Robust Inference for Differentiated Product Demand Systems*

Xian Li†

Job Market Paper - Latest Version

November 21, 2017

Abstract

This paper provides robust inference for differentiated product demand with measurement error in market shares. Market shares have been used as estimators for choice probabilities generated from a (random coefficient) discrete choice model. However, there are situations in which market shares are inaccurately measured, such in the cases of unobserved choice set variations (e.g., stock-out events), sampling error and measurement error in market sizes. The existing point identification approaches to address measurement error introduced by stock-out events do not allow for endogenous price. The partial identification approach by moment inequalities (e.g., Gandhi et al. (2013)), in general, does not characterize a sharp identified set and the demand function can only be estimated based on market level variations. This paper gives a sharp characterization of the identified set using moment equalities with latent variables. A feasible estimation of such sets requires reducing the dimension of an optimization problem. The existing duality approach does not have a natural generalization to the demand estimation environment due to the special dependence structure inside the market. This paper further contributes to the existing econometric literature by providing a new method to reduce the dimension that exploits the convexity nature of this problem to accommodate within market dependence. Theoretically, the method is proven to be robust to measurement error in market shares, and it is also verified by simulations and empirical studies.

*I am grateful to my advisor, Andrés Aradillas-López, and Ronald Gallant, Daniel Grodzicki, Patrik Guggenberger, Marc Henry, Sung Jae Jun, Keisuke Hirano, Peter Newberry, Joris Pinkse for valuable comments and discussions.

†PSU Department of Economics. Email: xxl155@psu.edu.

1 Introduction

Many recent works in economics and marketing have estimated the differentiated product demand systems using the method developed by Berry (1994) and Berry et al. (1995) (henceforth BLP). BLP generates choice probabilities from a (random coefficient) discrete choice model and then matches them with market shares. A key assumption in this approach is that the observed market share of a product is equal to the choice probability generated from the true parameter value.

However, this assumption does not always hold. First, sometimes consumers face a more restricted choice set rather than the entire set of products considered by econometricians. For example, if a consumer's preferred choice stocks-out, and the exact timing of the stock-out is not observed, the market share of the stock-out product will be capped at its capacity and, therefore, is less than its actual choice probability; however, the market share of its available substitutes will be higher than their choice probabilities. Such choice set variation cannot be fixed by changing the discrete choice model because variations of choice set are typically unobserved. In addition, when the market size is small, the market share will not be an accurate representation of the corresponding choice probability due to sampling error[1]. In some cases, even zero market share will be observed. Lastly, it is often the case that market size is not properly measured and researchers choose a set of numbers based on experience as a robustness check. If we ignore the above issues and proceed to estimation, the estimators can be inconsistent and the confidence sets can undercover the true parameter values, thereby invalidating counterfactual analysis based on the demand estimation.

This paper seeks to solve the problems that result from any type of measurement error found in actual applications. It adopts a partial identification approach that introduces a vector of latent variables to represent the measurement error. For example, in the stock-out situation, I introduce a vector of latent variables that represent the unfulfilled demand for stocked-out products and the spillover demand for their substitutes. Now instead of replacing the choice probabilities with the market shares suggested by BLP, this paper proposes to replace the choice probabilities with the market shares plus the latent variables, and to

[1] Berry et al. (2004) is aware of the sampling error in BLP framework and imposes regularity assumptions to let the market size grow at a much higher rate than the sample size to make BLP estimator consistent.

subsequently generate moment conditions based on it.

The model is partially identified because each conditional distribution of the latent variables corresponds to a vector of parameter values. The parameter values in the identified set are such that one can find a conditional distribution for the latent variables to satisfy the moment restrictions. The key idea to bound the identified set is to bound the support of these latent variables under different measurement error situations. This will result in moment equalities with latent variables in moment functions, which is different from the traditional moment inequalities method. The set is naturally sharp by its definition. This paper also proposes a novel method to estimate this type of model, which accounts for the special dependence structure within a market[2], as the existing methods in the literature are designed to handle only independent or weakly dependent data and cannot be easily generalized in the context of demand estimation. Hence, this method makes it possible to identify and estimate the demand function with either market- or product-level variations.

As far as I know, this is the first paper in the literature to attempt to solve the stock-out problem in demand estimation under the BLP framework, which is known to work with the price endogeneity issue while allowing for rich substitution patterns. The industrial organization and marketing literature have been aware of the fact that consumers sometimes face a limited choice set due to stock-out events. A natural response to remedy this issue would be to consider a different model other than BLP that can account for stock-out events. The problem is that the timing of a stock-out event can never be pinned down to an exact time point, therefore the variations of some consumers' choice sets are unobserved. Various approaches have been developed to address this problem. Conlon and Mortimer (2013) proposes a model to estimate an experimental data set on vending machines with observed inventory levels such that the timing of stock-out events can be partially pinned down. Musalem et al. (2010) develops a method to simulate consumers' choice set using a scanner data set acquired from grocery stores. Other papers of interest in this literature include those by Bruno and Vilcassim (2008), Che et al. (2012), Campo et al. (2004), Ching et al. (2015), and Matsa (2011), which address different out-of-stock environments and their

[2] For example, if the firms compete in price, then the price of each product will depend on the prices of all the products in that market.

implications. However, to point identify the demand function, these methods either fail to account for price endogeneity, which plays a critical role in demand estimation, or hinge on strong assumptions related to the stock-out process (e.g., a uniform consumer arrival process that assumes consumers with different tastes arrive at the store with equal probability). Besides, some of the proposed methods rely on simulating consumers' arrival processes to approximate the timing of an out-of-stock event. Such simulations always assume a specific parametric family of distributions and, thus, restrict consumers' choice patterns and substitution patterns.

This paper deals with the unobserved choice set variation by introducing the latent variables discussed above. I only require econometricians to observe a binary variable that indicates whether a product has experienced stock-out during a given observation period[3]. In addition, the latent variables are modeled in a way that is agnostic about their dependence on the observed/unobserved variables and any possible consumer arrival process. Hence, it can allow for very rich consumer choice and substitution patterns. Besides, this treatment is under the BLP framework, which makes it possible to deal with the endogenous prices by using a set of instrumental variables.

The measurement error of the market shares falls under the general situation in which there is measurement error in dependent variables like censoring. However, the traditional way of point identification using a quantile independence assumption[4], cannot be applied here as the unobserved shocks are multi-dimensional and buried inside a nonlinear system. Another common treatment is to simply drop the contaminated observations, however, it will suffer from the problem of sample selection bias. When the market size is small and sampling error is large, Gandhi et al. (2013) [5] (henceforth GLS) proposes a moment inequality method to characterize the identified set. Their method, however, does not deliver a sharp identified set, as the bound of the moment inequalities can be attained by parameter values outside the sharp identified set due to dependence within a market. The moment inequalities

[3] It does not have to be an exact time. For example, if one has weekly data, then I require researchers to know if a product has ever experienced stock-out during a week.

[4] See, e.g., Powell (1984) and the literature thereafter.

[5] In their latest draft, Gandhi et al. (2017), they abandon the partial identification approach and adopt a new assumption to point identify the demand function. They still claim that the old draft is applicable when the point identification assumption does not hold.

approach can also only be used to estimate demand function based on market-level variation by aggregating the product-level moments. This is because the random variables within the same market are usually dependent in a complicated way, which is not independent nor any form of weak dependence due to the strategic competition within a market, and none of the current estimation procedures presented in the moment inequality literature can deal with such a dependence structure. Under the same set of assumptions as in GLS, this paper provides a sharp characterization of the identified set and is able to handle within market dependence.

The price to pay for using the method provided in this paper is that the identified set is not defined by moment inequalities but moment equalities with latent variables. Unlike the sets that are defined by moment inequalities, there are very few methods that aim to estimate the sets defined by moment equalities. The major problem is that the optimization required in the model is performed over the space of conditional distributions of the latent variables, which is an infinite dimensional space and, hence, infeasible to compute. Ekeland et al. (2010), Galichon and Henry (2013), and Schennach (2014) develop methods to transform the infinite dimensional optimization problem into a finite dimensional dual problem. However, their methods are designed for independent or weakly dependent data and do not fit the problem considered here due to the special dependence structure within a market.

This paper proposes a new method to reduce the dimension of the original problem by exploiting the fact that the support of the conditional distribution of the latent variables is always a compact and convex set. The technique hinges on the fact that the space of probability measures over a compact support is a convex set and compact in the weak* topology and, therefore, the optimization can be attained by only searching over the set of extreme points in this space, which is finite dimensional. After the objective function is turned into a feasible finite dimensional optimization problem, the construction of a set estimator and confidence region can be performed using the methods suggested by Chernozhukov et al. (2007), Romano and Shaikh (2008), and Romano and Shaikh (2010).

I conducted two simulation studies on the proposed method, both of which focused on the out-of-stock situation. The first of these is a simple logit setup in which there is only one product in each market, and I can obtain an analytical form of the identified set. The simulation suggests that, even with a very low stock-out rate across markets, the BLP confidence interval severely undercovers the true parameter value. For example, if 3 out of 100 markets observed a stock-out, then the coverage probability of a 95% confidence interval only covers the true parameter value at a rate of 47.1%. However, the method provided in this paper always has the correct coverage probability for both the identified set and the true parameter value. In the second simulation I generate data similar to the situation in the actual empirical application considered in this paper. The confidence region for the true parameter value of this paper still provides correct coverage probability while the BLP estimator is inconsistent.

I apply the method developed in this paper to a scanner data set consisting of weekly sales data related to shampoo product category which was obtained from multiple grocery stores in different regions. The scanner data set itself does not contain information about stock-out events. However, it is very clear from the data set that, at some point, one product is experiencing an abnormally low volume of sales before returning to standard levels in the following week. I apply a method described in the existing marketing literature (Gruen et al. (2002), Gruen and Corsten (2007), and Grubor and Milicevic (2015)) to determine stock-out events. The empirical results suggest that the procedure developed in this paper results in a significant correction for price elasticity in comparison to the BLP estimators, thereby exhibiting a similar pattern to that identified in the simulation study. This provides a more robust counterfactual analysis from the demand side.

The remainder of this paper is organized as follows. In Section 2, I introduce a simple two-product example to highlight the primary contribution and ideas of this paper. In Section 3, I formally introduce the discrete choice model I will be using throughout the paper. In Section 4, I discuss the new identification strategy and the identified set introduced in this paper. In Section 5, I provide a method of conducting estimation and inference on the identified set and the true parameter value. In Section 6, I describe the finite sample performance of the new estimator and confidence set via Monte Carlo simulations. In Section 7, I conduct the method on a scanner data set acquired from the retail industry. Section 8 concludes the paper.

2 Motivating Example

Suppose there are two products $j = 1, 2$ and an outside option $j = 0$ in each market $t \in \{1, 2, \ldots, T\}$, the consumers in each market choose to buy one and only one product or the outside option. The utility of consumer i choosing product j in market t is represented as:

$$u_{ijt} = \theta + \xi_{jt} + \varepsilon_{ijt}$$

where $\xi_{jt} \in \mathbb{R}$ is product j's unobserved characteristic in market t with mean zero, $\varepsilon_{ijt} \in \mathbb{R}$ is consumer idiosyncratic shock, and $\theta \in \Theta \subset \mathbb{R}$ is the parameter of interest. The utility of the outside option is normalized to zero: $u_{i0t} = 0$.

Assume ε_{ijt} is independent across i, j and t, and follows a Type I extreme value distribution. The consumers would choose the option that gives them the highest utility. Since we do not observe a consumer's idiosyncratic shock, a canonical result says that we can integrate it out to obtain the choice probability of product j in market t in the following closed form:

$$\sigma_{jt} = \frac{\exp(\theta + \xi_{jt})}{1 + \exp(\theta + \xi_{1t}) + \exp(\theta + \xi_{2t})}, \ j = 1, 2 \tag{1}$$

and

$$1 - \sigma_{1t} - \sigma_{2t} = \sigma_{0t} = \frac{1}{1 + \exp(\theta + \xi_{1t}) + \exp(\theta + \xi_{2t})} \tag{2}$$

We can then solve the system of equations (1) and (2) for ξ_{jt}, and get

$$\xi_{jt} = \log\big(\frac{\sigma_{jt}}{1 - \sigma_{1t} - \sigma_{2t}}\big) - \theta$$

Since $\mathbb{E}[\xi_{jt}] = 0$, let the true value of θ be θ_0, then we have

$$\mathbb{E}[\xi_{jt}] = \mathbb{E}\Big[\log\big(\frac{\sigma_{jt}}{1 - \sigma_{1t} - \sigma_{2t}}\big) - \theta\Big] = 0 \tag{3}$$

when $\theta = \theta_0$.

Assume we observe the market share of product j in market t, s_{jt}, and there is no measurement error in the market share ($\sigma_{jt} = s_{jt}$), then we can replace σ_{jt} with s_{jt} in equation (3) and get

$$\mathbb{E}\Big[\log\big(\frac{s_{jt}}{1 - s_{1t} - s_{2t}}\big) - \theta\Big] = 0 \tag{4}$$

when $\theta = \theta_0$. Equation (4) can be used to identify and estimate θ_0.

However, equating market shares to choice probabilities can introduce measurement error. For example, if at some point in market t, product 1 is out-of-stock, then the market share of product 1 will always be no larger than the choice probability evaluated at the true parameter values, i.e., $s_{1t} \leqslant \sigma_{1t}$. On the other hand, due to the forced substitution from the stock-out of product 1, the market share of product 2 will always be no smaller than the choice probability evaluated at the true parameter values, i.e., $s_{2t} \geqslant \sigma_{2t}$.

There can be other types of measurement error in the absence of stock-out events. Notice that, market share is the sample mean while choice probability is the population mean of a multinomial distribution. There is always sampling error that may not be negligible when the market size is small. Lastly, if researchers are uncertain of the market size when computing the market shares, the market shares will always be measured with error.

Hence the traditional identification strategy discussed above no longer works when there is measurement error. If a researcher estimates the demand function with the same strategy outlined above, the estimators will generally be inconsistent.

Let r_{jt} be the measurement error such that $s_{jt}+r_{jt} = \sigma_{jt}$, π_0 be the true joint distribution of (s_{1t}, s_{2t}), and μ be a conditional distribution of (r_{1t}, r_{2t}) conditional on (s_{1t}, s_{2t}). Notice that since the products are modeled in a symmetric way[6], we have that π_0 is symmetric in its arguments (s_{1t}, s_{2t}). Similarly, the marginal distribution of σ_{1t} should be the same as the marginal distribution of σ_{2t}, which implies, the marginal distribution of $(s_{1t} + r_{1t}, s_{2t} + r_{2t})$ under $\pi_0 \times \mu$ should also be symmetric.

Let $\mathcal{C}_t(s_{1t}, s_{2t})$ be the collection of μ that satisfies the above symmetric restriction. The

[6]Switching the subscript j will not affect anything.

sharp identified set is characterized as

$$\Big\{\theta \in \Theta : \exists \mu \in \mathcal{C}_t(s_{1t}, s_{2t}), \mathbb{E}_{\pi_0 \times \mu}\Big[\log\big(\frac{s_{jt} + r_{jt}}{1 - s_{1t} - s_{2t} - r_{1t} - r_{2t}}\big) - \theta\Big] = 0\Big\} \tag{5}$$

Notice that since we do not observe the true value of r_{jt}, and for each possible μ, there will be one θ to satisfy the above moment constraints, we would have a set of θ that satisfy the above moment constraints in addition to θ_0[7], the parameters are generally not point identified.

We could further restrict $\mathcal{C}_t(s_{1t}, s_{2t})$ based on different measurement error environments as discussed in later sections, but for this simple example, I am just focusing on the identified set characterized in (5).

Providing a set estimator for the identified set defined in (5) is not an easy task, as $\mathcal{C}_t(s_{1t}, s_{2t})$ is infinite dimensional. Before I explain my method to estimate it, let us look at a moment inequality approach to characterize an identified set (hence easier to estimate).

In the case when the market size is small and there is a large sampling error, GLS proposes the following moment inequalities:

$$\mathbb{E}\Big[\log\big(\frac{\tilde{s}_{1t} + \eta}{1 - \tilde{s}_{1t} - \tilde{s}_{2t} - \eta}\big) - \theta\Big] \quad \geqslant \quad 0 \tag{6}$$

where $\eta = \max_{\sigma_{1t}, \sigma_{2t}} \eta_{1t}(\sigma_{1t}, \sigma_{2t})$ is a constant, and $\eta_{1t}(\sigma_{1t}, \sigma_{2t})$ is the solution to

$$\mathbb{E}\Big[\log\big(\frac{\tilde{s}_{1t} + \eta_{1t}}{1 - \tilde{s}_{1t} - \tilde{s}_{2t} - \eta_{jt}}\big)|\sigma_{1t}, \sigma_{2t}\Big] = \log\big(\frac{\sigma_{1t}}{1 - \sigma_{1t} - \sigma_{2t}}\big)$$

Here $\tilde{s}_t$ is introduced to solve the issue that log is not defined at zero and has the following representation:

$$\tilde{s}_t = \frac{n_t s_t + 1}{n_t + 3} \tag{7}$$

[7]Let the true conditional distribution be μ_0, then θ_0 is the unique parameter value to satisfy

$$\mathbb{E}_{\pi_0 \times \mu_0}\Big[\log\big(\frac{s_{jt} + r_{jt}}{1 - s_{1t} - s_{2t} - r_{1t} - r_{2t}}\big) - \theta\Big] = 0$$

where n_t is the market size of market t.

Let $\tilde{\theta}$ be such that

$$\mathbb{E}\Big[\log\big(\frac{\tilde{s}_{1t}+\eta}{1-\tilde{s}_{1t}-\tilde{s}_{2t}-\eta}\big)-\tilde{\theta}\Big]=0 \tag{8}$$

which satisfies the inequality in (6).

Substituting (7) into (8) and rearranging we can get

$$\mathbb{E}\Big[\log\big(\frac{s_{1t}+\frac{1}{n_t}+\frac{n_t+3}{n_t}\eta}{1-s_{1t}-s_{2t}+\frac{1}{n_t}-\frac{n_t+3}{n_t}\eta}\big)-\tilde{\theta}\Big]=0$$

To match it with (5), $\tilde{\theta}$ is such that, $r_{1t}=\frac{1}{n_t}+\frac{n_t+3}{n_t}\eta$ and $r_{2t}=-\frac{2}{n_t}$ regardless of the value of (s_{1t},s_{2t}), which fails the symmetric requirement on $\pi_0\times\mu$. This implies that, $\tilde{\theta}$, which is in the identified set characterized by the moment inequality, does not belong to the sharp identified set.

GLS provides a way to characterize an identified set via moment inequalities, and we can see that it is not the sharp set from the above discussion. One can propose other ways of moment inequality characterizations, however, none of the moment inequality characterizations would lead to a sharp identified set because they will always fail the symmetric requirement in a way similar to GLS (unless there is only one product in each market).

Though the identified set characterized by moment inequalities is easier to estimate, due to the above reason, this paper estimates the set in (5) directly without using moment inequalities.

To solve the infinite dimensional problem, this paper proves that, it is sufficient to consider the set of Dirac measures[8] in $\mathcal{C}_t(s_{1t},s_{2t})$, which is finite dimensional. The later sections will formally discuss this idea in a more general setup.

[8]A probability measure that assigns probability one to a single point.

3 Model

Following discrete choice model literature, I assume consumer i can choose one from J products in market $t \in \{1, 2, \ldots, T\}$. Let the utility of consumer i of consuming product $j \in \{1, 2, \ldots, J\}$ be

$$u_{ijt} = u(x_{jt}, \xi_{jt}, \varepsilon_{ijt}; \beta)$$

with known parametric form[9], where $x_{jt} \in \mathbb{R}^{d_x}$ and $\xi_{jt} \in \mathbb{R}$ are observed and unobserved product and market characteristics, $\varepsilon_{ijt} \in \mathbb{R}$ is unobserved consumer idiosyncratic shock, and $\beta \in \mathbb{R}^{d_x}$ is a finite dimensional parameter vector. The unobserved consumer idiosyncratic shock ε_{ijt} is assumed to follow a distribution $F(\cdot; \eta)$, which is known up to a finite dimensional parameter vector $\eta \in \mathbb{R}^{d_\eta}$. The utility of choosing the outside option is normalized to $u_{i0t} = 0$.

The consumers always choose the product with the highest utility. Therefore by integrating out the unobserved consumer idiosyncratic shock we can get the conditional choice probability of product j:

$$\sigma_{jt} = P(u_{ijt} > u_{ij't},\ j' \neq\ j | x_{jt}, \xi_{jt}) = \sigma_j(x_t, \xi_t; \beta, \eta).$$

Assume that the market share of product j, s_{jt} is equal to σ_{jt}[10]. Berry et al. (2013) shows that under mild conditions, $s_{jt} = \sigma_{jt}(x_{jt}, \xi_{jt}; \beta, \eta)$ can be inverted as $\xi_{jt} = \sigma_j^{-1}(x_t, \sigma_t; \beta, \eta) = \sigma_j^{-1}(x_t, s_t; \beta, \eta)$. Given a set of instruments z_t such that ξ_{jt} is conditional mean independent of z_t, we have,

$$\mathbb{E}[\xi_{jt} | z_{jt}] = \mathbb{E}\big[\sigma_j^{-1}(x_t, s_t; \beta, \eta) | z_{jt}\big] = 0,\ a.s.$$

when (β, η) is equal to the true parameter value (β_0, η_0). Then the above conditional moment restrictions can be used to construct estimators for β_0 and η_0.

We have introduced a simple example under this framework in Section 2. The following example presents the random coefficient logit model. This is the most frequently used model in the industrial organization literature.

Example 3.1. *(Random Coefficient Logit with Endogenous Price) Let the utility of consumer*

[9] The most common choice is the linear form.

[10] See Berry et al. (2004) for a set of conditions that warrants this.

i of consuming product $j \in \{1, 2, \ldots, J\}$ in market t be:

$$u_{ijt} = x'_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \sum_k \eta_k x_{jtk}\nu_{itk} + \eta_p p_{jt}\nu_{itp} + \varepsilon_{ijt}$$

where $x_{jt} \in \mathbb{R}^{d_x}$ and $\xi_{jt} \in \mathbb{R}$ are observed/unobserved product characteristics, $p_{jt} \in \mathbb{R}$ is the price of product j in market t which is correlated with ξ_{jt}. The variables ν_{itk}, ν_{itp} and ε_{ijt} are unobserved consumer idiosyncratic shock. The variables ν_{itk} and ν_{itp} are distributed as standard normal distribution and the variable ε_{ijt} is distributed as Type I extreme value distribution. Also let $u_{i0t} = 0$.

The parameters β, α, and $\eta = (\eta_1, \ldots, \eta_k, \eta_p)$ are to be estimated.

We can compute the conditional choice probability of this model as

$$\sigma_{jt} = \int_\nu \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \sum_k \eta_k x_{jtk}\nu_{itk} + \eta_p p_{jt}\nu_{itp})}{1 + \sum_{l=1}^J \exp(x_{lt}\beta - \alpha p_{lt} + \xi_{lt} + \sum_k \eta_k x_{ltk}\nu_{itk} + \eta_p p_{lt}\nu_{itp})} d\Phi(\nu)$$

This system does not have an analytical solution but it can be shown that the right hand side when viewed as a mapping in ξ_{jt} is a contraction mapping. Therefore it can still be easily inverted to represent ξ_{jt} as a function of observed random variables and parameters.

$$\xi_{jt} = \sigma_j^{-1}(x_t, p_t, \sigma_t; \beta, \alpha, \eta)$$

Thus, if $\sigma_t = s_t$, then we have

$$\xi_{jt} = \sigma_j^{-1}(x_t, p_t, s_t; \beta, \alpha, \eta)$$

Let z_t be a set of instruments which also includes x_t, then we have

$$\mathbb{E}\big[\sigma_j^{-1}(x_t, p_t, s_t; \beta, \alpha, \eta)|z_{jt}\big] = 0$$

when (β, α, η) is equal to the true parameter value $(\beta_0, \alpha_0, \eta_0)$.

The above conditional moment equalities can thus be used for constructing estimators for $(\beta_0, \alpha_0, \eta_0)$.

However, sometimes $\sigma_{jt} \neq s_{jt}$. For example, GLS shows that, since s_{jt} is the realization of a multinomial distribution, if the market size in a given market is small, the sampling error of s_{jt} is too large to ignore.

Another example that $\sigma_{jt} \neq s_{jt}$ is when there is unobserved consumer choice set variation. Suppose there are product availability issues, when product j stocks-out, generally $s_{jt} < \sigma_{jt}$ and its available substitutes j' will show $s_{j't} > \sigma_{j't}$.

Let r_{jt} be the measurement error in the above mentioned examples, i.e., $\sigma_{jt} = s_{jt} + r_{jt}$. If one proceeds to use the above mentioned method, the ideal way would be to invert

$$s_{jt} + r_{jt} = \sigma_{jt} = \sigma_{jt}(x_{jt}, \xi_{jt}; \beta, \eta)$$

and represent ξ_{jt} as a function of observed random variables, r_{jt}, and parameters.

Unfortunately, r_{jt} is usually unobserved to econometricians. Ignoring this and still using the stardard BLP estimators assuming $r_{jt} = 0$ when it is actually not, will result in inconsistent estimators and the constructed confidence set will typically under cover the true parameter value.

One possible way to solve this is to drop the contaminated observations with $r_{jt} \neq 0$. For example, drop the observations with relatively small market sizes, or drop the observations with stock-out events. This is, however, not a panacea. For example, in the stock-out situation, this requires that the stock-out events are independent of the unobserved shock ξ_{jt}. This is a strong condition, as the unobserved shock can be the direct cause of stock-out events, e.g.,, unobserved promotions. This will create the classic sample selection bias problem. My simulation results suggests that the standard BLP estimators will severely underestimate the price elasticity and dropping observations will only partially fix the issue by reporting a number between BLP estimators and the true value.

Other methods in the literature rely on a special data structure or strong parametric assumptions to point identify parameters of interest. This paper, on the other hand, explores the partial identification approach, while still maintaining minimum assumptions on the model and the data.

Lastly, I do not specify a supply side model in this paper, as different measurement error environments will usually require different supply side models. Besides the Nash price competition model, for example, Aguirregabiria (2005) presents a model where firms compete in prices and inventories (probabilities of stock-out). Quan and Williams (2017) considers the situation when stores tend to keep low volume of products when the demand is low in that area.

4 Identification

4.1 Main Result

Following Section 3, suppose measurement error exists in market t, we have

$$s_{jt} + r_{jt} = \sigma_{jt} = \sigma_{jt}(x_{jt}, \xi_{jt}; \beta, \eta) \tag{9}$$

where r_{jt} is the the unobserved measurement error. For market t' when $s_{jt'} = \sigma_{jt'}$, we have $r_{jt'} = 0$, otherwise, $r_{jt'} \neq 0$. We can still invert equation (9) to represent ξ_{jt} as a function of other variables and get

$$\mathbb{E}[\xi_{jt}|z_{jt}] = \mathbb{E}\big[\sigma_j^{-1}(x_t, \sigma_t; \beta_0, \eta_0)|z_{jt}\big] = \mathbb{E}\big[\sigma_j^{-1}(x_t, s_t + r_t; \beta_0, \eta_0)|z_{jt}\big] = 0. \tag{10}$$

Apparently some restrictions have to be imposed on r_t. First of all there are some trivial bounds on the support of the conditional distribution of r_t, because $s_{jt} + r_{jt} = \sigma_{jt}$ is the choice probability which has to lie between 0 and 1 and they sum up to one. I specify this bound in the following lemma:

Lemma 4.1. *We have the following restrictions on r_{jt}:*

(1) For any j, $-s_{jt} \leq r_{jt} \leq 1 - s_{jt}$.

(2) $\sum_j r_{jt} = 0$.

Proof. See Appendix.

However, the following proposition shows that, without any further restrictions on r_t, any parameter values can satisfy equation (10) under any given joint distribution of x_t and z_t.

Proposition 4.1. *(Non-identification) Suppose r_t satisfies the bounds (1) and (2) in Lemma 4.1, then for any joint distribution of s_t, z_t, x_t, and any (β, η), there exists a conditional distribution of r_t conditional on s_t, z_t, x_t, such that equation (10) holds.*

Proof. See Appendix.

I present a simple example to illustrate the idea of Proposition 4.1.

Example 4.1. *(Simple Logit) Based on*

$$\mathbb{E}\Big[\log\big(\frac{s_{1t}+r_{1t}}{s_{0t}+r_{0t}}\big) - \alpha_0 - \beta_0 x_{1t}\Big] = 0$$

we have

$$\alpha_0 = \mathbb{E}\Big[\log\big(\frac{s_{1t}+r_{1t}}{s_{0t}+r_{0t}}\big)\Big] - \beta_0 \mathbb{E}[x_{1t}]$$

With only the bounds in Lemma 4.1, it is easy to see that no matter what information we have on β_0, the identified set for α_0 is $\mathbb{R}$ since we can push $s_{0t} + r_{0t}$ to either 0 or 1.

In order to provide informative bounds on the parameters of interests, we have to impose an additional requirement on data and model in order to shrink the support of r_t. To attain this, I require that econometricians observe a set of new variables d_{jt} under different measurement error environments. I will specify d_{jt} and how to use them to get a smaller identified set in different situations as follows.

4.1.1 Identification under Out-of-Stock Events

Let $\mathcal{U}_t \subset \{0, 1, 2, \ldots, J\}$ be the set of products that stock-out during an observation period, and $\mathcal{A}_t \subset \{0, 1, 2, \ldots, J\}$ be the set of products that do not stock-out in market t. Let s_{jt} be the market share of product j in market t, then there exists $j \in \mathcal{U}_t$, $s_{jt} \leq \sigma_{jt}(x_{jt}, \xi_{jt}; \beta_0, \eta_0)$, and for any $j \in \mathcal{A}_t$, $s_{jt} \geq \sigma_{jt}(x_{jt}, \xi_{jt}; \beta_0, \eta_0)$, where (η_0, β_0) is the true value of parameters.

First I require that stock-out events are observed. Specifically, let $d_{jt} = 1$ if $j \in \mathcal{U}_t$ and $d_{jt} = 0$ if $j \in \mathcal{A}_t$. I require that d_{jt} is observed by econometricians. This is due to the following reasons.

In many applications the researchers actually observe stock-out events. For example, in Che et al. (2012)'s study of one national chain grocery store, the staff members of the store inspect the shelf at approximated 6pm every day and report stock-out events. In Musalem et al. (2010)'s study of supermarkets in Spain, shelves are recorded at the beginning and the end of each day.

Even if stock-outs are not directly observed in the data, there is a small literature in marketing that aims to detect stock-out events from point-of-sale data. In the scanner data set considered by the empirical application of this paper, it is frequently observed that some of the products experience an abnormal low sales during a particular period before they go back to normal later. Such pattern can be attributed to stock-out events after ruling out the possibility of (1) other products are on sale, (2) stores are introducing a similar product that is strictly better (e.g., same brand, same price, larger size), (3) seasonality factors that cause a negative shock in demand. Thus a method described by Gruen and Corsten (2007)[11] to actually detect stock-out events is applied to this scanner data set[12].

Notice that here I only require very minimal information on stock-out: the researcher only has to know which product stocks-out in one period, one does not have to know the exact time when the stock-out event occurs, nor the order of the stock-out events.

If stock out did not happen but was marked as happened, the method in this paper is still feasible but will lead to a larger identified set. But if stock out happened and was marked as not, then the method in this paper no longer applies and will result in an inconsistent estimator.

Based on the additional information on stock-out events, we have the following lemma.

Lemma 4.2. *For any $j \in \mathcal{A}_t$, $-s_{jt} \leq r_{jt} \leq 0$.*

Proof. See Appendix.

Lemma 4.2 is important. With the help of it we can divide r_{jt} into two different groups and further shrink the support of those when $j \in \mathcal{A}_t$. Essentially, if we do not observe which product stocks-out then any substitution pattern can be the case.

[11]It is claimed by Gruen and Corsten (2007) that this method has 80%-90% accuracy in determining stock-out events.

[12]Notice here if stock-out did not happen but was marked as happened, the method in this paper is still feasible but will lead to a larger identified set. But if stock-out happened and was marked as not happened, the method in this paper no longer applies and will also result in an inconsistent estimator in a way similar to standard BLP estimator.

Notice that Lemma 4.1 and Lemma 4.2 together implies that there exists at least one $j \in \mathcal{U}_t$, such that $0 \leq r_{jt} \leq 1 - s_{jt}$.

Now we proceed to shrink the support of r_{jt} further when $j \in \mathcal{U}_t$.

The following requirement hinges on the outside option, specifically, I require that the outside option is always available and its choice probability is bounded away from 0.

Assumption 4.1. *For any t, $0 \in \mathcal{A}_t$ and $\sigma_{0t} \geq \zeta > 0$.*

With the help of Assumption 4.1, we can provide a tighter bound on r_{jt} when $j \in \mathcal{U}_t$.

Lemma 4.3. *For any $j \in \mathcal{U}_t$, $-s_{jt} \leq r_{jt} \leq 1 - s_{jt} - \zeta$.*

Proof. *See Appendix.*

It is clear that ζ will affect the bound of r_{jt} therefore affect the bound for the identified set. In an application, ζ cannot be consistently estimated without further assumptions and one has to manually choose one. However, some information can be obtained, for example, we can assume that $\zeta \leqslant \min_t s_{0t}$. Notice that the larger ζ is, the smaller the identified set will be.

Lemma 4.2 and Lemma 4.3 impose strong restrictions on the support of r_t, therefore, as we can see in the simulation and empirical studies, provide informative bounds on the identified set.

For example, suppose there are two products and an outside option. In market t we observe product 1 stocks-out. The bound of r_{1t} and r_{2t} provided above is the red area shown in Figure 1. Notice that, this set is convex and compact, which is an important feature I will be exploiting later in estimation and inference.

There are certain assumptions that are strong in general so I do not discuss them here. But if one is willing to impose such assumptions, a smaller identified set can be obtained. For example, if one is willing to assume independent of irrelevant alternatives (IIA), the support of r_{jt} can be made to be a line segment.

4.1.2 Identification under Small Market Size

The idea of market share inversion hinges on the fact that market size is large enough so the sampling error is negligible. Essentially, Berry et al. (2004) provides conditions for which

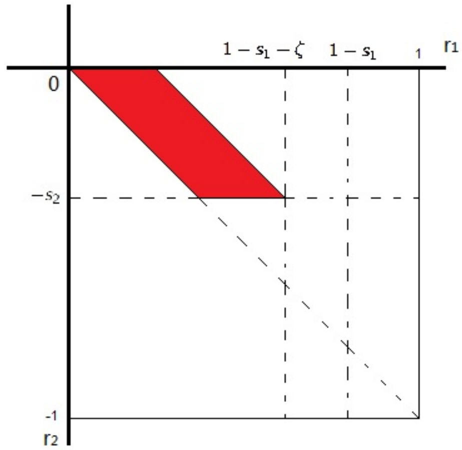

Figure 1: Graphical Illustration of Bounds

market share inversion will yield consistent estimators. However, if the market size is small enough, the asymptotic results in Berry et al. (2004) does not hold and we have to take sampling error into account. GLS demonstrates the failure of identification when market size is small and adopted a moment inequality approach.

In GLS, the key assumption to deliver an informative identified set is that σ_{jt} is bounded away from zero for any product j (including outside option) in any market t. In the application, this lower bound cannot be estimated either. In their paper, the reciprocal of the market size was used to proxy for this lower bound.

The method in this paper can also be applied to the environment they are considering. Specifically, I impose a similar assumption as follow.

Assumption 4.2. *For any j, t, $\sigma_{jt} \geqslant \frac{1}{d_t}$, where d_t is the size of market t.*

The above assumption will result in the following bound on r_t.

Lemma 4.4. *For any j, $\frac{1}{d_t} - s_{jt} \leq r_{jt} \leq 1 - \frac{J+1}{d_t} - s_{jt}$.*

Proof. *See Appendix.*

4.1.3 Identification under Unobserved Market Size

Sometimes the definition of market size is not very clear, and can be inaccurately measured. Since market shares are defined to be the sales in that market over market size, this will lead market shares to be inaccurately measured as well. A commonly used solution to this problem is to try different values of market size.

In this paper, I assume the econometricians observe the following.

Assumption 4.3. *The actual market size lies in an interval $[d_{1t}, d_{2t}]$, where d_{1t} and d_{2t} are observed.*

Let the market size computed according to d_{2t} be s_t, then the above assumption will result in the following bound on r_t.

Lemma 4.5. *For any $j \neq 0$, $0 \leqslant r_{jt} \leqslant s_{jt}(\frac{d_{2t}}{d_{1t}} - 1)$ and $\frac{r_{1t}}{s_1 t} = \frac{r_{2t}}{s_{2t}} = \cdots = \frac{r_{Jt}}{s_{Jt}}$.*

Proof. *See Appendix.*

4.2 Sharp Characterization of the Identified Set

For notation simplicity, let $w_t = (x_t, z_t, s_t, d_t)$, $\theta = (\beta, \eta)$. First I introduce the concept of symmetric distribution function:

Definition 4.1. *Let $F(\cdot)$ be the joint distribution of a random vector, then $F(\cdot)$ is symmetric if its value does not change given arbitrary permutation of its arguments.*

Next I impose the symmetric assumption on the distribution of the random variables associate with the products in a market:

Assumption 4.4. *Let the true joint distribution of w_t be π_0, then π_0 is symmetric.*

This is because before the realization w_t, each product j does not contain any information other than the index j.

Then I introduce a class of conditional distribution functions:

Definition 4.2. *Let the support of r_t conditional on w_t be $\mathcal{R}_t(w_t)$, the class $\mathcal{C}_t(w_t)$ is defined as the set of all conditional distributions of r_t conditional on w_t with support $\mathcal{R}_t(w_t)$, such that, for any $\mu \in \mathcal{C}_t(w_t)$, the marginal distribution of $s_t + r_t$ under $\pi_0 \times \mu$ is symmetric.*

Here $\mathcal{R}_t(w_t)$ under different measurement error environments is defined in the lemmas from Section 4.1. Thus, as we can see from the lemmas, $\mathcal{R}_t(w_t)$ only depends on s_t and d_t and can be written as $\mathcal{R}_t(s_t, d_t)$.

Now with the help of the lemmas in Section 4.1, I will be able to characterize the identified set in the following definition.

Definition 4.3. *Suppose the parameter space is Θ, let the true joint distribution of (x_t, s_t, d_t) conditional on z_t be π_z*[13]*. The identified set Θ_0 is*

$$\Theta_0 = \Big\{\theta \in \Theta : \exists \mu \in \mathcal{C}_t(w_t), \mathbb{E}_{\pi_z \times \mu}\big[\sigma_j^{-1}(x_t, s_t + r_t; \theta)|z_{jt}\big] = 0\Big\}. \tag{11}$$

The above set is a sharp identified set in the sense that it exploits all the information available. The index j is not necessary here as it does not contain any more information than the observed variables already provide (it just serves as an indicator that it is the jth product in the market), hence moment restrictions at the product level suffice and adding moment conditions based on other products in the market will result in the same identified set.

For estimation and inference purposes we have to transform the above conditional moment conditions into unconditional moment conditions. One can find a set of optimal instruments $f(z_{jt}) \in \mathbb{R}^{d_z}$ from a rich enough class based on an existing procedure (e.g., Gandhi and Houde (2016)) and consider an equivalent representation of Θ_0:

Definition 4.4. *Suppose the parameter space is Θ. The identified set Θ_0 can also be represented as*

$$\Theta_0 = \Big\{\theta \in \Theta : \exists \mu \in \mathcal{C}_t(w_t), \mathbb{E}_{\pi_0 \times \mu}\big[f_j(z_{jt})\sigma_j^{-1}(x_t, s_t + r_t; \theta)\big] = 0\Big\}. \tag{12}$$

In GLS, they characterize the identified set as

$$\Theta_{GLS} = \Big\{\theta \in \Theta : \mathbb{E}\big[f_j(z_t)\underline{\sigma}_j^{-1}(x_t, s_t; \theta)\big] \leqslant 0 \leqslant \mathbb{E}\big[f_j(z_t)\bar{\sigma}_j^{-1}(x_t, s_t; \theta)\big]\Big\} \tag{13}$$

where $\underline{\sigma}_j^{-1}(\cdot)$ and $\bar{\sigma}_j^{-1}(\cdot)$ are some known functions. As discussed in Section 2, if there are

[13]Notice that, π_z can degenerate to fewer dimensions if z_t includes some of x_t

more than two products in the market, the identified set characterized by such moment inequalities is not sharp.

To see this from another point of view, as pointed out by Theorem 2.2 in Schennach (2014), the identified set characterized by (12) has an equivalent representation that consists of infinite number of moment inequalities, which implies that any finite number of moment inequalities will result in a larger identified set.

5 Estimation and Inference

5.1 Constructing Sample Objective Function

Let

$$g_j(w_t, r_t; \theta) = f_j(z_{jt})\sigma_j^{-1}(x_t, s_t + r_t; \theta)$$

the above identification strategy boils down to conduct estimation and inference for the identified set:

$$\Theta_0 = \{\theta \in \Theta : \exists \mu \in \mathcal{C}_t(w_t), \mathbb{E}_{\pi_0 \times \mu}[g_j(w_t, r_t; \theta)] = 0\} \tag{14}$$

where w_t is observed and r_t is a vector of latent variables.

We assume that, econometricians observe market $1, 2, \ldots, T$, each with product $1, 2, \ldots, J$. The observations across markets are i.i.d..

In order to obtain an extreme type estimator, notice that Θ_0 has the following equivalent representation:

$$\Theta_0 = \{\theta \in \Theta : \theta \in \arg\min_{\theta \in \Theta} \min_{\mu \in \mathcal{C}_t(w_t)} \|\mathbb{E}_{\pi_0 \times \mu}[g_j(w_t, r_t; \theta)]\|\} \tag{15}$$

The construction of an estimator then hinges on finding a sample objective function which uniformly converges to

$$\min_{\mu \in \mathcal{C}_t(w_t)} \|\mathbb{E}_{\pi_0 \times \mu}[g_j(w_t, r_t; \theta)]\|$$

However, the identified set is not characterized by moment inequalities, but moment equalities with latent variables in moment functions. Therefore the well developed moment inequality literature[14] cannot help us with this case. Also, notice that $\mathcal{C}_t(w_t)$ is an infinite dimensional space, therefore, searching over it in an actual application is not feasible and a method to reduce the dimension of the original problem is required.

Before introducing of the dimension reduction technique developed in this paper, I discuss several approaches in the literature that try to solve this problem.

The first method to deal with distributions of latent variable is introduced by Pakes and Pollard (1989) in which they assume a parametric family of conditional distributions $\tilde{\mathcal{C}}_t$ with finite dimension and estimate via a simulated method of moments objective function as below:

$$\min_{\mu \in \tilde{\mathcal{C}}_t} \|\mathbb{E}_{\pi \times \mu}[g_j(w_t, r_t; \theta)]\|$$

It is apparent that this method considers a much smaller class than the original problem, and thus will usually lead to an inconsistent estimator for the identified set.

Another approach is to exploit the fact that a duality of the above infinite dimensional optimization problem can be finite. There are two papers in the existing literature, Ekeland et al. (2010) and Schennach (2014), that consider this method. They propose two different dual formulations of the original problem which are finite dimensional. Using their methods, the identified set in (12) can be shown to have the following equivalent representation

$$\Theta_0 = \left\{\theta \in \Theta : \inf_{\gamma \in \mathbb{R}^{d_z}} \|\mathbb{E}_{\pi_0}[\tilde{g}_j(w_t; \theta, \gamma)]\| = 0\right\}$$

where $\tilde{g}_j(\cdot)$ is a known function.

Here we transformed the original infinite dimensional problem into one with finite dimension. We can therefore construct a sample objective function based on the population objective function as below:

$$\inf_{\gamma \in \mathbb{R}^G} \| \sum_{t=1}^{T} \sum_{j=1}^{J} \tilde{g}_j(w_t; \theta, \gamma)\|$$

[14]See Andrews and Soares (2010) for inference procedure with unconditional moment inequalities, Andrews and Shi (2013) for inference procedure with conditional moment inequalities, Andrews and Guggenberger (2009) for uniform valid inference procedure, and Bugni (2010) for a procedure using bootstrap method to determine critical values.

Both papers mentioned above introduced a way to transform the original objective function into one that is easy to handle. However, they both require i.i.d. sample. In the demand estimation problem, however, the sample in one market are always dependent. One way to get around with this is to treat each market t (instead of (j, t)) as one observation and only estimate using the market level variation by aggregating the moments across j in each market.

However, if the data has relatively small number of markets with large number of products in each market, i.e., a large J but a small T, then researchers have to treat (j, t) as an individual observation and have to consider asymptotic properties when J goes to infinity. In such a situation, if one proceeds with the above method without considering the possible dependence in one market, it generally will produce a very large confidence set.

In this paper, I propose a novel method to reduce the dimension of the above problem by taking advantage of the special property that the support of the latent variables $\mathcal{R}(w_t)$ is compact and convex[15]. The reduction consists of two steps, first step uses the fact that $\mathcal{R}(w_t)$ is compact and reduces the dimension from infinite to finite. Second step uses the fact that $\mathcal{R}(w_t)$ is convex and its image under $g(w_t, \cdot; \theta)$ is a convex set and reduces the dimension further.

Assumption 5.1. $\mathcal{R}(w_t)$ *is compact and convex.*

Assumption 5.1 holds under all environments I am considering in this paper, as can be seen from Section 4.

Now we consider the space of all signed measures $\mathcal{M}(\mathcal{R}(w_t))$ on the Borel σ-algebra $\mathcal{B}(\mathcal{R}(w_t))$ on set $\mathcal{R}(w_t)$. For any $\mu_1,\ \mu_2 \in \mathcal{M}(\mathcal{R}(w_t))$, and $a \in \mathbb{R}$, I use the following rule:

Definition 5.1. *For any* $B \in \mathcal{B}(\mathcal{R}(w_t))$, $(\mu_1 + \mu_2)(B) = \mu_1(B) + \mu_2(B)$ *and* $(a \cdot \mu_1)(B) = a \cdot \mu_1(B)$.

Under Definition 5.1, $\mathcal{M}(\mathcal{R}(w_t))$ is a linear space and the set of all probability measures $\mathcal{P}(\mathcal{R}(w_t))$ is a convex subset. Since $\mathcal{R}(w_t)$ is also compact, I have the following result:

[15] Ekeland et al. (2010) and Schennach (2014) consider a more general support which allows for nonconvex or noncompact set.

Lemma 5.1. *The set $\mathcal{P}(\mathcal{R}(w_t))$ is compact in its weak* topology and convex.*

Proof. See Appendix.

To apply my dimension reduction method, I do a transformation of the population objective function by using the law of iterated expectation:

$$\min_{\mu \in \mathcal{C}_t(w_t)} \|\mathbb{E}_{\pi_0 \times \mu}[g_j(w_t, r_t; \theta)]\| = \min_{\mu \in \mathcal{C}_t(w_t)} \|\mathbb{E}_{\pi_0}\left[\mathbb{E}_\mu[g_j(w_t, r_t; \theta)|w_t]\right]\| \tag{16}$$

It now suffices to characterize the set of all values of $\mathbb{E}_\mu[g_j(w_t, r_t; \theta)]$ for any $\mu \in \mathcal{P}(\mathcal{R}(w_t))$. Define a continuous linear operator $L : \mathcal{P}(\mathcal{R}(w_t)) \to \mathbb{R}^{d_z}$ such that for any $\mu \in \mathcal{P}(\mathcal{R}(w_t))$, $L\mu = \mathbb{E}_\mu[g_j(w_t, r_t; \theta)|w_t]$. Notice that the set I want to characterize is exactly $L(\mathcal{P}(\mathcal{R}(w_t)))$.

By Lemma 5.1 we have the following result about the image of $\mathcal{P}(\mathcal{R}(w_t))$: $L(\mathcal{P}(\mathcal{R}(w_t)))$.

Lemma 5.2. *The set $L(\mathcal{P}(\mathcal{R}(w_t)))$ is compact and convex.*

Proof. See Appendix.

Now we have the main result of the first step of the dimension reduction:

Proposition 5.1. *The population objective function in (16) has the the following equal representation:*

$$\min_{\mu \in \mathcal{E}_t(w_t)} \|\mathbb{E}_{\pi_0 \times \mu}[g_j(w_t, r_t; \theta)]\| = \min_{\mu \in \mathcal{E}_t(w_t)} \|\mathbb{E}_{\pi_0}\left[\mathbb{E}_\mu[g_j(w_t, r_t; \theta)|w_t]\right]\| \tag{17}$$

where $\mathcal{E}_t(w_t)$ is the set of all probability measures in $\mathcal{C}_t(w_t)$ with finite support $supp(\mu) \subset \mathcal{R}(w_t)$ and the number of points in $supp(\mu)$ is less than $d_z + 1$.

The idea is to apply Krein-Milman theorem which states a compact convex subset of a vector space can be represented as the closed convex hull of its extreme points. Here the extreme points of the set can be shown to be the points that correspond to all probability measures that have only a single point in their support (Dirac measure). Then according to Carathéodory's theorem each convex combination only has to contain at most $d_z + 1$ points, which completes the proof. The details of the proof can be found in appendix.

Without assumptions on the function $g_j(w_t, r_t; \theta)$ this is the furthest I can go. However, notice that

$$g_j(w_t, r_t; \theta) = f_j(z_{jt})\sigma_j^{-1}(x_t, s_t + r_t; \beta, \eta)$$

We can further exploit this structure to attain our second step dimension reduction. Notice that, for each $\mu \in \mathcal{E}_t(w_t)$, let $supp(\mu) = \{r_{t0}, r_{t1}, \ldots, r_{td_z}\}$ with assigned probability $\{p_0, p_1, \ldots, p_{d_z}\}$. Then

$$\mathbb{E}_\mu[g_j(w_t, r_t; \theta)|w_t] = \sum_{k=0}^{d_z} g_j(w_t, r_{tk}; \theta)$$

which is essentially a convex combination of points in $supp(\mu)$. Then it suffices to prove that the image of $g_j : \mathcal{R}(w_t) \to \mathbb{R}^{d_z}$, denoted as $g_j(\mathcal{R}(w_t))$, is a convex set.

Lemma 5.3. *$g_j(\mathcal{R}(w_t))$ is a convex set.*

The proof simply exploits the fact that g_j consists of two parts: a one dimensional mapping $\sigma_j^{-1}(x_t, s_t + r_t; \beta, \eta)$ and an affine transformation by multiplying a vector $f_j(z_{jt})$. The details of proof can be found in appendix. Now I introduce a new class of functions:

Definition 5.2. *Let $\mathcal{F}$ be the collection of functions that map w_t to $r_t \in \mathcal{R}(w_t)$ that satisfies, for any $r(\cdot) \in \mathcal{F}$, we have $\mu \in \mathcal{C}_t(w_t)$. Here μ is defined as: $\mu(r_t) = 1$ if $r_t = r(w_t)$ and $\mu(r_t) = 0$ otherwise.*

Then I have the main result of second step of dimension reduction:

Proposition 5.2. *The population objective function in (16) has the following equal representation:*

$$Q(\theta) = \min_{r(\cdot) \in \mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|. \tag{18}$$

The above dimension reduction technique fully exploits the fact that the image of $g_j(w_t, \cdot; \theta)$ is convex and compact under the setup considered in this paper. It does not have a general extension to other environments. For example, suppose the image of $g_j(w_t, \cdot; \theta)$ is not convex, as shown in Figure 2, then the image of the linear operator $\mathbb{E}_\mu g_j(w_t, r_t; \theta)$ taken μ as input can be shown in Figure 3 and clearly just considering the set in Figure 2 is not enough.

Based on the above population objective function this paper proposes the following sample objective function:

$$\hat{Q}_{JT}(\theta) = \min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t; \theta)\| \tag{19}$$

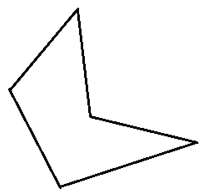

Figure 2: Image of $g_j(w_t, \cdot; \theta)$

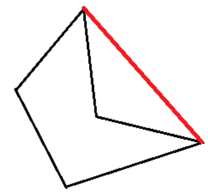

Figure 3: image of $\mathbb{E}_\mu[g_j(w_t, r_t; \theta)]$

5.2 Estimation

Define

$$\hat{\Theta} = \{\theta \in \Theta : \hat{Q}_{JT}(\theta) \leqslant C\}$$

where C is a constant.

In this section, we will prove given a choice of C, $d_H(\hat{\Theta}, \Theta_0) \xrightarrow{p} 0$, where d_H is the Hausdorff distance defined as:

$$d_H(\hat{\Theta}, \Theta_0) = \max\Big\{\sup_{\theta \in \hat{\Theta}} \inf_{\theta' \in \Theta_0} \|\theta - \theta'\|, \sup_{\theta' \in \Theta_0} \inf_{\theta \in \hat{\Theta}} \|\theta - \theta'\|\Big\} \tag{20}$$

We first discuss assumptions on data, then move on to different asymptotic results. Suppose econometricians observe $(x_{jt}, z_t, s_{jt}, d_{jt})$, $j = 1, \dots, J$, $t = 1, \dots, T$. I have the following assumption:

Assumption 5.2. *$(x_{jt}, z_t, s_{jt}, d_{jt}, \xi_{jt})$ are independent and identically distributed across t.*

Assumption 5.2 asks that the product characteristics (observed and unobserved), instrumental variables and market shares together with indicator of measurement error be independent across markets. However, it can be shown that the following analysis under an extension that allows weak dependence across markets would also hold.

We establish the results of this section in two types of asymptotic, $T \to \infty$ and $J \to \infty$. The reason is that researchers typically encounter two types of data structures: large number of markets, or small number of markets with large number of products in each market. In the former one, T is much larger than J therefore fixing J and let T goes to infinity is a suitable way to characterize asymptotic result and vice versa. In what follows, if the asymptotics are in T, I use $\hat{Q}_T$ to denote $\hat{Q}_{JT}$, and if the asymptotics are in J, I use $\hat{Q}_J$ to denote $\hat{Q}_{JT}$.

Lemma 5.4. *For any $\theta \in \Theta$, fixing J and when $T \to \infty$, we have*

$$\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t; \theta)\| \xrightarrow{p} \min_{r(\cdot) \in \mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|$$

Proof. *See Appendix.*

To obtain uniform convergence we need the following stochastic equicontinuity condition.

Assumption 5.3. $\{\hat{Q}_T(\theta) : T \geqslant 1\}$ *is stochastic equicontinuous on* Θ*, i.e.,* $\forall \varepsilon > 0$*,* $\exists \delta > 0$ *such that*

$$\limsup_{T \to \infty} P(\sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} |\hat{Q}_T(\theta) - \hat{Q}_T(\theta')| > \varepsilon) < \varepsilon$$

Assumption 5.3 is a high level assumption, I verify it for the logit model in the appendix, which is the most widely used model.

Now I apply the result from Chernozhukov et al. (2007) and construct the set estimator as:

$$\hat{\Theta} = \{\theta \in \Theta : Q_T(\theta) \leqslant \hat{c}\}$$

where $\hat{c} \geqslant \sup_{\theta \in \Theta_0} Q_T(\theta)$ with probability approaching 1 and $\hat{c}/T \xrightarrow{p} 0$ is a data dependent number.

We can then apply Theorem 3.1 from Chernozhukov et al. (2007) to establish the following proposition:

Proposition 5.3. *Assume (1) The parameter space* Θ *is a nonempty compact subset of* $\mathbb{R}^{d_\Theta}$*, (2)* $g_{jt}(\cdot, \cdot; \theta)$ *is a measurable function, and (3) Assumptions 5.1, 5.2, and 5.3 hold.*

Then $\Theta_0 \subset \hat{\Theta}$ *with probability approaching 1 when* $T \to \infty$ *and*

$$d_H(\hat{\Theta}, \Theta_0) = o_p(1).$$

where d_H *is the Hausdorff metric.*

Proof. *See Appendix.*

In the actual application, I use the following sample objective function for efficiency:

$$\hat{Q}_T(\theta) = \|W_T(\theta)^{-\frac{1}{2}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t^*; \theta)\| - \inf_{\theta \in \Theta} \|W_T(\theta)^{-\frac{1}{2}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t^*; \theta)\|$$

where

$$r_t^* \in \arg\min_{r_t \in \mathcal{R}_t} \|\frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t; \theta)\|$$

and

$$W_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \Big(\sum_{j=1}^{J} g_j(w_t, r_t^*; \theta) \Big) \Big(\sum_{j=1}^{J} g_j(w_t, r_t^*; \theta) \Big)'.$$

The case for asymptotic theory when $J \to \infty$ and T is fixed is much more complicated and demands more regularity assumptions[16]. In what follows, I adapt the assumptions and results from Berry et al. (2004) to the partial identification setting.

We first establish uniform convergence results for $\theta \in \Theta_0$ and extend the the assumptions from Berry et al. (2004) into the set identification setup about behavior of $\hat{Q}_J(\theta)$ outside of the identified set.

Assumption 5.4. *ξ_{jt} is independent and identically distributed across j.*

Assumption 5.4 can be extended to a more general setup allowing for week dependence, this is because the estimation hinges on the exclusion restrictions $\mathbb{E}[\xi_{jt}|z_{jt}] = 0$.

Notice that the inversion $\sigma_j^{-1}(x_t, s_t + r_t; \theta)$ is now possibly dependent across j. We know that when evaluated at true parameter value θ_0, together with the true distribution μ_0 of r_t, we can regenerate

$$\mathbb{E}_{\mu_0}[\sigma_j^{-1}(x_t, s_t + r_t; \beta, \eta)|w_t] = \xi_{jt}$$

almost surely, which is an independent sequence based on Assumption 5.4.

Assumption 5.4 introduces new information into the data generating process, which can help us to make the identified set smaller. Now I define the new identified set as follows.

[16] Armstrong (2016) shows that some supply side models can make BLP instruments lose their identification power when J goes to infinity. Since this paper does not specify a supply side model, I assume that BLP instruments considered here do not suffer from this situation.

Definition 5.3. *Suppose the parameter space is* Θ*. Consider the following set* $\tilde{\Theta}$

$$\tilde{\Theta} = \left\{ \theta \in \Theta : \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t);\theta)]\| = 0 \right\}$$

For any $\theta \in \tilde{\Theta}$*, let* S^* *be the set of solutions to*

$$\min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t);\theta)]\|$$

The the identified Θ_0 *is defined as:*

$$\Theta_0 = \left\{ \theta \in \tilde{\Theta} : \exists r(\cdot) \in S^*, \text{ such that, } g_j(w_t, r(w_t);\theta) \text{ are mutually independent across } j \right\}.$$

Now we have the following result which states the convergence inside the identified set.

Lemma 5.5. *For any* $\theta \in \Theta_0$*, fixing* T *and when* $J \to \infty$*, we have*

$$\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J} g_j(w_t, r_t;\theta)\| \xrightarrow{p} \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t);\theta)]\|$$

Proof. *See Appendix.*

Similar to the case when T goes to infinity, we need the following condition to restrict the richness of the empirical process to ensure uniform convergence inside identified set.

Assumption 5.5. $\{\hat{Q}_J(\theta) : J \geqslant 1\}$ *is stochastic equicontinuous on* Θ*:* $\forall \varepsilon > 0$*,* $\exists \delta > 0$ *such that*

$$\limsup_{J\to\infty} P(\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |\hat{Q}_J(\theta) - \hat{Q}_J(\theta')| > \varepsilon) < \varepsilon$$

Assumption 5.5 is a high level assumption, similar to Assumption 5.3, I verify this for logit model in the appendix.

The behavior of $\sigma_j^{-1}(x_t, s_t + r_t;\theta)$ when $\theta \notin \Theta_0$ clearly depends on the joint distribution of w_t (this includes endogenous variables such as price), which is impossible to discuss without a fully specified supply side model. Similar to Berry et al. (2004), I impose high level regularity assumptions without specifying a supply side model to ensure consistency. Specifically, since it maybe too strong to assume that $\hat{Q}_J(\theta)$ to even converge outside of Θ_0, I adopt assumption A6 in Berry et al. (2004) into this paper to make sure it is asymptotically bounded away from zero.

Assumption 5.6. *Let $\mathcal{N}(\Theta_0, \varepsilon) = \{\theta \in \Theta, \inf_{\theta' \in \Theta_0} \|\theta - \theta'\| \leqslant \varepsilon\}$. For all $\varepsilon > 0$, there exists $C(\varepsilon)$ such that*

$$\lim_{J \to \infty} P(\inf_{\theta \notin \mathcal{N}(\Theta_0, \varepsilon), \theta' \in \Theta_0} \|\hat{Q}_J(\theta) - \hat{Q}_J(\theta)\| \geqslant C(\varepsilon)) = 1$$

The above condition relates to the behavior of $\hat{Q}_J(\theta)$ for θ outside of Θ_0, which is impossible to verify without fully specifying a supply side model. As mentioned earlier, a supply side model on how endogenous variables (e.g., prices) are generated is not specified in this paper because different measurement error environment requires different supply side model. Hence, this paper does not provide verification of this high level assumption regarding a specific supply side model, similar to Berry et al. (2004).

Now the consistency of a set estimator

$$\hat{\Theta} = \{\theta \in \Theta : Q_J(\theta) \leqslant \hat{c}\}$$

follows by slightly adjusting the proofs in Chernozhukov et al. (2007), where $\hat{c} \geqslant \sup_{\theta \in \Theta_0} Q_J(\theta)$ with probability approaching 1 and $\hat{c}/J \xrightarrow{p} 0$ is a data dependent number.

Proposition 5.4. *Assume (1) The parameter space Θ is a nonempty compact subset of $\mathbb{R}^{d_\Theta}$, (2) $g_{jt}(\cdot, \cdot; \theta)$ is a measurable function, and (3) Assumptions 5.1, 5.2, 5.4, 5.5 and 5.6 hold.*

Then $\Theta_0 \subset \hat{\Theta}$ with probability approaching 1 when $J \to \infty$ and

$$d_H(\hat{\Theta}, \Theta_0) = o_p(1).$$

where d_H is the Hausdorff metric.

Proof. *See Appendix.*

Similarly, I use the following sample objective function for efficiency:

$$\hat{Q}_J(\theta) = \|W_J(\theta)^{-\frac{1}{2}} \frac{1}{JT} \sum_{j=1}^{J} \sum_{t=1}^{T} g_j(w_t, r_t^*; \theta)\| - \inf_{\theta \in \Theta} \|W_J(\theta)^{-\frac{1}{2}} \frac{1}{JT} \sum_{j=1}^{J} \sum_{t=1}^{T} g_j(w_t, r_t^*; \theta)\|$$

where

$$r_t^* \in \arg\min_{r_t \in \mathcal{R}_t} \| \frac{1}{JT} \sum_{j=1}^{J} \sum_{t=1}^{T} g_j(w_t, r_t; \theta) \|$$

and

$$W_J(\theta) = \frac{1}{J} \sum_{j=1}^{J} \Big(\sum_{t=1}^{T} g_j(w_t, r_t^*; \theta) \Big) \Big(\sum_{t=1}^{T} g_j(w_t, r_t^*; \theta) \Big)'.$$

5.3 Inference

Now we consider the confidence set for the identified set Θ_0 and true value θ_0. Here I use QLR test statistics

$$QLR(\theta) = \min_{r_t \in \mathcal{R}_t(w_t)} \| W_T^{-\frac{1}{2}}(\theta) \frac{1}{\sqrt{JT}} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t; \theta) \|^2. \tag{21}$$

The confidence set for θ_0 is constructed as

$$CR(\Theta_0) = \{\theta \in \Theta : QLR(\theta) \leqslant \hat{C}(1-\alpha)\} \tag{22}$$

We seek to find $\hat{C}(1-\alpha)$ such that,

$$\liminf_{T \to \infty} P(\Theta_0 \subset CR(\Theta_0)) \geqslant 1 - \alpha \tag{23}$$

and

$$\liminf_{J \to \infty} P(\Theta_0 \subset CR(\Theta_0)) \geqslant 1 - \alpha \tag{24}$$

where $1-\alpha$ is the desired coverage probability.

The idea is to apply procedure developed by Chernozhukov et al. (2007) and Romano and Shaikh (2010) for the choice of $\hat{C}(1-\alpha)$ via subsampling.

We have the convergence result followed directly from the estimation part.

Lemma 5.6. $\sup_{\theta \in \Theta_0} QLR_{TJ}(\theta)$ *converges in distribution to a nondegenerate and continuous distribution when either* $T \to \infty$ *or* $J \to \infty$*.*

Now consider the following subsampling procedure similar to Romano and Shaikh (2008):

1. Let $S = \Theta$. If $\sup_{\theta \in S} QLR_{JT}(\theta) \leqslant \hat{C}(S, 1-\alpha)$, then accept all hypotheses and stop, here $\hat{C}(S_1, 1-\alpha)$ is an estimator of $1-\alpha$ quantile of $\sup_{\theta \in S} QLR_{JT}(\theta)$, which will be discussed later.

2. If not, then set $S = \{\theta \in \Theta : QLR_{JT}(\theta) \leqslant \hat{C}(S, 1-\alpha)\}$ and repeat step 1.

The above procedure may be computationally infeasible in practice, then one can always use the $1-\alpha$ quantile of $\sup_{\theta \in \hat{\Theta}} QLR_{JT}(\theta)$ as $\hat{C}$ similar to Chernozhukov et al. (2007), where $\hat{\Theta}$ is the set estimator in Section 5.2.

Applying Theorem 2.1 from Romano and Shaikh (2010) we have that (23) and (24) holds for the above procedure.

Now I introduce a subsampling procedure[17] to estimate $\hat{C}(S, 1-\alpha)$. To do this, we need some more notations:

For each θ, let

$$r_t(\theta) \in \arg \min_{r_t \in \mathcal{R}_t(w_t)} \| \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t; \theta) \|$$

and let

$$QLR^*_{JT}(\theta) = \| \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} g_j(w_t, r_t(\theta); \theta) \| = QLR(\theta)$$

Let $b = b_T < T$ or $b = b_J < J$ be

For the case when $T \to \infty$, let $b_T < T$ be a sequence of positive integers tending to infinity with $b_T/T \to 0$ as $T \to \infty$. Let $N_T = \binom{T}{b_T}$ and let

$$QLR^*_{JT,b_T,i}(\theta) = \| \frac{1}{Jb_T} \sum_{t=1}^{b_T} \sum_{j=1}^{J} g_j(w_t, r_t(\theta); \theta) \|$$

be the QLR statistic evaluated at the ith subset of data with size b_T. Notice that such subsampling draw is at market level. For any set $S \subset \Theta$, the estimator is defined as

$$\hat{C}(S, 1-\alpha) = \inf \left\{ x : \frac{1}{N_T} \sum_{i=1}^{N_T} 1\{\sup_{\theta \in S} QLR^*_{JT,b_T,i}(\theta) \leqslant x\} \geqslant 1-\alpha \right\}$$

[17] Bugni (2010) pointed out that naive bootstrap does not work for moment inequality type objective function and proposed a fix that hinges on the special structure of moment inequalities. It is however not clear that his method can be generalized to our environment, which can be left as future research.

For the case when $J \to \infty$ but still with significant size in T, it is still recommended to draw at market level according to the above procedure. The reason is that, observation at market level is always assumed to be independent while the observation at product level may not. We had assumptions in the previous section to warrant the correlation between products inside a market going to zero when J tends to infinity, however in finite sample this may still behave worse than drawing at market level. If T is not large enough, then one could proceed to draw subsample at product level based on the procedure below.

For the case when $J \to \infty$, let $b_{JT} < JT$ be a sequence of positive integers tending to infinity with $b_{JT}/JT \to 0$ as $J \to \infty$. Let $N_{JT} = \binom{JT}{b_{JT}}$ and let

$$QLR^*_{JT,b_{JT},i}(\theta) = \| \frac{1}{b_{JT}} \sum_{i=1}^{b_{JT}} g_i(w_t, r_t(\theta); \theta) \|$$

be the QLR statistic evaluated at the ith subset of data with size b_T. Notice that such subsampling draw is at market level. For any set $S \subset \Theta$, the estimator is defined as

$$\hat{C}(S, 1-\alpha) = \inf \left\{ x : \frac{1}{N_T} \sum_{i=1}^{N_T} 1\{\sup_{\theta \in S} QLR^*_{JT,b_{JT},i}(\theta) \leqslant x\} \geqslant 1-\alpha \right\}$$

We can also construct confidence set for the true parameter value θ_0 following Romano and Shaikh (2008), the idea is to exploit the duality of hypothesis testing and confidence set. I still use the above QLR statistics and the confidence region is defined as

$$CR(\theta) = \{\theta \in \Theta : H_0 : \ \theta_0 = \theta \text{ cannot be rejected}\} = \{\theta \in \Theta : QLR(\theta) \leqslant \hat{C}(1-\alpha)\}$$

The estimator for the $1-\alpha$ quantile is similar as above via a subsampling procedure:

For the case when $T \to \infty$, let $b_T < T$ be a sequence of positive integers tending to infinity with $b_T/T \to 0$ as $T \to \infty$. Let $N_T = \binom{T}{b_T}$ and let

$$QLR^*_{JT,b_T,i}(\theta) = \| \frac{1}{Jb_T} \sum_{t=1}^{b_T} \sum_{j=1}^{J} g_j(w_t, r_t(\theta); \theta) \|$$

be the QLR statistic evaluated at the ith subset of data with size b_T. Notice that such subsampling draw is at market level. For any set $S \subset \Theta$, the estimator is defined as

$$\hat{C}(1-\alpha) = \inf\left\{x : \frac{1}{N_T}\sum_{i=1}^{N_T} 1\{QLR^*_{JT,b_T,i}(\theta) \leqslant x\} \geqslant 1-\alpha\right\}$$

For the case when $J \to \infty$, let $b_{JT} < JT$ be a sequence of positive integers tending to infinity with $b_{JT}/JT \to 0$ as $J \to \infty$. Let $N_{JT} = \binom{JT}{b_{JT}}$ and let

$$QLR^*_{JT,b_{JT},i}(\theta) = \|\frac{1}{b_{JT}}\sum_{i=1}^{b_{JT}} g_i(w_t, r_t(\theta); \theta)\|$$

be the QLR statistic evaluated at the ith subset of data with size b_T. Notice that such subsampling draw is at market level. For any set $S \subset \Theta$, the estimator is defined as

$$\hat{C}(1-\alpha) = \inf\left\{x : \frac{1}{N_T}\sum_{i=1}^{N_T} 1\{QLR^*_{JT,b_{JT},i}(\theta) \leqslant x\} \geqslant 1-\alpha\right\}$$

In practice, researchers are often interested in some particular parameters than full vector of them. In such case, I use the projection of the full confidence region to the parameters of interest to determine the confidence interval for that parameter.

6 Monte-Carlo Simulation

To see the performance of the method under finite sample, I do two simulation studies. Only measurement error caused by out-of-stock event is considered here. In the first one I set up a very simple environment such that I can compute the identified set analytically to determine the coverage probability of the confidence set. In the second simulation I set up an environment which is very similar to the empirical research I conducted later, the identified set is difficult to compute but the coverage probability of confidence set for true parameter value is provided. At the same time it can shed some light on the results from the empirical research.

6.1 One Product in Each Market

Suppose there is only one product in each market. The utility function of consumer i in market t is

$$u_{it} = -5 + x_{1t} + \xi_{1t} + \varepsilon_{i1t}$$

where x_{1t} and ξ_{1t} are independent with each other and i.i.d. normally distributed with mean zero and variance one. The distribution for ε_{i1t} is assumed to be independent across i and t and following a Type I extreme value distribution. The choice probability for the product is then

$$\sigma_{1t} = \frac{\exp(-5 + x_{1t} + \xi_{1t})}{1 + \exp(-5 + x_{1t} + \xi_{1t})}$$

and the choice probability for the out side option is

$$\sigma_{0t} = \frac{1}{1 + \exp(-5 + x_{1t} + \xi_{1t})}$$

If there is no out-of-stock event observed in market t, let $s_t = \sigma_t$. In the situation of an out-of-stock event, I let $s_{1t} = 0.1 \times \sigma_{1t}$ and $s_{0t} = 1 - s_{1t}$.

I generated 1,000 markets according to the above data generating process with different out-of-stock rate. For example, if out-of-stock rate is set at 5%, then a total of 50 markets are observed stock-out and has $d_{1t} = 1$.

Here we are interested in the coefficient of x_{1t}. The true value is 1 as can be seen from the data generating process, the identified set can be computed analytically as $[0.95, 1]$ when out-of-stock rate is 5% and $[0.9, 1]$ when out-of-stock rate is 10%.

We generate BLP confidence interval for the true parameter value and robust confidence interval for the identified set based on 5% and 10% out-of-stock rate, the coverage probability can be seen from table 1. Here the first column presents the coverage probability of the BLP confidence interval for the true parameter value, and the second column presents the coverage probability of the robust confidence interval proposed by this paper for the identified set.

The simulation results imply that, BLP confidence interval only covers true parameter value 86.8% and 67.2% of the time when out-of-stock rate is 10% and 5% respectively, while the correct coverage probability should be controlled at 95%.

OOS Rate	CI of BLP Estimator	CI for Identified Set
5%	86.8%	94.5%
10%	67.3%	95.2%

Table 1: Coverage probability of 95% confidence region

However, the robust inference procedure suggested by this paper provides the correct coverage probability for the identified set.

6.2 Twenty Products in Each Market

In this subsection I simulate an environment which is close to the empirical study I will be presenting later. Suppose there are twenty products in each market. The utility function of consumer i for product j in market t is

$$u_{ijt} = -5 - p_{jt} + q_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

Here p_{jt} plays the role of an endogenous variable that is possibly correlated with ξ_{jt} and q_{jt} is exogenous. In addition to the above variables, I generate an instrumental variable z_{jt} that does not enter consumer's utility function. q_{jt}, z_{jt} and ξ_{jt} are mutually independent and i.i.d. normal distributed with mean zero and variance one. The endogenous variable p_{jt} is generated according to

$$p_j = 5z_j + \xi_j$$

The choice probability for product j is

$$\sigma_{jt} = \frac{\exp(-5 - p_{jt} + q_{jt} + \xi_{jt})}{1 + \sum_{j=1}^{J} \exp(-5 + p_{jt} + q_{jt} + \xi_{jt})}$$

If there is no out-of-stock event observed in market t, let $s_t = \sigma_t$. In the situation of an out-of-stock event, if product j is out-of-stock, $s_{jt} = 0.1 \times \sigma_{jt}$, let

$$W = \sum_{j \text{ out-of-stock}} 0.9\sigma_{jt}$$

and for product j' that is not out-of-stock, $s_{jt} = \sigma_{jt}(1 + W)$.

OOS Rate	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
CI of BLP Estimator	88.9%	71.7%	47.1%	30.1%	14.4%	5.3%	2.3%	0.6%	0.1%	0.1%
CI of Robust Inference	96.4%	94.7%	96.2%	97.4%	99.8%	100%	100%	100%	100%	100%

Table 2: Coverage probability of 95% confidence interval

I generated 100 markets according to the above data generating process with different out-of-stock rate. For example, if out-of-stock rate is set at 5%, then a total of 5 markets are observed stock-out and has $d_{1t} = 1$. In each of such market, I make 5 out of 20 products stock-out.

Here we are interested in the coefficient of p_{jt}. The true value is 1 as can be seen from the data generating process. The identified set cannot be obtained analytically here and is therefore omitted from this study. However I use the confidence region for the true parameter value instead to see the validity of the method.

We generate BLP confidence interval for the coefficient and robust confidence region for the true parameter value, the coverage probability can be seen from table 2.

Here the first row presents the coverage probability of the BLP confidence interval for the true parameter value, and the second column presents the coverage probability of the robust confidence interval proposed by this paper for the true parameter value.

The simulation results imply that BLP confidence interval severely undercovers the true parameter value even at very low out-of-stock rate (for example, at 2% out-of-stock rate it should provide 95% coverage probability while now the coverage probability is only 71.7%). However, the robust inference procedure suggested by this paper provides the correct coverage probability for the true parameter value.

The reason that the confidence interval for the robust inference procedure always covers the true parameter value when out-of-stock rate is large is that, under this specific setup of data generating process, the true parameter value is always far away from the boundary of the identified set when out-of-stock rate is large. While the we require the confidence region to cover each point in the identified set with at least 95% probability, the true parameter value will always be covered by such set I constructed (in fact, such set always converges to the identified set).

7 Empirical Study

This paper applies the above identification and inference procedure to the Nielsen scanner data set. The scanner data set contains weekly pricing, sales volume generated by point-of-sale systems from more than 90 participating retail chains across all US markets. It also includes a specific store's geographic information as well as product characteristics for each product (brand, size, etc).

I choose 78 stores from the data set with relatively large volume of sales throughout the year therefore sampling error due to small market size is not a concern. In addition, I choose them such that there are least number of competitive stores around, which means that the stores are the only place to serve the entire market. I hence define each store as a single market.

The data set contains weekly sales from 2006-2015, here I randomly choose one week in 2011 for the empirical study to simplify the construction of moment conditions and avoid possible dependence between markets.

Among all the product categories I choose shampoo as the interest of study. It is due to the following reasons: First, it has a very simple category: shampoo, which means basically consumers can only substitute inside this category. If the desired choice of shampoo stocks-out, a consumer will most likely choose other brands of shampoo, while for example, if a consumer wants to buy orange juice, and the best choice stocks-out, he or she may switch to tea other than different brands of orange juice. Second, it has relatively large volume of sales and it does not subject to seasonality, which means the sales are relatively stable throughout the year and makes it easier to pin down out-of-stock events. Third, based on a survey data by Efficient Consumer Response (2013), many consumers' decision on shampoo when their first best choice runs out-of-stock is to get a second one rather than waiting for it to restock.

Many of the 78 stores hold more than 50 different types of shampoo throughout the year 2011, here I choose the top 20 brands in that store in year 2011, which account for 30% to 70% shares of shampoo in that market. The rest of the products not considered here are treated as outside option.

Some of the top 20 products are completely out-of-stock for some of the weeks so they

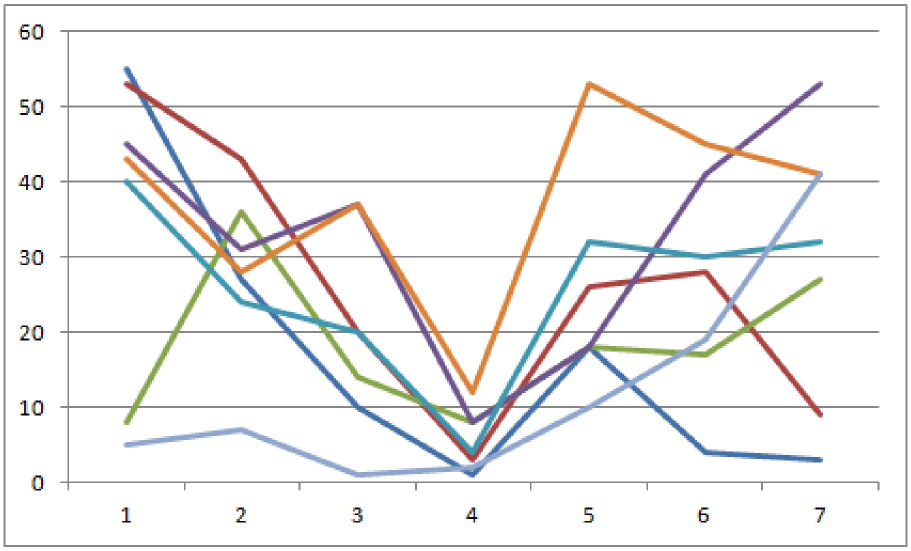

Figure 4: Weekly sales of OOS products three weeks before and after

are excluded from the choice set as well. As a result the sample size is 1,758 at product level. Among the 1,758 observations, some of the products can be treated as out-of-stock in that week due to very simple reasons. For example, the weekly sales of one product in 2011 generally ranges from 40-50, but in that specific week it drops to 5, and the week before or after its still 40-50. Such patterns can be seen in figure 4, in which I selected several products that display such phenomenon. The vertical axle represents weekly sales, and horizontal axle represents weeks. Each colored line represents one product, and week 4 is the one considered in the estimation. It can be seen that the products considered here has an abnormal low sales in week 4. It is also unlikely that another product is on sale that causes this, because the rest of the products in that market does not display such an abnormal pattern.

To build a more rigorous out-of-stock detection rule, I used a method called POS estimation technology summarized by Gruen et al. (2002) and Gruen and Corsten (2007), which according to their paper, is widely used by commercial ventures, consultants and academics. The idea is to find the sales pattern of one product from past sales, and generate prediction for sales of a particular point, and if it does not match the lower bound then it will be marked as out-of-stock. The full version of the algorithm takes many factors into account, for example, seasonality. Here due to the simple data structure and data type I have chosen, I use the simplified version by Grubor and Milicevic (2015): Compute the mean and standard

RESULTS FROM STANDARD ESTIMATORS

		OLS		IV		Random Coefficient	
	Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Means	Constant	-4.4557	-4.5709	-4.2755	-4.3227	-3.7583	-4.2297
		(0.0564)	(0.1020)	(0.0700)	(0.1244)	(0.1458)	(0.1783)
	Price	-0.0352	-0.0511	-0.3125	-0.4608	-1.8613	-0.5161
		(0.0203)	(0.0284)	(0.0837)	(0.1009)	(0.1332)	(0.1238)
	Size	0.0205	0.0260	0.0495	0.0831	-0.1467	0.0801
		(0.0031)	(0.0043)	(0.0109)	(0.0163)	(0.0206)	(0.0267)
Standard	Price	-	-	-	-	1.1170	0.1469
Deviations		-	-	-	-	(0.9294)	(0.1320)
	Size	-	-	-	-	0.6012	0.2174
		-	-	-	-	(0.5523)	(0.1721)

(i),(iii),(v) are results without brand dummies.
(ii),(iv),(vi) are results with brand dummies.

Table 3: Standard Estimators

deviation of the weekly sales in 2011, an out-of-stock event is marked if and only if the sales of that week is lower than mean minus 2 times standard deviation.

It is however worth noting that, there might be false positives and negatives. A product may not be out-of-stock and marked as out-of-stock, and a product that is actually out-of-stock may be omitted. Gruen and Corsten (2007) claims that this method was proved to have 80%-90% accuracy in most of the studies. Hence it is not a big problem here.

After applying the above out-of-stock detection method, I found 23 products in 6 markets are out-of-stock during the week. This implies we have to apply the robust inference procedure to all the products in the 6 markets. This is because even for products without out-of-stock issue, their market shares are contaminated by spillover demand. In total, the contaminated sample consists 7.7% of the total sample.

The results from standard estimator can be found in table 3, where I compare OLS, IV and random coefficient logit estimator. It can be seen that IV estimator provides significant bias correction.

I then construct confidence intervals for the true parameter value of price elasticity as described in Section 5 to compare it with confidence intervals constructed from standard estimators. The results can be found in table 4. The robust version of confidence interval covers the the random coefficient interval because it can always be the case that last consumer

Comparing Standard BLP and Robust CI

Variable	OLS	IV	Random Coefficient	Robust
Constant	[-4.7709,-4.3710]	[-4.5666,-4.0788]	[-4.5792,-3.8802]	[-4.62,-3.53]
Price	[-0.1068,0.0045]]	[-0.6585,-0.2630]	[-0.7587,-0.2735]	[-0.79,-0.13]
Size	[0.0176,0.0344]	[0.0510,0.1151]	[0.0278,0.1324]	[0.01,0.15]

Table 4: 95% Confidence Interval

who wants the item takes the last one on the shelf. If one thinks this is not the case in the actual application, it is always feasible to adjust set $\mathcal{R}(w_t)$ to reflect it and shrink the confidence interval.

8 Conclusion

Most modern demand estimation approaches involving aggregate data depend on a process by which observed market shares are equated with the choice probabilities generated via a discrete choice model. For example, the most commonly used method suggested by Berry et al. (1995) assumes a random coefficient logit model and reverts the above equality to construct moment conditions.

However, in many situations, there are significant measurement errors in market shares. This paper discusses three possible situations: An unobserved choice set variation caused by out-of-stock events, sampling error, and unobserved market size. It can be shown that the estimators are very sensitive to the measurement error in market shares such that even small errors can lead to significant bias in estimation and inference.

The measurement error in demand estimation is not a traditional measurement error in dependent variables situation because the unobserved shock is multi-dimensional and there is always special dependence structure among variables within the same market. Due to the above reason, point identification is not an easy task. The past studies on out-of-stock literature have typically relied on strong assumptions relating to the discrete choice model to point identify the demand function. However, although such approaches have eliminated the bias caused by stock-outs, all of the existing methods have failed to address the endogenous price problem within their modified models, which is a critical aspect of demand estimation.

This paper proposes to employ the (random coefficient) logit model while acknowledging measurement error exist which leads to the failure of point identification. Instead, I look for an identified set under different types of measurement error, as suggested by this paper. One way to achieve this is to use the classical characterization of (conditional) moment inequalities. Gandhi et al. (2013) develops moment inequalities restrictions to deal with the measurement error caused by sampling error. However, as shown in this paper, a finite number of moment inequalities can never characterize a sharp identified set.

This paper proposes a new method that uses moment equalities with latent variables to characterize the sharp identified set. Such a set is not easy to estimate as the question requires to solve an infinite dimensional optimization problem. Existing methods in the literature (e.g., Schennach (2014)) propose to consider a finite duality; however, this cannot be readily generalized in the demand estimation setup due to the special dependence structure within a market. This paper proposes another dimension reduction technique that exploits the convexity nature of the model and works with such dependence.

The results of the Monte Carlo simulation and an empirical study on data acquired from the retail industry suggested the standard estimator, which ignores the measurement error, is inconsistent and severely undercovers the true parameter value. However the robust inference procedure achieves the correct coverage probability and provides a significant correction to the estimators.

References

Aguirregabiria, V. (2005). Retail stockouts and manufacturer brand competition. *Working Paper*.

Andrews, D. and P. Guggenberger (2009). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory 25*, 669–709.

Andrews, D. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica 81*, 609–666.

Andrews, D. and G. Soares (2010). Inference for parameters defined by moment ineqaulities using generalized moment selection. *Econometrica 78*, 119–157.

Armstrong, T. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica 84*, 1961–1980.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand Journal of Economics 25*, 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 63*, 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.

Bruno, H. and N. Vilcassim (2008). Research note-structural demand estimation with varying product avalability. *Marketing Science 27*(6), 1126–1131.

Bugni, F. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica 78*, 735–753.

Campo, K., E. Gijsbrechts, and P. Nisol (2004). Dynamics in consumer response to product unavailability: Do stock-out reastions signal response to permanent assortment reductions? *Journal of Business Research 57*, 834–843.

Che, H., X. Chen, and Y. Chen (2012). Investigating effects of out-of-stock on consumer stockkeeping unit choice. *Journal of Marketing Research 49*, 502–513.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometrics models. *Econometrica 75*, 1243–1284.

Ching, A., F. Hayashi, and H. Wang (2015). Quantifying the impacts of limited supply: the case of nursing homes. *International Economic Review 56*, 1291–1322.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economics Journal: Microeconomics 5*, 1–30.

Efficient Consumer Response (2013). *Optimal Shelf Availability*.

Ekeland, I., A. Galichon, and M. Henry (2010). Optimal transportation and the falsifiability of incompletely specified economic models. *Economic Theory 42*, 355–374.

Galichon, A. and M. Henry (2013). Dilation bootstrap. *Journal of Econometrics 177*, 109–115.

Gandhi, A. and J. Houde (2016). Measuring substituion patterns in differentiated product industries. *Working Paper*.

Gandhi, A., Z. Lu, and X. Shi (2013). Estimating demand for differentiated products with error in market shares. *Working Paper*.

Gandhi, A., Z. Lu, and X. Shi (2017). Estimating demand for differentiated products with zeroes in market share data. *Working Paper*.

Grubor, A. and N. Milicevic (2015). Measuring on-shelf availability of fmcg products. *Industrija 43*, 53–71.

Gruen, T. W. and D. Corsten (2007). A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry. *Grocery Manufacturers of America*.

Gruen, T. W., D. Corsten, and S. Bharadwaj (2002). Retail out of stocks: A worldwide examination of causes, rates, and consumer responses. *Grocery Manufacturers of America*.

Matsa, D. (2011). Competition and product quality in the supermarket industry. *The Quarterly Journal of Economics 126*, 1539–1591.

Musalem, A., M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten (2010). Structural estimation of the effect of out-of-stocks. *Management Science 56*, 1180–1197.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–1057.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics 25*, 303–325.

Quan, T. W. and K. R. Williams (2017). Product variety, across-market demand heterogeneity, and the value of online retail. *Working Paper*.

Romano, J. and A. Shaikh (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference 138*, 2786–2807.

Romano, J. and A. Shaikh (2010). Inference for identified set in partially identified econometric models. *Econometrica 78*, 169–211.

Schennach, S. (2014). Entropic latent variable integration via simulation. *Econometrica 82*, 345–385.

Appendix

A Preliminaries

We state some well known theorems in this section, the proofs of such theorems can be found in many textbooks and therefore omitted.

Krein-Milman Theorem

Let X be a locally convex topological vector space, and Let K be a compact convex subset of X. Then K is the closed convex hull of its extreme points.

Carathéodory's Theorem

If a point x of $\mathbb{R}^d$ lies in the convex hull of a set P, there is a subset P' consisting of $d+1$ or fewer points such that x lies in the convex hull of P'.

B Extreme Points

We need the use of the characterization of set of all extreme points of space of probability measures in the following proofs.

Definition of Extreme Points. An extreme point of a convex set in a real vector space is a point in the set which does not lie in any open line segment joining two points of the set.

Lemma B.1. *The set of extreme points of convex set $\mathcal{P}(\mathcal{R}(w_t))$ is*

$$ext(\mathcal{P}(\mathcal{R}(w_t))) = \{\mu \in \mathcal{P}(\mathcal{R}(w_t)) : \exists r_t \in \mathcal{R}(w_t),\ \mu(\{r_t\}) = 1\}$$

which is the collection of all Dirac measures.

Proof. *First, for any μ in the above set, suppose there are two points μ_1 and μ_2 in $\mathcal{P}(\mathcal{R}(w_t))$ such that μ lies in the open line segment of μ_1 and μ_2, this means either $\mu_1(\{r_t\}) > 1$ or $\mu_2(\{r_t\}) > 1$, which contradicts the fact that μ_1 and μ_2 are probability measures.*

Second, for any μ not in the above set with support $supp(\mu)$, let A be a Borel set in $supp(\mu)$ such that $0 < \mu(A) < 1$. Define $\mu_1 \in \mathcal{P}(\mathcal{R}(w_t))$, such that $\mu_1(B) = \frac{\mu(B)}{1-\mu(A)}$ for any set B in the Borel σ-algebra restricted to set $supp(\mu) - A$, and $\mu_1(B) = 0$ for any set B otherwise.

Define $\mu_2 \in \mathcal{P}(\mathcal{R}(w_t))$, such that $\mu_2(B) = \frac{\mu(B)}{\mu(A)}$ for any set B in the Borel σ-algebra restricted to set A, and $\mu_2(B) = 0$ for any set B otherwise. Then $\mu = (1-\mu(A)) \cdot \mu_1 + \mu(A) \cdot \mu_2$ which lies in the open line segment of μ_1 and μ_2.

C Proofs

In this section I provide all proofs for lemmas and propositions in this paper.

Proof of Lemma 4.1

Notice that $0 \leq s_{jt} + r_{jt} \leq 1$ and $\sum_j s_{jt} + \sum_j r_{jt} = 1$.

Proof of Proposition 4.1

It is sufficient to prove, the range of $\sigma_j^{-1}(x_t, s_t + r_t; \beta, \eta)$ is $\mathbb{R}$. Then we can just pick r_t such that the generated randon variable is conditional mean independent of z_t.

Indeed, for any $\xi_{jt} \in \mathbb{R}$, let $r_{jt} = \sigma_{jt}(x_{jt}, \xi_{jt}; \beta, \eta) - s_{jt}$. Notice that $\sigma_{jt}(x_{jt}, \xi_{jt}; \beta, \eta) \in [0, 1]$, therefore, $r_{jt} \in [-s_{jt}, 1 - s_{jt}]$, which falls into the interval specified in Lemma 4.1.

Proof of Lemma 4.2

Notice that for any $j \in \mathcal{A}_t$, $s_{jt} + r_{jt} = \sigma_{jt} \in [0, s_{jt}]$.

Proof of Lemma 4.3

$0 \leq r_{jt} + s_{jt} \leq 1 - s_{0t} \leq 1 - \zeta$.

Proof of Lemma 4.4

Notice that $s_{jt} + r_{jt} = \sigma_{jt} \in [\frac{1}{d_t}, 1 - \frac{J+1}{d_t}]$.

Proof of Lemma 4.5

Notice that $s_{jt} + r_{jt} = \sigma_{jt} \in [s_{jt}, s_{jt}\frac{d_{2t}}{d_{1t}}]$.

Proof of Lemma 5.1

Denote the space of all continuous real function over $\mathcal{R}(w_t)$ as $C(\mathcal{R}(w_t))$.

We know that the space of all continuous real value function over a compact (and hence also bounded) set is separable. Therefore, for any $\varepsilon > 0$, there exists a countable dense set $\{f_k\}$, $f_k \in C(\mathcal{R}(w_t))$, $k = 1, 2, \dots$, such that, for any $f \in C(\mathcal{R}(w_t))$, there exists f_k

$$\|f_k - f\|_\infty < \frac{\varepsilon}{3} \tag{25}$$

For any sequence $\{\mu_k\}$ in $\mathcal{P}(\mathcal{R}(w_t))$, consider $\mathbb{E}_{\mu_k} f_1$. Since $|\mathbb{E}_{\mu_k} f_1| < \|f_1\|_\infty$ is a bounded real sequence, we can find a convergent subsequence $\mathbb{E}_{\mu_{k_{n_1}}} f_1$. Now repeat the above argument to $\mathbb{E}_{\mu_{k_{n_{m-1}}}} f_m$, we can find another convergent subsequence $\mathbb{E}_{\mu_{k_{nm}}} f_m$.

Consider the diagonal sequence $\mu_{k_{n_k}}$, $\mathbb{E}_{\mu_{k_{n_k}}} f_k$ converges for any k, which implies for any k, there exists sufficient large K_1 and K_2, such that, $\|\mathbb{E}_{\mu_{k_{n_{K_1}}}} f_k - \mathbb{E}_{\mu_{k_{n_{K_2}}}} f_k\| < \frac{\varepsilon}{3}$.

Hence, for any $f \in C(\mathcal{R}(w_t))$, let f_k be satisfying (25). We have

$$\begin{aligned}
& \|\mathbb{E}_{\mu_{k_{n_{K_1}}}} f - \mathbb{E}_{\mu_{k_{n_{K_2}}}} f\| \\
\leqslant\ & \|\mathbb{E}_{\mu_{k_{n_{K_1}}}} f - \mathbb{E}_{\mu_{k_{n_{K_1}}}} f_k\| + \|\mathbb{E}_{\mu_{k_{n_{K_1}}}} f_k - \mathbb{E}_{\mu_{k_{n_{K_2}}}} f_k\| + \|\mathbb{E}_{\mu_{k_{n_{K_2}}}} f_k - \mathbb{E}_{\mu_{k_{n_{K_2}}}} f\| \\
<\ & \|f - f_k\|_\infty + \frac{\varepsilon}{3} + \|f_k - f\|_\infty \\
<\ & \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\
=\ & \varepsilon
\end{aligned}$$

which implies $\mathbb{E}_{\mu_{k_{n_k}}} f$ converges since ε is arbitrary. Notice the choice of f is also arbitrary, we have proved for any sequence $\{\mu_k\}$, there exists a convergent subsequence $\{\mu_{k_{n_k}}\}$ in the weak* topology. It suffices to prove it converges to a point in $\mathcal{P}(\mathcal{R}(w_t))$.

Indeed, let u be an operator on $C(\mathcal{R}(w_t))$ such that, for any $f \in C(\mathcal{R}(w_t))$, $u \circ f = \lim_{k\to\infty} \mathbb{E}_{\mu_{k_{n_k}}} f$. It is easy to verify that u satisfies the conditions for Riesz representation theorem. Hence there exists $\mu \in \mathcal{P}(\mathcal{R}(w_t))$, such that $\mathbb{E}_\mu f = u \circ f = \lim_{k\to\infty} \mathbb{E}_{\mu_{k_{n_k}}} f$, which completes the proof.

Proof of Lemma 5.2

Let $\{v_k\}$ be a sequence of vectors in $L(\mathcal{P}(\mathcal{R}(w_t)))$, then there exists a sequence of probability measures $\{\mu_k\}$ in $\mathcal{P}(\mathcal{R}(w_t))$, such that, $v_k = L\mu_k$. Due to Lemma 5.1, $\mathcal{P}(\mathcal{R}(w_t))$ is weak* compact, which implies there exists a subsequence $\{mu_{k_n}\}$, and a probability measure μ in $\mathcal{P}(\mathcal{R}(w_t))$, such that, $L(\mu_{k_n}) \to L\mu$. Notice that $v_{k_n} = L(\mu_{k_n})$, hence there exists a subsequence $\{v_{k_n}\}$ and a point $v = L\mu$, such that $v_{k_n} \to v$. Therefore $L(\mathcal{P}(\mathcal{R}(w_t)))$ is compact.

Let $\{v_k\}$ be a finite set of vectors in $L(\mathcal{P}(\mathcal{R}(w_t)))$, then there exists a finite set of probability measures $\{\mu_k\}$ in $\mathcal{P}(\mathcal{R}(w_t))$, such that, $v_k = L\mu_k$. Due to Lemma 5.1, $\mathcal{P}(\mathcal{R}(w_t))$ is convex, which implies for any λ_k, $\sum_k \lambda_k = 1$ and $\lambda_k \geqslant 0$, we have $\sum_k \lambda_k \mu_k \in \mathcal{P}(\mathcal{R}(w_t))$. Hence $L(\sum_k \lambda_k \mu_k \in \mathcal{P}(\mathcal{R}(w_t))) = \sum_k \lambda_k L(\mu_k) = \sum_k \lambda_k v_k \in L(\mathcal{P}(\mathcal{R}(w_t)))$. Therefore $L(\mathcal{P}(\mathcal{R}(w_t)))$ is convex.

Proof of Proposition 5.1

Based on Lemma 5.1 we know that $\mathcal{P}(\mathcal{R}(w_t))$ is compact in its weak* topology and convex, therefore according to Krein-Milman theorem we can write $\mathcal{P}(\mathcal{R}(w_t))$ as

$$\mathcal{P}(\mathcal{R}(w_t)) = cl^*(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))$$

where cl^* denotes the closure of a set in the weak* topology and co is the convex hull. $ext(\mathcal{P}(\mathcal{R}(w_t))))$ denotes the set of extreme points in $\mathcal{P}(\mathcal{R}(w_t)))$. Therefore we have

$$L(\mathcal{P}(\mathcal{R}(w_t))) = L(cl^*(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))) \tag{26}$$

Next we show that

$$L(cl^*(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))) \subset cl(L(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))) \tag{27}$$

where cl denotes the closure in $\mathbb{R}^{d_z}$. Indeed, let $v \in L(cl^*(co(ext(\mathcal{P}(\mathcal{R}(w_t))))))$, then we can find a $\mu \in cl^*(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))$, such that $v = L\mu$. Then we can find a sequence $\{\mu_k\} \in co(ext(\mathcal{P}(\mathcal{R}(w_t))))$ such that $L\mu_k \to L\mu$, since L is a linear mapping from $\mathcal{R}(w_t)$ to $\mathbb{R}^{d_z}$. Notice that $\{L\mu_k\}$ is a sequence in $L(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))$, this proves that $v \in cl(L(co(ext(\mathcal{P}(\mathcal{R}(w_t))))))$.

Now that since $co(ext(\mathcal{P}(\mathcal{R}(w_t)))) \subset \mathcal{P}(\mathcal{R}(w_t))$, and $L(\mathcal{P}(\mathcal{R}(w_t)))$ is compact due to

Lemma 5.2, we have

$$cl(L(co(\mathcal{P}(\mathcal{R}(w_t))))) \subset cl(L(\mathcal{P}(\mathcal{R}(w_t)))) = L(\mathcal{P}(\mathcal{R}(w_t))) \tag{28}$$

Now combine equation (26), (27) and (28) we have

$$L(\mathcal{P}(\mathcal{R}(w_t))) = cl(L(co(ext(\mathcal{P}(\mathcal{R}(w_t))))))$$

Next we show that

$$L(co(ext(\mathcal{P}(\mathcal{R}(w_t))))) = co(L(ext(\mathcal{P}(\mathcal{R}(w_t)))))$$

Let $v \in L(co(ext(\mathcal{P}(\mathcal{R}(w_t)))))$, which can be represented as $v = L(\sum_k \lambda_k \mu_k)$, where $\{\mu_k\}$ is a finite sequence in $ext(\mathcal{P}(\mathcal{R}(w_t)))$ and $\sum_k \lambda_k = 1$ with $\lambda_k \geqslant 0$. Therefore $v = \sum_k \lambda_k (L\mu_k) \in co(L(ext(\mathcal{P}(\mathcal{R}(w_t)))))$. The reverse direction is similar.

Finally we have

$$L(\mathcal{P}(\mathcal{R}(w_t))) = cl(co(L(ext(\mathcal{P}(\mathcal{R}(w_t))))))$$

which boils down to characterize $ext(\mathcal{P}(\mathcal{R}(w_t)))$, which due to Lemma B.1, is the set of all Dirac measures. Hence $L(ext(\mathcal{P}(\mathcal{R}(w_t)))) = G(\mathcal{R}(w_t))$, where $G : \mathcal{R}(w_t) \to \mathbb{R}^{d_z}$ and $G(r_t) = g_j(w_t, r_t; \theta)$ is a continuous mapping. Notice that $\mathcal{R}(w_t)$ is compact, and G is continuous, we have $co(G(\mathcal{R}(w_t)))$ is also compact, therefore we have

$$cl(co(L(ext(\mathcal{P}(\mathcal{R}(w_t)))))) = cl(co(G(\mathcal{R}(w_t)))) = co(G(\mathcal{R}(w_t)))$$

By Carathéodory's theorem, since $co(G(\mathcal{R}(w_t))) \in \mathbb{R}^{d_z}$, we have $co(G(\mathcal{R}(w_t))) = L(\mathcal{E}_t(w_t))$, which completes the proof.

Proof of Lemma 5.3

Since $\mathcal{R}(w_t)$ is convex, and $\sigma_j^{-1}(x_t, s_t + r_t; \theta)$ is a continuous mapping from $\mathcal{R}(w_t)$ to $\mathbb{R}$, therefore, the image of $\mathcal{R}(w_t)$ under $\sigma_j^{-1}(x_t, s_t + r_t; \theta)$ is convex. Notice that a scalar multiplied by $f_j(z_{jt})$ is an affine transformation which keeps the convexity of sets, hence

$g_j(\mathcal{R}(w_t))$ is convex. $\square$

Proof of Proposition 5.2

Since $g_j(\mathcal{R}(w_t))$ is convex, for any $\mu \in \mathcal{E}(w_t)$, there exists $r_t^* \in \mathcal{R}(w_t)$, such that

$$\mathbb{E}_\mu[g_j(w_t, r_t; \theta)|w_t] = g_j(w_t, r_t^*; \theta)$$

This implies that searching over $\mathcal{E}(w_t)$ and searching over $\mathcal{R}_t(w_t)$ is the same thing.

Proof of Lemma 5.4

Denote r_t^* as the solution to

$$\arg \min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} g_j(w_t, r_t; \theta)\|$$

and $r^*(\cdot)$ as the solution to

$$\arg \min_{r(\cdot) \in \mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|$$

Then we have

$$\begin{aligned}
& \min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} g_j(w_t, r_t; \theta)\| \\
= \quad & \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} g_j(w_t, r_t^*; \theta)\| \\
\leqslant \quad & \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} g_j(w_t, r^*(w_t); \theta)\| \\
\xrightarrow{p} \quad & \|\mathbb{E}_{\pi_0}\frac{1}{J}\sum_{j=1}^{J}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \frac{1}{J}\sum_{j=1}^{J}\|\mathbb{E}_{\pi_0}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \|\mathbb{E}_{\pi_0}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \min_{r(\cdot) \in \mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|
\end{aligned}$$

Here the inequality is because r_t^* is the one that minimizes

$$\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r_t; \theta)\|$$

This implies for any $\varepsilon > 0$, there exists a sequence of sets $A_1, A_2, \ldots, A_T$, such that $\lim_{T\to\infty} P(A_T) = 1$ and for any $\omega \in A_T$,

$$\begin{aligned}
\min_{r_t\in\mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r_t; \theta)\| &\leqslant \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r^*(w_t); \theta)\| \\
&\leqslant \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\| + \varepsilon
\end{aligned} \tag{29}$$

Also for any $r(\cdot) \in \mathcal{F}$

$$\begin{aligned}
& \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r(w_t); \theta)\| \\
\xrightarrow{p} \; & \|\mathbb{E}_{\pi_0}\sum_{j=1}^{J}[g_j(w_t, r(w_t); \theta)]\| \\
= \; & \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)] \\
\geqslant \; & \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|
\end{aligned}$$

This implies for any $\varepsilon > 0$, there exists a sequence of sets $B_1, B_2, \ldots, B_T$, such that $\lim_{T\to\infty} P(B_T) = 1$ and for any $\omega \in B_T$, there exists $r(\cdot) \in \mathcal{F}$, such that

$$\begin{aligned}
\min_{r_t\in\mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r_t; \theta)\| &\geqslant \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\| - \varepsilon \\
&\geqslant \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\| - \varepsilon
\end{aligned} \tag{30}$$

Combine (29) and (30) we have for any $\omega \in A_T \cap B_T$,

$$|\min_{r_t\in\mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}g_j(w_t, r_t; \theta)\| - \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|| < \varepsilon$$

Notice that

$$P(A_T \cap B_T) = 1 - P(\bar{A}_T \cup \bar{B}_T) \geqslant 1 - P(\bar{A}_T) - P(\bar{B}_T) \to 1$$

as $T \to \infty$, hence $\lim_{T\to\infty} P(A_T \cap B_T) = 1$, which implies

$$\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} g_j(w_t, r_t; \theta)\| \xrightarrow{p} \min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|$$

Proof of Proposition 5.3

Since Θ is compact, together with Lemma 5.4 and Assumption 5.3 we have

$$\sup_{\theta\in\Theta} \|\hat{Q}_T(\theta) - Q(\theta)\| \xrightarrow{p} 0$$

Thus, the conditions for Theorem 3.1 in Chernozhukov et al. (2007) are satisfied and hence

$$d_H(\hat{\Theta}, \Theta_0) \xrightarrow{p} 0$$

Proof of Lemma 5.5

For any $\theta \in \Theta_0$, denote r_t^* as the solution to

$$\arg\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t; \theta)\|$$

and $r^*(\cdot)$ as the solution to

$$\arg\min_{r(\cdot)\in\mathcal{F}} \|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|$$

Then we have

$$
\begin{aligned}
& \min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t; \theta)\| \\
= \quad & \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t^*; \theta)\| \\
\leqslant \quad & \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t^*(w_t); \theta)\| \\
\xrightarrow{p} \quad & \|\mathbb{E}_{\pi_0}\frac{1}{T}\sum_{t=1}^{T}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \frac{1}{T}\sum_{t=1}^{T}\|\mathbb{E}_{\pi_0}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \|\mathbb{E}_{\pi_0}[g_j(w_t, r^*(w_t); \theta)]\| \\
= \quad & \min_{r(\cdot)\in\mathcal{F}}\|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\| \\
= \quad & 0
\end{aligned}
$$

Here the inequality is because r_t^* is the one which minimizes

$$
\frac{1}{J}\sum_{j=1}^{J}\|\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t; \theta)\|
$$

the convergence in probability is due to Definition 5.3, and the last equality is because of $\theta \in \Theta_0$.

Since

$$
\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t; \theta)\| \geqslant 0
$$

we have

$$
\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{J}\sum_{j=1}^{J}\frac{1}{T}\sum_{t=1}^{T} g_j(w_t, r_t; \theta)\| \xrightarrow{p} \min_{r(\cdot)\in\mathcal{F}}\|\mathbb{E}_{\pi_0}[g_j(w_t, r(w_t); \theta)]\|
$$

when $\theta \in \Theta_0$.

Proof of Proposition 5.4

Since Θ is compact, together with Lemma 5.1 and Assumption 5.5 we have

$$\sup_{\theta \in \Theta_0} \|\hat{Q}_J(\theta)\| \xrightarrow{p} 0 \tag{31}$$

For any $\theta \in \Theta_0$, we have

$$\hat{Q}_J(\theta) \leqslant \sup_{\theta \in \Theta_0} Q_J(\theta) \leqslant \hat{c}$$

with probability approaching 1, which implies with probability approaching 1, $\theta \in \hat{\Theta}$, and $\Theta_0 \subset \hat{\Theta}$. Hence

$$\sup_{\theta \in \theta_0} d(\theta, \hat{\Theta}) \xrightarrow{p} 0 \tag{32}$$

For any $\varepsilon > 0$, we have

$$\inf_{\theta \notin \mathcal{N}(\Theta_0, \varepsilon)} \hat{Q}_J(\theta) \geqslant \sup_{\theta \in \Theta_0} \hat{Q}_J(\theta) + C(\varepsilon) = o_p(1) + C(\varepsilon)$$

where the inequality is from Assumption 5.6 and the equality is from (31).

Now we have

$$\sup_{\theta \in \hat{\Theta}} \hat{Q}_J(\theta) \leqslant \hat{c} = o_p(1) < o_p(1) + C(\varepsilon) \leqslant \inf_{\theta \notin \mathcal{N}(\Theta_0, \varepsilon)} \hat{Q}_J(\theta)$$

which implies $\hat{\Theta} \cap (\Theta - \mathcal{N}(\Theta_0, \varepsilon))$ is empty with probability approaching 1.

Hence, $\hat{\Theta} \subset \mathcal{N}(\Theta_0, \varepsilon)$ with probability approaching 1 and

$$\sup_{\theta \in \hat{\theta}} d(\theta, \Theta_0) \leqslant \varepsilon \tag{33}$$

with probability approaching 1.

Combine (32) and (33) we have

$$d_H(\hat{\Theta}, \Theta_0) \leqslant \varepsilon$$

with probability approaching 1, hence completes the proof.

D Verification of Regularity Conditions

We verify the stochastic equicontinuity condition: Assumption 5.3 for logit model, i.e., a model with following sample objective function:

$$\hat{Q}_T(\theta) = \min_{r_t \in \mathcal{R}_t(w_t)} \| \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt} - r_{jt}) - \log(s_{0t} - r_{0t}) - x'_{jt}\theta) \| \tag{34}$$

The following analysis can be extended to a random coefficient logit model with same idea but more complicated analysis. The key idea is to notice that the solution to the optimization problem is uniform continuous in θ.

We first made some further assumptions about the data generating process.

Assumption D.1. *The choice probably σ_j is bounded away from zero for each j, namely, there exists a positive number $\zeta_j > 0$, such that, $\sigma \geqslant \zeta_j > 0$.*

Assumption D.1 is similar to Condition S in Berry et al. (2004), which is necessary for any uniform results regarding logit assumptions. This is because we have to deal with objects which are close to $\log \sigma_j$ frequently. If σ_j is not bounded away from zero, then $\log \sigma_j$ can be made arbitrarily large and thus fails any proofs of uniformity.

Assumption D.2. *The following convergence result holds:*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{J} \sum_{j=1}^{J} f_j(z_{jt}) x'_{jt} \xrightarrow{p} \mathbb{E} \frac{1}{J} \sum_{j=1}^{J} f_j(z_{jt}) x'_{jt}$$

and $\|\mathbb{E} \sum_{j=1}^{J} f_j(z_{jt}) x'_{jt}\| < \infty$

Assumption D.2 is a standard convergence assumption. Now we prove the stochastic equicontinuous property.

Proposition D.1. *The sample objective function in 34 satisfies the condition in Assumption 5.3.*

Proof. *Fix $\varepsilon > 0$ throughout proof, for any θ, $\theta' \in \Theta$, let $r_t(\theta)$ and $r_t(\theta')$ be the solution* *to the optimization problem*

$$\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt})-\log(s_{0t}-r_{0t})-x'_{jt}\theta)\|$$

and

$$\min_{r_t \in \mathcal{R}_t(w_t)} \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt})-\log(s_{0t}-r_{0t})-x'_{jt}\theta')\|$$

respectively.

Since the sample objective function is in norm form, it suffices to bound

$$\begin{aligned}&\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt}(\theta))-\log(s_{0t}-r_{0t}(\theta))-x'_{jt}\theta)\\ &-\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt}(\theta'))-\log(s_{0t}-r_{0t}(\theta'))-x'_{jt}\theta')\|\end{aligned} \tag{35}$$

Rewrite (35) such that

$$\begin{aligned}&\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt}(\theta))-\log(s_{0t}-r_{0t}(\theta))-x'_{jt}\theta)\\ -\;&\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt}(\theta'))-\log(s_{0t}-r_{0t}(\theta'))-x'_{jt}\theta')\|\\ =\;&\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})((\log(s_{jt}-r_{jt}(\theta))-\log(s_{jt}-r_{jt}(\theta')))\\ &-(\log(s_{0t}-r_{0t}(\theta))-\log(s_{0t}-r_{0t}(\theta')))\\ &-x'_{jt}(\theta-\theta'))\|\\ \leqslant\;&(a)+(b)+(c)\end{aligned}$$

where

$$(a)=\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{jt}-r_{jt}(\theta))-\log(s_{jt}-r_{jt}(\theta'))\|$$

and

$$(b) = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})(\log(s_{0t} - r_{0t}(\theta)) - \log(s_{0t} - r_{0t}(\theta'))\|$$

and

$$(c) = \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})x'_{jt}(\theta - \theta')\|$$

Notice that $r_{jt}(\theta)$ is the solution to the minimization problem in (34), which is continuous in both r_t and θ. By Maximum Theorem, we have that $r_{jt}(\theta)$ is continuous in θ. Since Θ is compact, we have that $r_{jt}(\theta)$ is uniform continuous in θ, which implies there exists $\delta > 0$, for all $\|\theta - \theta'\| < \delta < \varepsilon$, we have $\|r_{jt}(\theta) - r_{jt}(\theta')\| < \varepsilon$ for $j = 1, \ldots, J$.

Hence due to Assumption D.1 and D.2, there exists C_1 depending on T, such that

$$(a) \leqslant \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})|\log(\zeta_j)|\|\|r_{jt}(\theta) - r_{jt}(\theta')\| \leqslant |\log(\zeta_j)|\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})\|\varepsilon < C_1\varepsilon$$

with probability approaching one.

Similarly

$$(b) \leqslant \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})|\log(\zeta_j)|\|\|r_{0t}(\theta) - r_{0t}(\theta')\| \leqslant |\log(\zeta_j)|\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})\|\varepsilon < C_1\varepsilon$$

with probability approaching one.

Due to Assumption D.2, there exists C_2 depending on T, such that

$$(c) \leqslant \|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J} f_j(z_{jt})x'_{jt}\|\|\theta - \theta'\| < C_2\varepsilon$$

with probability approaching one.

Since the choice of ε is arbitrary, we complete the proof.

The verification for Assumption 5.5 is similar and omitted.