k_J,t})$, where $\nu^\dagger_{\psi,k_J,t} = -\omega^\dagger_{J,k_J,t}\nu^\dagger_{J,\beta}$. Let $\nu^\dagger_{J,k_J} = (\nu^\dagger_{J,\beta}, \nu^\dagger_{\psi,k_J,1}, \ldots, \nu^\dagger_{\psi,k_J,T})$. It is easy to see that $\bar{\mathcal{W}}_{k_J} \subseteq \bar{\mathcal{W}}$ because $\Psi_{k_J} \subseteq \Psi$. Similar to the weak norm defined on Ψ above, we can define the weak norm on $\bar{\mathcal{W}}$ as follows: for ω_t, $\omega^*_t \in \bar{\mathcal{W}}$, let

$$\|\omega_t - \omega^*_t\|^2_{\mathsf{w}} = \frac{1}{J} \sum_{j=1}^{J} E\left[\Sigma^{-1}_{j,J,o} \left\|E\left[\omega_t(X_{2,jt}) - \omega^*_t(X_{2,jt})\right| Z_{jt}\right]\right\|^2\right].$$

Let $\Theta_t = \mathcal{B} \times \Psi$ and $\Theta_{k_J,t} = \mathcal{B} \times \Psi_{k_J}$.

Assumption 10. (i) $\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \inf_{\pi_{\varsigma_J}\in\mathbb{R}^{\varsigma_J}} \sum_{j=1}^J (I^{\varsigma_J}(Z_{jt})'\pi_{\varsigma_J} - E[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}])^2 = o_p(J^{3/4})$ *for all* t.

(ii) $\|\psi^0_{J,k_J,t} - \psi^0_{J,t}\|_{\mathsf{w}} = o_p(J^{-1/4})$ *and* $\|\omega^\dagger_{J,k_J,t} - \omega^\dagger_{J,t}\|_{\mathsf{w}} = o(J^{-1/4})$.

For each t, let

$$\Omega_{J,t} := \frac{1}{J}\sum_{j=1}^J E[\Sigma_{j,J,o} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})'].$$

For notational simplicity, define

$$I^{\varsigma_J}(Z_{jt}) = (I_1(Z_{jt}), \ldots, I_s(Z_{jt}), \ldots, I_{\varsigma_J}(Z_{jt}))',$$
$$\mathbb{I}^{\varsigma_J}(Z_{jt}) = \Sigma^{1/2}_{j,J,o} I^{\varsigma_J}(Z_{jt}),\ \mathbb{I}_{z,t} = (\mathbb{I}^{\varsigma_J}(Z_{1t}), \ldots, \mathbb{I}^{\varsigma_J}(Z_{Jt}))'.$$

For any random variable W_{jt}, let $\mathbb{G}_J(W_{jt}) = \frac{1}{\sqrt{J}}\sum_{j=1}^J (W_{jt} - E(W_{jt}))$. For $\theta_t := (\beta, \psi_t)$, $\theta^*_t := (\beta^*, \psi^*_t) \in \Theta_t$, define the following quantities,

$$\rho_{jt}(\theta_t|\theta^*_t) = \xi_{jt} + X'_{1,jt}(\beta^* - \beta) + (\psi^*_t(X_{2,jt}) - \psi_t(X_{2,jt})),$$
$$\hat{\Pi}_{jt}(\theta_t|\theta^*_t) = \mathbb{I}^{\varsigma_J}(Z_{jt})' \left(\mathbb{I}'_{z,t}\mathbb{I}_{z,t}\right)^{-1} \sum_{j'=1}^J \mathbb{I}^{\varsigma_J}(Z_{j't}) \left(\frac{\rho_{j't}(\theta_t|\theta^*_t)}{\Sigma^{1/2}_{j',J,o}}\right),$$
$$\Pi_{jt}(\theta_t|\theta^*_t) = \Sigma^{-1/2}_{j,J,o} E\left[\rho_{jt}(\theta_t|\theta^*_t)|Z_{jt}\right].$$

Given the definitions above, it is easy to calculate the pathwise derivatives:

$$\frac{d\rho_{jt}(\theta_t|\theta^*_t)}{d\theta_t}[\nu_t] = -X'_{1,jt}\nu_\beta - \nu_{\psi,t}X_{2,jt},$$
$$\frac{d\Pi_{jt}(\theta_t|\theta^*_t)}{d\theta_t}[\nu_t] = \Sigma^{-1/2}_{j,J,o} E\left[(-X'_{1,jt}\nu_\beta - \nu_{\psi,t}(X_{2,jt}))|Z_{jt}\right]. \tag{S44}$$

Note that these pathwise derivatives do not depend on either θ_t or θ^*_t.

Assumption 11. (i) *For the* ς_J*-dimensional identity matrix* I_{ς_J}*, we have*

$$\left\| \sum_{t=1}^{T} \left(\frac{1}{J} \sum_{j=1}^{J} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})' - I_{\varsigma_J} \right) \right\| = o_p(J^{-1/4})$$

and for each fixed t*,*

$$\left\| \frac{1}{J} \sum_{j=1}^{J} \left(\xi_{jt}^2 - \Sigma_{j,J,o} \right) I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})' \right\| = o_p(J^{-1/4}),$$

$$\left\| \left(\frac{1}{J} \sum_{j=1}^{J} \Sigma_{j,J,o} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})' \right) - \left(\frac{1}{J} \sum_{j=1}^{J} E\left[\Sigma_{j,J,o} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})' \right] \right) \right\| = o_p(J^{-1/4});$$

(ii) *for each fixed* t*, and for* $\mathcal{N}_{o,J,t} = \{\theta_{k_J,t} \in \Theta_{k_J,t} : \|\theta_{k_J,t} - \theta_{J,t}^0\|_s = o_p(1), \|\theta_{k_J,t} - \theta_{J,t}^0\|_{\mathsf{w}} = o_p(J^{-1/4})\}$*, we have*

$$\sup_{\theta_{k_J,t} \in \mathcal{N}_{o,J,t}} \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left\{ \frac{d\Pi_{jt}(\theta_{J,t}^0 | \theta_{J,t}^0)}{d\theta} \left[\nu_{J,t}^{\dagger} \right] \left[\hat{\Pi}_{jt}(\theta_{k_J,t} | \theta_{J,t}^0) - \hat{\Pi}_{jt}(\theta_{J,t}^0 | \theta_{J,t}^0) - \Pi_{jt}(\theta_{k_J,t} | \theta_{J,t}^0) \right] \right\} = o_p(1),$$

$$\sup_{\theta_{k_J,t} \in \mathcal{N}_{o,J,t}} \mathbb{G}_J \left\{ \frac{d\Pi_{jt}(\theta_{J,t}^0 | \theta_{J,t}^0)}{d\theta} \left[\nu_{J,t}^{\dagger} \right] \Pi_{jt}(\theta_{k_J,t} | \theta_{J,t}^0) \right\} = o_p(1), \text{ and}$$

$$\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left\{ \frac{d\Pi_{jt}(\theta_{J,t}^0 | \theta_{J,t}^0)}{d\theta} \left[\nu_{J,t}^{\dagger} \right] \left[\hat{\Pi}_{jt}(\theta_{J,t}^0 | \theta_{J,t}^0) - \frac{\xi_{jt}}{\Sigma_{j,J,o}^{1/2}} \right] \right\} = o_p(1);$$

(iii) *for each* t*, we have that uniformly over* $\theta_{k_J,t} \in \Theta_{k_J,t}$ *and* $\theta_t \in \Theta_t$*,*

$$\frac{\frac{1}{J} \sum_{j=1}^{J} \left[\Sigma_{j,J,o}^{-1/2} E\left[\rho_{jt}(\theta_{k_J,t} | \theta_t) | Z_{jt} \right] \right]^2}{\|\theta_{k_J,t} - \theta_t\|_{\mathsf{w}}^2} = 1 + o_p(1).$$

(iv) *for a sequence of positive real values* $\{a_J\}_{J=1}^{\infty}$ *such that* $a_J^2 \varsigma_J / \sqrt{J} = o(1)$*, we have*

$$E\left[\max_t \max_{1 \le s \le \varsigma_J} \sup_{\theta_{k_J,t} \in \Theta_{k_J,t}, \theta_t \in \Theta_t} \left(\frac{1}{\sqrt{J}} \sum_{j=1}^{J} I_s(Z_{jt}) \left\{ \rho_{jt}(\theta_{k_J,t} | \theta_t) - E\left[\rho_{jt}(\theta_{k_J,t} | \theta_t) | Z_{jt} \right] \right\} \right)^2 \right] \le a_J^2,$$

and

$$E\left[\max_t \max_{1\le s\le \varsigma_J}\left(\frac{1}{\sqrt{J}}\sum_{j=1}^J I_s(Z_{jt})\left\{\frac{d\rho_{jt}(\theta_t|\theta_t)}{d\theta}\left[\nu^\dagger_{J,t}\right]-E\left[\frac{d\rho_{jt}(\theta_t|\theta_t)}{d\theta}\left[\nu^\dagger_{J,t}\right]|Z_{jt}\right]\right\}\right)^2\right]\le a_J^2,$$

where the θ_t *in the second line is arbitrary since the pathwise derivative depends only on* $\nu^\dagger_{J,t}$*.*

Remark. Assumptions 10(i) and (ii) are about the approximation errors of the basis functions. Assumption 11 is a high-level assumption on the convergence of functions related to Z_{jt}. It can be verified by using empirical process theory if we assume a specific data structure of $\{Z_{jt}\}_{j=1,t=1}^{J,T}$. For instance, when $\{X_{jt}, Z_{jt}\}_{j=1,t=1}^{J,T}$ are independent, Assumption 11(i) and (ii) are verified in Chen and Christensen (2018) and Ai and Chen (2003), respectively; Assumption 11(iii)-(iv) can be verified by Theorem 2.7.11 and Theorem 2.14.9 in van der Vaart and Wellner (1996), respectively.

Proposition 1. *Suppose that Assumptions 1-3, and Assumptions 10-11 hold, that the preliminary estimator* $\tilde\theta_{J,t}$ *satisfies the equation* $\|\hat\Omega_t(\tilde\theta_{J,t})-\frac{1}{J}\sum_{j=1}^J \xi_{jt}^2 I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})'\|=o_p(1)$ *for all* t*, and that the smallest eigenvalue of the matrix* $\frac{1}{J}\sum_{j=1}^J E[D_{j,J,t}(Z_{jt})D_{j,J,t}(Z_{jt})']$ *is bounded away from zero uniformly over* J *for all* t*. Then Assumptions* 4*,* 5(ii)*, and* 6 *hold.*

Proof. Step 1. We first verify Assumption 4. When Assumption 3 is satisfied, we assume that $\tilde I^{\varsigma_J}(z)=I^{\varsigma_J}(z)$ and $E\left[I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})'\right]=I_{\varsigma_J}$ without loss of generality, where I_{ς_J} is an $\varsigma_J\times\varsigma_J$ identity matrix.

For Assumption 4(i), note that for $\bar g_t(\theta^0_{J,k_J,t})=\frac{1}{J}\sum_{j=1}^J g_{jt}(\theta^0_{J,k_J,t})$, by triangular inequality,

$$\left\|\bar g_t(\theta^0_{J,k_J,t})\right\|\le\left\|\bar g_t(\theta^0_{J,t})\right\|+\left\|\bar g_t(\theta^0_{J,k_J,t})-\bar g_t\left(\theta^0_{J,t}\right)\right\|. \tag{S45}$$

For the first term on the RHS of (S45), note that $\left\|\bar g_t(\theta^0_{J,t})\right\|=\|\frac{1}{J}\sum_{j=1}^J \xi_{jt}I^{\varsigma_J}(Z_{jt})\|$. Since

$$\frac{1}{J}\sum_{j=1}^J E[\|I^{\varsigma_J}(Z_{jt})\|^2]=\frac{1}{J}\sum_{j=1}^J tr\left(E[I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})']\right)=tr(I_{\varsigma_J})=\varsigma_J, \tag{S46}$$

we have,

$$E\left[\|\bar{g}_t(\theta^0_{J,t})\|^2\right] \le \frac{C}{J^2}\sum_{j=1}^J E\left[\|I^{\varsigma_J}(Z_{jt})\|^2\right] = C\varsigma_J/J,$$

where the inequality follows from Assumption 2(iv) and the last equality follows from (S46). Then $\frac{1}{J}\sum_{j=1}^J \|\bar{g}_t(\theta^0_{J,t})\| = O_p\left(\sqrt{\varsigma_J/J}\right) = o_p(1)$ by Markov inequality and Assumption 3(ii).

For the second term on the RHS of (S45), we first consider the derivation:

$$\begin{aligned}
&\|\bar{g}_t(\theta^0_{J,k_J,t}) - \bar{g}_t(\theta^0_{J,t})\|^2 \\
&= \sum_{s=1}^{\varsigma_J}\left(\frac{1}{J}\sum_{j=1}^J I_s(Z_{jt})\left(\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})\right)\right)^2 \\
&\le \sum_{s=1}^{\varsigma_J}\left(\frac{1}{J}\sum_{j=1}^J I_s(Z_{jt})(\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt}) - E_X[\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})|Z_{jt}])\right)^2 + \\
&\quad \left\|\frac{1}{J}\sum_{j=1}^J I^{\varsigma_J}(Z_{jt})E_X[\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})|Z_{jt}]\right\|^2. \qquad \text{(S47)}
\end{aligned}$$

The first summand on the RHS is bounded by

$$\begin{aligned}
&\frac{\varsigma_J}{J}\max_{1\le s\le\varsigma_J}\left(\frac{1}{\sqrt{J}}\sum_{j=1}^J I_s(Z_{jt})(\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt}) - E_X[\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})|Z_{jt}])\right)^2 \\
&\le \frac{\varsigma_J}{J}\max_{1\le s\le\varsigma_J}\max_{\theta_t\in\Theta_t,\theta_{k_J,t}\in\Theta_{k_J,t}}\left(\frac{1}{\sqrt{J}}\sum_{j=1}^J I_s(Z_{jt})(\rho_{jt}(\theta_{k_J,t}|\theta_t) - E[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}])\right)^2 \\
&= O_p(\varsigma_J a_J^2 J^{-1}) = o_p(1), \qquad \text{(S48)}
\end{aligned}$$

where the inequality holds because $(\beta^0, \psi^0_{J,t}) \in \Theta_t$ and $(\beta^0, \psi^0_{J,k_J,t}) \in \Theta_{k_J,t}$, the first equality holds by Markov's inequality and Assumption 11(iv), and the second equality holds by Assumption 11(iv). Furthermore, by the Cauchy-Schwarz inequality, the second summand on the RHS of (S47) is bounded by

$$\left(\frac{1}{J}\sum_{j=1}^J \|I^{\varsigma_J}(Z_{jt})\|^2\Sigma_{j,J,o}\right) \times \left(\frac{1}{J}\sum_{j=1}^J E_X[\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})|Z_{jt}]^2\Sigma^{-1}_{j,J,o}\right). \qquad \text{(S49)}$$

By Assumption 2(iv), we have

$$\left(\frac{1}{J}\sum_{j=1}^{J}\|I^{\varsigma_J}(Z_{jt})\|^2\Sigma_{j,J,o}\right) \leq C\left(\frac{1}{J}\sum_{j=1}^{J}\|I^{\varsigma_J}(Z_{jt})\|^2\right) = O_p(\varsigma_J), \tag{S50}$$

where the equality holds by (S46). Moreover, by Assumption 11(iii), we have

$$\frac{1}{J}\sum_{j=1}^{J}E_X[\psi^0_{J,k_J,t}(X_{2,jt}) - \psi^0_{J,t}(X_{2,jt})|Z_{jt}]^2\Sigma^{-1}_{j,J,o} = O_p(\|\psi^0_{J,k_J,t} - \psi^0_{J,t}\|^2_{\mathsf{w}}) = o_p(J^{-1/2}), \tag{S51}$$

where the second equality holds by Assumption 10(ii). Therefore, the expression in (S49) is $o_p(\varsigma_J J^{-1/2})$ which is $o_p(1)$. This and (S47) and (S48) together imply that the second term on the RHS of (S45) is $o_p(1)$. Therefore, both terms on the RHS of (S45) are $o_p(1)$. Thus, Assumption 4(i) is verified.

For Assumption 4 (ii), recall that $\hat{\Omega}_t(\tilde{\theta}_{J,t}) = \frac{1}{J}\sum_{j=1}^{J} g_{jt}(\tilde{\theta}_{J,t})g_{jt}(\tilde{\theta}_{J,t})'$, we first show that for $\Omega_{J,t} = \frac{1}{J}\sum_{j=1}^{J}E[\Sigma_{j,J,o}I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})']$,

$$\|\hat{\Omega}_t\left(\tilde{\theta}_{J,t}\right) - \Omega_{J,t}\| = o_p(1). \tag{S52}$$

By the triangular inequality,

$$\|\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \Omega_{J,t}\| \leq \|\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \frac{1}{J}\sum_{j=1}^{J}\xi^2_{jt}I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})'\| + \|\frac{1}{J}\sum_{j=1}^{J}\xi^2_{jt}I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})' - \Omega_{J,t}\|. \tag{S53}$$

For the first term on the right-hand side of (S53), by the assumption of $\tilde{\theta}_{J,t}$ in the proposition, it is $o_p(1)$. For the second term on the right-hand side of (S53), by Assumption 11(i),

$$\|\frac{1}{J}\sum_{j=1}^{J}\xi^2_{jt}I^{\varsigma_J}(Z_{jt})I^{\varsigma_J}(Z_{jt})' - \Omega_{J,t}\| = o_p(1). \tag{S54}$$

Then (S53) implies that $\|\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \Omega_{J,t}\| = o_p(1)$. Thus, we have

$$\left|eig_{\min}\left(\hat{\Omega}_t\left(\tilde{\theta}_{J,t}\right)\right) - eig_{\min}\left(\Omega_{J,t}\right)\right| \xrightarrow{p} 0, \ \ and \ \ \left|eig_{\max}\left(\hat{\Omega}_t\left(\tilde{\theta}\right)\right) - eig_{\max}\left(\Omega_{J,t}\right)\right| \xrightarrow{p} 0, \quad \text{(S55)}$$

We next show that

$$0 < \inf_{J,t} eig_{min}(\Omega_{J,t}) \leq \sup_{J,t} eig_{\max}(\Omega_{J,t}) < \infty. \quad \text{(S56)}$$

Recall that $\Omega_{J,t} = \frac{1}{J}\sum_{j=1}^{J} E\left[\Sigma_{j,J,o} I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})'\right]$, and $\Sigma_{j,J,o} = E\left[\xi_{jt}^2 | \left\{Z_{jt}\right\}_{j=1}^{J}\right]$. Then

$$\sup_{J,t}\left(eig_{\max}\left(\Omega_{J,t}\right)\right) \leq \sup_{j,J,t}\sup_{\mathbf{z}\in\mathcal{Z}_t}\left[\xi_{jt}^2 | \left\{Z_{j't}\right\}_{j'=1}^{J} = \mathbf{z}\right] \sup_{J,t} eig_{\max}\left(\frac{1}{J}\sum_{j=1}^{J} E\left[I^{\varsigma_J}(Z_{jt}) I^{\varsigma_J}(Z_{jt})'\right]\right) < \infty$$

because of Assumption 2 (iv) and Assumption 3 (i). Similar argument also implies that $\inf_{J,t} eig_{min}(\Omega_{J,t}) > 0$. Equations (S55) and (S56) together imply that $1/C \leq eig_{\min}(\hat{\Omega}_t(\tilde{\theta}_{J,t})) \leq eig_{\max}(\hat{\Omega}_t(\tilde{\theta}_{J,t})) \leq C$ w.p.a.1 for some $C \in (0,\infty)$.

Similarly, Assumption 11(i) implies that

$$\left\|W_t^{2SLS} - I_{\varsigma_J}\right\| = o_p(1).$$

Because the eigenvalues of an identity matrix are bounded away from zero and above, we have for some constant $C' \in (0,\infty)$,

$$1/C' \leq eig_{\min}(W_t^{2SLS}) \leq eig_{\max}(W_t^{2SLS}) \leq C', \quad \text{(S57)}$$

and Assumption 4 (ii) follows immediately.

Step 2. Next, we verify $\sup_{\theta_{k_J}\in\Theta_{k_J}}\left|\tilde{\mathcal{L}}_J(\theta_{k_J}) - \bar{\mathcal{L}}_J(\theta_{k_J}|\theta_J^0)\right| \xrightarrow{p} 0$ in Assumption 5(ii).

Let $\hat{\rho}_t(\theta_t|\theta_t^*) = (\rho_{1t}(\theta_t|\theta_t^*), \ldots, \rho_{1t}(\theta_t|\theta_t^*))'$, and let

$$I_{z,t} = (I^{\varsigma_J}(Z_{1t}), \ldots, I^{\varsigma_J}(Z_{Jt}))'.$$

Note that

$$\tilde{\mathcal{L}}_J(\theta_{k_J}) = \frac{1}{J^2}\sum_{t=1}^{T} \hat{\rho}_t(\theta_{k_J,t}|\theta_{J,t}^0)' I_{z,t} W_t^{2SLS} I_{z,t}' \hat{\rho}_t(\theta_{k_J,t}|\theta_{J,t}^0).$$

For arbitrary $\theta_{k_J,t} \in \Theta_{k_J,t}$ and $\theta_t \in \Theta_t$, let $\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) := \frac{1}{J} I^{\varsigma_J}(Z_{jt})' W_t^{2SLS} I_{z,t}' \hat{\rho}_t(\theta_{k_J,t}|\theta_t)$. Consider the derivation

$$\begin{aligned}
&\left| \frac{1}{J^2} \hat{\rho}_t(\theta_{k_J,t}|\theta_t)' I_{z,t} W_t^{2SLS} I_{z,t}' \hat{\rho}_t(\theta_{k_J,t}|\theta_t) - \frac{1}{J}\sum_{j=1}^{J} E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]^2 \right| \\
=& \left| \frac{1}{J}\sum_{j=1}^{J} \left(\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t)\right)^2 - \frac{1}{J}\sum_{j=1}^{J} E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]^2 \right| \\
\leq& \sqrt{\frac{1}{J}\sum_{j=1}^{J} \left(\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) - E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]\right)^2} \\
&\times \left(\sqrt{\frac{1}{J}\sum_{j=1}^{J} E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]^2} + \sqrt{\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} \check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t)^2} \right).
\end{aligned} \tag{S58}$$

Consider each terms on the RHS of (S58). First, we show that for each t,

$$\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \frac{1}{J}\sum_{j=1}^{J} \left(\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) - E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]\right)^2 = o_p(J^{-1/2}). \tag{S59}$$

Let $\tilde{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) = \frac{1}{J} I^{\varsigma_J}(Z_{jt})' W_t^{2SLS} \sum_{j'=1}^{J} I^{\varsigma_J}(Z_{j't}) E\left[\rho_{j't}(\theta_{k_J,t}|\theta_t)|Z_{j't}\right]$. Then

$$\begin{aligned}
&\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \frac{1}{J}\sum_{j=1}^{J} \left(\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) - \tilde{\Pi}_{jt}(\theta_{k_J,t}|\theta_t\right)^2 \\
\leq& \sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \left(\frac{1}{J}\sum_{j=1}^{J} \left\{\rho_{jt}(\theta_{k_J,t}|\theta_t) - E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]\right\} I^{\varsigma_J}(Z_{jt}) \right)'
\end{aligned}$$

$$\times \left(I_{z,t}I'_{z,t}/J\right)^{-1}\left(\frac{1}{J}\sum_{j=1}^{J}\left\{\rho_{jt}(\theta_t|\theta_t)-E\left[\rho_{jt}(\theta_t|\theta_t)|Z_{jt}\right]\right\}I^{\varsigma_J}(Z_{jt})\right). \tag{S60}$$

The RHS of (S60) is bounded by

$$eig_{\min}\left(I_{z,t}I'_{z,t}/J\right)^{-1}\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\sum_{s=1}^{\varsigma_J}\left(\left|\frac{1}{J}\sum_{j=1}^{J}\{\rho_{jt}(\theta_{k_J,t}|\theta_t)-E[\rho_{jt}(\theta_{k_J,t}|\theta_t))]\}I_s(Z_{jt})\right|\right)^2$$

where $eig_{\min}\left(I_{z,t}I'_{z,t}/J\right)^{-1}=O_p(1)$ by (S57) in Step 1 of the proof. Consider that

$$\begin{aligned}
&\Pr\left(\sum_{s=1}^{\varsigma_J}\left(\max_t\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\sqrt{J}\left|\frac{1}{J}\sum_{j=1}^{J}\left\{\rho_{jt}(\theta_{k_J,t}|\theta_t)-E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]\right\}I_s(Z_{jt})\right|\right)^2>\epsilon\right)\\
&\leq\frac{\varsigma_J}{\sqrt{J}\epsilon}E\left[\max_t\max_{1\leq s\leq\varsigma_J}\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\left(\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\rho_{jt}(\theta_{k_J,t}|\theta_t)-E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]\right\}I_s(Z_{jt})\right)^2\right]\\
&\to 0,
\end{aligned}\tag{S61}$$

where the inequality holds by Markov's inequality and the convergence holds by Assumption 11(iv).

Combining (S60)-(S61), we have

$$\max_{1\leq t\leq T}\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\frac{1}{J}\sum_{j=1}^{J}\left(\check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t)-\tilde{\Pi}_{jt}(\theta_{k_J,t}|\theta_t)\right)^2=o_p(J^{-1/2}). \tag{S62}$$

Furthermore, using the definition of $\tilde{\Pi}_{jt}(\theta_{k_J,t},\theta_t)$ (above equation(S60)), we have

$$\begin{aligned}
&\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\Pi}_{jt}(\theta_{k_J,t}|\theta_t)-E[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}]\right)^2\\
&\qquad=\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t}\sum_{t=1}^{T}\inf_{\pi_{\varsigma_J}\in\mathbb{R}^{\varsigma_J}}\frac{1}{J}\sum_{j=1}^{J}(I^{\varsigma_J}(Z_{jt})'\pi_{\varsigma_J}-E[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}])^2\\
&\qquad=o_p(J^{-1/2}),
\end{aligned}\tag{S63}$$

where the second equality holds by Assumption 10(i). Equations (S62) and (S63) together imply (S59) by triangular inequality.

Next, we show that

$$\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \frac{1}{J}\sum_{j=1}^{J} E\left[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}\right]^2 = O_p(1). \tag{S64}$$

This is because

$$\begin{aligned} E\left[\rho_{jt}\left(\theta_t|\theta_t^*\right)|Z_{jt}\right]^2 =& E\left[\xi_{jt} + X_{1,jt}'\left(\beta^* - \beta\right) + \left(\psi_t^*\left(X_{2,jt}\right) - \psi_t(X_{2,jt})\right)|Z_{jt}\right]^2 \\ \leq& 2E\left[X_{1,jt}'(\beta^* - \beta)|Z_{jt}\right]^2 + 2E\left[\{\psi_t^*(X_{2,jt}) - \psi_t(X_{2,jt})\}|Z_{jt}\right]^2. \end{aligned}$$

Hence, (S64) holds by Assumption 2 (i)-(iii) and (v) (also see a similar argument in the proof of Lemma S2).

Note that (S64) and (S59) together imply that

$$\sup_{\theta_{k_J,t}\in\Theta_{k_J,t},\theta_t\in\Theta_t} \frac{1}{J}\sum_{j=1}^{J} \check{\Pi}_{jt}(\theta_{k_J,t}|\theta_t) = O_p(1). \tag{S65}$$

Now we can come back to (S58) and conclude that

$$\sup_{\theta_{k_J}\in\Theta_{k_J}} \left|\tilde{\mathcal{L}}_J(\theta_{k_J}) - \bar{\mathcal{L}}(\theta_{k_J}|\theta_J^0)\right| \xrightarrow{p} 0. \tag{S66}$$

Step 3.1. We now verify Assumption 6(i). To do so, we use a similar idea to the one in the partially linear IV example of Ai and Chen (2003) and first show that $\|\widehat{\theta}_J - \theta_J^0\|_{\mathsf{w}} = o_p\left(J^{-1/4}\right)$.

First, the condition on the preliminary estimator $\tilde{\theta}_{J,t}$ in the proposition and the second statement of Assumption 11(i) together imply that,

$$\left\|\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right\| = o_p(1). \tag{S67}$$

Therefore, uniformly over $\theta_{k_J,t} \in \Theta_{k_J,t}$,

$$\begin{aligned}
&\left|\hat{\mathcal{L}}_t(\theta_{k_J,t}) - \bar{g}_t(\theta_{k_J,t})'\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\bar{g}_t(\theta_{k_J,t})\right| \\
=&\left|\bar{g}_t(\theta_{k_J,t})'\left\{\hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1} - \left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\right\}\bar{g}_t(\theta_{k_J,t})\right| \\
=&\left|\bar{g}_t(\theta_{k_J,t})'\hat{\Omega}_t(\tilde{\theta}_{J,t})^{-1}\left\{\hat{\Omega}_t(\tilde{\theta}_{J,t}) - \frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right\}\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}'\right)^{-1}\bar{g}_t(\theta_{k_J,t})\right| \\
=&o_p\left(\bar{g}_t(\theta_{k_J,t})'\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\bar{g}_t(\theta_{k_J,t})\right), \qquad \text{(S68)}
\end{aligned}$$

where the last equality follows from Assumption 4(ii) and (S67).

Recall that $\hat{\Pi}_{jt}(\theta_{k_J,t}|\theta_{J,t}^0) = \mathbb{I}^{\varsigma_J}(Z_{jt})'\left(\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\sum_{j'=1}^{J}\frac{\mathbb{I}^{\varsigma_J}(Z_{j't})\rho_{j't}(\theta_{k_J,t}|\theta_{J,t}^0)}{\Sigma_{j',J,o}(Z_{j't})^{1/2}}$. Then

$$\bar{g}_t(\theta_{k_J,t})'\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\bar{g}_t(\theta_{k_J,t}) = \frac{1}{J}\sum_{j=1}^{J}\left[\hat{\Pi}_{jt}(\theta_{k_J,t}|\theta_{J,t}^0)\right]^2. \qquad \text{(S69)}$$

By similar arguments as those for (S58), (S59), (S64), and (S65) (where we replace W_t^{2SLS}, $I^{\varsigma_J}(Z_{jt})$, and $\rho_{jt}(\theta_{k_J,t}|\theta_{J,t}^0)$ with $\frac{1}{J}\sum_{j=1}^{J}\mathbb{I}^{\varsigma_J}(Z_{jt})\mathbb{I}^{\varsigma_J}(Z_{jt})'$, $\mathbb{I}^{\varsigma_J}(Z_{jt})$, $\Sigma_{j,J,o}^{-1/2}\rho_{jt}(\theta_{k_J,t}|\theta_{J,t}^0)$ respectively), one can show that

$$\sup_{\theta_{k_J,t}\in\Theta_{k_J,t}}\left|\frac{1}{J}\sum_{j=1}^{J}\left(\hat{\Pi}_{jt}(\theta_{k_J,t}|\theta_{J,t}^0)\right)^2 - \frac{1}{J}\sum_{j=1}^{J}\Sigma_{j,J,o}^{-1}(E[\rho_{jt}(\theta_{k_J,t}|\theta_t)|Z_{jt}]|_{\theta_t=\theta_{J,t}^0})^2\right| = o_p(J^{-1/2}), \qquad \text{(S70)}$$

which together with (S69) and Assumption 11(iii) imply that

$$\sup_{\theta_{k_J,t}\in\Theta_{k_J,t}}\left|\bar{g}_t(\theta_{k_J,t})'\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\bar{g}_t(\theta_{k_J,t}) - \|\theta_{k_J,t} - \theta_{J,t}^0\|_{\mathsf{w}}^2\right| = o_p(J^{-1/2}). \qquad \text{(S71)}$$

Therefore, we have uniformly over $\theta_{k_J,t} \in \Theta_{k_J,t}$,

$$\hat{\mathcal{L}}_t(\theta_{k_J,t}) = \bar{g}_t(\theta_{k_J,t})'\left(\frac{1}{J}\mathbb{I}_{z,t}'\mathbb{I}_{z,t}\right)^{-1}\bar{g}_t(\theta_{k_J,t}) + o_p(J^{-1/2}) + o_p(\|\theta_{k_J,t} - \theta_{J,t}^0\|_{\mathsf{w}}^2)$$

$$= \|\theta_{k_J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}} + o_p(\|\theta_{k_J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}) + o_p(J^{-1/2}), \tag{S72}$$

where the first equality holds by (S68) and (S71), and the second equality holds by applying (S71) again. In particular, we have

$$\hat{\mathcal{L}}_t(\widehat{\theta}_{J,t}) = \|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}} + o_p(\|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}) + o_p(J^{-1/2}) \tag{S73}$$

$$\hat{\mathcal{L}}_t(\theta^0_{J,k_J,t}) = \|\theta^0_{J,k_J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}} + o_p(\|\theta^0_{J,k_J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}) + o_p(J^{-1/2}) = o_p(J^{-1/2}), \tag{S74}$$

where the second equality in (S74) follows from Assumption 10(ii).

Moreover, by the definition of $\widehat{\theta}_{J,t}$, we have $\hat{\mathcal{L}}_t(\widehat{\theta}_{J,t}) \leq \hat{\mathcal{L}}_t(\theta^0_{J,k_J,t})$, which together with (S73) and (S74) imply that the following derivation goes through:

$$0 \leq \hat{\mathcal{L}}_t(\theta^0_{J,k_J,t}) - \hat{\mathcal{L}}_t(\widehat{\theta}_{J,t}) = o_p(J^{-1/2}) + o_p(\|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}) - \|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}, \tag{S75}$$

which implies that $\|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|_{\mathsf{w}} = o_p(J^{-1/4})$. Thus, $\widehat{\theta}_{J,t} \in \mathcal{N}_{o,J,t}$ for $\mathcal{N}_{o,J,t}$ defined in Assumption 11(ii).

Step 3.2. To continue verifying Assumption 6(i), we first establish a nonparametric first-order condition for the GMM problem. To begin for some $\tau \in [0,1]$, let $\nu^\dagger_{J,t} = (\nu^\dagger_{J,\beta}, \nu^\dagger_{J,\psi,t})$, $\nu^\dagger_{J,k_J,t} = (\nu^\dagger_{J,\beta}, \nu^\dagger_{\psi,k_J,t})$, $u^\dagger_t = \pm\nu^\dagger_{J,t}$, $u^\dagger_{k_J,t} = \pm\nu^\dagger_{J,k_J,t}$, and $u^\dagger_{k_J} = \pm\nu^\dagger_{J,k_J} = \pm(\nu^\dagger_{J,\beta}, \nu^\dagger_{\psi,k_J,1}, \ldots, \nu_{\psi,k_J,T})$. Let $\theta(0) = \widehat{\theta}_J$, $\theta(1) = \widehat{\theta}_J + \epsilon_J u^\dagger_{k_J}$ for $0 < \epsilon_J = o(J^{-1/2})$, and $\theta(\tau) = \widehat{\theta}_J + \tau\epsilon_J u^\dagger_{k_J}$. Since $\widehat{\theta}_J = \arg\min_{\theta\in\Theta_{k_J}} \hat{\mathcal{L}}_J(\theta)$, a second-order Taylor expansion implies that for some $s \in [0,1]$,

$$\begin{aligned} 0 &\leq \hat{\mathcal{L}}_J\left(\widehat{\theta}_J + \epsilon_J u^\dagger_{k_J}\right) - \hat{\mathcal{L}}_J\left(\widehat{\theta}_J\right) = \hat{\mathcal{L}}_J\left(\theta(1)\right) - \hat{\mathcal{L}}_J(\theta(0)) \\ &= \left.\frac{d\hat{\mathcal{L}}_J\left(\theta(\tau)\right)}{d\tau}\right|_{\tau=0} + \frac{1}{2}\left.\frac{d^2\hat{\mathcal{L}}_J(\theta(\tau))}{d\tau^2}\right|_{\tau=s} \\ &= \sum_{t=1}^T \left.\frac{d\hat{\mathcal{L}}_{J,t}\left(\theta_t(\tau)\right)}{d\tau}\right|_{\tau=0} + \frac{1}{2}\sum_{t=1}^T \left.\frac{d^2\hat{\mathcal{L}}_{J,t}(\theta_t(\tau))}{d\tau^2}\right|_{\tau=s} \end{aligned} \tag{S76}$$

For each t, for the first term on the RHS of (S76), by direct calculation, we obtain

$$\left.\frac{d\hat{\mathcal{L}}_{J,t}\left(\theta_t(\tau)\right)}{d\tau}\right|_{\tau=0}$$
$$=2\left(\frac{1}{J}\sum_{j=1}^{J}\frac{d\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}\left[\epsilon_J u^{\dagger}_{k_J,t}\right]I^{\varsigma_J}(Z_{jt})\right)'\hat{\Omega}^{-1}_t(\tilde{\theta}_{J,t})\left(\frac{1}{J}\sum_{j=1}^{J}I^{\varsigma_J}(Z_{jt})\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right).$$

Then by the same arguments as those for (S68), we have

$$\left.\frac{d\hat{\mathcal{L}}_{J,t}\left(\theta_t(\tau)\right)}{d\tau}\right|_{\tau=0}=2\epsilon_J\times(\pm A_{J,t})+o_p(\epsilon_J\times(\pm A_{J,t})), \tag{S77}$$

where

$$A_{J,t}=\left(\frac{1}{J}\sum_{j=1}^{J}\frac{d\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}\left[\nu^{\dagger}_{J,k_J,t}\right]I^{\varsigma_J}(Z_{jt})'\right)\left(\frac{\mathbb{I}'_{z,t}\mathbb{I}_{z,t}}{J}\right)^{-1}\left(\frac{1}{J}\sum_{j=1}^{J}I^{\varsigma_J}(Z_{jt})\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right). \tag{S78}$$

Recall that $\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})=\mathbb{I}^{\varsigma_J}(Z_{jt})'\left(\mathbb{I}'_{z,t}\mathbb{I}_{z,t}\right)^{-1}\sum_{j'=1}^{J}\frac{\mathbb{I}^{\varsigma_J}(Z_{j't})\rho_{j't}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{\Sigma_{j,J,o}(Z_{j't})^{1/2}}$, we can write

$$A_{J,t}=\frac{1}{J}\sum_{j=1}^{J}\frac{d\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}[\nu^{\dagger}_{J,k_J,t}]'\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t}). \tag{S79}$$

Next consider the second-order term on the RHS of (S76), $\left.\frac{d^2\hat{\mathcal{L}}_{J,t}(\theta(\tau))}{d\tau^2}\right|_{\tau=s}$. Due to the linearity of $\rho_{jt}(\theta_t|\theta^0_{J,t})$, we have that $\frac{d\rho_{jt}(\theta_t|\theta^0_{J,t})}{d\theta}[u]$ does not depend on θ_t and that $\frac{d^2\rho_{jt}(\theta_t|\theta^0_{J,t})}{d\theta d\theta}[u]=0$. These imply that for all $s\in[0,1]$,

$$\left.\frac{d^2\hat{\mathcal{L}}_{J,t}(\theta(\tau))}{d\tau}\right|_{\tau=s}=2\left(\frac{1}{J}\sum_{j=1}^{J}\frac{d\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}\left[\epsilon_J u^{\dagger}_{k_J,t}\right]I^{\varsigma_J}(Z_{jt})\right)'\hat{\Omega}^{-1}_t(\tilde{\theta}_{J,t})\times$$
$$\left(\frac{1}{J}\sum_{j=1}^{J}\frac{d\rho_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}\left[\epsilon_J u^{\dagger}_{k_J,t}\right]I^{\varsigma_J}(Z_{jt})\right).$$

Then by similar arguments as those for (S68), we have

$$\left.\frac{d^2\hat{\mathcal{L}}_{J,t}(\theta(\tau))}{d\tau}\right|_{\tau=s} = 2\epsilon_J^2 \times (\pm B_{J,t}) + o_p(\epsilon_J^2 \times (\pm B_{J,t})), \tag{S80}$$

where

$$B_{J,t} = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{d\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}[\nu^{\dagger}_{J,k_J,t}]\right)^2.$$

Similar arguments as those for (S59) (replacing W_t^{2SLS} by $\left(J^{-1}\sum_{j=1}^{J}\mathbb{I}^{\varsigma_J}(Z_{jt})\mathbb{I}^{\varsigma_J}(Z_{jt})'\right)^{-1}$ and $\rho_{jt}(\theta_{k_J,t}|\theta^0_{J,t})$ by $\Sigma_{j,J,o}^{-1/2}\frac{d\rho_{jt}(\theta_{k_J,t}|\theta^0_{J,t})}{d\theta}[\nu^{\dagger}_{J,k_J,t}]$), we can show that for each t, uniformly over $\theta_{k_J,t} \in \mathcal{B} \times \ominus_{k_J}$,

$$\frac{1}{J}\sum_{j=1}^{J}\left(\frac{d\hat{\Pi}_{jt}(\theta_{k_J,t}|\theta^0_{J,t})}{d\theta}[\nu^{\dagger}_{J,k_J,t}] - \Sigma_{j,J,o}^{-1/2}E\left[\frac{d\rho(\theta_{k_J,t}|\theta^0_{J,t})}{d\theta}\left[\nu^{\dagger}_{J,k_J,t}\right]\middle| Z_{jt}\right]\right)^2 = o_p(J^{-1/2}). \tag{S81}$$

Also note that

$$\begin{aligned}
&\frac{1}{J}\sum_{j=1}^{J}E\left(\Sigma_{j,J,o}^{-1}E\left[\frac{d\rho(\theta_{k_J,t}|\theta^0_{J,t})}{d\theta}\left[\nu^{\dagger}_{J,k_J,t}\right]\middle| Z_{jt}\right]^2\right)\\
=&\frac{1}{J}\sum_{j=1}^{J}E\left(\Sigma_{j,J,o}^{-1}E(X'_{1,jt}\nu^{\dagger}_{J,\beta} - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})'\nu^{\dagger}_{J,\beta}|Z_{jt})^2\right)\\
=&\frac{1}{J}\sum_{j=1}^{J}E\left(\Sigma_{j,J,o}^{-1}\left(\lambda'_J V_{J,\beta}E[X_{1,jt} - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})|Z_{jt}]\right)^2\right)\\
\leq&\|\lambda_J\|^2 eig_{\max}(V_{J,\beta})\frac{1}{J}\sum_{j=1}^{J}E\left(\|E[X_{1,jt} - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})|Z_{jt}]\|^2\Sigma_{j,J,o}^{-1}\right)\\
\leq&\|\lambda_J\|^2 eig_{\max}(V_{J,\beta})\frac{1}{J}\sum_{j=1}^{J}E\left(\|E[X_{1,jt}|Z_{jt}]\|^2\Sigma_{j,J,o}^{-1}\right)\\
\leq&\sup_{J,t}(\inf_{\mathbf{z}\in\mathcal{Z}_t}E[\xi_{jt}^2|\{Z_{jt}\}_{j=1}^{J}=\mathbf{z}])^{-1}\|\lambda_J\|^2 eig_{\max}(V_{J,\beta})\frac{1}{J}\sum_{j=1}^{J}E(\|X_{1,jt}\|^2)\\
<&\infty,
\end{aligned}$$

where the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds by the definition of $\omega^{\dagger}_{J,k_J,t}(X_{2,jt})$, and the fact that $0 \in \bar{\mathcal{W}}_{k_J}$, the third inequality holds by Hölder's inequality, and the law of iterated expectations, and the last inequality holds by Assumptions 2(iv)-(v) and the rank condition regarding $\frac{1}{J}\sum_{j=1}^{J} E[D_{j,J,t}(Z_{jt})D_{j,J,t}(Z_{jt})']$ in the proposition. Note that the left-hand side does not depend on $\theta_{k_J,t}$, which implies that the inequalities hold uniformly over $\theta_{k_J,t} \in \Theta_{k_J}$. This and (S81) together imply that

$$B_{J,t} = O_p(1). \tag{S82}$$

Combining (S76), (S77), (S80), and (S82), we obtain

$$0 \leq 2\epsilon_J \times (\pm A_{J,t}) + o_p(\epsilon_J \times (\pm A_{J,t})) + 2\epsilon_J^2 \times O_p(1) + o_p(\epsilon_J^2)$$

Since this holds both when $\pm$ is $+$ and when $\pm$ is $-$ and $\epsilon_J > 0$ and $\epsilon_J = o_p(J^{-1/2})$ by construction, we have

$$A_{J,t} = O_p(\epsilon_J) = o_p(J^{-1/2}). \tag{S83}$$

Now we have established the nonparametric first-order condition.

Next, we write $A_{J,t}$ in terms of $\lambda_J'(\hat{\beta}_J - \beta)$. By (S70) using the fact that $\Pi_{jt}(\theta_t|\theta^0_{J,t}) = \Sigma_{j,J,o}^{-1/2}\left(E\left[\rho_{jt}(\theta_t|\theta_t^*)|Z_{jt}\right]|_{\theta_t^*=\theta^0_{J,t}}\right)$, we have

$$\left|\frac{1}{J}\sum_{j=1}^{J}\left(\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right)^2 - \frac{1}{J}\sum_{j=1}^{J}\left(\Pi_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right)^2\right| = o_p(J^{-1/2}). \tag{S84}$$

Also by Assumptions 11(iii), we have

$$\frac{1}{J}\sum_{j=1}^{J}\left(\Pi_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right)^2 = \|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}} + o_p(\|\widehat{\theta}_{J,t} - \theta^0_{J,t}\|^2_{\mathsf{w}}) = o_p(J^{-1/2}),$$

where the second equality holds by Step 3.1 above. Combining this with (S84), we have

$$\frac{1}{J}\sum_{j=1}^{J}\left(\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\right)^2 = o_p(J^{-1/2}). \tag{S85}$$

This and (S81) together imply that

$$\begin{aligned}
\sqrt{J}A_{J,t} =& \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})}{d\theta}\left[\nu^{\dagger}_{J,k_J,t}\right]\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t}) + o_p(1)\\
=& \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma^{-1}_{j,J,o}E(X'_{1,jt} - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t}) + o_p(1)\\
=& \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma^{-1}_{j,J,o}E(\omega^{\dagger}_{J,t}(X_{2,jt})' - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t})\\
&+ \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma^{-1}_{j,J,o}E(X'_{1,jt} - \omega^{\dagger}_{J,t}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta^0_{J,t}) + o_p(1),
\end{aligned} \tag{S86}$$

where the second equality is derived by writing out the pathwise derivative.

Consider the derivation

$$\begin{aligned}
&\frac{1}{J}\sum_{j=1}^{J}E\left\{\Sigma^{-1}_{j,J,o}E(\omega^{\dagger}_{J,t}(X_{2,jt})' - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda_J\right\}^2\\
\leq&\|\lambda'_J V_{J,\beta}\|^2\|\omega^{\dagger}_{J,t} - \omega^{\dagger}_{J,k_J,t}\|^2_{\mathsf{w}}\\
\leq&\|\lambda_J\|^2 eig_{\max}(V_{J,\beta})\|\omega^{\dagger}_{J,t} - \omega^{\dagger}_{J,k_J,t}\|^2_{\mathsf{w}}\\
=&o(J^{-1/2}),
\end{aligned}$$

where the equality holds by the definition of λ_J (which guarantees that $\|\lambda_J\| = O(1)$) and the rank condition regarding $\frac{1}{J}\sum_{j=1}^{J}E[D_{j,J,t}(Z_{jt})D_{j,J,t}(Z_{jt})']$ in the proposition (which guarantees that $eig_{\max}(V_{J,\beta}) = O(1)$), and by Assumption 10(ii). This implies that

$$\frac{1}{J}\sum_{j=1}^{J}\left\{\Sigma^{-1}_{j,J,o}E(\omega^{\dagger}_{J,t}(X_{2,jt})' - \omega^{\dagger}_{J,k_J,t}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}^2 = o_p(J^{-1/2}).$$

This combined with (S85) imply that

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma_{j,J,o}^{-1}E(\omega_{J,t}^{\dagger}(X_{2,jt})'-\omega_{J,k_J,t}^{\dagger}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0})=o_p(1).$$

Therefore,

$$\begin{aligned}\sqrt{J}A_{J,t}&=\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma_{j,J,o}^{-1}E(X_{1,jt}'-\omega_{J,t}^{\dagger}(X_{2,jt})'|Z_{jt})V_{J,\beta}\lambda\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0})+o_p(1)\\&=\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0})+o_p(1)\end{aligned}\tag{S87}$$

where the second equality holds by the definition of the pathwise derivative. Note that the θ arguments in the pathwise derivative do not matter due to the linearity of ρ_{jt}.

Now write $\sqrt{J}A_{J,t}$ as

$$\begin{aligned}&\sqrt{J}A_{J,t}\\&=\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}(\hat{\Pi}_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0})-\hat{\Pi}_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})-\Pi_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0}))+\\&\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}(\hat{\Pi}_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})-\xi_{jt}/\Sigma_{j,J,o}^{1/2})+\\&\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}\xi_{jt}/\Sigma_{j,J,o}^{1/2}+\\&\quad\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}(\Pi_{jt}(\widehat{\theta}_{J,t}|\theta_{J,t}^{0}))+o_p(1).\end{aligned}\tag{S88}$$

By Assumption 11(ii), the first and the second summand on the right-hand side of (S88) is $o_p(1)$, and the fourth summand, summed up over t, equals

$$\begin{aligned}&\frac{1}{\sqrt{J}}\sum_{t=1}^{T}\sum_{j=1}^{J}E\left[\left\{\frac{d\Pi_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\right\}(\Pi_{jt}(\theta_{J,t}|\theta_{J,t}^{*}))\right]\Bigg|_{\theta_{J,t}=\widehat{\theta}_{J,t},\theta_{J,t}^{*}=\theta_{J,t}^{0}}+o_p(1)\\=&\sqrt{J}\left\langle\nu_{J}^{\dagger},\widehat{\theta}_{J}-\theta_{J}^{0}\right\rangle_{\mathsf{w}}+o_p(1).\end{aligned}\tag{S89}$$

The third summand on the right-hand side of (S88) can be written as

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left\{\Sigma_{j,J,o}^{-1}E\left[\frac{d\rho_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\middle|\, Z_{jt}\right]\right\}\xi_{jt}.$$

Therefore

$$\sqrt{J}\sum_{t=1}^{T}A_{J,t}=\sqrt{J}\left\langle \nu_{J}^{\dagger},\widehat{\theta}_{J}-\theta_{J}^{0}\right\rangle_{\mathsf{w}}+$$
$$\frac{1}{\sqrt{J}}\sum_{t=1}^{T}\sum_{j=1}^{J}\left\{\Sigma_{j,J,o}^{-1}E\left[\frac{d\rho_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]\middle|\, Z_{jt}\right]\right\}\xi_{jt}+o_{p}(1). \tag{S90}$$

Combining this with (S83), we get

$$\begin{aligned}
&\sqrt{J}\lambda_{J}'\left(\hat{\beta}_{J}-\beta^{0}\right)\\
&=\sqrt{J}\left\langle \nu_{J}^{\dagger},\widehat{\theta}_{J}-\theta_{J}^{0}\right\rangle_{\mathsf{w}}\\
&=-\frac{1}{\sqrt{J}}\sum_{t=1}^{T}\sum_{j=1}^{J}E\left[\frac{d\rho_{jt}(\theta_{J,t}^{0}|\theta_{J,t}^{0})}{d\theta}\left[\nu_{J,t}^{\dagger}\right]|Z_{jt}\right]\Sigma_{j,J,o}^{-1}\xi_{jt}+o_{p}(1)\\
&=-\frac{1}{\sqrt{J}}\sum_{t=1}^{T}\sum_{j=1}^{J}\lambda_{J}'V_{J,\beta}E\left[X_{1,jt}-\omega_{J,t}^{\dagger}(X_{2,jt})|Z_{jt}\right]\Sigma_{j,J,o}^{-1}\xi_{jt}+o_{p}(1)\\
&=-\frac{\lambda_{J}'}{\sqrt{J}}V_{J,\beta}\sum_{t=1}^{T}\sum_{j=1}^{J}D_{j,J,t}(Z_{jt})\Sigma_{j,J,o}^{-1}\xi_{jt}+o_{p}(1)
\end{aligned}\tag{S91}$$

where the first equality follows from (S43), the second equality follows from (S89) and (S90), the third equality follows from the two lines below (S43), and the fourth equality follows from the definition of $D_{j,J,t}(Z_{jt})$ (above Assumption 6). We thus have verified Assumption 6(i).

Finally, to verify Assumption 6(ii), since

$$\hat{V}_{J,t}^{-1}=\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\hat{\omega}_{J,t}^{\dagger}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right)'\hat{\Omega}_{J,t}(\hat{\theta}_{t})^{-1}\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\hat{\omega}_{J,t}^{\dagger}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right), \tag{S92}$$

it sufficies to show that $\sum_{t=1}^{T}\hat{V}_{J,t}^{-1}-\sum_{t=1}^{T}V_{J,t}^{-1}\xrightarrow{p}0$.

The argument is similar to the one in Step 3.1 and Step 3.2 so we skip the details below. By using a similar argument in the proof of Theorem 1, one can show that

$$\|\hat{\omega}_{J,t}-\omega_{J,t}^{\dagger}\|_{s}=o_{p}(1) \tag{S93}$$

for each $t=1,\ldots,T$. Then similar to the argument for (S81), we have

$$\begin{aligned}
&\sum_{t=1}^{T}\hat{V}_{J,t}^{-1}\\
&=\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\hat{\omega}_{J,t}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right)'\hat{\Omega}_{J,t}(\hat{\theta}_{t})^{-1}\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\hat{\omega}_{J,t}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right)\\
&=\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\omega_{J,t}^{\dagger}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right)'\hat{\Omega}_{J,t}(\hat{\theta}_{t})^{-1}\left(\frac{1}{J}\sum_{j=1}^{J}\left(X_{1,jt}-\omega_{J,t}^{\dagger}(X_{2,jt})\right)I^{\varsigma_J}(Z_{jt})\right)+o_{p}(1)\\
&=\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J}E\left(\Sigma_{j,J,o}(Z_{jt})^{-1}E\left[X_{1,jt}-\omega_{J,t}^{\dagger}(X_{2,jt})|Z_{jt}\right]\right)^{2}\right)+o_{p}(1)=\sum_{t=1}^{T}V_{J,t}^{-1}+o_{p}(1),
\end{aligned}$$

which implies the condition in Assumption 6(ii). □

S2.3 Verification of Assumption 8(ii)

The following proposition verifies the first condition in Assumption 8(ii). The second condition in Assumption 8 (ii) follows similar arguments.

Proposition 2. *Suppose that $X_{2,jt}$ is independent across j and exogenous. Conditions in Theorem 1 imply that*

$$\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J}\left[\hat{\psi}_{k_J,t}(X_{2,jt})-\psi_{J,t}^{0}(X_{2,jt})\right]^{2}\xrightarrow{p}0.$$

Proof. We first introduce some generic notation. Let $\mathbb{P}_J=J^{-1}\sum_{j=1}^{J}\delta_{X_j}$ of the Dirac measures at the observations, and let Ph denote $J^{-1}\sum_{j=1}^{J}E_X h(X_{2,jt})$ for any measurable function $h(\cdot)$. For a function $h(\cdot) \in \Psi$, let $\|h\|_{L_2(\mathbb{P}_J)} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} h^2(X_j)}$ and $\|h\|_{L_1(\mathbb{P}_J)} = \sqrt{\frac{1}{J}\sum_{j=1}^{J} |h(X_j)|}$. Given a collection $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$, let $\|Q\|_{\mathcal{G}} = \sup\{|Qg| : g \in \mathcal{G}\}$ for signed measure Q. To simplify notation and make the presentation clearer, we assume one single market without loss of generality and omit the index t.

Recall that $\Psi = \{\psi \in \mathscr{W}_{m_0,\delta_0}(\mathcal{X}_2) : \|\psi\|_{m_0,2,\delta_0} \leq B\}$. Theorem 1 implies that we can find a sequence ε_J such that $\varepsilon_J = o_p(1)$ and $\|\hat{\psi}_{k_J} - \psi_J^0\|_{0,2,\delta_0} < \varepsilon_J$. Let

$$\Psi_\varepsilon = \{\Delta\psi_J := \psi_J - \tilde{\psi}_J : \forall \psi_J, \tilde{\psi}_J \in \Psi \text{ such that } \|\psi_J - \tilde{\psi}_J\|_{0,2,\delta_0} < \varepsilon_J\}.$$

Since $\hat{\psi}_{k_J}(\cdot)$ and $\psi_{J,t}^0$ both belong to Ψ with probability 1, it suffices to show that $\|\mathbb{P}_J \Delta\psi_J^2\|_{\Psi_\varepsilon} := \sup_{\Delta\psi_J \in \Psi_\varepsilon} |\mathbb{P}_J \Delta\psi_J^2| = o_p(1)$.

Note that for any $\Delta\psi_J \in \Psi_\varepsilon$, we have

$$\begin{aligned} P\Delta\psi_J^2 = \frac{1}{J}\sum_{j=1}^{J} E_X\left[\Delta\psi_J(X_{2,j})\right]^2 =& \frac{1}{J}\sum_{j=1}^{J}\int \left[\Delta\psi_J(x_2)\right]^2 f_{X_{2,j}}(x_2)dx_2 \\ \leq& \int \left[\Delta\psi_J(x_2)\right]^2 (1 + x_2'x_2)^{\delta_0} dx_2 \\ =& \|\Delta\psi_J\|_{0,2,\delta_0}^2 \leq \varepsilon_J. \end{aligned} \tag{S94}$$

where the first inequality follows from Assumption 2(vi). This implies that $\|P\Delta\psi_J^2\|_{\Psi_\varepsilon} \to 0$ for $\Delta\psi_J \in \Psi_\varepsilon$. It then suffices to show that $\|\mathbb{P}_J - P\|_{\Psi_\varepsilon} = o_p(1)$.

Note that $\Psi_\varepsilon \subset \Psi \cdot \Psi := \{\psi_1\psi_2 : \psi_1 \in \Psi, \psi_2 \in \Psi\}$, it suffices to show that $\|\mathbb{P}_J - P\|_{\Psi\cdot\Psi} := \sup_{h \in \Psi\cdot\Psi} |(\mathbb{P}_J - P)h| = o_p(1)$, which we show next by verifying the conditions for the Glivenko-Cantelli Theorem for triangular arrays in Theorem 22 of Pollard (1990).

Let the notation $a \lesssim b$ denote $a \leq cb$ for some constant $c > 0$ that does not depend on J. To verify $\|\mathbb{P}_J - P\|_{\Psi\cdot\Psi} = o_p(1)$, note that for any $\psi \in \Psi$, $\|\psi\|_{m_0,2,\delta_0} \leq B$, combining with Assumption 2(vi) implies the existence of an envelope function $\bar{\psi}$ for the class Ψ such that $\|\bar{\psi}\|_{0,2,\delta_0} < \infty$. Then the class $\Psi \cdot \Psi$ has $\|\cdot\|_{0,2,\delta_0}-$ integrable envelope $(2\bar{\psi})^2$.

Let $N(\epsilon, \mathcal{G}, \|\cdot\|)$ denote the covering number of $\mathcal{G}$ under $\|\cdot\|$, that is the smallest number of balls of radius ϵ to cover $\mathcal{G}$. Lemma A.3 in Santos (2012) shows that if $\mathcal{X}_2$ is unbounded, $\log N(\epsilon, \Psi, \|\cdot\|_\infty) \lesssim (1/\epsilon)^{(m_0+\delta_0)d_{x_2}/(\delta_0 m_0)}$ while when $\mathcal{X}_2$ is bounded, $\log N(\epsilon, \Psi, \|\cdot\|_\infty) \lesssim (1/\epsilon)^{d_{x_2}/m_0}$. Thus, if $\mathcal{X}_2$ is unbounded, Theorem 2.7.11 in van der Vaart and Wellner (1996) implies that

$$N(\epsilon\|\bar{\psi}\|_{L_2(\mathbb{P}_J)}, \Psi, \|\cdot\|_{L_2(\mathbb{P}_J)}) \leq N_{[]}(2\epsilon\|\bar{\psi}\|_{L_2(\mathbb{P}_J)}, \Psi, \|\cdot\|_{L_2(\mathbb{P}_J)}) \leq N(\epsilon, \Psi, \|\cdot\|_\infty) \lesssim (1/\epsilon)^{(m_0+\delta_0)d_{x_2}/(\delta_0 m_0)},$$

where $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the bracketing number of size ϵ under $\|\cdot\|$. Similarly, if $\mathcal{X}_2$ is bounded

$$N(\epsilon\|\bar{\psi}\|_{L_2(\mathbb{P}_J)}, \Psi, \|\cdot\|_{L_2(\mathbb{P}_J)}) \lesssim (1/\epsilon)^{d_{x_2}/m_0}.$$

Therefore for ϵ sufficiently small,

$$\log N(\epsilon\|\bar{\psi}\|^2_{L_1(\mathbb{P}_J)}, \Psi \cdot \Psi, \|\cdot\|_{L_1(\mathbb{P}_J)}) \leq N(\epsilon\|\bar{\psi}\|_{L_2(\mathbb{P}_J)}, \Psi, \|\cdot\|_{L_2(\mathbb{P}_J)}) = o(J),$$

where the inequality follows from Cauchy-Schwartz inequality and the equality follows from Assumption 2(iii). Then Theorem 22 in Pollard (1990) (Glivenko-Cantelli Theorem for triangular array) implies that $\|\mathbb{P}_J - P\|_{\Psi\cdot\Psi} = o_p(1)$. The claimed result follows.

□

S3 Cross-Product Elasticity in Random Coefficients Logit

What happens to cross-product elasticity in the random coefficients logit model when the number of products grows to infinity? The answer may depend on the distribution of the random coefficients. We show below that if the random coefficients have Gaussian tails, cross-product elasticity converges to zero at the rate of J^{-1}. On the other hand, if the random coefficients have an exponential or thicker tail, cross-product elasticity may stay positive in the limit as illustrated in Section 2. To illustrate the main idea, we focus on one single market by dropping the index t for simplicity.

S3.1 Cross-product elasticity in Gaussian random coefficients logit model

Consider a simple random coefficient logit model with one random coefficient:

$$U_j = X_j\beta + vW_j + \varepsilon_j, \tag{S95}$$

where $\varepsilon_j : j = 0, \ldots, J$ are independent type-I extreme value distributed random variables, and v is a random coefficient with density f. Then the choice probability is

$$s_j = \int_R \frac{\exp(X_j\beta + vW_j)}{1 + \sum_{j'=1}^{J} \exp(X_{j'}\beta + vW_{j'})} f(v)dv \tag{S96}$$

For j and k such that $j \neq k$, we have (let $\delta_k = X_k\beta$)

$$\begin{aligned} \frac{\partial s_j}{\partial \delta_k} &= \int_R \frac{\exp(X_j\beta + vW_j)\exp(X_k\beta + vW_k)}{(1 + \sum_{j'} \exp(X_{j'}\beta + vW_{j'}))^2} f(v)dv \\ \frac{\partial s_j}{\partial W_k} &= \int_R \frac{\exp(X_j\beta + vW_j)\exp(X_k\beta + vW_k)v}{(1 + \sum_{j'} \exp(X_{j'}\beta + vW_{j'}))^2} f(v)dv \end{aligned} \tag{S97}$$

Lemma S3. *Suppose that* $\left(J^{-1}\sum_{j=1}^{J}\exp(X_j\beta)\right)^{-1} = O_p(1)$ *as* $J \to \infty$, *and that* W_j *is a bounded random variable taking values in the interval* $[m, M]$.

(i) *If* $\int \exp(2(M-m)v)f(v)dv < \infty$, *then* $\partial s_j/\partial \delta_k = O_p(J^{-2})$.

(ii) *If* $\int \exp(2(M-m)v)vf(v)dv < \infty$, *then* $\partial s_j/\partial W_k = O_p(J^{-2})$.

Proof. We show part (i) only since part (ii) is analogous. To begin, consider:

$$J^2\frac{\partial s_j}{\partial \delta_k} = \exp(X_j\beta + X_k\beta)\int_R \frac{\exp(v(W_j + W_k))}{(J^{-1} + J^{-1}\sum_{j'=1}^{J}\exp(X_{j'}\beta + vW_{j'}))^2} f(v)dv$$

$$ \tag{S98} $$

The integral equals

$$
\begin{aligned}
&\int_0^\infty \frac{\exp(v(W_j + W_k))}{(J^{-1} + J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta + vW_{j'}))^2} f(v)dv \\
&+ \int_{-\infty}^0 \frac{\exp(v(W_j + W_k))}{(J^{-1} + J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta + vW_{j'}))^2} f(v)dv \\
&\leq \int_0^\infty \frac{\exp(2(M-m)v)}{(J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta))^2} f(v)dv + \int_{-\infty}^0 \frac{\exp(2(M-m)v)}{(J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta))^2} f(v)dv \\
&= \frac{\int_0^\infty \exp(2(M-m)v)f(v)dv}{(J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta))^2} + \frac{\int_{-\infty}^0 f(v)dv}{(J^{-1}\sum_{j'=1}^J \exp(X_{j'}\beta))^2} = O_p(1).
\end{aligned} \tag{S99}
$$

□

Note that s_k converges to zero at the rate of $1/J$. Thus, the cross-product elasticity must converge to zero as $J \to \infty$. Similar derivation holds when W_j is a vector. The condition $\int \exp(2(M-m)v)f(v)dv < \infty$ holds if v is normal. It is violated if v has the same tail decay rate as the logistic distribution which is exponential. There is an intuitive explanation for this: In the random coefficients logit model, the logit error represents product-based preference while the random coefficients represent the characteristic-based preference. The logit error tends to spread cross-product elasticity around, while the random coefficients generate more localized substitution. Intuitively, as the number of products gets large, substitution should localize among products with similar characteristics, rather than spread thin uniformly across all products. To generate this pattern in the model, we need the characteristic-based preference to be at least as strong (dispersed) as the product-based preference. When the distribution of random coefficients has a thinner tail than the logit error, this intuitive pattern of localized substitution cannot be generated.

In general, characterizing the limit of cross-product elasticity when the integrability condition above does not hold is difficult because dominated convergence can no longer be used. In some special models, however, this can be done, because a closed-form solution for the share exists, as we illustrate below.

S3.2 Non-Gaussian Random Coefficient–An Example

Consider a slight variation of the model above:

$$U_j = X_j\beta + v_1 D_j + v_0(1 - D_j) + \varepsilon_j, \tag{S100}$$

where D_j is a dummy variable, $\varepsilon_j : j = 0, \ldots, J$ are independent type-I extreme value distributed random variables. Let $\lambda \in (0, 1]$ be a parameter, and let λv_0 and λv_1 be independent with each other and both be jointly independent with the ε_j's, and each follow the $C(\lambda)$ distribution in Cardell (1997): $f_\lambda(v) = \lambda^{-1}\sum_{n=1}^{\infty}\frac{(-1)^n e^{-nv}}{n!\Gamma(-\lambda n)}$. Then, this random coefficient logit model, according to Cardell (1997), is a nested logit model with two nests, where all j's with $D_j = 0$ are in one nest, and the rest of the products are in the other nest. Then it follows that when for j such that $D_j = 1$:

$$s_j = \frac{\exp(X_j\beta)}{\sum_{j'=1}^{J} D_{j'}\exp(X_{j'}\beta)} \frac{\left(\sum_{j'=1}^{J} D_{j'}\exp(X_{j'}\beta)\right)^{\lambda}}{\left(\sum_{j'=1}^{J} D_j\exp(X_{j'}\beta)\right)^{\lambda} + \left(\sum_{j'=1}^{J}(1 - D_j)\exp(X_{j'}\beta)\right)^{\lambda}} \tag{S101}$$

It can be derived that, for k such that $D_k = D_j = 1$,

$$\partial s_j/\partial\delta_k = -(1-\lambda)s_j\frac{s_k}{\sum_{j':D_{j'}=1} s_{j'}} - \lambda s_j s_k, \tag{S102}$$

and the cross-product elasticity is

$$\frac{\delta_k}{s_j}\frac{\partial s_j}{\partial\delta_k} = -\lambda\delta_k s_k - (1-\lambda)\delta_k\frac{s_k}{\sum_{j':D_{j'}=1} s_{j'}}. \tag{S103}$$

(Recall that $\delta_k = X_k\beta$). This cross-product elasticity does not converge to zero as $J \to \infty$ as long as the relative share of k within its nest stays positive in the limit. The latter happens if the nest size stays constant as $J \to \infty$.

S4 Additional Monte Carlo Results

In this section, we collect additional Monte Carlo results. These complement those reported in Section 5.3.

S4.1 Additional Tables on The Single Random Coefficient Specification

In this subsection, we collect the additional tables, Tables 13, 14, 15, 16, and 17, that were mentioned in Section 5.3 in the main text.

Table 13: Monte Carlo Results: Fixed Coefficients

		$T = 10$				20			40		
Parameter	Estimator	J	25	50	100	25	50	100	25	50	100
β	SN	RtMSE	.0249	.0169	.0116	.0176	.0119	.0080	.0125	.0085	.0059
		Bias	-.0024	5.90E-4	.0013	-.0019	7.44E-4	.0018	-.0012	.0014	.0015
	BLP	RtMSE	.0218	.0156	.0106	.0176	.0117	.0077	.0143	.0091	.0061
		Bias	-.0100	-.0061	-.0037	-.0104	-.0062	-.0034	-.0104	-.0059	-.0037
α	SN	RtMSE	.0330	.0285	.0234	.0234	.0215	.0167	.0180	.0155	.0120
		Bias	.0080	9.30E-4	4.48E-4	.0075	1.14E-4	-4.34E-4	.0072	1.28E-4	5.97E-4
	BLP	RtMSE	.0260	.0228	.0184	.0236	.0196	.0167	.0184	.0172	.0160
		Bias	.0100	.0141	.0132	.0159	.0150	.0138	.0138	.0146	.0146

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$.
2. The distribution of random coefficient for the BLP estimator is misspecified.

Table 14: Monte Carlo Results: Inference on Fixed Coefficients

		$T = 10$			20			40		
Parameter	J	25	50	100	25	50	100	25	50	100
β	True S.D.	.0248	.0169	.0115	.0175	.0119	.0078	.0125	.0084	.0058
	Ave. S.E.	.0191	.0152	.0116	.0140	.0111	.0084	.0100	.0080	.0061
α	True S.D.	.0321	.0285	.0234	.0222	.0215	.0167	.0165	.0155	.0119
	Ave. S.E.	.0254	.0268	.0222	.0181	.0192	.0158	.0129	.0137	.0113

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $.5 \times N(-1, .2^2) + .5 \times N(-2, .5^2)$.
2. The distribution of random coefficient for the BLP estimator is mis-specified.

S4.2 Additional Table on Multiple Independent Random Coefficients

In this subsection, Table 18 shows the Monte Carlo results for the fixed coefficients in the case of 3 independent random coefficients described in Section 5 in the main text.

Table 15: Monte Carlo Results: Mean of Random Coefficients

		$T=10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-I-MD	RtMSE	.0396	.0312	.0194	.0313	.0241	.0171	.0251	.0183	.0182
	Bias	-.0083	-.0126	-.0083	-.0123	-.0116	-.0116	-.0131	-.0112	-.0151
SN-I-GMM1	RtMSE	.0335	.0260	.0194	.0263	.0207	.0181	.0212	.0172	.0189
	Bias	-.0046	-.0109	-.0102	-.0110	-.0107	-.0144	-.0108	-.0129	-.0165
SN-I-GMM2	RtMSE	.0337	.0236	.0180	.0244	.0189	.0162	.0192	.0150	.0172
	Bias	-.0036	-.0068	-.0086	-.0075	-.0073	-.0117	-.0071	-.0094	-.0146
SN-II-MD	RtMSE	.0328	.0293	.0194	.0308	.0224	.0180	.0238	.0167	.0524
	Bias	.0038	-.0064	-.0033	-.0060	-.0094	-.0075	-.0084	-.0097	.0045
SN-II-GMM1	RtMSE	.0288	.0235	.0182	.0254	.0366	.0401	.0268	.0148	.0353
	Bias	.0036	-.0030	-.0023	-.0018	-.0037	-.0126	-.0023	-.0064	.0067
SN-II-GMM2	RtMSE	.0288	.0229	.0173	.0249	.0280	.0172	.0393	.0372	.0423
	Bias	.0063	-3.72E-4	-9.09E-5	.0012	-.0031	-.0066	.0115	.0096	.0197
SN-III-MD	RtMSE	.0459	.0297	.0196	.0305	.0232	.0159	.0252	.0178	.0154
	Bias	-.0076	-.0091	-.0076	-.0031	-.0094	-.0094	-.0082	-.0086	-.0115
SN-III-GMM1	RtMSE	.0386	.0249	.0199	.0256	.0205	.0170	.0216	.0162	.0161
	Bias	-.0070	-.0068	-.0091	-.0013	-.0091	-.0112	-.0065	-.0097	-.0126
SN-III-GMM2	RtMSE	.0379	.0240	.0186	.0255	.0189	.0153	.0204	.0144	.0143
	Bias	-.0035	-.0033	-.0069	.0015	-.0061	-.0087	-.0039	-.0067	-.0103
SN-Para-MD	RtMSE	.0313	.0174	.0113	.0248	.0122	.0073	.0207	.0097	.0053
	Bias	.0242	.0099	.0011	.0206	.0071	7.16E-4	.0182	.0071	6.34E-4
SN-Para-GMM1	RtMSE	.0303	.0168	.0111	.0239	.0134	.0072	.0235	.0116	.0054
	Bias	.0238	.0101	.0017	.0202	.0097	.0029	.0217	.0099	.0026
SN-Para-GMM2	RtMSE	.0318	.0177	.0105	.0263	.0127	.0075	.0214	.0098	.0054
	Bias	.0267	.0126	.0036	.0234	.0068	2.35E-5	.0187	.0068	-1.73E-4
BLP	RtMSE	.0330	.0192	.0118	.0277	.0150	.0092	.0249	.0136	.0080
	Bias	.0280	.0147	.0072	.0249	.0118	.0068	.0232	.0122	.0066

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, F^0 is $N(-2,.5)$.
2. The distribution of random coefficient for the BLP estimator and SN-Para estimators is correctly specified.

S4.3 Correlated Random Coefficients

Now we turn to a simulation design with two correlated random coefficients. Specifically, we add one exogenous characteristics $X_{2,jt}$ to the data generating process in Section 5.1 and let the random coefficients of P_{jt} and $X_{2,jt}$ be correlated. So the market shares can be written as

$$\sigma_j\left(\delta_t, P_t, X_{2,t}; F\right) = \int \frac{\exp\left(\delta_{jt} + \upsilon_1 P_{jt} + \upsilon_2 X_{2,jt}\right)}{1+\sum_{k=1}^{J}\exp\left(\delta_{kt} + \upsilon_1 P_{kt} + \upsilon_2 X_{2,kt}\right)} dF^0\left(\upsilon_1, \upsilon_2\right),$$

where P_{jt} is generated as in Section 5.1, $X_{2,jt}$ is drawn from a standard normal distribution, $F^0\left(\upsilon_1,\upsilon_2\right)$ is a bivariate normal distribution with means (μ_1,μ_2), standard deviations (σ_1,σ_2) and correlation coefficient ρ.

The parametric estimation of F^0, including BLP and SN with parametric F^0, is straightforward.

Table 16: Monte Carlo Results: Standard Deviation of Random Coefficient

		$T = 10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-I-MD	RtMSE	.0396	.0312	.0194	.0313	.0241	.0171	.0251	.0183	.0182
	Bias	-.0083	-.0126	-.0083	-.0123	-.0116	-.0116	-.0131	-.0112	-.0151
SN-I-GMM1	RtMSE	.0335	.0260	.0194	.0263	.0207	.0181	.0212	.0172	.0189
	Bias	-.0046	-.0109	-.0102	-.0110	-.0107	-.0144	-.0108	-.0129	-.0165
SN-I-GMM2	RtMSE	.0337	.0236	.0180	.0244	.0189	.0162	.0192	.0150	.0172
	Bias	-.0036	-.0068	-.0086	-.0075	-.0073	-.0117	-.0071	-.0094	-.0146
SN-II-MD	RtMSE	.0328	.0293	.0194	.0308	.0224	.0180	.0238	.0167	.0524
	Bias	.0038	-.0064	-.0033	-.0060	-.0094	-.0075	-.0084	-.0097	.0045
SN-II-GMM1	RtMSE	.0288	.0235	.0182	.0254	.0366	.0401	.0268	.0148	.0353
	Bias	.0036	-.0030	-.0023	-.0018	-.0037	-.0126	-.0023	-.0064	.0067
SN-II-GMM2	RtMSE	.0288	.0229	.0173	.0249	.0280	.0172	.0393	.0372	.0423
	Bias	.0063	-3.72E-4	-9.09E-5	.0012	-.0031	-.0066	.0115	.0096	.0197
SN-III-MD	RtMSE	.0459	.0297	.0196	.0305	.0232	.0159	.0252	.0178	.0154
	Bias	-.0076	-.0091	-.0076	-.0031	-.0094	-.0094	-.0082	-.0086	-.0115
SN-III-GMM1	RtMSE	.0386	.0249	.0199	.0256	.0205	.0170	.0216	.0162	.0161
	Bias	-.0070	-.0068	-.0091	-.0013	-.0091	-.0112	-.0065	-.0097	-.0126
SN-III-GMM2	RtMSE	.0379	.0240	.0186	.0255	.0189	.0153	.0204	.0144	.0143
	Bias	-.0035	-.0033	-.0069	.0015	-.0061	-.0087	-.0039	-.0067	-.0103
SN-Para-MD	RtMSE	.0313	.0174	.0113	.0248	.0122	.0073	.0207	.0097	.0053
	Bias	.0242	.0099	.0011	.0206	.0071	7.16E-4	.0182	.0071	6.34E-4
SN-Para-GMM1	RtMSE	.0303	.0168	.0111	.0239	.0134	.0072	.0235	.0116	.0054
	Bias	.0238	.0101	.0017	.0202	.0097	.0029	.0217	.0099	.0026
SN-Para-GMM2	RtMSE	.0318	.0177	.0105	.0263	.0127	.0075	.0214	.0098	.0054
	Bias	.0267	.0126	.0036	.0234	.0068	2.35E-5	.0187	.0068	-1.73E-4
BLP	RtMSE	.0330	.0192	.0118	.0277	.0150	.0092	.0249	.0136	.0080
	Bias	.0280	.0147	.0072	.0249	.0118	.0068	.0232	.0122	.0066

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, F^0 is $N(-2, .5)$.
2. The distribution of random coefficient for the BLP estimator and SN-Para estimators is correctly specified.

But the sieve approximation of F^0 needs to be adjusted to account for the dependence between the two random coefficients. In particular, we consider two alternative sieve approximations of F^0.

The first one is a Gaussian copula with non-parametric marginals, i.e.,

$$F^0_{M_J}(\upsilon_1, \upsilon_2) = C\left(F_{1,M_J}(\upsilon_1), F_{2,M_J}(\upsilon_2), \rho\right),$$

where ρ is the correlation coefficient in the Gaussian copula $C(\cdot)$, F_{1,M_J}, and F_{2,M_J} are sieve approximations of the two marginal distributions. To obtain random draws from this distribution, we first draw (u_1, u_2) from copula $C(\cdot, \cdot, \rho)$ for a given ρ, and then transform them using $(\upsilon_1, \upsilon_2) = \left(F^{-1}_{1,M_J}(u_1), F^{-1}_{2,M_J}(u_2)\right)$, where F^{-1}_{l,M_J} $(l = 1, 2)$ is approximated by

Table 17: Monte Carlo Results: Skewness of Random Coefficients

		$T=10$			20			40		
Estimator	J	25	50	100	25	50	100	25	50	100
SN-I-MD	RtMSE	.5300	.4607	.2715	.5104	.4034	.2384	.5119	.3867	.2070
	Bias	.2936	.2891	.0717	.3906	.3011	.1115	.4513	.3387	.1220
SN-I-GMM1	RtMSE	.4445	.3756	.2612	.4257	.3573	.2759	.4344	.3657	.2619
	Bias	.2589	.2549	.1296	.3294	.2825	.2093	.3803	.3306	.2217
SN-I-GMM2	RtMSE	.4461	.3773	.2600	.4300	.3541	.2692	.4308	.3571	.2602
	Bias	.2704	.2357	.1390	.3306	.2788	.2030	.3781	.3215	.2191
SN-II-MD	RtMSE	.5167	.4539	.2703	.4829	.3686	.2121	.4433	.3157	.1751
	Bias	.3812	.3048	.1057	.3841	.2812	.0851	.3871	.2708	.0909
SN-II-GMM1	RtMSE	.4464	.3552	.2417	.4000	.3005	.1896	.3534	.2537	.1746
	Bias	.2465	.2521	.1416	.3196	.2345	.1177	.3038	.2202	.1282
SN-II-GMM2	RtMSE	.4433	.3498	.2330	.4028	.2990	.1824	.3573	.2364	.1600
	Bias	.3395	.2393	.1356	.3233	.2339	.1034	.3102	.1954	.1019
SN-III-MD	RtMSE	.5487	.4525	.3424	.4883	.3762	.3203	.4850	.3068	.2919
	Bias	.3376	.2319	-.0310	.3415	.2398	-.0610	.4047	.2169	-.0748
SN-III-GMM1	RtMSE	.4553	.3587	.2764	.3922	.3168	.2243	.3894	.2856	.1964
	Bias	.2738	.1772	.0789	.2704	.2158	.0903	.3171	.2202	.0951
SN-III-GMM2	RtMSE	.4567	.3507	.2675	.3986	.3124	.2183	.3881	.2795	.1926
	Bias	.2787	.1609	.0794	.2820	.2066	.0840	.3192	.2111	.0970
SN-Para, BLP	RtMSE		-			-			-	
	Bias		.6429			.6429			.6429	

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, F^0 is $.5\times N(-1,.2^2)+.5\times N(-2,.5^2)$.
2. The distribution of random coefficient for the BLP and SN-Para estimators is misspecified.

Table 18: Monte Carlo Results with Independent Random Coefficients: Fixed Coefficients

			$T=10$		20	
Parameter	Estimator		J =50	100	50	100
β	BLP	RtMSE	.0212	.0127	.0190	.0105
		Bias	-.0161	-.0082	-.0165	-.0081
	SN	RtMSE	.0738	.0169	.0751	.0151
		Bias	-.0017	.0098	-.0031	-.0059
α	BLP	RtMSE	.0281	.0253	.0199	.0186
		Bias	.0217	-.0047	.0242	.0111
	SN	RtMSE	.5067	.1127	.5535	.0860
		Bias	-.0079	.0072	-.0098	.0037

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, $\mu_1=-2$, $\mu_2=\mu_3=1$, $\sigma_1=\sigma_2=\sigma_3=.5$.
2. The distribution of random coefficients for the BLP estimator is correctly specified.

Fosgerau and Mabit (2013)'s approach as mentioned in Section 5.2. This approximation of F^0 is "semi-nonparametric" because the copula is parametric and marginals are nonparametric, see Chen et al. (2006) for more discussions on this approximation strategy.

The second one is a multivariate extension of the first sieve approximation in Section 5.2 (see Fosgerau and Mabit (2013)), i.e., we draw two uniformly distributed random variables, u_1 and u_2, and transform them using power polynomials

$$\upsilon_l = \sum_{k_1,k_2 \geq 0, k_1+k_2 \leq M_J} b_{l,k_1,k_2} u_1^{k_1} u_2^{k_2},\ l = 1, 2,$$

where b_{l,k_1,k_2}'s are unknown coefficients to be estimated. Here υ_1 and υ_2 are dependent because they share common polynomial terms of u_1 and u_2 (with different coefficients though).

Note that, in both cases, (υ_1, υ_2) is a draw from the unknown, bivariate distribution $F^0_{M_J}$. Once we obtain these random draws from $F^0_{M_J}$, the market shares can be calculated in the same way as in (29). The results for fixed coefficients are similar to the previous cases, as shown in Table 20. Table 19 shows the Monte Carlo results for random coefficients. In the table, we label estimators based on the above two sieve approximations as "SN-IV-GMM2" and "SN-V-GMM2" (as before "GMM2" refers to the estimator defined by (16)), respectively.[24] The results show that: 1) again the SN estimator with a parametric F achieves a very similar performance to the parametric BLP estimator; 2) compared with the parametric estimators (with correct specification), the two alternative sieve approximations both works quite well for the key parameters in the distribution, i.e., mean, standard deviations and correlation coefficients of the two random coefficients.

[24] We set $M_J = 3$ for all the cases shown in Table 19.

Table 19: Monte Carlo Results: Correlated Random Coefficients

			$T=10$		20					$T=10$		20	
Parameter	Estimator		$J=50$	100	50	100	Parameter	Estimator		$J=50$	100	50	100
μ_1	BLP	RtMSE	.0193	.0119	.0176	.0093	σ_1	BLP	RtMSE	.0228	.0193	.0152	.0140
		Bias	.0147	.0066	.0150	.0066			Bias	.0024	.0040	.0033	.0042
	SN-Para-GMM2	RtMSE	.0184	.0113	.0167	.0078		SN-Para-GMM2	RtMSE	.0316	.0209	.0276	.0154
		Bias	.0120	.0036	.0127	.0037			Bias	.0015	.0073	9.26E-5	.0074
	SN-IV-GMM2	RtMSE	.0269	.0186	.0214	.0144		SN-IV-GMM2	RtMSE	.0511	.0453	.0437	.0396
		Bias	-.0047	-.0040	-.0045	-.0038			Bias	.0050	.0104	-.0047	.0057
	SN-V-GMM2	RtMSE	.0262	.0187	.0209	.0139		SN-V-GMM2	RtMSE	.0489	.0330	.0424	.0252
		Bias	-.0077	-.0072	-.0078	-.0073			Bias	-.0226	-.0090	-.0253	-.0100
μ_2	BLP	RtMSE	.0128	.0094	.0092	.0064	σ_2	BLP	RtMSE	.0221	.0177	.0155	.0122
		Bias	-8.00E-4	2.09E-4	1.31E-4	1.91E-4			Bias	8.26E-4	.0023	.0018	.0026
	SN-Para-GMM2	RtMSE	.0135	.0095	.0099	.0065		SN-Para-GMM2	RtMSE	.0282	.0183	.0227	.0129
		Bias	-.0011	2.26E-5	-3.12E-5	1.25E-5			Bias	-8.90E-4	.0043	-4.79E-4	.0045
	SN-IV-GMM2	RtMSE	.0276	.0183	.0213	.0136		SN-IV-GMM2	RtMSE	.0483	.0449	.0417	.0419
		Bias	-.0114	-.0048	-.0099	-.0058			Bias	.0118	.0150	.0040	.0136
	SN-V-GMM2	RtMSE	.0254	.0167	.0188	.0128		SN-V-GMM2	RtMSE	.0461	.0332	.0379	.0246
		Bias	-.0081	-.0031	-.0060	-.0040			Bias	-.0156	-.0063	-.0152	-.0032
ρ	BLP	RtMSE	.0644	.0512	.0461	.0353		SN-IV-GMM2	RtMSE	.0796	.0614	.0646	.0466
		Bias	-7.35E-4	-.0064	-.0059	-.0059			Bias	.0011	-.0077	-.0033	-.0034
	SN-Para-GMM2	RtMSE	.0704	.0523	.0563	.0364		SN-V-GMM2	RtMSE	.0886	.0678	.0708	.0471
		Bias	-9.44E-4	-.0115	-.0032	-.0046			Bias	.0213	.0092	.0258	.0139

Note: 1. True parameter values in DGP: $\alpha=-10$, $\beta=1$, $\mu_1=-2$, $\mu_2=1$, $\sigma_1=\sigma_2=.5$, $\rho=.5$.
2. The distribution of random coefficients for the BLP and SN-Para estimators is correctly specified.

S4.4 Misspecification of Logit Errors

In this subsection, we explore how the SN estimator performs when the logit assumption is violated. In particular, we modify the DGP of "Design I: F^0 is Normal" into an RC multinomial probit model, i.e., replacing the logit errors (Gumbel distributed) with normal ones. We use a simple accept-reject method to simulate market shares and focus on a small $J = 10$ case to reduce simulation errors in the generated shares (e.g., to avoid zeroes) with a manageable number of consumer draws (i.e., 10,000).

With the simulated dataset, we implement the BLP and SN estimators that are based on the misspecified RC logit specification. Since the coefficients on X and P from the misspecified model are not directly comparable to the ones in the DGP, we compare their implied elasticities. Table 21 summarizes the X- and P-elasticities evaluated at several quantiles of X and P, averaged across simulation repetitions. The misspecified logit errors indeed lead to biased estimates of the elasticities, but the biases are not large, especially around the medians of X and P. On the other hand, the SN estimator with nonparametric RC does not outperform the ones that assume a normal RC, suggesting that the logit assumption is substantial and a flexible specification of RC does not address the misspecfication of the idiosyncratic error.

Table 20: Monte Carlo Results with Correlated Random Coefficients: Fixed Coefficients

			$T = 10$		20	
Parameter	Estimator		J =50	100	50	100
β	BLP	RtMSE	.0124	.0105	.0159	.0077
		Bias	-.0078	-.0037	-.0073	-.0037
	SN	RtMSE	.0291	.0131	.0330	.0102
		Bias	.0155	.0056	.0128	.0062
α	BLP	RtMSE	.0222	.0276	.0324	.0199
		Bias	-.0046	-.0058	-.0032	-.0056
	SN	RtMSE	.1917	.0630	.2081	.0470
		Bias	.0035	.0020	.0136	.0014

Note: 1. True parameter values in DGP: $\alpha = -10$, $\beta = 1$, $\mu_1 = -2$, $\mu_2 = 1$, $\sigma_1 = \sigma_2 = .5$ and $\rho = .5$.
2. The distribution of random coefficient for the BLP estimator is correctly specified.

Table 21: Monte Carlo Results: Misspecified Logit Errors

		$T = 10$					20					40				
Evaluated at X's Percentiles		10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
X-Elasticity	True (RC Probit)	1.24	.75	.15	.73	1.54	1.26	.69	.09	.72	1.57	1.25	.74	.06	.75	1.56
	BLP (RC Logit)	1.06	.57	.10	.57	1.11	1.02	.56	.07	.56	1.08	1.03	.56	.04	.58	1.11
	SN-Para-GMM2	1.10	.60	.11	.59	1.15	1.06	.58	.07	.59	1.13	1.07	.58	.05	.60	1.15
	SN-III-GMM2	1.07	.59	.11	.59	1.15	1.03	.57	.07	.58	1.12	1.04	.57	.05	.60	1.14
Evaluated at Price's Percentiles		10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Price-Elasticity	True (RC Probit)	-3.00	-2.00	-.39	-1.78	-2.44	-3.03	-2.01	-.25	-1.79	-2.62	-3.04	-2.00	-.17	-1.92	-2.42
	BLP (RC Logit)	-3.42	-1.92	-.31	-1.56	-2.66	-3.41	-1.88	-.20	-1.55	-2.63	-3.44	-1.90	-.14	-1.56	-2.64
	SN-Para-GMM2	-3.50	-1.97	-.32	-1.56	-2.63	-3.48	-1.92	-.20	-1.55	-2.60	-3.53	-1.94	-.14	-1.57	-2.61
	SN-III-GMM2	-3.45	-2.03	-.32	-1.58	-2.71	-3.48	-2.01	-.20	-1.56	-2.68	-3.53	-2.03	-.14	-1.57	-2.70

Note: 1. The DGP is the same as "Design I: F^0 is Normal", except that logit errors are replaced by probit errors (normally distributed).
2. Here $J = 10$, $k_J = \varsigma_J = 3$, and M_J is 3, 4, 5 for $T = 10$, 20, 40, respectively.
3. "True" elasticities are computed using numerical derivatives because there is no closed-form formula for elasticities in the probit model.

S4.5 Computational Time Comparison

Table 22 documents the average computational time (across repetitions) for the baseline Monte Carlo results in Subsection 5.3.1. Though different repetitions are run by different computational nodes (in a cloud environment) with different specifications (e.g., CPU, memory), the average computational time across repetitions is still informative about the relative computational burdens of different estimators. Compared to the BLP estimator, the SN-Para-GMM2, which has the same model specification as BLP, is much faster because of the avoidance of demand inversion. The SN-III-GMM2 is, as expected, slower because it involves more parameters due to the non-parametric specification of the random coefficient density. The BLP estimator with the same non-parametric specifications of the random coefficient is computationally prohibitive given our computational resources for the purpose of Monte Carlo simulations.

Table 22: Computational Time Comparison

		$T = 10$			20			40		
Estimator		$J = 25$	50	100	25	50	100	25	50	100
BLP		5.77	12.62	24.82	12.90	23.03	58.24	24.34	44.32	105.16
SN 1st Step		.10	.15	.20	.18	.25	.42	.53	.93	2.08
SN 2nd Step	SN-Para-GMM2	2.00	4.50	8.38	4.52	8.07	15.69	8.40	13.55	26.29
	SN-III-GMM2	12.78	37.87	135.37	44.83	127.99	397.85	133.98	337.51	984.67

Note: This table documents the computational time (in seconds) based on the simulations of "Design I: F^0 is Normal" and the reported numbers are the averages across repetitions performed on different computational nodes in a cloud environment. The nodes differ in specifications and typically have at least 1 CPU, 2G memory, and 8G disk space.

S5 Empirical Application to China's Auto Market

Now we apply our estimator to the Chinese auto market, where the data structure is quite different from the BLP auto application. In particular, we use data in one year, 2014, and markets are defined geographically, i.e., by province. A product is defined as a model-displacement pair, e.g., Accord 2.4L. The observables, price, quantity, and product characteristics, are similar to the BLP auto data, and there are in total 32 markets and on average each market has 864 products. Table 24 provides some summary statistics of the data for some provinces.

The data differs from BLP auto data in an important aspect: price and other characteristics do not vary across markets. So the only variation across markets comes from product compositions. This is not a concern here because our identification results depend only on variations within instead of across markets.

We present the estimation results in Table 24, which has the same structure as Table 11 in the main context. Note that BLP and SN estimators both yield an estimate of the standard deviation of random coefficient that is very close to zero.[25] Hence, with the current

[25]In this application, we treat price as an exogenous variable and thus do not use IVs. The main reason is that even the simple logit IV regression with standard BLP IVs yields an unreasonable, positive price coefficient. Moreover, finding better IVs, which is a non-trivial task, is not our focus of this empirical

Table 23: Summary Statistics for Selected Markets

Market (Province)	No. of HH (K)	Sales (K)	Price (K CNY)	Horsepower (KW)	Weight (KG)	Liters/100KM	Size (M^3)
Shanghai	8,251	2,314	189	112	1471	7.51	12.86
Yunnan	12,355	6,952	123	96	1347	7.08	12.54
Neimenggu	8,176	3,956	128	95	1328	7.06	12.26
Beijing	6,680	3,941	192	113	1480	7.52	12.96
Jilin	9,002	5,480	128	94	1329	7.04	12.27
Sichuan	25,802	10,453	135	98	1360	7.10	12.47
Tianjin	3,662	1,522	144	99	1348	7.08	12.29
Ningxia	1,842	674	128	97	1361	7.22	12.56
Anhui	18,308	7,592	126	96	1351	7.09	12.48
Shandong	30,105	28,541	113	91	1293	6.87	12.13

Note: All the product characteristics (including price) are quantity-weighted averages.

sample and model specification, both BLP and SN estimators yield the logit outcome, which means that there is little preference heterogeneity on price.[26]

Table 24: China Auto Market: 2014 Data

	BLP		SN
Fixed Coefficient	Logit	RC-Logit	
HP/Weight (log)	.33 (.15)	.26 (.16)	.35 (.15)
Size (log)	6.12 (.23)	6.01 (.24)	6.19 (.24)
Liters per 100km (log)	−4.31 (.16)	−4.34 (.17)	−4.28 (.16)
Brand Dummy	Yes	Yes	Yes
R.C. on Price (Log)			
Mean	−1.39 (.08)	−1.34 (.10)	-1.33
Std. Dev.	-	.000 (3.33)	.001
Ave. No. of Prod. per Mkt.		864.26	
No. of Mkt.		31	

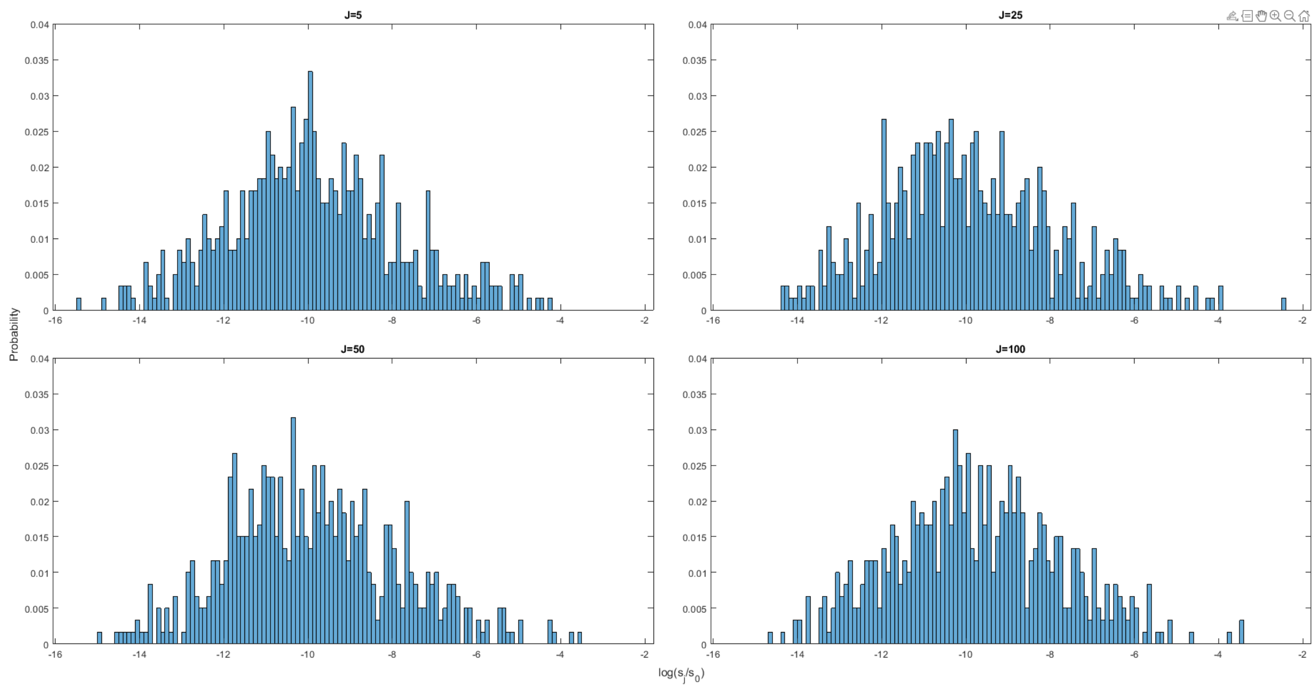
exercise, so we proceed by treating price as exogenous.

[26]In this case, the price elasticities/substitution pattern exhibit IIA property, which are not particularly interesting, so we do not show these results.

S6 Distribution of $\log(s_{jt}/s_{0t})$ as J Grows

In this section, we plot the histograms of log ratio $\log(s_{jt}/s_{0t})$ for different J's based on the DGP of our baseline Monte Carlo simulation design with one normally distributed random coefficient in Section 5.3.1. The histograms show that the distribution of $\log(s_{jt}/s_{0t})$ remains stable as J increases, suggesting that s_{jt} and s_{0t} converge to zero at about the same rate.

Figure 3: Histograms of $\log(s_{jt}/s_{0t})$ for Different J's



S7 A Simple Example of Estimating A Random Location

Since the idea of estimating a random parameter is unconventional, to illustrate the usefulness of it in the simplest setting possible, we now strip away all the complications of the demand model and present a simple example of location estimation.

Let $\xi_j : j = 1, \ldots, J$ be i.i.d. draws from a distribution F, and $E[\xi_j] = 0$. For any sample size J, let

$$Y_{Jj} = \theta_J + \xi_j \quad j = 1, \ldots, J,$$

where $\theta_J = \text{sign}\left(\sum_{j=1}^{J} \phi(\xi_j)\right)$ for some unknown (deterministic) function $\phi(\cdot)$. In this model, θ_J is a random parameter. In general, it does not even have a deterministic limit as $J \to \infty$. How do we estimate θ_J based on the data set $\{Y_{J1}, \ldots, Y_{JJ}\}$?

We propose the estimator $\hat{\theta}_J = J^{-1} \sum_{j=1}^{J} Y_{Jj}$. This estimator is consistent:

$$|\hat{\theta}_J - \theta_J| \xrightarrow{p} 0, \text{as } J \to \infty,$$

as long as $E[|\xi_J|] < \infty$. This is simply because

$$|\hat{\theta}_J - \theta_J| = \left| J^{-1} \sum_{j=1}^{J} (Y_{Jj} - \theta_J) \right| = \left| J^{-1} \sum_{j=1}^{J} \xi_j \right| \xrightarrow{p} 0,$$

where the convergence holds by LLN.

Note that the consistency argument does not require θ_J to converge to a deterministic limit. Neither does it require $J^{-1} \sum_{j=1}^{J} \phi(\xi_j)$ to converge to a deterministic limit. In fact, it imposes no restriction on $\phi(\cdot)$ at all. The sign function defining θ_J may also be replaced with any finite valued function. The only things that are important are that θ_J does not vary with J and is additively separable from ξ_J.

References

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Alj, A., R. Azrak, and G. Mélard (2014): "On conditions in central limit theorems for martingale difference arrays," *Economics letters*, 123, 305–307.

Cardell, S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*.

Casella, G. and R. L. Berger (2001): *Statistical Inference (Second Edition)*, Cengage Learning.

Chen, X. and T. M. Christensen (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9, 39–84.

Chen, X., Y. Fan, and V. Tsyrennikov (2006): "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of the American Statistical Association*, 101, 1228–1240.

Fosgerau, M. and S. L. Mabit (2013): "Easy and flexible mixture distributions," *Economics Letters*.

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *The Annals of Statistics*, 33, 2904–2929.

Horowitz, J. L. (2009): *Semiparametric and nonparametric methods in econometrics*, vol. 12, Springer.

Pollard, D. (1990): *Empirical Process Theory and Application, NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. II.

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80, 213–275.

van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer.