Title
Essays in Econometrics

Author
Kim, Jeonghwan

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Essays in Econometrics

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Jeonghwan Kim

2021

ABSTRACT OF THE DISSERTATION

Essays in Econometrics

by

Jeonghwan Kim

Doctor of Philosophy in Economics

University of California, Los Angeles, 2021

Professor Andres Santos, Chair

This dissertation studies a few econometric theories potentially useful for applied economists. In the first two chapters, I study estimation and inference in a semi-parametric model under a monotonicity restriction on the non-parametric component. I develop a new semi-parametric estimator that can be implemented without choosing any smoothing parameters and construct a confidence band for the non-parametric component under monotonicity. The finite dimensional parametric estimator satisfies asymptotic normality. The asymptotic distribution of L_∞ - distance for the non-parametric component is Gumbel with the rate of convergence $O_p((\frac{\log n}{n})^{1/3})$. I apply the estimator to estimate the returns to schooling under the restriction that age has monotonic effect on wages. The confidence interval of the returns to schooling and the confidence band of the age effect on the log of wage under an assumed monotonic relationship are reported. I illustrate the confidence intervals of the semi-parametric estimator and the confidence band of the semi-nonparametric estimator using Monte Carlo simulations.

On the last chapter, my coauthors and I propose a pragmatic approach to the errors-in-variables and nonlinear panel models. These models are often deemed impossible to estimate in their most general forms. For example, the higher order moments approach to errors-in-variables model fails when there is conditional heteroscedasticity. Similarly, nonlinear panel models with fixed effects and small T are known to be problematic to estimate. We propose estimating these models using approximate moments, using a Taylor series approximation applied to Kadane's (1971) small sigma approach. Simulation results suggest that the approximation leads to reasonable sampling properties. Our proposal complements the newly resurgent literature on sensitivity analysis.

The dissertation of Jeonghwan Kim is approved.

Denis Chetverikov

Shuyang Sheng

Hyungsik Roger Moon

Andres Santos, Committee Chair

University of California, Los Angeles

2021

I dedicate my dissertation work to my family who always supports me.

I dedicate this work to all the faculty members, my friends, and the department staff.

Special thanks to Professor Andres Santos, Professor Jinyong Hahn, Professor Denis Chetverikov, Professor Jerry Hausman, Professor Shuyang Sheng, Professor Rosa Matzkin, Professor Zhipeng Liao, Professor Hyungsik Roger Moon, Professor Jay Lu, Professor Simon Board, Professor Ed McDevitt,

Changsu Ko, Daehyun Kim, Wookun Kim, Jonathan Gu, Dongook Eun, Seokho Bang, Hualei Sheng, Nan Liu, Wenyu Zhou, Akina Ikudo, Chiara Paz, Grace Fransisca,

and my love Hyunjin Kim.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am so delighted to have an opportunity to show my gratitude.

I greatly appreciate my advisor Professor Andres Santos. He has always been a great teacher and friend to me. He has prevented me from being distracted, guided me to the right track, and taught me how to be positive, considerate, and acute. Above all, I thank him a lot for his amount of patience.

I thank Professor Denis Chetverikov a lot. Whenever I have some difficulties in processing my research. He suggested many brilliant methods readily. It always helped me to widen my perspective on the subject of probability theory and statistics.

I am grateful for Professor Jerry Hausman and Professor Jinyong Hahn. I have been honored to have the chance to witness their discussions and join the project as a coauthor.

VITA

2007–2012	B.S. (Mathematics) and B.A. (Economics), Yonsei University, Seoul, South Korea.
2013–2015	Graduate School of Economics, Yonsei University, Seoul, South Korea.
2016	M.A. (Economics), UCLA, Los Angeles, California.

PUBLICATIONS

Public Transfer Income, Retirement, Living Alone and the Consumption Behavior of Older Households: Evidence from Korea using Panel Quantile Regression. (with Yeonha Jung Yonsei Graduate Student Association Vol.58, 271-292. (2013)

CHAPTER 1

Semiparametric Estimation and Inference under Monotonicity

1.1 Introduction

Monotonicity restrictions play an important role in economics. Imposing such restrictions motivated by economic theory allows us to obtain identification and estimation strategies. For example, Matzkin (1991, 1992, 2003), Imbens and Angrist (1994), and Chernozhukov and Hansen (2004) use a monotonicity restriction for identifying parameters of interests. In estimation related literature, Henderson et al (2012), Freyberger and Horowitz (2015), and Chetverikov and Wilhelm (2017) impose monotonic restriction on parameters. [1]

The semi-parametric models have been very popular due to their flexibility. They do a better job of fitting the data than simple regression models while still allowing for a relatively easy interpretation of the parameters of interest. In this chapter, I will focus mainly on a partially linear model.

$$Y = X'\theta_0 + g_0(W) + U, \qquad E[U|W,Z] = 0$$

where Y is a dependent variable, X is endogenous, W is exogenous, and Z is an instrumental variable. Assume that all the variables above are 1-dimensional random variables except X and Z. In many studies, g_0 is considered a nuisance parameter, which is non-parametrically pre-estimated to estimate θ_0 and most literature requires the smoothing parameters such as bandwidth, basis functions, kernels and so on. For example, in Hall and Huang (2001),

[1] There are good review papers for those who are not familiar with the literature. See Matzkin (1994), Chetverikov et al. (2018), and Guntuboyina and Sen (2018).

Racine et al. (2009), and Horowitz and Lee (2017), the authors develop many frameworks of estimation and inference under shape restriction if the practitioners could determine the bandwidth, kernel or series length. In this paper, I will provide the asymptotic distribution of the estimator of θ_0 under estimating g_0 without choosing any smoothing parameters. Often, the smoothing parameters have to be predetermined to estimate the non-parametric component, and they could affect the estimation procedure. If g_0 is assumed to be monotonically increasing or decreasing, it can be consistently estimated using isotonic regression which does not require any smoothing parameters and the estimator is uniquely determined. Of course the monotonic constraints limit the scope of the application. However, in much economics research, there are many settings in which a monotonic assumption makes sense. [2] A recent study, Liu and Yu (2019), is similar in spirit to this paper by suggesting a tuning parameter free method applying Groeneboom and Hendrickx (2018) to Heckman's sample selection model. They impose a monotonicity restriction on the selection correction function.

In this paper, I will explore the usefulness of the added monotonicity assumption on g_0 at the partially linear model above. Two main contributions are explored: we can obtain asymptotic normality of the finite-dimensional estimator without sample splitting, [3], the estimation procedure is tuning parameter-free.

There is extensive literature on the semi-parametric model in econometrics, so I note only a few papers here: Robinson (1988), Newey et al. (1999), Newey and Powell (2003), Ai and Chen (2003), Blundell et al. (2007), Chen and Pouzo (2009, 2012), Imbens and Newey (2009), Darolles et al. (2011), Wooldridge (2015), Blundell et al. (2017), and Chen and Christensen (2018). See also Powell (1994), and Bickel et al. (2005) for the literature. In statistics literature, isotonic regression was starting in the 1950s. I introduce a few papers related to my paper directly. Zhang (2002) derives the upper bound of the empirical L_p -distance of the isotonic regression under the fixed points design. Huang (2002) and Cheng

[2]Some examples are the relationship between income and consumption and education and wage.

[3]See Chernozhukov et al. (2018). The authors suggest sample splitting to guarantee the independence between two estimators, which are the finite-dimensional estimator and the estimator of the nuisance parameters.

(2009) study the semi-parametric isotonic regression under exogeneity. Mammen and Yu (2007) analyze the additive isotonic regression model that reduces model complexity under a separability assumption.

The paper consists mainly of three parts. In section 1.2, the new semi-parametric estimator is introduced with 4-step estimation procedure. In section 1.3, I use my methods to estimate the returns to schooling with the data set by Card (1993). In section 1.4, the MC simulation study is conducted. All the proofs of theorems and lemmas are attached on the appendix.

1.2 Model

Before I describe the model, let me briefly introduce the notations. Let X be a random variable. The L_2-norm is defined as $\|X\|_2 := (E[X^2])^{\frac{1}{2}}$. To discuss asymptotic properties, define $X = O_p(a_n)$ as $\frac{X}{a_n}$ is bounded in probability, and $X = o_p(b_n)$ as $\frac{X}{b_n}$ converges to 0 in probability. For simplicity, $X \lesssim O_p(a_n)$ means that there exists $C > 0$ such that $\frac{X}{a_n} \leq C$ as $n \to \infty$ with probability 1.

1.2.1 The Monotonic Non-Parametric Instrumental Variable Problem in a Semi-Parametric Model

A semi-parametric model with k-dimensional parameters of interest, θ_0 is as follows.

$$Y_i = X_i'\theta_0 + g_0(W_i) + U_i, \qquad E[U_i|W_i, Z_i] = 0 \tag{1.1}$$

$$Z_i = \phi_0(W_i) + V_i, \qquad E[V_i|W_i] = 0 \tag{1.2}$$

where $Y_i \in \mathbb{R}$, $X_i = (X_i^1, \cdots, X_i^k)' \in \mathbb{R}^k$, $W_i \in \mathbb{R}$, $Z_i = (Z_i^1, \cdots, Z_i^l)' \in \mathbb{R}^l$ and $\phi_0(\cdot) = (\phi_0^1(\cdot), \cdots, \phi_0^l(\cdot))'$. Y is a dependent variable, X is a vector of endogenous variables, W is exogenous, and Z is a vector of instrumental variables. Assume that all the variables above are 1-dimensional random variables except X and Z. Define $dim(X) = k$ and $dim(Z) = l$ where $l \geq k$ for rank condition. Also, assume that both non-parametric parameters, g_0 and

ϕ_0, are monotonic.[4] As a digression, Z is correlated with X via V, and U and V are independent each other. In estimation strategies, to estimate $\hat{\theta}$, we need incidental parameter estimators, $\hat{\eta}$ for (g_0, ϕ_0). However, again we need another consistent estimator for θ_0 to estimate $\hat{\eta}$. So, I propose an intermediate estimator for θ_0, $\tilde{\theta}$, in the estimation algorithm. I will provide 4 steps procedures as follows.

STEP 1 (1st Stage Estimation)

Estimate $\hat{\phi} = (\hat{\phi}^1, \cdots, \hat{\phi}^l)'$ from the 1st stage regression equation (1.2), which satisfies the rate of convergence $o_p(n^{-1/6})$.[5] As an example, it could be derived by isotonic regression if ϕ_0 is a monotonic function. The isotonic regression estimator is

$$\hat{\phi}^m(W_{(j)}) = \underset{a \leq j}{max} \quad \underset{b \geq j}{min} \quad \frac{1}{b-a+1}\sum_{i=a}^{b} Z_{(i)}^m$$

where $\{(W_{(i)}, Z_{(i)}^m)\}_{i=1}^n$ is the ordered data set with respect to $\{W_i\}_{i=1}^n$ for all $m = 1, \cdots, l$.

STEP 2 (Intermediate Estimator of θ_0)

Estimate $\tilde{\theta}$ by using GMM with the moment condition $E[(Y - X'\theta_0)(Z - \phi_0(W))] = 0$ given $\hat{\phi}$.

$$\tilde{\theta} = (\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\mathbb{X})^{-1}(\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\mathbb{Y})$$

where $\mathbb{Y} = (Y_1, \cdots, Y_n)'$, $\mathbb{X} = (X_1', \cdots, X_n')'$, $\tilde{\mathbb{V}} = ((Z_1 - \hat{\phi}(W_1))', \cdots, (Z_n - \hat{\phi}(W_n))')'$, and $\hat{\Omega}$ is a consistent estimator for $\Omega = E[\mathbb{V}\mathbb{V}'U^2]^{-1}$

Then one can show that $\tilde{\theta}$ is a consistent estimator for θ_0 by satisfying $\|\tilde{\theta} - \theta_0\| \lesssim O_p(n^{-1/3})$. See the proof of theorem 1.2.1 in the appendix A.

STEP 3 (Intermediate Estimator of g_0)

---

[4] In Chernozhukov et al. (2018), as far as $\|\hat{\phi} - \phi_0\|_2 \times \|g_0 - \hat{g}\|_2 = o_p(n^{-\frac{1}{2}})$, we could get the asymptotic normality of the parametric estimator. Hence, from the perspective of practitioners, any estimator that satisfies the condition above can be used to get $\sqrt{n}$ - consistency on the parametric estimator.

[5] This rate satisfies Chernozhukov et al. (2018)

Get $\tilde{g}$ by using isotonic regression:

$$\tilde{g}(W_{(j)}) = \underset{a\leq j}{max} \quad \underset{b\geq j}{min} \quad \frac{1}{b-a+1}\sum_{i=a}^{b}(Y_{(i)} - X'_{(i)}\tilde{\theta})$$

where $\{(W_{(i)}, Y_{(i)} - X'_{(i)}\tilde{\theta})\}_{i=1}^{n}$ is the ordered data set with respect to $\{W_i\}_{i=1}^{n}$.

STEP 4

Estimate θ_0 by using GMM with the moment condition $E[(Y - X'\theta_0 - g_0(W))(Z - \phi_0(W))] = 0$ given $(\tilde{\theta}, \hat{\phi}, \tilde{g})$.

$$\hat{\theta} = (\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\mathbb{X})^{-1}(\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\tilde{\mathbb{Y}})$$

where $\tilde{\mathbb{Y}} = (Y_1 - \tilde{g}(W_1), \cdots, Y_n - \tilde{g}(W_n))'$, $\mathbb{X}$, $\tilde{\mathbb{V}}$, and $\hat{\Omega}$ are defined in STEP 2,

Then $\hat{\theta}$ holds $\sqrt{n}$ - consistency, which is asymptotic normal.

From STEP 4, we have $\hat{\theta}$ and it will achieve asymptotic normality. For a brief sketch of the proof, define $T := (Y, X', W, Z')'$, and the score function as

$$\psi(T; \theta, \eta) = (Y - X'\theta - g(W))(Z - \phi(W)) \in \mathbb{R}^l \tag{1.3}$$

where $\eta = (\phi, g)$. Note that $\psi(T; \theta, \eta) = (\psi^1(T; \theta, \eta), \cdots, \psi^l(T; \theta, \eta))'$. For all $m = 1, \cdots, l$, sample counterpart of the moment condition is,

$$\frac{1}{n}\sum_{i=1}^{n}\psi^m(T_i; \hat{\theta}, \hat{\eta}) = o_p(1)$$

Under some regularity conditions, we can derive,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^m(T_i; \hat{\theta}, \hat{\eta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^m(T_i; \theta_0, \hat{\eta}) + o_p(1) \tag{1.4}$$

If η is Donsker and $\hat{\eta}$ is included in some parameter space, then

$$\nu_n\psi^m(T; \theta_0, \hat{\eta}) = \nu_n\psi^m(T; \theta_0, \eta_0) + \sqrt{n}\bigg(\int \psi^m(\cdot; \theta_0, \hat{\eta})dP(\cdot) - \int \psi^m(\cdot; \theta_0, \eta_0)dP(\cdot)\bigg) + o_p(1) \tag{1.5}$$

where ν_n is the notation for the empirical process such that $\nu_n f := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{f(X_i) - \int f(x)dP(x)\}$. The first term on the right-hand side of the equation (1.5) can be approximated by Gaussian distribution by a classical central limit theorem. The second term will disappear by Neyman-orthogonality introduced in Chernozhukov et al. (2018). The detailed proofs are in the appendix.

Theorem 1.2.1 ***(Asymptotic Normality of $\hat{\theta}$)***
Suppose that the model is given as equation (1.1) and (1.2) with $\theta_0 \in \Theta$ where Θ is bounded, and ϕ, g are monotonic. Assume that (a) $\{T_i\}_{i=1}^n$ have bounded supports. (b) $\{T_i\}_{i=1}^n$ are independent of each other and identically distrubuted. (c) $E[U^4], E[(V^m)^4] < \infty$ for all $m = 1, \cdots, l$. (d) $E[\psi^m(T; \theta, \eta)]$ is twice Gateaux differentiable with respect to η for all $(\theta, \eta) \in (\Theta, \mathcal{F}_\eta)$ where Θ is bounded and $\mathcal{F}_\eta$ is Donsker class. Then, there exists $\hat{\Sigma}^{-1}$, a consistent estimator for $\Sigma^{-1} = E[X'V]\Omega E[V'X]$, and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

Assumption $(a), (b)$, and (c) are regularity conditions and (d) is crucial for applying Neyman Orthogonality. If we impose monotone restriction on the incidental parameters, we can still get asymptotic normality of the estimator because the monotone functions are included in Donsker class, which has the entropy numbers easily controlled by the empirical process theories. Assume that $\eta = (\phi_0, g_0)$ are monotonic functions. Zhang (2002) proposes that L_2 risk bound of the isotonic regression estimator is $O_p(n^{-\frac{1}{3}})$ if the function has non-zero derivatives on the designed non-random points. Under economic friendly assumptions such as the conditional mean is equal to zero, the rate of convergence of the L_2 risk can still be derived as $O_p(n^{-\frac{1}{3}})$. See lemma A.2.5 in the appendix A. This rate is sufficient to control the estimation errors from the non-parametric components. Then the asymptotic normality of $\hat{\theta}$ holds. When it comes to dimension expansion, practitioners can use multivariate exogenous variables, W. However, there are some issues if $dim(W) \geq 2$. We will discuss this in remark 1.2.3. For further discussion on the multivariate isotonic regression, we need to define multi-dimensional monotonicity as below.

Definition 1.2.1 ***Monotonic increasing (decreasing) function in multidimensional space***

Let $x, y \in \mathbb{R}^k$. A function $f : \mathbb{R}^k \to \mathbb{R}$ is monotonic increasing (decreasing) if $x \leq y$, then $f(x) \leq f(y)$ $(f(x) \geq f(y))$ where $x \leq y$ is defined as $\forall i = 1, \cdots, k$, $x_i \leq y_i$.

Definition 1.2.1 is called as co-monotonicity. Multi-dimensional co-monotonic function has higher entropy bracketing numbers than 1-dimensional monotonic function. As a result, it could cause problems getting the $\sqrt{n}$ - consistency of $\hat{\theta}$. Remark 1.2.1 gives us the idea of overcoming the curse of the dimensionality by assuming the additive separable structure on the co-monotonic function.

Remark 1.2.1 *(Multivariate W Case)*
The multi-dimensional monotonic function is not in the class of Donsker since the entropy bracketing numbers of the class,

$$\mathcal{F}_{M(k)} := \{f : [0,1]^k \to [0,1] \quad \textit{such that} \quad x \leq y \implies f(x) \leq f(y)\}$$

is derived by Gao, Wellner (2007) proposition 3.1 as $logN_{[\,]}(\epsilon, \mathcal{F}_{M(k)}, \|\cdot\|_2) \lesssim \epsilon^{-k} log(\frac{1}{\epsilon})^k$ where $k \geq 2$. In result, theorem 1.2.1 can not be applied if we estimate both $\hat{\phi}$ and $\tilde{g}$ by multivariate isotonic regression with $dim(W) \geq 2$. To detour this problem, the additive separable monotonic function space is required. Define

$$\mathcal{F}^{add}_{M(k)} := \{f \in \mathcal{F}_{M(k)} : f(x) = \sum_{j=1}^{k} f_j(x_j), \forall j = 1, \cdots, k, f_j \in \mathcal{F}_M(1)\}$$

$\mathcal{F}^{add}_{M(k)}$ is Donsker, and Mammen and Yu (2007) show that the additive isotonic estimator has a $O_p(n^{-\frac{1}{3}})$ convergence rate under some regularity assumptions. Under the additive structure, $\hat{\theta}$ satisfies theorem 1.2.1.

In some empirical studies, the interaction effect between explanatory variables should be closely considered. Unfortunately, one can only use a multivariate isotonic regression estimator in estimating either $\hat{\phi}$ or $\tilde{g}$ if $dim(W) \leq 5$. The idea is described in remark 1.2.2.

Remark 1.2.2 *(Multivariate Isotonic Regression without Additive Separability)*
If we do not use additive separable structures on function g_0 or ϕ_0. In Han et al. (2019)

theorem 4, they study L_2 - empirical risk bound for the isotonic regression in multivariate case assuming that the error term has a normal distribution. In their proof, they use the statistical dimension[6] *to obtain the risk bound. Let the dimension of W be s. Once we can approximate the statistical dimension by using non-Gaussian error, then we can argue that $\|\hat{\phi}_I - \phi_0\|_2 \lesssim n^{-1/s}(\log n)^{\gamma_s}$ for $s \geq 2$ where ϕ_I is a multivariate isotonic regression estimator, and γ_s is a constant depending only on s. Define s_1 to be the number of exogenous variables in g_0 and s_2 to be the number of exogenous variables in ϕ_0. Then the asymptotic normality of θ_0 is still valid as far as $\frac{1}{s_1} + \frac{1}{s_2} < \frac{1}{2}$ followed by Chernozhukov et al. (2018).*

1.3 Returns to Education

Card (1993) gives several models for estimating the returns to schooling where the years of education are correlated with the individual's productivity by using some instruments, such as college proximity, parent's education, IQ score, and so on. I use the data which is available on his website. [7] The sample size is $n = 2,962$, and the model is constructed as follows.

$$Y_i = X_i\theta_0 + g_0(W_i) + U_i, \qquad E[U_i|W_i, Z_i] = 0$$

where Y is log of wage, X is years of education, W is age, and Z is 'Knowledge of the world' (KWW) score.[8] In Card (1993), the author defends why the 'KKW' score is a proper instrument theoretically and statistically. Note that ϕ_0 is the conditional expectation of the 'KWW' score given a certain age level. It could be considered as a proxy of experience, so it is reasonable to assume that ϕ_0 is monotonic increasing with respect to W. Similarly, g_0 also measures the effect of experience on wage. Hence, we can impose the monotonic restriction which is actually less restrictive than the linear regression model. The brief description of

[6] *See Amelunxen et al. (2014) for more detail.*

[7] There are missing data for individuals in 1976 and I omit those.

[8] The KWW score variable was administered to young men who represent the civilian, noninstitutional population of males 14-24 years of age in 1966 in the U.S. It was a part of the initial interview in the National Longitudinal Surveys (NLS).

Variables	Mean	Median	Variance	Inter-Quartile
Log Wage	6.264	6.287	0.195	[5.984, 6.565]
Yrs of Educ	13.292	13	7.029	[12, 16]
Age	28.115	28	9.824	[25, 31]
KWW Score	33.546	34	74.043	[28, 40]

Table 1.1: Summary Statistics

Estimator	95% Confidence Interval
OLS	[0.0511, 0.0513]
2SLS	[0.0987, 0.1209]
NPIV	[0.1113, 0.1340]

Table 1.2: The Confidence Interval of θ_0

the summary statistics for each variable is in table 1.1.

On the 1st stage regression, we get the isotonic regression estimator of $E[Z|W]$. Taking the residual of the 1st stage regression, $\hat{V} = Z - \hat{\phi}(W)$, as instrument, we can estimate the parameters of interest by 4 steps. Comparing to the OLS and 2SLS estimator, the confidence interval of the returns to schooling is in table 1.2. In OLS regression, Y is regressed on intercept, X, and W. In the 2SLS estimation, 3 variables, intercept, X, and W, are used as explanatory variables and intercept, Z, and W are used as instruments.

The 2SLS and NPIV estimates of the return to schooling are higher than the OLS estimate and are consistent with the result in Card (1993). Comparing confidence intervals, NPIV estimator performs as good as 2SLS even if it has non-parametric estimator which is notorious for its huge variance.

1.4 Simulation

1.4.1 DGP

In the simulation studies, the DGP is as below for the semi-parametric estimator $\hat{\theta}$.

$$
\begin{aligned}
V &\sim N(0,1) \\
U &\sim N(0,1) \\
X &= 0.5U + 0.5V + N(0,1) \\
W &\sim N(3,1) \\
Y &= X\theta_0 + g_0(W) + U \\
Z &= \phi_0(W) + V
\end{aligned}
$$

where $\theta_0 = 1$. The number of replications is $M = 1,000$ and the sample size is $n = 1,000$.

1.4.2 Simulation on θ_0

In this section, I compare the monotonic NPIV estimator with the 2SLS and polynomial series estimator. The moment condition of the 2SLS estimator is

$$E\Big[(Y - \alpha_0 - X\theta_0 - W\gamma_0)\mathbb{T}\Big] = 0 \qquad \text{where} \qquad \mathbb{T} = \begin{bmatrix} 1 \\ W \\ Z \end{bmatrix}$$

When estimating the polynomial series estimator, I use $(1, W, W^2, W^3, W^4, W^5)$ as basis functions to estimate both $\tilde{g}$ in step 3 and $\hat{\phi}$ in step 1 of section 1.2.1. In table 1.3, g_0 and ϕ_0 are chosen by linear functions, $g_0(W) = W$ and $\phi_0(W) = 2W$. 'ECR', 'Avg C.I Length', and 'Med C.I Length' represent 'Empirical Coverage Rate', 'Average Confidence Interval Length', and 'Median Confidence Interval Length' respectively. Confidence level is 0.95. In table 1.4, g_0 and ϕ_0 are chosen by monotonic increasing functions, $g_0(W) = \frac{1}{1-e^{-W}}$ and $\phi_0(W) = 1 - e^{-W}$.

Estimator	2SLS	Isotonic	Series
Mean Bias	0.0026	0.0031	-0.0014
Median Bias	0.0033	0.0039	-0.0006
RMSE	0.0635	0.0632	0.0634
ECR	0.948	0.946	0.949
Avg C.I Length	0.2506	0.2485	0.2499
Med C.I Length	0.2485	0.2462	0.2476

Table 1.3: MC Result on θ_0 (Linear g_0 and ϕ_0)

Estimator	2SLS	Isotonic	Series
Mean Bias	-0.0036	-0.0034	-0.0044
Median Bias	-0.0026	-0.0017	-0.0026
RMSE	0.0642	0.0658	0.0655
ECR	0.944	0.930	0.933
Avg C.I Length	0.2509	0.2482	0.2506
Med C.I Length	0.2485	0.2466	0.2488

Table 1.4: MC Result on θ_0 : (Monotonic g_0 and ϕ_0)

1.5 Conclusion

In this chapter, I show the asymptotic normality of the semi-parametric estimator under monotonicity assumption on the semi-nonparametric component. The isotonic regression estimator allows us to avoid predetermining smoothing parameters such as bandwidth and basis functions. However, there are some limitations: the curse of the dimensionality caused by the multivariate comonotonic functions without applying additive separability and the confidence band depending on the choice of the smoothing parameters. When it comes to estimating returns to education, comparing to the OLS and 2SLS estimates, NPIV estimate is bigger. In addition, the confidence band of the level effect of the experience on log wage is available. In simulation studies, if the non-parametric components are monotonic, then NPIV semi-parametric estimator works as good as the other candidates such as 2SLS and NPIV by a series estimation.

REFERENCES

[1] Ai, C., Chen, X. (2003). *"Efficient estimation of models with conditional moment restrictions containing unknown functions." Econometrica, 71(6), 1795-1843.*

[2] Amelunxen, D., Lotz, M., McCoy, M. B., Tropp, J. A. (2014). *"Living on the edge: Phase transitions in convex programs with random data." Information and Inference: A Journal of the IMA, 3(3), 224-294.*

[3] Andrews, D. W. (1994). *"Empirical process methods in econometrics." Handbook of econometrics, 4, 2247-2294.*

[4] Angrist, J. D., Imbens, G. W. (1995). *"Identification and estimation of local average treatment effects." (No. t0118). National Bureau of Economic Research.*

[5] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. A., Wellner, J. A. (2005). *"Semiparametric inference and models." Mimeo, University of California, Berkeley.*

[6] Blundell, R., Chen, X., Kristensen, D. (2007). *"Semi-nonparametric IV estimation of shape-invariant Engel curves." Econometrica, 75(6), 1613-1669.*

[7] Blundell, R., Horowitz, J., Parey, M. (2017). *"Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction." Review of Economics and Statistics, 99(2), 291-304.*

[8] Card, D. (1993). *"Using geographic variation in college proximity to estimate the return to schooling." (No. w4483). National Bureau of Economic Research.*

[9] Chen, X., Christensen, T. M. (2018). *"Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression." Quantitative Economics, 9(1), 39-84.*

[10] Chen, X., Pouzo, D. (2009). *"Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals." Journal of Econometrics, 152(1), 46-60.*

[11] Chen, X., Pouzo, D. (2012). *"Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals." Econometrica, 80(1), 277-321.*

[12] Cheng, G. (2009). *"Semiparametric additive isotonic regression." Journal of Statistical Planning and Inference, 139(6), 1980-1991.*

[13] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018) *"Double/debiased machine learning for treatment and structural parameters." The Econometrics Journal, 21(1), C1-C68.*

[14] Chernozhukov, V., Hansen, C. (2004). *"The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis." Review of Economics and statistics, 86(3), 735-751.*

[15] Chernozhukov, V., Newey, W. K., Santos, A. (2020). *"Constrained conditional moment restriction models." Available at "http://www.econ.ucla.edu/andres/pdfs/ccmmv2.pdf" (The old version is published in 2015)*

[16] Chetverikov, D., Wilhelm, D. (2017). *"Nonparametric instrumental variable estimation under monotonicity." Econometrica, 85(4), 1303-1320.*

[17] Darolles, S., Fan, Y., Florens, J. P., Renault, E. (2011). *" Nonparametric instrumental regression." Econometrica, 79(5), 1541-1565.*

[18] Freyberger, J., Horowitz, J. L. (2015). *"Identification and shape restrictions in nonparametric instrumental variables estimation." Journal of Econometrics, 189(1), 41-53.*

[19] Groeneboom, P., Hendrickx, K. (2018). *"Current status linear regression." The Annals of Statistics, 46(4), 1415-1444.*

[20] Hall, P., Huang, L. S. (2001). *"Nonparametric kernel regression subject to monotonicity constraints." Annals of Statistics, 624-647.*

[21] Han, Q., Wang, T., Chatterjee, S., Samworth, R. J. (2019). *"Isotonic regression in general dimensions." The Annals of Statistics, 47(5), 2440-2471.*

[22] Henderson, D. J., List, J. A., Millimet, D. L., Parmeter, C. F., Price, M. K. (2012). *"Empirical implementation of nonparametric first-price auction models." Journal of Econometrics, 168(1), 17-28.*

[23] Horowitz, J. L., Lee, S. (2017). *"Nonparametric estimation and inference under shape restrictions." Journal of Econometrics, 201(1), 108-126.*

[24] Huang, J. (2002). *"A note on estimating a partly linear model under monotonicity constraints." Journal of Statistical Planning and Inference, 107(1-2), 343-351.*

[25] Imbens, G. W., Newey, W. K. (2009). *"Identification and estimation of triangular simultaneous equations models without additivity." Econometrica, 77(5), 1481-1512.*

[26] Ledoux, Michel, and Michel Talagrand (1991).*"Probability in Banach Spaces: isoperimetry and processes." Springer Science & Business Media.*

[27] Liu, R., Yu, Z. (2019). *" Sample Selection Models with Monotone Control Functions." Available at SSRN 3420924.*

[28] Mammen, E., Yu, K. (2007). *"Additive isotone regression." In Asymptotics: particles, processes and inverse problems (pp. 179-195). Institute of Mathem*

[29] Matzkin, R. L. (1991). *"Semiparametric estimation of monotone and concave utility functions for polychotomous choice models." Econometrica: Journal of the Econometric Society, 1315-1327.*

[30] Matzkin, R. L. (1992). *"Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models." Econometrica: Journal of the Econometric Society, 239-270.*

[31] Matzkin, R. L. (2003). *" Nonparametric estimation of nonadditive random functions." Econometrica, 71(5), 1339-1375.*

[32] Newey, W. K., Powell, J. L., Vella, F. (1999). *"Nonparametric estimation of triangular simultaneous equations models." Econometrica, 67(3), 565-603.*

[33] Newey, W. K., Powell, J. L. (2003). *"Instrumental variable estimation of nonparametric models." Econometrica, 71(5), 1565-1578.*

[34] Powell, J. L. (1994). *"Estimation of semiparametric models." Handbook of econometrics, 4, 2443-2521.*

[35] Racine, J. S., Parmeter, C. F., Du, P. (2009). *" Constrained nonparametric kernel regression: Estimation and inference." In Working paper.*

[36] Robinson, P. M. (1988). *" Root-N-consistent semiparametric regression." Econometrica: Journal of the Econometric Society, 931-954.*

[37] Van Der Vaart, Aad W., and Jon A. Wellner. (1996) *"Weak convergence and empirical processes." Springer, New York, NY, 1996. 16-28.*

[38] Wooldridge, J. M. (2015). *"Control function methods in applied econometrics." Journal of Human Resources, 50(2), 420-445.*

[39] Zhang, C. H. (2002). *"Risk bounds in isotonic regression." The Annals of Statistics, 30(2), 528-555.*

CHAPTER 2

A Confidence Band for the Isotonic Regression Estimator and Application to Semi-parametric Model under Monotonicity

2.1 Introduction

Continuing from the first chapter, I will focus mainly on a partially linear model.

$$Y = X'\theta_0 + g_0(W) + U, \qquad E[U|W, Z] = 0$$

where Y is a dependent variable, X is endogenous, W is exogenous, and Z is an instrumental variable. Assume that all the variables above are 1-dimensional random variables except X and Z. In previous studies, g_0 is considered a nuisance parameter, which is non-parametrically pre-estimated to estimate θ_0 and most literature requires the smoothing parameters such as bandwidth, basis functions, kernels and so on. However, in this chapter, both θ_0 and g_0 are the parameters of interest, and I will provide the asymptotic distribution of L_∞ - distance of the estimator of g_0 under monotonicity. Under assuming that we have $\sqrt{n} - consistent$ semiparametric estimator, we can build a confidence band for the semi-nonparametric component. When it comes to building the confidence band of g_0, the key is to derive the asymptotic distribution of L_∞ - distance of the isotonic regression estimator, which result in Gumbel distribution with the rate $O_p((\frac{\log n}{n})^{1/3})$.

In statistics literature, Durot (2007) deals with the L_p errors on the isotonic regression estimator under the fixed point design. Durot et al. (2012) covers the L_∞ - distance of Grenander's estimator, which is potentially useful for building the confidence band of the isotonic regression.

The chapter consists mainly of 3 parts. In section 2.2, I start the discussion with the confidence band of the isotonic regression estimator and apply it to the semi-nonparametric estimator. In section 2.3, I apply my methods to estimate the returns to schooling with the data set by Card (1993) as in chapter 1. In section 2.4, the MC simulation study is conducted. All the proofs of theorems and lemmas are attached on the appendix.

2.2 Model

2.2.1 Confidence Band of the Monotone Increasing Function

In many empirical works, the non-parametric component in a semi-parametric model is considered the nuisance parameter. However, it could also provide interesting economic implications since it has a direct level effect on the dependent variable. Hence, in this section, we will focus on inferencing g_0 by building the confidence band. In Durot et al. (2012), the authors build the L_∞-distance of Grenander's estimator in general setup. Their work can also be applied to the isotonic regression estimator with some modification. See Durot and Lophuaa (2018) for more detail. However, some of their assumptions are too strong to use in economics literature. On top of their framework, I try to weaken a few assumptions and find the low-level assumptions to build a valid confidence band. Let us assume W to be a 1-dimensional random variable in this entire section. [1]

2.2.1.1 Confidence Band of ϕ_0

Before building the confidence band of g_0, we need to know how to build the confidence band with the isotonic regression estimator under random design. Recall that equation (2)

[1] To the best of my knowledge, the confidence band of the additive isotonic regression and the multivariate isotonic regression are still open questions.

in chapter 1, the non-parametric regression model, is

$$Z_i = \phi_0(W_i) + V_i, \qquad E[V_i|W_i] = 0$$

Let us assume as follows.

Assumption 2.2.1 *(Confidence Band for Monotone Increasing Function $\phi_0(\cdot)$)*
Assume equation (2) as our model with $dim(Z) = dim(W) = 1$. In addition, we assume the following.
(a) The function $\phi_0(\cdot)$ is strictly increasing and differentiable on all over its domain, $[-M, M]$ with

$$\inf_{t\in[-M,M]} \phi_0'(t) > 0, \qquad \sup_{t\in[-M,M]} \phi_0'(t) < \infty.$$

(b) W is a continuous random variable which has the bounded support, $[-M, M]$.
(c) ϕ_0' is Lipschitz continuous. i.e. there exists $C_0 > 0$ such that for all $u, t \in [-M, M]$,

$$|\phi'(u) - \phi'(t)| \leq C_0|u - t|$$

(d) There exists a universal constant c such that $E[V^4] < c < \infty$.

Durot et al.(2012) assume that there exists a decreasing function f, and estimate it by Grenander-type estimator. They suppose that (A1) f is strictly decreasing with the bounded derivative, (A2) there exists a real-valued function $L : [0, 1] \to \mathbb{R}$, which determines the asymptotic distribution of the cumulative sum diagram process distance $F_n - F$ and the distance between the process and the asymptotic distribution should be small enough, (A3) the modulus of the continuity of the cumulative sum diagram process should be small enough, and (A4) the function L in (A2) is twice differentiable, and f and L' are Lipschitz continuous.[2] Durot and Lophuaa (2018) show that the isotonic regression estimator can be represented by the framework by Durot et al.(2012) if f is assumed to be $\phi_0 \circ G^{-1}$ where G is the cumulative distribution of W. F_n is defined differently from Durot et al.(2012) since ϕ_0

[2]Durot et al. (2012) assume more general conditions in (A2) and (A4). See Durot et al.(2012) for the details.

is strictly increasing. Moreover, because of the randomness of W, the L function is more complex. However, the proof strategies are very similar to the previous paper.

Assumption 2.2.1 (a) and (b) are the sufficient condition for (A1) in Durot et al. (2012).[3] Assumption 2.2.1 (c) is sufficient for (A4) in Durot et al. (2012). The most challenging parts are to derive (A2) and (A3). Unfortunately, I could not get the exactly same condition of (A2) since we need to deal with the random W instead of the fixed designed points framework in Durot et al. (2012). However, I could show that the alternative condition is sufficient to get the valid confidence band. (A3) can be also derived by applying the alternative condition. Before the further discussion, we need new notations and definitions which are different from Durot et al. (2012) since I will use the setup for estimating the monotonic increasing function instead of Grenander's estimator. Note that the isotonic estimator of ϕ_0 is

$$\hat{\phi}(W_{(j)}) = \underset{k\leq j}{max} \quad \underset{l\geq j}{min} \quad \frac{1}{l-k+1}\sum_{i=k}^{l} Z_{(i)}$$

where $\{(W_{(i)}, Z_{(i)})\}_{i=1}^{n}$ is the ordered data set with respect to $\{W_i\}_{i=1}^{n}$. Durot and Lophuaa (2018) introduce another representation for the isotonic regression estimator where the function is decreasing. If we consider the increasing function estimation, one can also derive the similar representation. Define

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} Z_i 1_{\{W_i<t\}}$$

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{W_i\leq t\}}$$

$F_n(t)$ is a pseudo cumulative sum diagram. In Durot and Lophuaa (2018), the indicator function is $1_{\{X_i\leq t\}}$. However, the right end point should be skipped when the parameter is an increasing function. $G_n(t)$ is the standard empirical distribution of W. If $\{(W_i, Z_i)\}_{i=1}^{n}$ is given, then we can get n numbers of points as t chooses the value as following. $t \in \mathbb{W} = \{W_1, W_2, \cdots, W_n\}$. Let $W_{(k)}$ denote the k-th order statistic of $(W_1, \cdots, W_n)$. If we get the data sample, so the realization of the order statistic is given as $w_{(k)} = t_k$, then $G_n(t_k) = \frac{k}{n}$

[3] If we define $f = \phi_0 \circ G^{-1}$ where $G(t)$ is the cumulative distribution function of X, then the derivative of f is bounded away from zero and bounded above. In this paper, f is strictly increasing.

and $F_n(t_k) = \frac{1}{n}\sum_{i=1}^{k-1} z_{(i)}$. Draw all the points, such as $\{(G_n(t_k), F_n(t_k)\}_{k=1}^{n}$ and build the greatest convex minorant (GCM) of these points. Lastly, take the right derivative of GCM on each kink to get the isotonic regression estimator.

For detailed analysis, define (pseudo) quantile $G_n^{-1}(s) = \underset{t\in[-M,M]}{sup}\{t : G_n(t) < s\}$. Let $\Lambda_n : [0,1] \to \mathbb{R}$ be a real-valued function defined by $\Lambda_n = F_n \circ G_n^{-1}$. Strictly speaking, I derive the uniform confidence band for $f := \phi_0 \circ G^{-1}$ which correcponds to the confidence band for the quantile transformed function of the parameter of interest. The asymptotic distribution of the L_∞ distance of f can be derived by using the inverse process's asymptotic distribution.

$$U_n(a) = \underset{s\in[0,1]}{argmin}\{\Lambda_n(s) - as\} \tag{2.1}$$

Note that $U_n(a)$ is an estimator for $f^{-1}(a)$. Defining $L(s) = Var(Z1_{\{W<G^{-1}(s)\}} - \phi_0(G^{-1}(s))1_{\{W\leq G^{-1}(s)\}})$, by Koltchinskii coupling, one can show that there exists a real valued function $L : [0,1] \to \mathbb{R}$ which is positive and increasing with respect to s, such that it satisfies:

$$P(n^{1/4}\|(\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_\infty > x) \lesssim x^{-4}$$

where $\Lambda = F \circ G^{-1}$ and B_n denote standard Brownian motion and $x > K \log n$ for some $K > 0$. See lemma 6.6 for the detailed proof in the appendix. Then the asymptotic distribution of U_n can be derived by the following theorem.

Theorem 2.2.1 ***(Asymptotic Distribution of*** $U_n(a)$***)***

Suppose that we have equation (2) as the model. Let assumption 2.2.1 be satisfied. For all $a \in \mathbb{R}$*, define the normalizing function* $A(a)$

$$A(a) = \frac{f'(f^{-1}(a))^{2/3}}{\{4L'(f^{-1}(a))\}^{1/3}}$$

Let $0 \leq u < v \leq 1$ *be fixed, and let* $(\alpha_n)_n$ *and* $(\beta_n)_n$ *be sequences such that* $\alpha_n \to 0$*,* $\beta_n \to 0$ *and* $0 \leq u + \alpha_n < v - \beta_n \leq 1$ *for* n *sufficiently large. Define*

$$S_n = n^{1/3} \sup_{a\in[f(u+\alpha_n), f(v-\beta_n)]} A(a)|U_n(a) - f^{-1}(a)|$$

Then

$$P(S_n \le u_n) \to \exp\Big\{ - 2\tau \int_u^v \frac{f'(t)^{2/3}}{\{4L'(t)\}^{1/3}} dt \Big\}$$

for any sequence $(u_n)_n$ *such that* $u_n \to \infty$ *in such a way that* $n^{1/3}\mu(u_n) \to \tau > 0$, *where* μ *denotes the density of* $\zeta(0) = \underset{t\in\mathbb{R}}{argmax}\{B_n(t) - t^2\}$.

Moreover, for all $x \in \mathbb{R}$,

$$P\bigg(\log n\Big\{ \Big(\frac{2}{\log n}\Big)^{1/3} S_n - \mu_n \Big\} \le x \bigg) \to \exp\Big\{ - e^{-x} \Big\}$$

Once we get the asymptotic distribution of the inverse process, we can connect it to L_∞-distance between $\hat{f} = \hat{\phi} \circ G_n^{-1}$ and $f = \phi_0 \circ G^{-1}$ by lemma 6.21 in the appendix, such as

$$\begin{aligned} &\sup_{u\in(s,1-t]} B(u)|\hat{f}(u) - f(u)| \\ &\quad = \sup_{a\in[f(s),f(1-s)]} A(a)\Big|U_n(a) - g(a)\Big| + O_p\Big(\Big(\frac{\log n}{n}\Big)^{2/3}\Big) \end{aligned} \tag{2.2}$$

Now we can build the confidence band for f by the following theorem.

Theorem 2.2.2 *(Confidence Band for the Isotonic Regression) Suppose that we have equation (2.2) as the model. All the assumptions in theorem 2.2.1 hold. Then for fixed u and v such that* $0 \le u < v \le 1$, *for any sequence* $\alpha_n \to 0$, $\beta_n \to 0$ *such that* $1 - v + \beta_n, u + \alpha_n > n^{-1/3}(\log n)^{-2/3}$, *we have that for any* $x \in \mathbb{R}$,

$$P\bigg(\log n\Big\{ \Big(\frac{n}{\log n}\Big)^{1/3} \sup_{s\in(u+\alpha_n, v-\beta_n} \frac{|\hat{f}(s) - f(s)|}{\{2f'(s)L'(s)\}^{1/3}} - \mu_n \Big\} \le x \bigg) \to \exp\{-e^{-x}\}$$

as $n \to \infty$, *where*

$$\mu_n = 1 - \frac{\kappa}{2^{1/3}(\log n)^{2/3}} + \frac{1}{\log n}\Big[\frac{1}{3}\log\log n + \log(\lambda C_{f,L})\Big]$$

with

$$C_{f,L} = 2\int_u^v \Big(\frac{f'(s)^2}{L'(s)}\Big)^{1/3} ds$$

and $\lambda \approx 1.79425$ *and* $\kappa \approx 2.94582$.

Note that λ and κ are some constants calculated by Groeneboom (1989). The proof of theorem 2.2.2 is in the appendix. From the theorem, we can build the confidence band of $f(s) = \phi_0 \circ G^{-1}(s)$ as $[CBL_{\phi_0}(s), CBU_{\phi_0}(s)]$ where

$$CBL_{\phi_0}(s) = \hat{f}(s) - \left(\frac{\log n}{n}\right)^{1/3} \left\{2f'(s)L'(s)\right\}^{1/3} \left(\mu_n + \frac{x}{\log n}\right)$$

$$CBU_{\phi_0}(s) = \hat{f}(s) + \left(\frac{\log n}{n}\right)^{1/3} \left\{2f'(s)L'(s)\right\}^{1/3} \left(\mu_n + \frac{x}{\log n}\right)$$

with $\exp\{-e^{-x}\}$ confidence level. To get a feasible confidence band, we need the consistent estimators for L' and f'. By some algebra, we get $L'(s)$ as

$$\begin{aligned} L'(s) &= Var(V) \\ &+ 2\phi_0'(G^{-1}(s))\left(\frac{1-s}{p(G^{-1}(s))}\right)\left\{\phi_0(G^{-1}(s)) - E\left[\phi_0(W)1_{\{W \leq G^{-1}(s)\}}\right]\right\} \end{aligned}$$

Then we can construct the estimators for $Var(V)$ and $E\left[\phi_0(W)1_{\{W \leq G^{-1}(s)\}}\right]$ by using a sample analog such as

$$\widehat{Var}(V) = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \hat{\phi}(W_i))^2$$

$$E_n\left[\phi_0(W)1_{\{W \leq G^{-1}(s)\}}\right] = \frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(W_i)1_{\{W \leq G_n^{-1}(s)\}}$$

The remaining parameters we need to estimate are ϕ_0', f' and $p(G^{-1}(s))$. In practice, one can estimate ϕ_0' and f' by using isotonic regression, but the estimator seems not appropriate for estimating the slope of the parameter since it is not smoothly increasing. Hence, we need to use another estimation method, such as a series estimator, which has advantages in estimating smooth functions. When it comes to estimating $p(G^{-1}(s))$, I use the kernel density estimator of W and plug in G_n^{-1} instead of G^{-1}.

2.2.1.2 The Confidence Band of g_0

Theorem 2.2.2 gives us the idea of how to build the confidence band of the isotonic regression estimator where the standard non-parametric regression model is assumed. To build the confidence band of the semi-nonparametric estimator, define $D_i = Y_i - X_i\theta_0$ and let the isotonic regression estimator of D on W be $\tilde{g}_0$. Then by triangular inequality, we have $\|\hat{g} - g_0\|_\infty \leq \|\tilde{g}_0 - g_0\|_\infty + \|\hat{g} - \tilde{g}_0\|_\infty$. Note that $\|\hat{g} - \tilde{g}_0\|_\infty = O_p(n^{-1/2})$ by similar procedure in the proof of theorem 2.2.2 Then $\|\hat{g} - g_0\|_\infty$ and $\|\hat{g} - \tilde{g}_0\|_\infty$ have the same asymptotic property since they are $O_p(n^{-1/3})$. As a result, we can build $100(1-\alpha)\%$ confidence band of $\xi := g_0 \circ G^{-1}$ as $[CBL_{g_0}(s), CBU_{g_0}(s)]$ where

$$CBL_{g_0}(s) = \hat{\xi}(s) - \left(\frac{\log n}{n}\right)^{1/3} \left\{2\xi'(s)L'_{g_0}(s)\right\}^{1/3} \left(\mu_n + \frac{x}{\log n}\right)$$

$$CBU_{g_0}(s) = \hat{\xi}(s) + \left(\frac{\log n}{n}\right)^{1/3} \left\{2\xi'(s)L'_{g_0}(s)\right\}^{1/3} \left(\mu_n + \frac{x}{\log n}\right)$$

where $\hat{\xi} = \hat{g} \circ G_n^{-1}$ with $\exp\{-e^{-x}\}$ confidence level. To get a feasible confidence band, we need the consistent estimators for L'_{g_0} and ξ'. By some algebra, we get $L'_{g_0}(s)$ as

$$\begin{aligned} L'_{g_0}(s) &= Var(U) \\ &\quad + 2g'_0(G^{-1}(s))\left(\frac{1-s}{p(G^{-1}(s))}\right)\left\{g_0(G^{-1}(s)) - E\left[g_0(W)1_{\{W \leq G^{-1}(s)\}}\right]\right\} \end{aligned}$$

Then we can construct the estimators for $Var(U)$ and $E\left[g_0(W)1_{\{W \leq G^{-1}(s)\}}\right]$ by using a sample analog such as

$$\widehat{Var}(U) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'\hat{\theta} - \hat{g}(W_i))^2$$

$$E_n\left[g_0(W)1_{\{W \leq G^{-1}(s)\}}\right] = \frac{1}{n}\sum_{i=1}^{n}\hat{g}(W_i)1_{\{W \leq G_n^{-1}(s)\}}$$

In simulation studies, I use a polynomial series to estimate g'_0, ξ' and a kernel estimator for $G^{-1}(s)$.

2.3 Returns to Education

Card (1993) gives several models for estimating the returns to schooling where the years of education are correlated with the individual's productivity by using some instruments, such as college proximity, parent's education, IQ score, and so on. I use the data which is

Variables	Mean	Median	Variance	Inter-Quartile
Log Wage	6.264	6.287	0.195	[5.984, 6.565]
Yrs of Educ	13.292	13	7.029	[12, 16]
Age	28.115	28	9.824	[25, 31]
KWW Score	33.546	34	74.043	[28, 40]

Table 2.1: Summary Statistics

available on his website. [4] The sample size is $n = 2,962$, and the model is constructed as follows.

$$Y_i = X_i\theta_0 + g_0(W_i) + U_i, \qquad E[U_i|W_i, Z_i] = 0$$

where Y is log of wage, X is years of education, W is age, and Z is 'Knowledge of the world' (KWW) score.[5] In Card (1993), the author defends why the 'KKW' score is a proper instrument theoretically and statistically. Note that ϕ_0 is the conditional expectation of the 'KWW' score given a certain age level. It could be considered as a proxy of experience, so it is reasonable to assume that ϕ_0 is monotonic increasing with respect to W. Similarly, g_0 also measures the effect of experience on wage. Hence, we can impose the monotonic restriction which is actually less restrictive than the linear regression model. The brief description of the summary statistics for each variable is in table 2.1.

On the 1st stage regression, we get the isotonic regression estimator of $E[Z|W]$. Figure 2.1 describes the confidence band of the conditional expectation of the 'KWW' score given the age quantile with setting the lower quantile 0.2 and the upper quantile 0.8.

The effect of age on the log of wage can be captured by g_0 which is derived by the isotonic regression. In figure 2.2, the isotonic regression estimator jumps around 0.4, 0.5, and 0.6

[4] There are missing data for individuals in 1976 and I omit those.

[5] The KWW score variable was administered to young men who represent the civilian, noninstitutional population of males 14-24 years of age in 1966 in the U.S. It was a part of the initial interview in the National Longitudinal Surveys (NLS).

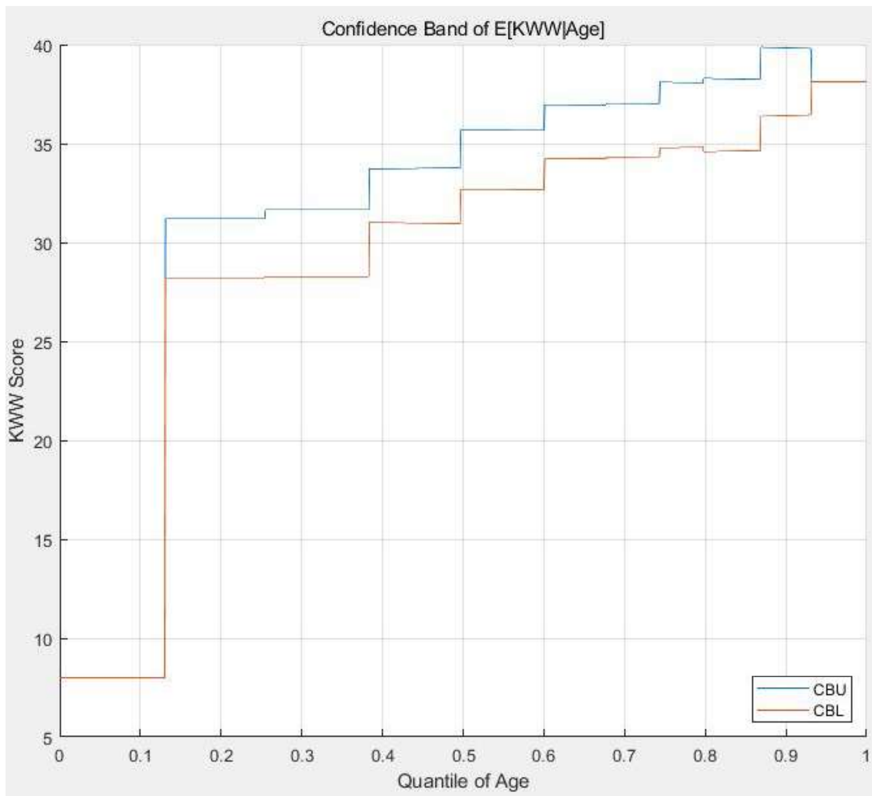

Figure 2.1: The Confidence Band of $E[Z|W]$

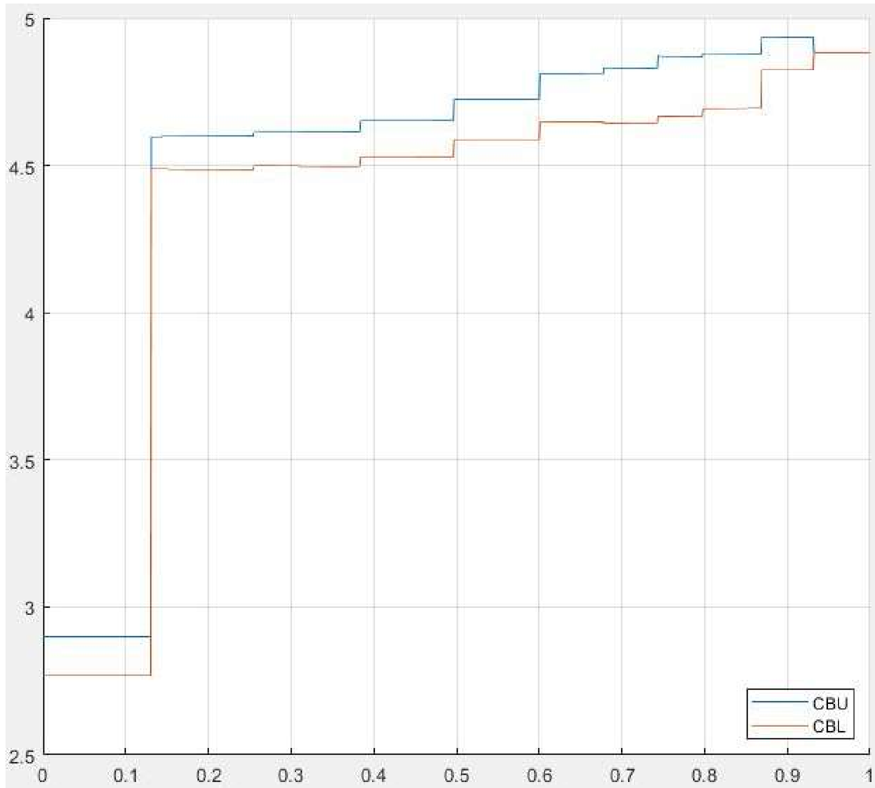

Figure 2.2: The Confidence Band of g_0

Quantile	[0.2, 0.4]	[0.4, 0.5]	[0.5, 0.6]	[0.6, 0.8]
Avg. CB	[4.51, 4.60]	[4.55, 4.64]	[4.60, 4.71]	[4.66, 4.82]

Table 2.2: The Confidence Band of g_0

quantile. The average confidence band length of each age quantile is in table 2.2.

Quantile	[0.05,0.95]	[0.10,0.90]	[0.25, 0.75]	[0.4, 0.6]
ECR	0.546	0.681	0.863	0.996
Avg C.B Length	1.3904	1.3429	1.2198	1.0360
Med C.B Length	1.3886	1.3419	1.2192	1.0349

Table 2.3: MC Result on $f = g_0 \circ G^{-1}$

2.4 Simulation

2.4.1 DGP

In the simulation studies, the DGP is as below for the semi-parametric estimator $\hat{\theta}$.

$$\begin{aligned}
V &\sim N(0,1) \\
U &\sim N(0,1) \\
X &= 0.5U + 0.5V + N(0,1) \\
W &\sim N(0,1) \\
Y &= X\theta_0 + g_0(W) + U \\
Z &= \phi_0(W) + V
\end{aligned}$$

where $\theta_0 = 1$. The number of replications is $M = 1,000$ and the sample size is $n = 1,000$.

2.4.2 Simulation on g_0

I set $g_0(W) = \phi_0(W) = 1 - e^{-W}$. The isotonic regression estimator confidence band performs not well on the extreme quantile of g_0. In table 2.3, the lower quantile, v and the upper quantile, u for confidence band are chosen by $[0.05, 0.95]$, $[0.10, 0.90]$, $[0.25, 0.75]$, and $[0.4, 0.6]$. ECR (Empirical Coverage Rate), Avg C.B Length (Average Confidence Band Length), and Med C.B Length (Median Confidence Band Length) are reported with the confidence level 0.95. As the difference of lower and upper quantile decreases, the confidence band length also decreases since $C_{f,L}$ in theorem 2.2.2 decreases.

Quantile	[0.2,0.4]	[0.4,0.5]	[0.5, 0.6]	[0.6, 0.8]
ECR	0.718	0.991	0.994	1.000
Avg C.B Length	1.1390	0.9249	0.9115	1.0404
Med C.B Length	1.1374	0.9243	0.9118	1.0376

Table 2.4: MC Result on $f = g_0 \circ G^{-1}$

In table 2.4, we can check that the isotonic regression confidence band works well on the flatter part than the steep one. Set the lower quantile v and the upper quantile u for confidence band as $[0.2, 0.4], [0.4, 0.5], [0.5, 0.6]$, and $[0.6, 0.8]$. The empirical coverage rate is increasing on the upper quantile since f is concave and increasing.

2.5 Conclusion

In this paper, I show how to build a confidence band of the non-parametric component under monotonicity. When it comes to estimating returns to schooling model, the confidence band of the level effect of the experience on log wage is available. In simulation studies, if the non-parametric components are monotonic, then NPIV semi-parametric estimator works as good as the other candidates such as 2SLS and NPIV by a series estimation. However, the performance of the confidence band of the non-parametric component is highly affected by the choice of the quantile and the steepness of the true parameter.

REFERENCES

[1] Amelunxen, D., Lotz, M., McCoy, M. B., Tropp, J. A. (2014). *"Living on the edge: Phase transitions in convex programs with random data." Information and Inference: A Journal of the IMA, 3(3), 224-294.*

[2] Andrews, D. W. (1994). *"Empirical process methods in econometrics." Handbook of econometrics, 4, 2247-2294.*

[3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. A., Wellner, J. A. (2005). *"Semiparametric inference and models." Mimeo, University of California, Berkeley.*

[4] Blundell, R., Chen, X., Kristensen, D. (2007). *"Semi-nonparametric IV estimation of shape-invariant Engel curves." Econometrica, 75(6), 1613-1669.*

[5] Blundell, R., Horowitz, J., Parey, M. (2017). *"Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction." Review of Economics and Statistics, 99(2), 291-304.*

[6] Card, D. (1993). *"Using geographic variation in college proximity to estimate the return to schooling." (No. w4483). National Bureau of Economic Research.*

[7] Chen, X., Christensen, T. M. (2018). *"Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression." Quantitative Economics, 9(1), 39-84.*

[8] Chen, X., Pouzo, D. (2009). *"Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals." Journal of Econometrics, 152(1), 46-60.*

[9] Chen, X., Pouzo, D. (2012). *"Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals." Econometrica, 80(1), 277-321.*

[10] Cheng, G. (2009). *"Semiparametric additive isotonic regression." Journal of Statistical Planning and Inference, 139(6), 1980-1991.*

[11] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J. (2018) *"Double/debiased machine learning for treatment and structural parameters." The Econometrics Journal, 21(1), C1-C68.*

[12] Chernozhukov, V., Newey, W. K., Santos, A. (2020). *"Constrained conditional moment restriction models." Available at "http://www.econ.ucla.edu/andres/pdfs/ccmmv2.pdf" (The old version is published in 2015)*

[13] Chetverikov, D., Wilhelm, D. (2017). *"Nonparametric instrumental variable estimation under monotonicity." Econometrica, 85(4), 1303-1320.*

[14] Darolles, S., Fan, Y., Florens, J. P., Renault, E. (2011). *" Nonparametric instrumental regression." Econometrica, 79(5), 1541-1565.*

[15] Durot, C. (2007) *"On the $\mathbb{L}_p$-error of monotonicity constrained estimators." The Annals of Statistics, 35(3), 1080-1104.*

[16] Durot, C., Kulikov, VN, Lopuhaä, HP (2012) *"The limit distribution of the L_∞-error of Grenander-type estimators." The Annals of Statistics , 40 (3), 1578-1608.*

[17] Durot, C., Kulikov, VN, Lopuhaä, HP (2012) *"The limit distribution of the L_∞-error of Grenander-type estimators", supplemental materials. The Annals of Statistics , 40 (3), 1578-1608.*

[18] Freyberger, J., Horowitz, J. L. (2015). *"Identification and shape restrictions in nonparametric instrumental variables estimation." Journal of Econometrics, 189(1), 41-53.*

[19] Groeneboom, P. (1989). *"Brownian motion with a parabolic drift and Airy functions." Probability theory and related fields, 81(1), 79-109.*

[20] Groeneboom, P., Hendrickx, K. (2018). *"Current status linear regression." The Annals of Statistics, 46(4), 1415-1444.*

[21] Hall, P., Huang, L. S. (2001). *"Nonparametric kernel regression subject to monotonicity constraints." Annals of Statistics, 624-647.*

[22] Horowitz, J. L., Lee, S. (2017). *"Nonparametric estimation and inference under shape restrictions." Journal of Econometrics, 201(1), 108-126.*

[23] Huang, J. (2002). *"A note on estimating a partly linear model under monotonicity constraints." Journal of Statistical Planning and Inference, 107(1-2), 343-351.*

[24] Imbens, G. W., Newey, W. K. (2009). *"Identification and estimation of triangular simultaneous equations models without additivity." Econometrica, 77(5), 1481-1512.*

[25] Ledoux, Michel, and Michel Talagrand (1991).*"Probability in Banach Spaces: isoperimetry and processes." Springer Science & Business Media.*

[26] Newey, W. K., Powell, J. L., Vella, F. (1999). *"Nonparametric estimation of triangular simultaneous equations models." Econometrica, 67(3), 565-603.*

[27] Newey, W. K., Powell, J. L. (2003). *"Instrumental variable estimation of nonparametric models." Econometrica, 71(5), 1565-1578.*

[28] Powell, J. L. (1994). *"Estimation of semiparametric models." Handbook of econometrics, 4, 2443-2521.*

[29] Racine, J. S., Parmeter, C. F., Du, P. (2009). *" Constrained nonparametric kernel regression: Estimation and inference." In Working paper.*

[30] Robinson, P. M. (1988). *" Root-N-consistent semiparametric regression." Econometrica: Journal of the Econometric Society, 931-954.*

[31] Van Der Vaart, Aad W., and Jon A. Wellner. (1996) *"Weak convergence and empirical processes." Springer, New York, NY, 1996. 16-28.*

[32] Wooldridge, J. M. (2015). *"Control function methods in applied econometrics." Journal of Human Resources, 50(2), 420-445.*

[33] Zhang, C. H. (2002). *"Risk bounds in isotonic regression." The Annals of Statistics, 30(2), 528-555.*

CHAPTER 3

A Small Sigma Approach to Certain Problems in Errors-in-Variables and Panel Data Models

3.1 Introduction

[1]Nonlinear panel data analysis is subject to fundamental difficulties presented by the incidental parameters problem. Except for a few cases, a reasonable theoretical approach that treats the individual fixed effect as an infinite dimensional component of the model often leads to unpalatable conclusions for practice that the parameters of interest are unidentified[2]. Such an impossibility was an important motivation behind the large n, large T asymptotic approximation[3], where the goal is to make an approximate inference on the parameters of interest. Despite the progress in this genre of literature, we should still contend with the fact that it is based on large T approximation, which is expected to make the approximation less appealing for empirical applications where the time series variation is small.

The purpose of this paper is to propose an alternative estimator that may be useful when T is small. Our approach is based on a synthesis of two seemingly unrelated literatures in the past. First, we adopt the panel data approach exploiting a certain exchangeability and a sufficient statistic. This approach began with Mundlak (1978), which was later adopted by Altonji and Matzkin (2005) in a nonparametric and nonseparable framework, and by Arkhangelsky and Imbens (2019) for estimation of treatment effects with clustered data.

---

[1]This papar is coauthored by Professor Jerry Hausman and Professor Jinyong Hahn.

[2]See Chamberlain (2010) or Hahn (2001).

[3]See Arellano and Hahn (2007) for a discussion.

This approach, which can be argued to be based on some 'sufficient statistic', was employed to justify the within estimator in linear models, but it has been less successful as a way of producing widely used methods of estimation for nonlinear models. Second, we adopt the small sigma approximation proposed by Kadane (1971) and Amemiya (1985). The small sigma asymptotics was originally proposed to analyze various existing estimators when the variances of errors are very small, but Amemiya (1990) extended the approach and used Taylor series expansions to develop a more refined estimator in nonlinear errors-in-variables models. Amemiya's (1990) approach continued the earlier approach of Kadane (1971), where the variance of the error is modeled as a function of the sample size. Although modeling of variances as functions of sample sizes was an elegant asymptotic framework, we believe that it complicated the analysis a bit, and ended up hiding the intuitive and simple nature of the underlying idea, which is to use the Taylor series approximation (up to a finite order) to develop an approximate moment that is easy to analyze and compute. We therefore adopt only the spirit of Amemiya's (1990) analysis, not the entire asymptotic framework.

Although development of a convenient panel data analysis was our original motivation, we believe that we have made independent contribution to the errors-in-variables (EIV) literature as well. Typical estimators in the literature take the form of IV estimators, which may be subject to the well-known finite sample problems and problematic validity of the instruments. Therefore, it may be interesting to develop estimators that do not require repeated measurements or IV, and the small sigma approach may provide a computationally convenient and statistically stable alternative. Based on the small sigma approach, we develop straightforward procedures, which allow for conditional heteroscedasticity (CH), an added benefit. Almost all of the higher order moment approaches to EIV depend on a conditional homoscedasticity assumption, yet in the last 40 years econometricians have become aware of the importance of CH in economic data. We demonstrate why the higher order moment approach does not work in the CH situation, and we propose a small sigma approach to the problem which works well in practice.

Susanne Schennach has considered higher order moment approaches in two recent survey papers Schennach (2016,2020). In both of these papers she considers many approaches to the EIV problem, but in the higher order moment situation she follows the classical approaches of Geary (1942) and Reiersol (1950) which assume independence of the measurement error from the true underlying unobserved variable. An independence assumption does not allow for conditional heteroscedasticity (CH). From the survey papers and other papers considered in the survey, we are unaware that CH is considered. As we discuss in this paper, the existence of CH causes the higher order moment estimators to be unidentified in the models we consider. We propose a pragmatic technique which allows approximate estimation in the situation of CH using higher order moments.

Our paper is closely related to Salanié and Wolak (2019), who recently calculated a Taylor series approximation of the moment for the commonly used demand system due to Berry, Levinsohn, and Pakes (1995, BLP hereafter). By the very nature of Taylor series approximation, our approximate moment are easy to code and implement in general, which is true in BLP as demonstrated by Salanié and Wolak (2019). While the approximate moment was calculated by Salanié and Wolak (2019), we make an additional contribution to the literature by recognizing that the BLP specification adopted in practice are in fact misspecified due to certain truncation problem. The number of products considered for estimation in BLP specification is often smaller than the actual number of products in the market, and therefore, such truncation bias is built into the BLP specification in application. It is therefore unclear whether the standard estimators for the BLP specification will dominate the small sigma approach. We compare the standard estimator and the small sigma estimator under such truncation through Monte Carlo simulation.

Our focus is development of convenient approximate moments and related point estimators. As for statistical inference such as confidence intervals, we refer to a recently emerging literature. This recent literature recognized that the moments adopted in empirical practice may not be exactly satisfied, and proposed to either understand its sensitivity to the violation of the moment restriction or develop a confidence intervals that accommodate such violations. See Hahn and Hausman (2005), Andrews, Gentzkow and Shapiro (2017), or Armstrong, and Kolesár (2019). In our view, this literature is still at its infancy, and further development and refinement are expected, which this paper is well-positioned to leverage.

3.2 Errors-in-Variables in Linear Models

In order to explain the basic intuition of our approach, it would be useful to start with a relatively simple model and examine how it would apply to basic linear errors-in-variables (EIV) models. The literature on measurement error is large and impossible to review in a few sentences, but it is probably safe to say that a typical analysis is predicated on the availability of some instruments including repeated measurements, at least in the econometric literature. A somewhat smaller literature does exist, where estimators are developed even when such additional variables are unavailable and instead uses higher order moments, including Geary (1942), Lewbel (1997), and Schennach and Hu (2013). These estimators are based on the assumption of the independence of the measurement error with the observed variables and, thus, do not allow for CH. We illustrate how the small sigma approach can be used to develop a convenient estimator.

The small sigma approach is based on an approximate moment condition. Consider a linear EIV model

$$\begin{aligned} y &= \alpha + \beta x_* + \varepsilon, \\ x &= x_* + v, \end{aligned} \tag{3.1}$$

where we observe (y, x). The x is a proxy for the true regressor x_*, and the measurement error v is assumed to be independent of (x_*, ε), i.e., classical EIV. We simplify notations by assuming that the (marginal) mean of v is zero, with similar assumption on ε. As for the relationship between x_*, ε, we allow for heteroscedasticity, i.e., we only assume that $E\left[\varepsilon \mid x_*\right] = 0$. Using higher order moments has a long history in the EIV literature, but without conditional heteroscedasticity (CH), including the recent paper by Schennach and Hu (2013).[4] However, it has become recognized that CH is important in applied work, so

[4]Schennach and Hu (2013) provide an excellent review of the literature.

extending the higher moments approach to allow for CH is an important topic since the usual methods do not work with CH.

Because $E[y] = \alpha + \beta E[x_*]$ and $E[x] = E[x_*]$, so we may write

$$\widetilde{y} = \beta \widetilde{x}_* + \varepsilon,$$
$$\widetilde{x} = \widetilde{x}_* + v,$$

where $\widetilde{y} \equiv y - E[y]$, $\widetilde{x} \equiv x - E[x]$, $\widetilde{x}_* \equiv x_* - E[x_*]$ all have mean zero by construction.

We see that the second moments are such that

$$E\left[\widetilde{x}^2\right] = E\left[\widetilde{x}_*^2\right] + E\left[v^2\right],$$
$$E\left[\widetilde{y}^2\right] = \beta^2 E\left[\widetilde{x}_*^2\right] + E\left[\varepsilon^2\right],$$
$$E\left[\widetilde{x}\widetilde{y}\right] = \beta E\left[\widetilde{x}_*^2\right].$$

If we are to use this system of equations as a basis of estimation, we recognize that we have three equations and four unknowns $(\beta, E[\widetilde{x}_*^2], E[v^2], E[\widetilde{\varepsilon}^2])$. Does this difficulty of identification change if we examine third moments in addition to the second moments? Using the assumption that (i) v is independent of (x_*, ε) and has mean zero; and (ii) $E[\varepsilon | x_*] = 0$, it is straightforward to show that

$$E\left[\widetilde{x}^3\right] = E\left[\widetilde{x}_*^3\right] + E\left[v^3\right],$$
$$E\left[\widetilde{x}^2\widetilde{y}\right] = E\left[\widetilde{x}_*^3\right]\beta,$$
$$E\left[\widetilde{x}\widetilde{y}^2\right] = E\left[\widetilde{x}_*^3\right]\beta^2 + E\left[\widetilde{x}_*\varepsilon^2\right],$$
$$E\left[\widetilde{y}^3\right] = E\left[\widetilde{x}_*^3\right]\beta^3 + 3E\left[\widetilde{x}_*\varepsilon^2\right]\beta + E\left[\varepsilon^3\right].$$

Relative to the second moments, we have four more equations, and four more unknowns including $(E[\widetilde{x}_*^3], E[v^3], E[\varepsilon^3], E[\widetilde{x}_*\varepsilon^2])$, so the problem continues.

We now discuss how the small sigma approach can be used. For this purpose, let's write with

$$x = x_* + \sigma v \tag{3.2}$$

and note that the second and third moments are

$$E\left[\widetilde{x}^2\right] = E\left[\widetilde{x}_*^2\right] + \sigma^2 E\left[v^2\right],$$
$$E\left[\widetilde{y}^2\right] = \beta^2 E\left[\widetilde{x}_*^2\right] + E\left[\varepsilon^2\right],$$
$$E\left[\widetilde{x}\widetilde{y}\right] = \beta E\left[\widetilde{x}_*^2\right].$$

and

$$E\left[\widetilde{x}^3\right] = E\left[\widetilde{x}_*^3\right] + \sigma^3 E\left[v^3\right],$$
$$E\left[\widetilde{x}^2\widetilde{y}\right] = E\left[\widetilde{x}_*^3\right]\beta,$$
$$E\left[\widetilde{x}\widetilde{y}^2\right] = E\left[\widetilde{x}_*^3\right]\beta^2 + E\left[\widetilde{x}_*\varepsilon^2\right],$$
$$E\left[\widetilde{y}^3\right] = E\left[\widetilde{x}_*^3\right]\beta^3 + 3E\left[\widetilde{x}_*\varepsilon^2\right]\beta + E\left[\varepsilon^3\right].$$

If we let $\sigma \to 0$, and ignore the smallest terms (i.e., $O\left(\sigma^3\right)$), we get

$$E\left[\widetilde{x}^2\right] = E\left[\widetilde{x}_*^2\right] + E\left[(\sigma v)^2\right],$$
$$E\left[\widetilde{y}^2\right] = \beta^2 E\left[\widetilde{x}_*^2\right] + E\left[\varepsilon^2\right],$$
$$E\left[\widetilde{x}\widetilde{y}\right] = \beta E\left[\widetilde{x}_*^2\right].$$

and

$$E\left[\widetilde{x}^3\right] \approx E\left[\widetilde{x}_*^3\right],$$
$$E\left[\widetilde{x}^2\widetilde{y}\right] = E\left[\widetilde{x}_*^3\right]\beta,$$
$$E\left[\widetilde{x}\widetilde{y}^2\right] = E\left[\widetilde{x}_*^3\right]\beta^2 + E\left[\widetilde{x}_*\varepsilon^2\right],$$
$$E\left[\widetilde{y}^3\right] = E\left[\widetilde{x}_*^3\right]\beta^3 + 3E\left[\widetilde{x}_*\varepsilon^2\right]\beta + E\left[\varepsilon^3\right].$$

We now have a system of seven (approximate) equations with seven unknowns, including β, $E\left[\widetilde{x}_*^2\right]$, $E\left[(\sigma v)^2\right]$, $E\left[\varepsilon^2\right]$, $E\left[\varepsilon^3\right]$, $E\left[\widetilde{x}_*^3\right]$, and $E\left[\widetilde{x}_*\varepsilon^2\right]$. Therefore, one may argue that the parameters are approximately identified.[5]

[5] Presumably it is possible to go to the fourth or even higher moments and obtain more sources of (approximate) identification. Using more moments leading to over-identification has previously been discussed in the part of the EIV literature which uses higher order moments for identification, but the discussion has been in the context of independence of the measurement error which rules out CH, e.g. Aigner et. al. (1984).

In fact, we can use the two (approximate) moments

$$E\left[\widetilde{x}^3\right] \approx E\left[\widetilde{x}_*^3\right],$$
$$E\left[\widetilde{x}^2\widetilde{y}\right] = E\left[\widetilde{x}_*^3\right]\beta,$$

and identify

$$\beta \approx \frac{E\left[\widetilde{x}^2\widetilde{y}\right]}{E\left[\widetilde{x}^3\right]} \tag{3.3}$$

The sample counterpart of this approximate identification strategy would be an IV estimator for the equation $\widetilde{y} = \beta\widetilde{x} +$ "error" using $\widetilde{x}^2$ as IV. Obviously the estimator is predicated on the assumption that $E\left[\widetilde{x}^3\right] \neq 0$, which rules out normal distribution as has been known in the literature since Reiersol (1950).

We conducted Monte Carlo simulations to examine the performance of the small sigma estimator relative to Geary's (1942) estimator.[6] In the Monte Carlo simulations, we specified $\log x_* \sim N(2,1)$. We considered several specifications of conditional distribution $\mathcal{L}(\varepsilon|\, x_*)$ of ε given x_* and the distribution of v, including (i) $\mathcal{L}(\varepsilon|\, x_*) \sim N(0, x_*^2/5)$ and $v \sim N(0, \sigma^2)$; (ii) $\mathcal{L}(\varepsilon|\, x_*) \sim ex^*\left(|x_*|/\sqrt{5}\right)$ and $v \sim \sigma \cdot ex^*(1)$;[7] and (iii) $\mathcal{L}(\varepsilon|\, x_*) \sim \exp^*\left(N\left(\ddot{\mu}, x_*^2/5\right)\right)$ and $v \sim \exp^*\left(N\left(\bar{\mu}, \bar{\sigma}^2\right)\right)$, where $\ddot{\mu}$, $\bar{\mu}$ and $\bar{\sigma}$ are chosen such that $\exp\left(N\left(\ddot{\mu}, x_*^2/5\right)\right)$ has mean 2, $\exp\left(N\left(\bar{\mu}, \bar{\sigma}^2\right)\right)$ has mean 2 and $\mathrm{Var}(v) = \sigma^2$.[8] All specifications were chosen such that $E[\varepsilon|x_*] = 0$, $\mathrm{Var}(\varepsilon) = 1$, and $\mathrm{Var}(v) = \sigma^2$. We considered $\sigma^2 = 1$, 0.5, 0.25 and 0.1, and the sample size equal to 1,000. The number of Monte Carlo simulation is $M = 10,000$.

Our simulation results are summarized in Tables C.1 - C.3. As is predicted by the theory, the small sigma estimator performs better when $\mathrm{Var}(v)$ is small. We find it encouraging that its performance often dominates Geary's estimator, as is predicted by theory. We note that

---

[6] We note that Lewbel (1997) generalizes Geary's (1943) estimator of the basic model to a more complicated model.

[7] We use the symbol $ex^*(\mu)$ to denotes the exponential distribution with mean μ subtracted by μ, i.e., the de-meaned exponential distribution. In the tables, we called it exponential distribution for simplicity.

[8] We use the symbol $\exp\left(N\left(\tilde{\mu}, \tilde{\sigma}^2\right)\right)$ to denote a log normal distribution, and $\exp^*\left(N\left(\tilde{\mu}, \tilde{\sigma}^2\right)\right)$ to denote a de-meaned log normal distribution. To be more specific, $\nu \sim \exp^*\left(N\left(\tilde{\mu}, \tilde{\sigma}^2\right)\right)$ implies that ν can be written as $\tilde{\nu} - E[\tilde{\nu}]$, where $\tilde{\nu} \sim \exp\left(N\left(\tilde{\mu}, \tilde{\sigma}^2\right)\right)$, i.e., $\log\tilde{\nu} \sim N\left(\tilde{\mu}, \tilde{\sigma}^2\right)$. In the tables, we called it log normal distribution for simplicity.

the moments leading to (3.3) are in fact exactly satisfied if the $\widetilde{v}$ is symmetrically distributed so the performance of the small sigma estimator when ν is normally distributed is not surprising, but we note that the small sigma estimator performs well even when symmetry is violated.

We note that the approximate identification strategy (3.3) can be viewed as the solution to the approximate moment equation

$$E\left[\widetilde{x}^2\left(\widetilde{y}-\widetilde{x}\beta\right)\right]\approx 0.$$

Throughout the rest of the paper, we develop estimators for various models based on approximate moments. We argue that approximate moments often leads to feasible estimators in models plagued with either theoretical or computational issues. Our focus is on development of point estimators. As for inference, we defer to an emerging literature recognizes that moment conditions are never exactly satisfied and proposes various methods of inference.

The small sigma approach to EIV goes back to Amemiya (1985, 1990). Our contribution is to recognize that the small sigma approach results in a very simple estimator for models characterized by CH. In order to understand how the small sigma approach can be applied in nonlinear models, we examine the model and the IV estimator considered by Amemiya (1990). We consider the model where

$$y = f\left(x_*;\beta\right)+\varepsilon \tag{3.4}$$

$$x = x_* + v$$

and an IV/moment of the form

$$0 = E\left[w\left(y - f\left(x_*;\beta\right)\right)\right].$$

If (w, y, x_*) were observed, we could have estimated β by solving

$$0 = n^{-1}\sum_{i=1}^{n} w_i\left(y_i - f\left(x_{*i};\hat{\beta}\right)\right). \tag{3.5}$$

(The equality above should be understood appropriately under overidentification, i.e., when the dimension of w exceeds that of β.) The feasible sample counterpart is not going to produce a reasonable estimator because $0 \neq E\left[w\left(y - f\left(x;\beta\right)\right)\right]$.

On the other hand, we have

$$\begin{aligned}E\left[w\left(y-f\left(x;\beta\right)\right)\right]-E\left[w\left(y-f\left(x_{*};\beta\right)\right)\right]&=-E\left[w\left(f\left(x;\beta\right)-f\left(x_{*};\beta\right)\right)\right]\\&=-E\left[w\left(f\left(x_{*}+v;\beta\right)-f\left(x_{*};\beta\right)\right)\right],\end{aligned}$$

so the small sigma approximation $x=x_{*}+\sigma v$ with $\sigma\to0$ suggests that we work with

$$\begin{aligned}&E\left[w\left(y-f\left(x;\beta\right)\right)\right]-E\left[w\left(y-f\left(x_{*};\beta\right)\right)\right]\\&=-E\left[w\left(f\left(x_{*}+\sigma v;\beta\right)-f\left(x_{*};\beta\right)\right)\right]\\&=-\sigma E\left[wf_{x}\left(x_{*};\beta\right)v\right]-\frac{\sigma^{2}}{2}E\left[wf_{xx}\left(x_{*};\beta\right)v^{2}\right]+o\left(\sigma^{2}\right),\end{aligned}\tag{3.6}$$

where f_x and f_{xx} denote the first and second derivatives with respect to x. If we assume that the measurement error v is independent of everything else and has a zero mean, we have

$$\begin{aligned}E\left[w\left(y-f\left(x;\beta\right)\right)\right]-E\left[w\left(y-f\left(x_{*};\beta\right)\right)\right]&=-\frac{\sigma^{2}}{2}E\left[wf_{xx}\left(x_{*};\beta\right)\right]E\left[v^{2}\right]+o\left(\sigma^{2}\right)\\&=-\frac{1}{2}E\left[wf_{xx}\left(x_{*};\beta\right)\right]E\left[\left(\sigma v\right)^{2}\right]+o\left(\sigma^{2}\right)\\&=-\frac{1}{2}E\left[wf_{xx}\left(x;\beta\right)\right]E\left[\left(\sigma v\right)^{2}\right]+o\left(\sigma^{2}\right).\end{aligned}\tag{3.7}$$

In other words, we can work with an approximate moment

$$0\approx E\left[w\left(y-f\left(x;\beta\right)\right)\right]+\frac{\sigma_{v}^{2}}{2}E\left[w\left(f_{xx}\left(x;\beta\right)\right)\right]=E\left[w\left(y-f\left(x;\beta\right)+\frac{\sigma_{v}^{2}}{2}f_{xx}\left(x;\beta\right)\right)\right]\tag{3.8}$$

where $\sigma_v^2=E\left[\left(\sigma v\right)^2\right]$. This is a way of fixing the moment itself. We note that Evdokimov and Zeleneev's (2020)[9] estimator boils down to (3.8) for Amemiya's (1990) model (3.4).

Our proposal is to use the approximate moment, and estimate β and σ_v^2 simultaneously, assuming that $\dim\left(w\right)$ is large enough to identify them. Amemiya's (1990) original proposal is a little different. He first proposes to obtain a preliminary estimator of β by solving (3.5). He then proposes to estimate σ_v^2 by noting that

$$\begin{aligned}E\left[\left.\left(y-f\left(x;\beta\right)\right)^{2}\right|x_{*}\right]&=E\left[\left.\left(\varepsilon-\left(f\left(x_{*}+\sigma v;\beta\right)-f\left(x_{*};\beta\right)\right)\right)^{2}\right|x_{*}\right]\\&\approx E\left[\left.\left(\varepsilon-f_{x}\left(x_{*};\beta\right)\sigma v\right)^{2}\right|x_{*}\right]\\&\approx\sigma_{\varepsilon}^{2}+f_{xx}\left(x_{*};\beta\right)^{2}\sigma_{v}^{2}\approx\sigma_{\varepsilon}^{2}+f_{xx}\left(x;\beta\right)^{2}\sigma_{v}^{2}\end{aligned}\tag{3.9}$$

---

[9] We thank Denis Chetverikov and Zhipeng Liao for bringing our attention to Evdokimov and Zeleneev's (2020) contribution.

and proposes to estimate σ_v^2 by $\hat{\sigma}_v^2$, which is obtained by regressing $\left(y - f\left(x; \hat{\beta}\right)\right)^2$ on a constant and $f_{xx}\left(x; \hat{\beta}\right)^2$. Finally, he proposes to plug $\hat{\sigma}_v^2$ into the sample counterpart of (3.8) to generate a bias corrected version $\tilde{\beta}$:

$$0 \approx n^{-1} \sum_{i=1}^{n} w_i \left(y_i - f\left(x_i; \tilde{\beta}\right) + \frac{\hat{\sigma}_v^2}{2} f_{xx}\left(x_i; \hat{\beta}\right) \right).$$

Amemiya (1990) worked with nonstochastic x and w, so we had to translate his framework into our IID framework, rendering the above analysis superficially different from his framework. Despite the superficial differences, it is immediate that the following two comparisons can be made. First, the small sigma approach applied to the moments estimates β and σ_v^2 simultaneously, whereas Amemiya (1990) proposes to estimate them sequentially. The simultaneous nature of the moment-based procedure implies that the dimension of w needs to be larger than that of β for implementation of (3.8), whereas Amemiya (1990) can work with w with the same dimension as β. Second, the second approximate equality in (3.9) makes it clear that his procedure is predicated on the independence of ε and x_*. In other words, his framework rules out conditional heteroscedasticity (CH) where $E\left[\varepsilon^2 \middle| x_*\right]$ may depend on x_*, while our small sigma approach allows for CH.

We conducted Monte Carlo simulations to evaluate the performance of the small sigma estimator (applied to moments) in comparison with Amemiya's estimator. We considered a nonlinear model

$$y = \beta x_*^2 + \varepsilon$$

where $\beta = 1$ and x_* is measured with error $x = x_* + v$. As for the instrument, we assumed that $w = \left(1, (x_* + r_*)^2\right)'$, where $r_* \sim N(0,1)$. In DGP below, x_* can have normal or log-normal distribution, ϵ is conditionally normal with respect to x_*, and v is a independent normal distribution random variable with different variances.[10] We assumed that (i) $x_* \sim N(0,1)$,

[10] In order to calculate Amemiya's estimator, we first estimate $\hat{\beta}$ by

$$\hat{\beta} = \left[\left(\sum_{i=1}^{n} x_i^2 w_i' \right) \left(\sum_{i=1}^{n} w_i w_i' \right)^{-1} \left(\sum_{i=1}^{n} w_i x_i^2 \right) \right]^{-1} \left(\sum_{i=1}^{n} x_i^2 w_i' \right) \left(\sum_{i=1}^{n} w_i w_i' \right)^{-1} \left(\sum_{i=1}^{n} w_i y_i \right),$$

$\mathcal{L}(\varepsilon | x_*) \sim N(0, x_*^2)$, and $v \sim N(0, \sigma^2)$; (ii) $x_* \sim N(0,1)$, $\mathcal{L}(\varepsilon | x_*) \sim \exp(N(\ddot{\mu}, x_*^2))$ and $v \sim \exp^*(N(\bar{\mu}, \bar{\sigma}^2))$, where $\ddot{\mu}$, $\bar{\mu}$ and $\bar{\sigma}$ are chosen such that $\exp(N(\ddot{\mu}, x_*^2))$ has mean 2, $\exp(N(\bar{\mu}, \bar{\sigma}^2))$ has mean 2 and $\mathrm{Var}(v) = \sigma^2$; (iii) $x_* \sim \exp^*(N(0,1))$, $\mathcal{L}(\varepsilon | x_*) \sim N(0, x_*^2/5)$, and $v \sim N(0, \sigma^2)$; and (iv) $x_* \sim \exp^*(N(0,1))$, $\mathcal{L}(\varepsilon | x_*) \sim N(0, x_*^2/5)$, $v \sim \exp^*(N(\bar{\mu}, \bar{\sigma}^2))$, where $\bar{\mu}$ and $\bar{\sigma}$ are chosen such that $\exp(N(\bar{\mu}, \bar{\sigma}^2))$ has mean 2 and $\mathrm{Var}(v) = \sigma^2$. All specifications were chosen such that $E[\varepsilon | x_*] = 0$, $\mathrm{Var}(\varepsilon) = 1$, $\mathrm{Var}(v) = \sigma^2$ with the sample size equal to 1,000. The number of simulation is 10,000. We considered $\sigma^2 = 1$, 0.5, 0.25 and 0.1. The Monte Carlo results are summarized in Tables C.4–C.7. As is predicted by the theory, the small sigma estimator performs better when $\mathrm{Var}(v)$ is small. The small sigma estimator often outperforms Amemiya's estimator.

3.3 Approximate Inference in Panel Data Models with Fixed Effects

We note that the small sigma approach (3.8) can be applied to as long as the moment

$$0 = E[w \cdot (y - f(x + v; \beta))]$$

is satisfied. It turns out that panel models can be understood to be a special case as long as certain assumptions are imposed, and we will discuss how to apply the small sigma approach there.

---

and estimate $\hat{\sigma}_{uu}$ by using OLS.

$$(y_i - \hat{\beta} x_i^2) = \hat{\sigma}_{\varepsilon\varepsilon} + 4\hat{\beta}^2 x_i^2 \hat{\sigma}_{uu},$$

and finally calculate his estimator by

$$\tilde{\beta} = \left[\left(\sum_{i=1}^n x_i^2 w_i' \right) \left(\sum_{i=1}^n w_i w_i' \right)^{-1} \left(\sum_{i=1}^n w_i x_i^2 \right) \right]^{-1} \left(\sum_{i=1}^n x_i^2 w_i' \right) \left(\sum_{i=1}^n w_i w_i' \right)^{-1} \left(\sum_{i=1}^n w_i \left(y_i + \hat{\beta} \hat{\sigma}_{uu} \right) \right).$$

To compare it with the small sigma estimator, we obtain the approximate moment

$$0 \approx E\left[w \left(y - \beta x^2 + \sigma_v^2 \beta \right) \right]$$

and we can estimate $\left(\beta, \sigma_v^2 \beta \right)$ from this moment equation as long as $\dim(w) = 2$. Then define the estimator of the coefficient of x^2 as the small sigma estimator.

We will consider a nonlinear panel model with fixed effects

$$y_{it} = f\left(x_{it}\beta + \alpha_i\right) + \eta_{it} \qquad t = 1, \ldots, T$$

where

$$E\left[\eta_{it} \middle| x_{i1}, \ldots, x_{iT}, \alpha_i\right] = 0.$$

The panel is short, i.e., the T is fixed. For simplicity, we will assume that $T = 2$ and that x_{it} is a scalar. The functional form of f is assumed to be known. We assume away any measurement error problem. If the distribution of the unobserved fixed effect α_i conditional on (x_{i1}, x_{i2}) is completely arbitrary, the parameter β is most often not identified.[11]

In order to make a progress, we assume the existence of a sufficient statistic, an approach pioneered by Mundlak (1978), and later adopted by Altonji and Matzkin (2005) and Arkhangelsky and Imbens (2019). To be more precise, we assume that the conditional distribution of α_i given (x_{i1}, x_{i2}) only depends on some function of (x_{i1}, x_{i2}), say $z_i = g\left(x_{i1}, x_{i2}\right)$. We then have $E\left[\alpha_i \middle| x_{i1}, x_{i2}\right] = \varphi\left(z_i\right)$, and $\operatorname{Var}\left[\alpha_i \middle| x_{i1}, x_{i2}\right] = \sigma^2\left(z_i\right)$ for some $\varphi\left(\cdot\right)$ and $\sigma^2\left(\cdot\right)$. Writing $u_i = \alpha_i - \varphi\left(z_i\right)$, we see that $E\left[u_i \middle| x_{i1}, x_{i2}\right] = 0$ and $E\left[u_i^2 \middle| x_{i1}, x_{i2}\right] = \sigma^2\left(z_i\right)$. We propose to consider the small sigma approach assuming that $\sigma^2\left(z_i\right)$ is small.

A small sigma approach would give rise to an approximate specification

$$\begin{aligned}
E\left[y_{it} \middle| x_{i1}, x_{i2}\right] &= E\left[f\left(x_{it}\beta + \varphi\left(z_i\right) + u_i\right) \middle| x_{i1}, x_{i2}\right] \\
&\approx f\left(x_{it}\beta + \varphi\left(z_i\right)\right) \\
&\quad + f^{(1)}\left(x_{it}\beta + \varphi\left(z_i\right)\right) E\left[u_i \middle| x_{i1}, x_{i2}\right] \\
&\quad + \frac{1}{2} f^{(2)}\left(x_{it}\beta + \varphi\left(z_i\right)\right) E\left[u_i^2 \middle| x_{i1}, x_{i2}\right] \\
&= f\left(x_{it}\beta + \varphi\left(z_i\right)\right) + \frac{1}{2} f^{(2)}\left(x_{it}\beta + \varphi\left(z_i\right)\right) \sigma^2\left(z_i\right),
\end{aligned} \tag{3.10}$$

which is still semiparametric and not convenient for immediate implementation.[12]

[11] See Chamberlain (2010) or Hahn (2001).

[12] The above approximation may be based on a model $\alpha_i = g\left(z_i, u_i\right)$, for some function g and that u_i is independent of z_i. Such a g can be motivated by the inverse quantile approach as discussed in Matzkin

In order to make it accessible for implementation, we propose to use a parametric specification, which can be interpreted to be a sieve approach as well. Specifically, we write

$$\alpha_i = z_i \gamma + u_i \tag{3.11}$$

with u_i independent of z_i. The small sigma approach would then give rise to an approximation

$$E\left[y_{it} \middle| x_{i1}, x_{i2} \right] \approx f\left(x_{it}\beta + z_i\gamma \right) + \frac{1}{2} f^{(2)}\left(x_{it}\beta + z_i\gamma \right) \sigma_u^2, \tag{3.12}$$

which is immediately implementable for nonlinear least squares. More precisely, we can estimate $(\beta, \gamma, \sigma_u^2)$ by nonlinear least squares of y_{it} on (x_{it}, z_i). The approximate specification (3.12) is inspired by a sieve approach to nonparametric specification, but a practitioner may feel uncomfortable with the implicit independence assumption between u_i independent of z_i. If so, a reasonable specification is to base the approximation to (3.10) and work with a more flexible specification along the line of

$$E\left[y_{it} \middle| x_{i1}, x_{i2} \right] \approx f\left(x_{it}\beta + z_i\gamma \right) + \frac{1}{2} f^{(2)}\left(x_{it}\beta + z_i\gamma \right) \sigma_u^2\left(z_i; \theta \right), \tag{3.13}$$

where $\sigma_u^2\left(z_i; \theta \right)$ is a parametric specification of the conditional variance $E\left[u_i^2 \middle| z_i \right]$.[13] As a practical matter, the heteroscdasticity assumption may be tested by any standard specification test.

If the linear specification (3.11) is not quite correct, and can only be interpreted to be the result of population regression of α_i on (x_{i1}, x_{i2}), we can only guarantee that $E\left[x_{it} u_i \right] = 0$,

(2007). The small sigma approach would approximate

$$\alpha_i = g\left(z_i, 0 \right) + g_z\left(z_i, 0 \right) u_i + g_{zz}\left(z_i, 0 \right) \frac{u_i^2}{2}$$

so the $\varphi\left(z_i \right)$ and $\sigma^2\left(z_i \right)$ above may be understood to be specifications of $g\left(z_i, 0 \right)$ and $g_{zz}\left(z_i, 0 \right)$, respectively.

[13] The $z_i\theta$ term may be replaced by a sufficiently rich polynomial function in z_i, thereby given a nearly nonparametric specification. Similar comment applies to $\sigma_u^2\left(z_i; \gamma \right)$.

and as such we can only write

$$
\begin{aligned}
E\left[y_{it}\middle|\, x_{i1}, x_{i2}\right] \approx\; & f\left(x_{it}\beta + z_i\gamma\right) \\
& + E\left[f^{(1)}\left(x_{it}\beta + z_i\gamma\right) u_i\middle|\, x_{i1}, x_{i2}\right] \\
& + E\left[\frac{1}{2} f^{(2)}\left(x_{it}\beta + z_i\gamma\right) u_i^2\middle|\, x_{i1}, x_{i2}\right],
\end{aligned}
$$

and therefore, a result along the line of (3.12) is not quite valid. On the other hand, if one has reasons to believe that $E\left[u_i\middle|\, x_{i1}, x_{i2}\right] \approx 0$ and $E\left[u_i^2\middle|\, x_{i1}, x_{i2}\right] \approx$ constant, which may not be such a bad approximation depending on the given application, the approximation (3.12) may be a good approach for applications. From a mathematical point of view, the idea $E\left[u_i\middle|\, x_{i1}, x_{i2}\right] \approx 0$ and $E\left[u_i^2\middle|\, x_{i1}, x_{i2}\right] \approx$ constant requires formalization of "approximate independence", which we have not been able to articulate.

3.3.1 Monte Carlo Simulations

We examined the performance of the small sigma estimator for several nonlinear panel models. First, we considered a panel cubic model with $T = 2$, where

$$
y_{it} = \left(x_{it} + \alpha_i\right)^3 + \eta_{it} \qquad t = 1, 2
$$

For simplicity, we will assume that x_{it} is a scalar. We assume that $\alpha_i = z_i\gamma + u_i$, where $z_i = \frac{x_{i1}+x_{i2}}{2}$. We assumed that $x_{i1}, x_{i2}, \eta_{i1}, \eta_{i2}$ are independent normal with zero mean. By Taylor expansion approximation, we have

$$
E\left[y_{it}\middle|\, x_{i1}, x_{i2}\right] \approx \left(x_{it}\beta + z_i\gamma\right)^3 + 3\sigma^2\left(x_{it}\beta + z_i\gamma\right),
$$

which we used to estimate parameters by nonlinear least squares. We set the true parameters, $\beta = 1, \gamma = 1$ and considered $\mathrm{Var}(u) = 1$, 0.5, 0.25 and 0.1 and the sample size equal to 1,000. The number of simulation is 10,000. Our results are summarized in Table C.8.

We also considered the panel quartic model with $T = 2$, where

$$
y_{it} = \left(x_{it} + \alpha_i\right)^4 + \eta_{it} \quad t = 1, 2
$$

with the same parameter combinations as in the panel cubic model. By Taylor expansion approximation

$$E\left[\left.y_{it}\right|x_{i1},x_{i2}\right]\approx\left(x_{it}\beta+z_i\gamma\right)^4+6\sigma^2\left(x_{it}\beta+z_i\gamma\right)^2,$$

which we used to estimate parameters by nonlinear least squares. Our results are summarized in Table C.9.

We also considered a panel probit model where

$$y_{it}=1\left(x_{it}\beta+\alpha_i-\varepsilon_{it}\geq 0\right)\qquad t=1,2$$

and $\varepsilon_{it}\sim N\left(0,1\right)$. We then have

$$E\left[\left.y_{it}\right|x_{i1},x_{i2},\alpha_i\right]=\Phi\left(x_{it}\beta+\alpha_i\right)$$

so the Taylor series approximation of interest is

$$E\left[\left.y_{it}\right|x_{i1},x_{i2}\right]\approx\Phi\left(x_{it}\beta+z_i\gamma\right)+\frac{\sigma^2}{2}\Phi^{(2)}\left(x_{it}\beta+z_i\gamma\right)$$

where

$$\Phi^{(2)}\left(t\right)=\frac{d}{dt}\Phi^{(1)}\left(t\right)=\frac{d}{dt}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)=-\frac{t}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right).$$

Therefore, the approximation adopted in the nonlinear least squares was

$$E\left[\left.y_{it}\right|x_{i1},x_{i2}\right]\approx\Phi\left(x_{it}\beta+z_i\gamma\right)-\frac{\sigma^2}{2\sqrt{2\pi}}\left(x_{it}\beta+z_i\gamma\right)\exp\left(-\frac{\left(x_{it}\beta+z_i\gamma\right)^2}{2}\right).$$

We set $\beta=1$. $\gamma=0$. We take the distribution of u_i, u ~uniform$\left(-\sqrt{3\sigma_u^2},\sqrt{3\sigma_u^2}\right)$ and considered $\sigma_u^2=1,0.5,0.25,0.1$. The number of simulation is 10,000. Our results are summarized in Table C.10.

Finally, we considered the logit counterpart, where we considered the performance of the conditional MLE (CMLE). Our results are summarized in Table C.11. In Tables C.8 - C.10, the small sigma estimator does not really have a competitor, but in the logit model, the CMLE is the standard estimator because it is known to be consistent and asymptotically normal. Our simulation results indicate that the small sigma estimator performs reasonably well, especially when Var (u) is small, as is predicted by the theory. The number of simulation is 10,000 as well.

3.3.2 More Flexible Specification

Our recommendation in the previous section to use the sufficient statistic approach is inspired by nonparametric consideration. Suppose that the x_{i1} and x_{i2} have multinomial distributions with identical support, consisting of M elements. Also suppose that the support of the joint distribution of (x_{i1}, x_{i2}) is "full" and contain M^2 elements. Without the sufficient statistic specification, we would write $E\left[\alpha_i \middle| x_{i1}, x_{i2}\right] = \varphi\left(x_{i1}, x_{i2}\right)$, and $\mathrm{Var}\left[\alpha_i \middle| x_{i1}, x_{i2}\right] = \sigma^2\left(x_{i1}, x_{i2}\right)$, and (3.10) would instead be written

$$E\left[y_{it} \middle| x_{i1}, x_{i2}\right] \approx f\left(x_{it}\beta + \varphi\left(x_{i1}, x_{i2}\right)\right) + \frac{1}{2} f^{(2)}\left(x_{it}\beta + \varphi\left(x_{i1}, x_{i2}\right)\right)\sigma^2\left(x_{i1}, x_{i2}\right). \tag{3.14}$$

Assuming as before that x_{it} is a scalar, the number of unknown parameters[14] is $2M^2 + 1$, including β and the values of $\varphi(\cdot, \cdot)$ and $\sigma^2(\cdot, \cdot)$ at M^2 support points of (x_{i1}, x_{i2}). The left hand side of (3.14) provides at most $2M^2$ conditional moments. In other words, the number of unknown parameters exceed the number of moments, and hence, there may be a problem of identification. On the other hand, if $E\left[\alpha_i \middle| x_{i1}, x_{i2}\right]$ and $\mathrm{Var}\left[\alpha_i \middle| x_{i1}, x_{i2}\right]$ depend only on z_i, and if the support of z_i consists of sufficiently small number of elements, we may be able to identify the parameters, which is a mathematical motivation for our proposal.

On the other hand, it may be appealing for some practitioners to adopt a flexible parametric perspective, while still relaxing the sufficient statistic assumption. For example, we may follow Chamberlain (1984), and assume

$$\alpha_i = x_{i1}\pi_1 + x_{i2}\pi_2 + u_i. \tag{3.15}$$

If we apply the small sigma approach here, we end up with

$$E\left[y_{i1} \middle| x_{i1}, x_{i2}\right] \approx f\left(x_{i1}\left(\beta + \pi_1\right) + x_{i2}\pi_2\right) + \frac{1}{2} f^{(2)}\left(x_{i1}\left(\beta + \pi_1\right) + x_{i2}\pi_2\right)\sigma_u^2,$$
$$E\left[y_{i2} \middle| x_{i1}, x_{i2}\right] \approx f\left(x_{i1}\pi_1 + x_{i2}\left(\beta + \pi_2\right)\right) + \frac{1}{2} f^{(2)}\left(x_{i1}\pi_1 + x_{i2}\left(\beta + \pi_2\right)\right)\sigma_u^2,$$

[14]Our proposed estimator is based on approximate moments, so these should really be called pseud-parameters, if we are to be semantically precise.

and β may be estimated by first estimating $(\beta + \pi_1, \pi_2, \sigma_u^2)$ from nonlinear least squares of y_{i1} on (x_{i1}, x_{i2}), $(\pi_1, \beta + \pi_2, \sigma_u^2)$ from nonlinear least squares of y_{i2} on (x_{i1}, x_{i2}), and finally by using an inter-equation restriction on the parameters.

The specification underlying (3.15) is sometimes called the Π matrix approach, and has been adopted only in linear models and panel probit models. We note that the small sigma approach can be applied to even further to arbitrary nonlinear models.

3.3.3 Conditional Heteroscedasticity of Measurement Error

Our discussion in the panel model includes a proposal to consider a parametric specification $\sigma_u^2(z_i;\theta)$ of the conditional variance $E[u_i^2 | z_i]$, which leads to the approximate moment (3.13). This approach may be reasonable for some empirical applications. We can consider a similar strategy in the nonlinear EIV models, and consider a parametric specification of the conditional variance of the measurement error. Suppose that. From (3.6), we see that

$$E[w(y - f(x;\beta))] - E[w(y - f(x_*;\beta))] \approx -\sigma E[w f_x(x_*;\beta) v] - \frac{\sigma^2}{2} E\left[w f_{xx}(x_*;\beta) v^2\right]$$

under the small sigma approximation $x = x_* + \sigma v$ with $\sigma \to 0$. Therefore, if we further assume that $E[\sigma v | w, x_*] = 0$ and $E\left[(\sigma v)^2 \middle| w, x_*\right] = \sigma_v^2(x_*;\theta)$, we would obtain

$$E[w(y - f(x;\beta))] - E[w(y - f(x_*;\beta))] \approx -\frac{1}{2} E\left[w f_{xx}(x_*;\beta) \frac{\sigma_v^2(x_*;\theta)}{2} f_{xx}(x;\beta)\right]$$

Under the further assumption that $\sigma_v^2(x_*;\theta)$ is continuous in x_*, we can replace the $\sigma_v^2(x_*;\theta)$ on the RHS by $\sigma_v^2(x;\theta)$, and obtain a feasible alternative to the approximate moment (3.8)

$$0 \approx E\left[w\left(y - f(x;\beta) + \frac{\sigma_v^2(x;\theta)}{2} f_{xx}(x;\beta)\right)\right].$$

3.4 Relationship with Salanié and Wolak (2019)

We noted in the beginning of the previous section that the small sigma approach applies to any model of the form

$$0 = E[w \cdot (y - f(x + v;\beta))].$$

It is straightforward to recognize that the approach can be applied to a slightly more complicated moment of the form

$$0 = E\left[w \cdot h\left(y, E_v\left[f\left(x, v; \beta\right)\right]\right)\right],$$

where $E_v[\cdot]$ denotes the expectation taken with respect to the marginal dsitribution of v, and h is a sufficiently smooth function. Berry, Levinsohn, and Pakes' (1995) moment is a special case. It is based on $g(x, v; \beta, \xi)$, which denotes the individual "shares" of demand for a given level of product characteristic ξ, and $E_v[g(x, v; \beta, \xi)] \equiv G(\xi; x, \beta)$, which denotes its aggregate counterpart. The v is a random variable that gives rise to the *mixed* logit model, and $E_v[\cdot]$ denotes the expectation taken with respect to the distribution of v, keeping every other variable fixed. Finally, the h denotes the "contraction mapping", which produces the value of ξ that equates the observed market share y with $E_v[g(x_*, v; \beta, \xi)]$, i.e., $G(h(y; x_*, \beta); x_*, \beta) = y$.

Under the small sigma approach, we work with σv and corresponding $h(y; x_*, \beta, \sigma)$. The goal is to let $\sigma \to 0$, and find an approximation

$$h(y; x, \beta, \sigma) \approx h(y; x, \beta, 0) + \sigma h_\sigma(y; x, \beta, 0) + \frac{\sigma^2}{2} h_{\sigma\sigma}(y; x, \beta, 0),$$

where h_σ and $h_{\sigma\sigma}$ denote the first and second derivatives with respect to σ. Salanié and Wolak (2019) contributes to the literature by calculating these derivatives h_σ and $h_{\sigma\sigma}$ explicitly.

We note that the small sigma approach cannot be dismissed as a mere approximation in applications of BLP. It is because the BLP specification adopted in practice can be argued to be subject to misspecification induced by truncation. The number of products tend to be fairly large in applications, and the BLP are often be applied to the truncated data focusing on a relatively small number of products that have reasonably large market shares. Assuming that the BLP specification is correct for the full set of products, it is straightforward to recognize that the truncated data set would not satisfy the premise of the BLP, unless the the Independence of Irrelevant Alternatives (IIA)[15] properties are satisfied. Because the

[15]See Hausman and McFadden (1984) or Hahn, Hausman and J. Lustig (2020) for further discussion.

IIA is unlikely to be satisfied in many applications, it would be reasonable to address the misspecification in BLP.

Given how the truncation induces misspecification, there is no reason to predict that the standard BLP estimator would outperform the small sigma estimator. It is therefore of interest to compare the performances of the small sigma estimator against the common estimators of BLP. We conducted such comparison through Monte Carlo simulations[16], which is summarized in Tables C.12 –C.17. We find that (i) MPEC outperforms small sigma estimator under correct specification, as expected; but (ii) their performances are comparable under the "misspecification" when only a few top products are used for BLP implementation.

3.5 Summary

We adopted a pragmatic perspective, and proposed using approximate moments to estimate various models. The approximation was based on the small sigma approach. Simulation results suggest that the approximation leads to reasonable sampling properties. Our proposal complements the newly resurgent literature on sensitivity analysis.

[16]Given the complexity of the BLP model, the description of the Monte Carlo is presented in the appendix. In order to avoid any weak IV problem interacting with the approximation issue, the price is an exogenous variable in our Monte Carlo.

REFERENCES

[1] Aigner, D.J., Hsiao, C., Kapteyn, A., and Wansbeek, T. (1984), "Latent Variable Models in Econometrics", Volume 2, eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, 1321–1393.

[2] Altonji, J.G., and Matzkin. R.L. (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors", Econometrica 73, 1053–1102.

[3] Amemiya, Y. (1985), "Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model", Journal of Econometrics 28, 273–289.

[4] Amemiya, Y. (1990), "Two-Stage Instrumental Variable Estimators for the Nonlinear Errors-in-Variables Model", Journal of Econometrics 44, 311-332.

[5] Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017), "Measuring the sensitivity of parameter estimates to sample statistics," Quarterly Journal of Economics 132, 1553–1592.

[6] Arkhangelsky, D., and Imbens, G.W. (2019), "The Role of the Propensity Score in Fixed Effect Models", Unpublished Manuscript.

[7] Arellano, M., and Hahn, J. (2007), "Understanding Bias in Nonlinear Panel Models: Some Recent Developments", in Advances in Economics and Econometrics, eds. Blundell, R. W.K. Newey, and T. Persson, Cambridge: Cambridge University Press.

[8] Armstrong, T. and Kolesár, M. (2019), "Sensitivity Analysis using Approximate Moment Condition Models", Unpublished Manuscript.

[9] Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium", Econometrica 60, 889-917.

[10] Chamberlain, G. (1984), "Panel Data," in Handbook of Econometrics, Volume 2, eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, 1247–1318.

[11] Chamberlain, G. (2010), "Binary Response Models for Panel Data: Identification and Information", Econometrica 78, 159–168.

[12] Dubé, J-P, Fox, J., and Su, C-L (2012), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," Econometrica 80, pp. 2231–2267.

[13] Evdokimov, K., and A. Zeleneev (2020), "Simple Estimation of Semiparametric Models with Measurement Errors," seminar slides.

[14] Fuller, W. A. (1987), Measurement Error Models, New York: Wiley.

[15] Geary, R. C. (1942), "Inherent Relations Between Random Variables," Proceedings of the Royal Irish Academy, 47A, 63–76.

[16] Hahn, J. (2001), "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects", Econometric Theory 17, 913-932.

[17] Hahn, J. and Hausman, J. A. (2005), "IV estimation with Valid and Invalid Instruments", Annales d'´Economie et de Statistique 79/80, 25–57.

[18] Hahn, J., Hausman, J.A., and J. Lustig (2020): "Specification Test on Mixed Logit Models", forthcoming in Journal of Econometrics.

[19] Hausman, J., and D. McFadden (1984): "Specification Tests for the Multinomial Logit Model," Econometrica 52, 1219–1240.

[20] Hausman, J.A., Newey, W.K., Ichimura, H., and Powell, J.L. (1991), "Identification and Estimation of Polynomial Errors-in-Variables Models", Journal of Econometrics 50, 273–295.

[21] Kadane, J. (1971), "Comparison of k-class Estimators when the Variance is Small", Econometrica 39, 723–737.

[22] Lewbel, A. (1997), "Constructing Instruments for Regressions With Measurement Error When no Additional Data are Available, with An Application to Patents and R&D", Econometrica 65, 1201–1213.

[23] Matzkin, R.L. (2007) "Nonparametric Identification," in Handbook of Econometrics, Vol. 6b, eds. J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

[24] Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data", Econometrica 46, 69–85.

[25] Reiersol, O. (1950): "Identifiability of a Linear Relation between Variables Which are Subject to Error". Econometrica 18, 375–389

[26] Salanié, B., and Wolak, F.A. (2019), "Fast, "Robust", and Approximately Correct: Estimating Mixed Demand Systems", NBER Working Paper 25726.

[27] Schennach, S.M. (2016), "Recent Advances in the Measurement Error Literature", Annual Review of Economics 8, 341–377.

[28] Schennach, S.M. (2020), "Mismeasured and Unobserved Variables", unpublished working paper.

[29] Schennach, S.M., and Hu, Y. (2013), "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models without Side Information", Journal of the American Statistical Association 108, 177–186.

APPENDIX A

Proofs of Theorems and Lemmas in Chapter 1 and 2

A.1 Proof of Theorems

Proof of Theorem 1.2.1

For the sake of the simplicity, let me prove the theorem under assuming X and Z be 1-dimensional random variables. To show the consistency of $\tilde{\theta}$, by applying lemma A.2.5, we have that $\|\hat{\phi}-\phi\|_2 \lesssim O_p(n^{-1/3})$. From the step 2 in section 1.2.1, we estimate $\tilde{\theta}$.

$$\tilde{\theta} = \frac{\sum_{i=1}^{n}(Z_i - \hat{\phi}(W_i))Y_i}{\sum_{i=1}^{n}(Z_i - \hat{\phi}(W_i))X_i} = \theta_0 + \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \hat{\phi}(W_i))(g(W_i)+V_i)}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \hat{\phi}(W_i))X_i}$$

Then the denominator of the bias converges to $E[VX]$ under assuming $\|\phi_0 - \hat{\phi}\|_{n,2}\|X\|_{n,2}$ is bounded. The numerator can be separated into 2 terms and the first one converges to $E[Vg(W)] = 0$ with the error term $\|\phi - \hat{\phi}\|_{n,2}\|g\|_{n,2} = O_p(n^{-1/3})$ under assuming g is bounded function and for the second one, it converges to $E[VU] = 0$ with the error term $\|\phi - \hat{\phi}\|_{n,2}\|U\|_{n,2} = O_p(n^{-1/3})$ under assuming U has finite 4th moment.

$$\frac{1}{n}\sum_{i=1}^{n} V_i g(W_i) + \frac{1}{n}\sum_{i=1}^{n}(\phi(W_i) - \hat{\phi}(W_i))g(W_i) \le E[Vg(W)] + O_p(n^{-1/3})$$

$$\frac{1}{n}\sum_{i=1}^{n} V_i U_i + \frac{1}{n}\sum_{i=1}^{n}(\phi(W_i) - \hat{\phi}(W_i))U_i = E[VU] + O_p(n^{-1/3})$$

Hence, $\tilde{\theta} = \theta_0 + O_p(n^{-1/3})$. From the step 3, estimate $\tilde{g}$ by applying isotonic regression of $Y - X\tilde{\theta}$ on W. Then the isotonic estimator $\tilde{g}(W_{(j)})$ is defined as

$$\tilde{g}(W_{(j)}) = \max_{k\le j}\min_{l\ge j}\frac{1}{l-k+1}\sum_{i=k}^{l}(Y_{(i)} - X_{(i)}\tilde{\theta})$$

Define $\tilde{g}_0(W_{(j)})$ be the isotonic regression estimator of $Y - X\theta_0$ on W.

$$\tilde{g}_0(W_{(j)}) = \max_{k\leq j}\min_{l\geq j}\frac{1}{l-k+1}\sum_{i=k}^{l}(Y_{(i)} - X_{(i)}\theta_0)$$

I borrow the idea from Huang(2002) page 348 to show that $|\tilde{g}(w) - g(w)| = O_p(n^{-1/3})$ for all w in the support of W. Note that for the fixed w in the support of W,

$$\begin{aligned}
&\tilde{g}(w) - \tilde{g}_0(w) \\
&= \max_{a\leq w}\min_{b\geq w}\frac{1}{N(W_i\in[a,b])}\sum_{i=1}^{n}1_{\{W_i\in[a,b]\}}(Y_i - X_i\tilde{\theta}) \\
&- \max_{a\leq w}\min_{b\geq w}\frac{1}{N(W_i\in[a,b])}\sum_{i=1}^{n}1_{\{W_i\in[a,b]\}}(Y_i - X_i\theta_0) \\
&\leq \sup_{a,b}\Bigg|\frac{1}{N(W_i\in[a,b])}\sum_{i=1}^{n}1_{\{W_i\in[a,b]\}}(Y_i - X_i\tilde{\theta}) \\
&- \frac{1}{N(W_i\in[a,b])}\sum_{i=1}^{n}1_{\{W_i\in[a,b]\}}(Y_i - X_i\theta_0)\Bigg| \\
&\leq |W||\tilde{\theta} - \theta_0| \lesssim O_p(n^{-1/3})
\end{aligned}$$

where $N(W_i\in[a,b])$ is the number of $W_i's$ included in $[a,b]$. Under assuming $|W|$ is bounded. Then we get for all w in the support of W,

$$\tilde{g}(w) - g(w) \leq |\tilde{g}(w) - \tilde{g}_0(w)| + |\tilde{g}_0(w) - g(w)|$$

The second term on the right hand side satisfies $O_p(n^{-1/3})$ by lemme A.2.5.

Then $\hat{\theta}$ is represented by

$$\hat{\theta} = \left[\frac{1}{n}\sum_{i=1}^{n}X_i(Z_i - \hat{\phi}(W_i))\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{g}(W_i))(Z_i - \hat{\phi}(W_i))\right]$$

where $\hat{\phi}$ and $\tilde{g}$ are isotonic regression estimators. By lemma A.2.5,

$$\|\phi_0 - \hat{\phi}\|_2 \lesssim O_p(n^{-1/3})$$

Note that:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[\frac{1}{n}\sum_{i=1}^{n}X_i(Z_i - \hat{\phi}(W_i))\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i + g_0(W_i) - \tilde{g}(W_i))(Z_i - \hat{\phi}(W_i))\right]$$

Now we are ready to prove the classical $\sqrt{n}$ consistency on $\hat{\theta}$. The strategies are as follows.
Step 1: (Glivinko-Cantelli) We need to show that

$$\sup_{\eta \in \mathcal{F}_\eta} \left| \frac{1}{n} \sum_{i=1}^{n} X_i(Z_i - \phi(W_i)) - E[X(Z - \phi(W))] \right| = o_P(1)$$

Proof of the step 1:
From the Theorem 2.4.1 of Van der Vaart and Wellner (1996), it suffices to show that $N_{[\,]}(\epsilon, \mathcal{F}_{xz}, \|\cdot\|_{P,1}) < \infty$ where $\mathcal{F}_{xz} := \{f \in \mathcal{F}_{xz} : f = X(Z - \phi(W))\}$.
Define $\mathcal{F}_x := \{f \in \mathcal{F}_x : f = X\}$, $\mathcal{F}_z := \{f \in \mathcal{F}_z : f = Z\}$, and $\mathcal{F}_\phi := \{f \in \mathcal{F}_\phi : f = \phi(W)\}$.
If $\|X\|_1 < \infty$, $\|Z\|_1 < \infty$, then

$$N_{[\,]}(\epsilon, \mathcal{F}_x, \|\cdot\|_{P,1}) \lesssim \frac{1}{\epsilon} < \infty$$

$$N_{[\,]}(\epsilon, \mathcal{F}_z, \|\cdot\|_{P,1}) \lesssim \frac{1}{\epsilon} < \infty$$

and

$$\log N_{[\,]}(\epsilon, \mathcal{F}_\phi, \|\cdot\|_{P,1}) \lesssim \frac{1}{\epsilon} < \infty$$

The last inequality can be derived by Theorem 2.7.5 in VW(1996) under assuming $\|\phi\|_\infty \leq M_\phi < \infty$ where M_ϕ is a big constant. In result, the entropy bracketing number of $\mathcal{F}_{xz}$ with L_1 norm is the rate of $\frac{1}{\epsilon}$ and it is finite for every $\epsilon > 0$. Hence, it is done by the theorem 2.4.1 of VW (1996).
Step 2: (Donsker) We need to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i + g_0(W_i) - \tilde{g}(W_i))(Z_i - \hat{\phi}(W_i)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i V_i + o_p(1)$$

One can show that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i + g_0(W_i) - \tilde{g}(W_i))(Z_i - \hat{\phi}(W_i)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i V_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(\phi_0(W_i) - \hat{\phi}(W_i)) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g_0(W_i) - \tilde{g}(W_i)) V_i \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g_0(W_i) - \tilde{g}(W_i))(\phi_0(W_i) - \hat{\phi}(W_i))
\end{aligned}$$

Note that the first term on the right-hand side converges to Gaussian distribution by the central limit theorem. All the other terms on the right hand side are $o_p(1)$ by Cauchy Schwarz inequality. Then one can show that the empirical L_2 norm can be bounded by the L_2 norm, and use lemma A.2.5 and $E[U^4], E[V^4]$ bounded. Combining step 1 and 2, then we have $\sqrt{n}(\hat{\theta} - \theta_0) = E[XV]^{-1}N(0, E[U^2V^2]) + +o_p(1)$. Recall that $\hat{\theta} = (\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\mathbb{X})^{-1}(\mathbb{X}'\tilde{\mathbb{V}}\hat{\Omega}\tilde{\mathbb{V}}'\tilde{\mathbb{Y}})$. One can also show that the each component of the matrix $\frac{1}{n}\mathbb{X}'\tilde{\mathbb{V}}$ converges to that of the matrix $E[VX']$. By similar logic of the 1-dimensional case, we have

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= \left(\frac{\mathbb{X}'\tilde{\mathbb{V}}}{n}\hat{\Omega}\frac{\tilde{\mathbb{V}}'\mathbb{X}}{n}\right)^{-1}\left(\frac{\mathbb{X}'\tilde{\mathbb{V}}}{n}\hat{\Omega}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_0(W_i) + U_i)(Z_i - \hat{\phi}(W_i))\right) \\ &= \left(E[V'X]E[U^2VV']^{-1}E[X'V]\right)^{-1}\left(E[V'X]E[U^2VV']^{-1}\right)N(0, E[U^2VV]) + o_p(1)\end{aligned}$$

Proof of Theorem 2.2.1

By lemma A.2.8,

$$\begin{aligned}&P\left(\sup_{a\in[f(0),f(1)]}|U_n(a) - f^{-1}(a)| > x\left(\frac{\log n}{n}\right)^{1/3}\right) \\ &\lesssim n^{1/3}\left(\frac{n^{-1/3}}{x^8(\log n)^{8/3}} + 2n^{-x^3}\right)\end{aligned}$$

Then there exist $C > 0$ such that for all $x > C$, the upper bound goes to zero as $n \to \infty$. Hence, we have

$$\sup_{a\in\mathbb{R}}|U_n(a) - f^{-1}(a)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$

Now we can apply lemma A.2.9,

$$\begin{aligned}S_n &= n^{1/3}\sup_{a\in f(u+\alpha_n),f(v-\beta_n)}|U_n(a) - f^{-1}(a)| \\ &= \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]}\frac{A(a)}{L'(f^{-1}(a))}|V_n(a)| + O_p(n^{-1/3}(\log n)^{-2/3})\end{aligned}$$

By applying lemma A.2.11,

$$\begin{aligned}S_n &= \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]}\frac{A(a)}{L'(g(a))}|V_n(a)| + O_p(n^{-1/3}(\log n)^{-2/3}) \\ &= \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]}\frac{A(a)}{L'(g(a))}|\tilde{V}_n(a)| + o_p((\log n)^{-2/3})\end{aligned}$$

By lemma A.2.12, we have

$$S_n \leq \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(b(a))}{L'(g(b(a)))} |\tilde{V}_n^+(a, b(a))| + o_p((\log n)^{-2/3})$$

and

$$S_n \geq \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(b(a))}{L'(g(b(a)))} |\tilde{V}_n^-(a, b(a))| + o_p((\log n)^{-2/3})$$

for any $b(a) \in \mathbb{R}$ that satisfies $|a - b(a)| \leq n^{-1/3}(\log n)^2$. Hence, we can use lemma A.2.13

$$S_B \leq \frac{S_n}{1 + O(\epsilon_n)} \leq \max\{S_B^{(1)}, S_A^{(2)}\} + o_p((\log n)^{-2/3})$$

Then the final result can be derived by applying Durot et al (2012) p.1592 $\sim$ 1595, the proof of the theorem 3.2.

Proof of Theorem 2.2.2

The proof is similar to the proof of theorem 2.2 in Durot et al (2012). Let $s_n = u + \alpha_n$ and $t_n = 1 - v + \beta_n$. By lemma A.2.21, we have

$$\sup_{u\in(u+\alpha_n,v-\beta_n]} B(u)|\hat{f}(u) - f(u)|$$

$$= \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} A(a)\Big|U_n(a) - f^{-1}(a)\Big| + O_p\left(\left(\frac{\log n}{n}\right)^{2/3}\right)$$

Define $a_n = n^{-1/3}(\log n)^{1/6}$. By using the monotonicity of f,

$$\sup_{u\in(u+\alpha_n,v-\beta_n]} |\hat{f}(u)-f(u)| \leq |\hat{f}(u+\alpha_n)-f(u+\alpha_n)|+|\hat{f}(u+a_n)-f(u+a_n)|+|f(u+\alpha_n)-f(a_n)|$$

See Durot et al (2012) supplement lemma 6.1 for detail. Then by lemma A.2.15 and lemma A.2.20, we have

$$\hat{f}(u + \alpha_n) - f(u + \alpha_n) = O_p((n\alpha_n)^{-1/2}) = o_p(n^{-1/3}(\log n)^{1/3})$$

$$\hat{f}(u + a_n) - f(u + a_n) = O_p(n^{-1/3})$$

In addition, we have

$$|f(u + \alpha_n) - f(a_n)| \leq \|f'\|_\infty|\alpha_n - a_n| = O(n^{-1/3}(\log n)^{1/6})$$

Since $B(s)$ is uniformly bounded, then we have

$$\sup_{s\in(u+\alpha_n,u+a_n]} B(s)|\hat{f}(s)-f(s)| = o_p\left(\left(\frac{\log n}{n})\right)^{1/3}\right)$$

$$\sup_{s\in(v-a_n,v-\beta_n]} B(s)|\hat{f}(s)-f(s)| = o_p\left(\left(\frac{\log n}{n})\right)^{1/3}\right)$$

Hence, by lemma A.2.21 again, the asymptotic distribution of $\sup_{s\in(u+\alpha_n,v-\beta_n]} B(s)|\hat{f}(s)-f(s)|$ and that of $\sup_{a\in[f(u+a_n),f(v-a_n)]} A(a)|U_n(a)-f^{-1}(a)|$ are the same.

A.2 Lemmas

A.2.1 Lemmas Related to Theorem 1.2.1

To prove L_2 bound of the isotonic regression estimator under conditional mean zero assumption (Lemma A.2.5), we need other lemmas from lemma A.2.1 to A.2.4. Suppose that the model is given as equation (1.2). From here to the end of the proof of Lemma A.2.5, define $\mathbf{W} := (W_1,\cdots,W_n)'$ and $(Z_{(i)},W_{(i)})_{i=1}^n$ is an ordered data set with respect to W. i.e. $W_{(1)}\le\cdots\le W_{(n)}$ and $\forall i=1,\cdots,n$, $Z_{(i)}=\phi_0(W_{(i)})+V_{(i)}$. Define the partial average of the function f with an ordered data set as $\bar{f}_{k,l} := \frac{1}{l-k+1}\sum_{i=k}^{l} f(W_{(i)})$. Similarly, the partial average of each random variable Z and V with the ordered data set is $\bar{Z}_{k,l} := \frac{1}{l-k+1}\sum_{i=k}^{l} Z_{(i)}$

Lemma A.2.1 *Given the model as equation (1.2), assume that ϕ_0 is monotonic increasing function. Then,*

$$E[(\hat{\phi}_j-\phi_{0j})^2|\mathbf{W}] \lesssim (\bar{\phi}_{0j,j+m}-\phi_{0j})^2 + E\left[\max_{1\le k\le j}\left(\bar{V}_{k,j+m}\right)_+^2\middle|\mathbf{W}\right]$$

where $\hat{\phi}_j := \hat{\phi}(W_{(j)})$ and $\phi_{0j} := \phi_0(W_{(j)})$.

By Brunk (1955), the isotonic regression has a closed form.

$$\hat{\phi}_j = \min_{l\ge j}\max_{k\le j}\frac{1}{l-k+1}\sum_{i=k}^{l} Z_{(i)}$$

$$= \min_{l\ge j}\max_{k\le j}(\bar{\phi}_{0k,l}+\bar{V}_{k,l})$$

Define the notation $X_+ := \max(X, 0)$ and m is a natural number such that $1 \le m \le n - j$. Then,

$$
\begin{aligned}
(\hat{\phi}_j - \phi_{0j})_+ &\le (\min_{l \ge j} \max_{k \le j}(\bar{\phi}_{0j,j+m} + \bar{V}_{k,l}) - \phi_{0j})_+ \\
&\le (\bar{\phi}_{0j,j+m} - \phi_{0j})_+ + (\min_{l \ge j} \max_{k \le j} \bar{V}_{k,l})_+ \\
&\le (\bar{\phi}_{0j,j+m} - \phi_{0j})_+ + (\max_{k \le j} \bar{V}_{k,j+m})_+
\end{aligned}
$$

The 1st inequality holds due to the monotonicity of ϕ_0. Then,

$$
E[(\hat{\phi}_j - \phi_{0j})_+^2 | \mathbf{W}] \le 2(\bar{\phi}_{j,j+m} - \phi_{0j})_+^2 + 2E\left[\max_{k \le j}(\bar{V}_{k,j+m})_+^2 \middle| \mathbf{W}\right]
$$

Define $X_- := \max(-X, 0)$. Then similarly,

$$
E[(\hat{\phi}_j - \phi_{0j})_-^2 | \mathbf{W}] \le 2(\bar{\phi}_{j-m,j} - \phi_{0j})_-^2 + 2E\left[\min_{l \ge j}(\bar{V}_{j-m,l})_-^2 \middle| \mathbf{W}\right]
$$

Under the assumption 1 and symmetry, the right hand side of the 2 inequalities above should have the same asymptotic convergence rate. Hence,

$$
\begin{aligned}
E[(\hat{\phi}_j - \phi_{0j})^2 | \mathbf{W}] &\le E[(\hat{\phi}_j - \phi_{0j})_+^2 | \mathbf{W}] + E[(\hat{\phi}_j - \phi_{0j})_-^2 | \mathbf{W}] \\
&\lesssim (\bar{\phi}_{0j,j+m} - \phi_{0j})^2 + E\left[\max_{1 \le k \le l}\left(\bar{V}_{k,j+m}\right)_+^2 \middle| \mathbf{W}\right]
\end{aligned}
$$

Lemma A.2.1 describes the upper bound of L_2 risk bound on j-th ordered explanatory variable. It consists of 2 main parts which are the drift explained by partial average process and the noise terms. This upper bound will be used to derive the upper bound of $E[(\hat{\phi}(W) - \phi_0(W))^2]$ by using the law of iterated expectation.

Lemma A.2.2 *Assume that W has a bounded support $[W_{(0)}, W_{(n+1)}]$ where $W_{(0)}$ is some constant such that $W_{(0)} \le W_{(1)}$ almost surely, and $W_{(n+1)}$ is some constant such that $W_{(n)} \le W_{(n+1)}$ almost surely. Define conditional probability of W given $\mathbf{W}, \mathbf{Z}$ as $p_{W|\mathbf{W},\mathbf{Z}}(w|\mathbf{W}, \mathbf{Z})$ and assume that it is bounded and away from 0 for all $w \in [W_{(0)}, W_{(n+1)}]$. Then,*

$$
E[(\hat{\phi}(W) - \phi_0(W))^2 | \mathbf{W}, \mathbf{Z}] \lesssim \sum_{j=0}^{n} (\hat{\phi}_j - \phi_{0j})^2 (W_{(j+1)} - W_{(j)}) + \sum_{j=0}^{n} (\phi_{0j+1} - \phi_{0j})^2 (W_{(j+1)} - W_{(j)})
$$

Moreover, if we assume that ϕ_{0j} *be bounded Lipchitz function. i.e.* $\forall x, y$, $\|\phi_0(x) - \phi_0(y)\| \leq L\|x - y\|$ *where* L *is universial constant and* $\sup_w |\phi_0(w)| < \infty$*, then,*

$$E[(\hat{\phi}(W) - \phi_0(W))^2|\mathbf{W}, \mathbf{Z}] \lesssim \sum_{j=0}^{n}(\hat{\phi}_j - \phi_{0j})^2(W_{(j+1)} - W_{(j)}) + \sum_{j=0}^{n}(W_{(j+1)} - W_{(j)})^3$$

Let $w_j(W) \in [0, 1]$ be the weight on $\hat{\phi}_j$ and $\hat{\phi}_{j+1}$ such that satisfies

$$\hat{\phi}(W) = w_j(W)\hat{\phi}_j + (1 - w_j(W))\hat{\phi}_{j+1}$$

where $W \in [W_{(j)}, W_{(j+1)}]$ conditioning on $\mathbf{W}$ and $\mathbf{Z}$. Then,

$$\begin{aligned}
E[(\hat{\phi}(W) - \phi_0(W))^2|\mathbf{W}, \mathbf{Z}] &= \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}(\hat{\phi}(w) - \phi_0(w))^2 p_{W|\mathbf{W},\mathbf{Z}}(w|\mathbf{W}, \mathbf{Z})dw \\
&\lesssim \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}(\hat{\phi}(w) - \phi_0(w))^2 dw \\
&\lesssim \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}[w_j(w)(\hat{\phi}_j - \phi_{0j}) + (1 - w_j(W))(\hat{\phi}_{j+1} - \phi_{0j+1})]^2 dw \\
&+ \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}[w_j(w)(\phi_{0j} - \phi_0(w)) + (1 - w_j(w))(\phi_{j+1} - \phi_0(w))]^2 dw \\
&\lesssim \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}(\hat{\phi}_j - \phi_{0j})^2 dw + \sum_{j=0}^{n}\int_{W_{(j)}}^{W_{(j+1)}}(\phi_{0j+1} - \phi_{0j})^2 dw \\
&= \sum_{j=0}^{n}(\hat{\phi}_j - \phi_{0j})^2(W_{(j+1)} - W_{(j)}) + \sum_{j=0}^{n}(\phi_{0j+1} - \phi_{0j})^2(W_{(j+1)} - W_{(j)})
\end{aligned}$$

Lemma A.2.2 shows that L_2 risk bound conditioning on the data set is depending on two main factors which are L_2 risk bound on j-th ordered explanatory variable and the variation of ordered statistics. The former part can be simplified by lemma A.2.1. Now we are ready to get the upper bound of L_2 risk bound.

Lemma A.2.3 *Assume that* W *has a bounded probability distribution function. (i.e.* $f_W(w) \leq \bar{M}, \forall w \in [W_{(0)}, W_{(n+1)}]$ *). Then,*

$$E[W_{(j+1)} - W_j] = O\left(\frac{1}{n}\right)$$

Define $F_W(w)$ as the CDF of W.

$$E[W_{(j)}] = \int_{W_{(0)}}^{W_{(n+1)}} w\left[\frac{n!}{(j-1)!(n-j)!}\right][F_W(w)]^{j-1}[1-F_W(w)]^{n-j}f_W(w)dw$$
$$= \int_0^1 F_W^{-1}(u)\left[\frac{n!}{(j-1)!(n-j)!}\right]u^{j-1}(1-u)^{n-j}du$$

Hence,

$$\begin{aligned}
E[W_{(j+1)} - W_{(j)}] &= \int_0^1 F_W^{-1}(u)\left[\frac{n!}{j!(n-j-1)!}\right]u^j(1-u)^{n-j-1}du \\
&- \int_0^1 F_W^{-1}(u)\left[\frac{n!}{(j-1)!(n-j)!}\right]u^{j-1}(1-u)^{n-j}du \\
&= \int_0^1 F_W^{-1}(u)\left[\frac{n!}{(j-1)!(n-j-1)!}\right]u^{j-1}(1-u)^{n-j-1}\left[\frac{nu-j}{j(n-j)}\right]du \\
&\lesssim \int_0^1 \left[\frac{n!}{j!(n-j-1)!}\right]u^j(1-u)^{n-j-1}\left[\frac{nu-j}{n-j}\right]du \\
&= \left(\frac{n}{n-j}\right)\left(\frac{j+1}{n+1}\right) - \frac{j}{n-j} = \frac{1}{n+1}
\end{aligned}$$

Lemma A.2.4 *(Conditional version of Doob's Inequality)*

Assume that $Var(V_i|\mathbf{W}) \leq \bar{\sigma}^2 < \infty$ *a.s. for all* $i = 1, \cdots, n$*. Then,*

$$E\left[\max_{1\leq k\leq j}\left(\bar{V}_{k,j+m}\right)_+^2\middle|\mathbf{W}\right] \lesssim O_p\left(\frac{1}{m}\right) + \sup_{s\neq t} E\left[V_tV_s\middle|\mathbf{W}\right]$$

almost surely.

Define $\eta := \max_{1\leq k\leq j}(\bar{V}_{k,j+m})_+$ and stopping time $\tau := sup\{1 \leq k \leq j : (\bar{V}_{k,j+m})_+ \geq t\} \vee 1$

with respect to filtration $\zeta_{k,j+m} := \sigma\Big(\bar{V}_{s,j+m} : j \geq s \geq k\Big)$. Then,

$$
\begin{aligned}
E\bigg[\max_{1\leq k\leq j}\Big(\bar{V}_{k,j+m}\Big)_+^2\bigg|\mathbf{W}\bigg] &= \int_\Omega\int_0^\infty 2t1_{[0,\eta(\omega)](t)}dtdP_{\eta|\mathbf{W}(\omega)} \\
&= \int_0^\infty 2tP(\eta \geq t|\mathbf{W})dt \\
&\leq 2\int_0^\infty E[\eta 1_{\eta\geq t}|\mathbf{W}]dt \\
&= 2\int_\Omega (\bar{V}_{\tau,j+m}(\omega))_+\int_0^\infty 1_{\eta(\omega)\geq t}dtdP_{\eta|\mathbf{W}}(\omega) \\
&= 2E[(\bar{V}_{\tau,j+m})_+\eta|\mathbf{W}] \\
&\leq 2(E[(\bar{V}_{\tau,j+m})_+^2|\mathbf{W}])^{\frac{1}{2}}(E[\eta^2|\mathbf{W}])^{\frac{1}{2}} \qquad a.s.
\end{aligned}
$$

It can be done by using Fubini-Tonelli theorem at the 2nd and 4th line, Markov inequality at the 3rd line and Caushy Schwarz inequality at the last line. Hence,

$$
\begin{aligned}
E\bigg[\max_{k\leq j}\Big(\bar{V}_{k,j+m}\Big)_+^2\bigg|\mathbf{W}\bigg] &\leq 4E[(\bar{V}_{\tau,j+m})_+^2|\mathbf{W}] \\
&\leq 4E[(\bar{V}_{1,j+m})^2|\mathbf{W}] \\
&= 4\Big(\frac{1}{j+m}\Big)^2E\bigg[\sum_{i=1}^{j+m}V_i^2 + 2\sum_{s<t}V_sV_t\bigg|\mathbf{W}\bigg] \\
&\lesssim \frac{\bar{\sigma}^2}{j+m} + \sup_{s\neq t}E[V_sV_t|\mathbf{W}] \\
&\leq \frac{\bar{\sigma}^2}{m+1} + \sup_{s\neq t}E[V_sV_t|\mathbf{W}]
\end{aligned}
$$

Lemma A.2.5 ***(L_2 Risk Bound of the Isotonic Regression)***

Suppose that the model is given as equation (2). Assume that (a) $\{W_i\}_{i=1}^n$ has bounded support. (b) $\{V_i\}_{i=1}^n$ are independent each other. (c) $|\phi_0'|$ is bounded below by zero and above by a large constant $C_{\phi'}$. (d) Probability distribution function of $\{W_i\}_{i=1}^n$ is away from 0 and bounded above on all the support of W. (e) ϕ_0 is bounded Lipschitz continuous strictly *increasing (or decreasing) function. Then, the isotonic regression estimator $\hat{\phi}$ satisfies,*

$$\|\hat{\phi} - \phi_0\|_2 = O_p(n^{-\frac{1}{3}})$$

By combining with lemmas A.2.1, A.2.2, A.2.3 and A.2.4,

$$\begin{aligned}
E[(\hat{\phi}(W) - \phi_0(W))^2] &\lesssim E\bigg[\sum_{j=0}^{n}(\hat{\phi}_j - \phi_{0j})^2(W_{(j+1)} - W_{(j)}) + \sum_{j=0}^{n}(\phi_{0j+1} - \phi_{0j})^2(W_{(j+1)} - W_{(j)})\bigg] \\
&\lesssim E\bigg[\sum_{j=0}^{n}(\bar{\phi}_{0j,j+m} - \phi_{0j})^2(W_{(j+1)} - W_{(j)})\bigg] \\
&+ E\bigg[E\bigg[\sum_{j=0}^{n}\max_{1\le k\le j}(\bar{V}_{k,j+m})_+^2\bigg|\mathbf{W}\bigg](W_{(j+1)} - W_{(j)})\bigg] \\
&+ E\bigg[\sum_{j=0}^{n}(\phi_{0j+1} - \phi_{0j})^2(W_{(j+1)} - W_{(j)})\bigg] \\
&\lesssim E\bigg[\sum_{j=0}^{n}O_p\bigg(\frac{m^2}{n^2}\bigg)O_p\bigg(\frac{1}{n}\bigg)\bigg] \\
&+ E\bigg[\sum_{j=0}^{n}O_p\bigg(\frac{1}{m}\bigg)O_p\bigg(\frac{1}{n}\bigg)\bigg] + E\bigg[\sum_{j=0}^{n}\sup_{s\neq t}E\bigg[V_tV_s\bigg|\mathbf{W}\bigg]O_p\bigg(\frac{1}{n}\bigg)\bigg] \\
&+ E\bigg[\sum_{j=0}^{n}O_p\bigg(\frac{1}{n^3}\bigg)\bigg] \\
&\lesssim O\bigg(\frac{m^2}{n^2}\bigg) + O\bigg(\frac{1}{m}\bigg) + 0 + O\bigg(\frac{1}{n^2}\bigg)
\end{aligned}$$

Note that

$$E\bigg[E\bigg[V_tV_s\bigg|\mathbf{W}\bigg]\bigg] = E[V_tV_s] = E[E[V_t|W_s, V_s]V_s] = E[E[V_t|W_t]V_s] = 0$$

By using independence. The lemma satisfies the sufficient condition in Chernozhukov et al (2018). Under assumptions given in the lemmas A.2.1, A.2.2, A.2.3, and A.2.4,

$$E[(\hat{\phi}(W) - \phi_0(W))^2] \lesssim O\bigg(\frac{m^2}{n^2}\bigg) + O\bigg(\frac{1}{m}\bigg) + O\bigg(\frac{1}{n^2}\bigg)$$

Moreover, choose $m = n^{\frac{2}{3}}$. Then,

$$E[(\hat{\phi}(W) - \phi_0(W))^2] \lesssim O\bigg(n^{-\frac{2}{3}}\bigg)$$

A.2.2 Lemmas Related to Theorem 2.2.1 and 2.2.2

From now on, I introduce lemmas for establishing the confidence band of the monotone functions. Most lemmas have the similar proof strategies in Durot et al (2012) and Durot et al (2012) supplement. The main difference between their paper and this paper is that (a) we need to deal with the random explanatory variables W instead of fixed design points, and (b) the parameter is increasing, so the cumulative sum diagram should be re-defined to get the valid isotonic regression estimator.

Lemma A.2.6 *(Koltchinskii Coupling Lemma (Chernozhukov, Newey, Santos (2020)) Assume the followings.*

(a) The joint probability density function of (W, V) is bounded away from 0 and above, ($0 < P(W = w, V = v) < \infty$ for all $w \in [-M, M], v \in [-M, M]$ for some constant M.)

(b) There is a continuously differentiable bijection $T : [0, 1]^2 \to \Omega$ and the Jacobian JT and its determinant $|JT|$ satisfy $\inf\limits_{v\in[0,1]^2} |JT(v)| > 0$ and $\sup\limits_{v\in[0,1]^2} \|JT(v)\|_o < \infty$ where $\|\cdot\|_o$ is the operator norm endowed with Euclidean norm.

(c) Let $\mathcal{F} = \{h \in \mathcal{F} : h(W, v) = (\phi_0(W) + V)1_{\{W<t\}} + \phi_0(t)1_{\{W\le t\}}\}$ where $t = G^{-1}(s)$ and $\phi_0(\cdot)$ is defined in equation (2). Define the integral modulus of the continuity,

$$\omega(h, b) = \sup_{\|s\|\le b}\left(\int_\Omega \{(h(w + s_1, v + s_2) - h(w, v)\}^2 1_{\{(w,v)+s\in\Omega\}} dP(w, v)\right)^{1/2}$$

Then $\mathcal{F}$ satisfy (c-1) $\sup\limits_{h\in\mathcal{F}} \omega(h, b) \le \varphi(b)$ for some $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\varphi(Cb) \le C^\kappa\varphi(b)$ for all $C > 0$ and some $\kappa > 0$ and (c-2) $\sup\limits_{h\in\mathcal{F}} \|h\|_\infty \le K$ for some $K > 0$.

Then there exist $K > 0$ such that for $x > K \log n$,

$$P(n^{3/4}\|\Lambda_n - \Lambda - n^{-1/2}B_n \circ L\|_\infty > x) \lesssim x^{-4} \tag{A.1}$$

where B_n is Brownian motion and $L(s) = Var(Z1_{\{W<G^{-1}(s)\}} - \mu(G^{-1}(s))1_{\{W\le G^{-1}(s)\}})$. Moreover,

$$\|\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_\infty = O_p\left(\frac{\log n}{\sqrt{n}}\right)$$

To verify the lemma, the first step is to simplify the empirical process, $\sqrt{n}(\Lambda_n - \Lambda)$.

$$
\begin{aligned}
&\sqrt{n}(\Lambda_n(s) - \Lambda(s)) \\
&\quad = \sqrt{n}(F_n \circ G_n^{-1}(s) - F \circ G_n^{-1}(s)) + \sqrt{n}(F \circ G_n^{-1}(s) - F \circ G^{-1}(s)) \\
&\quad = \sqrt{n}(F_n \circ G^{-1}(s) - F \circ G^{-1}(s)) + \sqrt{n}\bigg(\frac{dF}{dG^{-1}(s)}(G^{-1}(s))\bigg)(G_n^{-1}(s) - G^{-1}(s)) + o_p(1) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Z_i 1_{\{W_i < G^{-1}(s)\}} - E[Z 1_{\{W < G^{-1}(s)\}}]\} \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mu(G^{-1}(s))\{1_{\{W_i \leq G^{-1}(s)\}} - E[1_{\{W \leq G^{-1}(s)\}}]\} + o_p(1)
\end{aligned}
$$

To get the first part after second equality, I apply Andrews (1994) p.2265 (3.34). To show this, we need to show 3 conditions as mentioned in (3.36) in Andrews (1994). Let $\hat{t} = G_n^{-1}(s)$ and $t_0 = G^{-1}(s_0)$. Define $\nu_n(\cdot) = \sqrt{n}(F_n(\cdot) - F(\cdot))$. Then $\nu_n(\cdot)$ satisfies Donsker class because there exist envelope function for Z such that $\|Z\|_2^2 \leq \|\phi_0\|_\infty^2 + Var(V) < \infty$, and the uniform entropy condition holds since the bracketing number of $\mathcal{F}_1 := \{f_1 \in \mathcal{F}_1 : f_1(t) = Z1_{\{W<t\}}\}$ is approximately $N_{[\,]}(\epsilon, \mathcal{F}_1, \|\cdot\|_2) \lesssim \epsilon^{-2}$. It implies that the condition (i) holds. The condition (ii) and (iii) hold since $\hat{t}$ is a consistent estimator of t and should exist between 0 and 1 always. The latter part of the third line holds by the delta method. The last equality can be derived by $[dF/dG^{-1}(s)](\cdot) = \phi_0(\cdot)p(\cdot)$ which comes from,

$$
F(s) = \int_0^s f(s)ds = \int_{-M}^{G^{-1}(s)} \phi_0(x)p(x)dx
$$

where $p(x)$ is the probability distribution function of W. Also, the empirical quantile estimator's asymptotic distribution,

$$
\begin{aligned}
&\sqrt{n}(G_n^{-1}(s) - G^{-1}(s)) \\
&\quad = \sqrt{n}(G^{-1} \circ G \circ G_n^{-1}(s) - G^{-1}(s)) \\
&\quad = \sqrt{n}\bigg(\frac{dG^{-1}(s)}{ds}(s)\bigg)(G \circ G_n^{-1}(s) - s) + o_p(1) \\
&\quad = \sqrt{n}\bigg(\frac{1}{p(G^{-1}(s))}\bigg)(G(G_n^{-1}(s)) - G_n(G_n^{-1}(s))) + o_p(1) \\
&\quad = \bigg(\frac{1}{p(t)}\bigg)\sqrt{n}(G(t) - G_n(t)) + o_p(1) \to N\bigg(0, \frac{G(t)(1-G(t))}{p^2(t)}\bigg)
\end{aligned}
$$

where $t = G^{-1}(s)$ at the last line. Then, $Var(\sqrt{n}(\Lambda_n(s) - \Lambda(s))$ can be approximated by,

$$\begin{aligned} &Var(\sqrt{n}(\Lambda_n(s) - \Lambda(s))) \\ &\qquad \approx Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{i1}(s) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{i2}(s)\right) = L(s) \end{aligned}$$

where the influence functions Φ_{i1} and Φ_{i2} are

$$\Phi_{i1}(s) = \{Z_i 1_{\{W_i < G^{-1}(s)\}} - E[Z 1_{\{W < G^{-1}(s)\}}]\}$$

$$\Phi_{i2}(s) = -\mu(G^{-1}(s))\{1_{\{W_i \le G^{-1}(s)\}} - E[1_{\{W \le G^{-1}(s)\}}]\}$$

By Chernozhukov, Newey, Santos (CNS) (2020) Chapter S.6 Coupling via Koltchinskii, we first calculate the integral modulus of the continuity of $h(W, V)$,

$$\begin{aligned} &|h(w+s_1, v+s_2) - h(w,v)| \\ \le&|h(w+s_1, v+s_2) - h(w+s_1, v)| + |h(w+s_1, v) - h(w,v)| \\ =&|(\phi_0(w+s_1)1_{\{w+s_1<t\}} + (v+s_2)1_{\{w+s_1<t\}} + \phi_0(t)1_{\{w+s_1<t\}}) \\ &- (\phi_0(w+s_1)1_{\{w+s_1<t\}} + v1_{\{w+s_1<t\}} + \phi_0(t)1_{\{w+s_1<t\}})| \\ &+ |(\phi_0(w+s_1)1_{\{w+s_1<t\}} + v1_{\{w+s_1<t\}} + \phi_0(t)1_{\{w+s_1<t\}}) - (\phi_0(w)1_{\{w<t\}} + v1_{\{x<t\}} + \phi_0(t)1_{\{w<t\}})| \\ &\lesssim s_2 + \|\phi'\|_\infty s_1 + (\|\phi_0\|_\infty + \|\phi_0'\|_\infty)(1_{\{w+s_1<t\}} - 1_{\{w\le t\}}) \\ &\lesssim s_2 + s_1 + (1_{\{w+s_1<t\}} - 1_{\{w\le t\}}) \end{aligned}$$

The last inequality is by the boundedness of ϕ_0 and ϕ_0'. Then,

$$\begin{aligned} &E[\{h(W+s_1, V+s_2) - h(W,V)\}^2] \\ &\qquad \lesssim s_2^2 + s_1^2 + E[|1_{\{W+s_1<t\}} - 1_{\{W\le t\}}|] \\ &\qquad \le s_2^2 + s_1^2 + (\sup_{w\in[-M,M]} p(w))s_1 \\ &\qquad \lesssim s_2^2 + s_1^2 + s_1 \lesssim s_1 \end{aligned}$$

The last inequality holds if s_1, s_2 are small enough. Hence, by the assumption (c) above, we can take $\varphi(b) = \sqrt{b}$. To apply theorem S.6.1 in CNS (2020), we need to derive $S_n^2 = \sum_{i=1}^{\lceil log_2 n\rceil} 2^i \varphi^2(2^{i/d})$ where d is the dimension of the random variables. As above, there are 2 random variables (W, V), so $d = 2$. Plug this in the definition of S_n, we easily get,

$$S_n^2 = \sum_{i=1}^{\lceil log_2 n\rceil} 2^i \varphi^2(2^{i/2})$$
$$\lesssim 1 + \int_1^{log_2 n} 2^{x/2} dx$$
$$\lesssim \sqrt{n}$$

Hence, $S_n \lesssim n^{1/4}$. Define both $J_n = J_{[\,]}(\delta_n, \mathcal{F}, \|\cdot\|_2) = \int_0^{\delta_n} \sqrt{1 + \log N_{[\,]}(u, \mathcal{F}, \|\cdot\|_2)} du$ and $N_n = N_{[\,]}(\delta_n, \mathcal{F}, \|\cdot\|_2)$. Note that h is empirical process indexed by t. Separate $h(t)$ into $h_1(t)$ and $h_2(t)$ such that

$$h_1(t) = Z1_{\{W<t\}}$$
$$h_2(t) = \mu(t)1_{\{W\leq t\}}$$

Then, $N_{[\,]}(u, \mathcal{F}, \|\cdot\|_2) \lesssim N_{[\,]}(u, \mathcal{F}_1, \|\cdot\|_2) + N_{[\,]}(u, \mathcal{F}_2, \|\cdot\|_2)$ where $\mathcal{F}_2 := \{f_2 \in \mathcal{F}_2 : f_2(t) = \phi_0(t)1_{\{X<t\}}\}$. As $E[Z^2] < \infty$, The bracketing number of $\mathcal{F}_1$ has the same rate of the indicator function, $N_{[\,]}(u, \mathcal{F}_1, \|\cdot\|_2) \lesssim u^{-2}$. Due to the fact that $\mu(\cdot)$ is smooth and bounded monotone function, the bracketing number of $\mathcal{F}_2$ has the maximum rate between the parametric model and the indicator function space, $N_{[\,]}(u, \mathcal{F}_2, \|\cdot\|_2) \lesssim u^{-1} + u^{-2}$. Hence,

$$N_{[\,]}(u, \mathcal{F}, \|\cdot\|_2) \lesssim u^{-2}$$

where u is small enough. Plug it in to derive J_n,

$$J_n = \int_0^{\delta_n} \sqrt{1 + \log N_{[\,]}(u, \mathcal{F}, \|\cdot\|_2)} du$$
$$\leq \int_0^{\delta_n} \sqrt{1 + \log\left(\frac{c_2}{u^2}\right)} du$$

where $c_2 > 0$ is large enough. If u is large enough to cover the diameter of $\mathcal{F}$, then $N_{[\,]}(u, \mathcal{F}, \|\cdot\|_2) = 1$. So there should be 2 cases that (i) δ_n is large or (ii) δ_n is small. For the case (i),

$$J_n \leq \int_0^{\delta_n} \sqrt{1 + \log\left(\frac{c_2}{u^2}\right)} du$$
$$= \int_0^{c_2} \sqrt{1 + \log\left(\frac{c_2}{u^2}\right)} du + \int_{c_2}^{\delta_n} \sqrt{1 + \log\left(\frac{c_2}{u^2}\right)} du$$
$$\lesssim \delta_n$$

In case (ii),

$$
\begin{aligned}
J_n &\le \int_0^{\delta_n} \sqrt{1+\log\left(\frac{c_2}{u^2}\right)} du \\
&= \lim_{t\to 0} \int_t^{\delta_n} \sqrt{1+\log\left(\frac{c_2}{u^2}\right)} du \\
&\lesssim \lim_{t\to 0} \int_{1/\delta_n}^{1/t} \sqrt{\log s}\frac{1}{s^2} ds \\
&\lesssim \lim_{t\to 0} \int_{\log(1/\delta_n)}^{\log(1/t)} \sqrt{x} e^{-x} dx \\
&\lesssim \delta_n \sqrt{\log(1/\delta_n)}
\end{aligned}
$$

To show the rate of the supremum norm between empirical process and Brownian motion as $O_p\left(\frac{\log n}{\sqrt{n}}\right)$, apply CNS (2020) equation (S.308).

$$
\|\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_\infty = O_p\left(\frac{\log(nN_n)}{\sqrt{n}} + \frac{\sqrt{\log(nN_n)\log(n)}S_n}{\sqrt{n}} + J_n\left(1+\frac{J_n}{\delta_n^2\sqrt{n}}\right)\right)
$$

Plug in $N_n = \delta_n^{-2}$, $S_n = n^{1/4}$, and $J_n = \delta_n\sqrt{\log(1/\delta_n)}$. Then set $\delta_n = n^{-1/2+n^{-1/2}}$. Finally, we can get

$$
\|\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_\infty \lesssim O_p\left(\frac{\log n}{\sqrt{n}}\right)
$$

Define a function $l(\eta_1, \eta_2, S_n) = \eta_1 + \sqrt{\eta_1}\sqrt{\eta_2}(C_1 S_n + 1)$. To show the inequality (8), let $S_{\delta_n} \subset [0,1]$ denote a finite δ_n-net of $[0,1]$ with respect to $\|\cdot\|_2$. Now, we can apply inequality (S.311) in CNS (2020). There exists $B_n \circ L$ such that for all $\eta_1 > 0, \eta_2 > 0$,

$$
\begin{aligned}
&P(\sqrt{n}\|\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_{S_{\delta_n}} \ge l(\eta_1, \eta_2, S_n)) \\
&\quad \lesssim N_n e^{-C_2\eta_1} + n e^{-C_2\eta_2}
\end{aligned}
$$

Choose $\eta_1 = \frac{\log(x^4 N_n)}{C_2}$ and $\eta_2 = \frac{\log(x^4 n)}{C_2}$. Note that $S_n \approx n^{1/4}$, and $N_n \approx \delta_n^{-2} \approx n^{1-2n^{-1/2}}$. Then there exist C_3 and C_4 such that satisfies

$$
\begin{aligned}
&P(\sqrt{n}\|\sqrt{n}(\Lambda_n - \Lambda) - B_n \circ L\|_{S_{\delta_n}} \ge \eta_1 + \sqrt{\eta_1}\sqrt{\eta_2}(C_1 S_n + 1)) \\
&= P(n^{1/4}\|M_n - B_n \circ L\|_{S_{\delta_n}} \ge C_3 n^{-1/4}\log(x^4 n^{1-2n^{-1/2}}) + C_4\sqrt{\log(x^4 n^{1-2n^{-1/2}})}\sqrt{log(x^4 n)}) \\
&\lesssim x^{-4}
\end{aligned}
$$

where $M_n = \sqrt{n}(\Lambda_n - \Lambda)$. Assume that there exists a large $K > 0$, such that $x > K \log n$. Then we can show that

$$x > K\log n > C_3 n^{-1/4}\log(x^4 n^{1-2n^{-1/2}}) + C_4\sqrt{\log(x^4 n^{1-2n^{-1/2})}}\sqrt{log(x^4 n)}$$

In result, for all $x > K\log n$,

$$\begin{aligned}
&P(n^{1/4}\|M_n - B_n \circ L\|_{S_{\delta_n}} \geq x)\\
&\leq P(n^{1/4}\|M_n - B_n \circ L\|_{S_{\delta_n}} \geq C_3 n^{-1/4}\log(x^4 n^{1-2n^{-1/2}}) + C_4\sqrt{\log(x^4 n^{1-2n^{-1/2})}}\sqrt{log(x^4 n)})\\
&\lesssim x^{-4}
\end{aligned}$$

Then the tail probability can be rewritten by

$$\begin{aligned}
&P(n^{1/4}\|M_n - B_n \circ L\|_\infty > x)\\
&\leq P(n^{1/4}\|M_n - B_n \circ L\|_{S_{\delta_n}} > x/3) + P(n^{1/4}\|M_n \circ \Gamma_n - M_n\|_S > x/3)\\
&+ P(n^{1/4}\|B_n \circ \Gamma_n - B_n\|_S > x/3)\\
&\lesssim x^{-4} + P(n^{1/4}\|M_n \circ \Gamma_n - M_n\|_S > x/3) + P(n^{1/4}\|B_n \circ \Gamma_n - B_n\|_S > x/3)
\end{aligned}$$

where $x > K\log n$. To control the third term on the last inequality, I apply Markov's inequlity , the proposition A.2.4 in VW(1996), and CNS(2020) equation (S.313),

$$\begin{aligned}
&P(n^{1/4}\|B_n \circ \Gamma_n - B_n\|_S > x/3)\\
&= P(n\|B_n \circ \Gamma_n - B_n\|_S^4 > x^4/81) \lesssim x^{-4} n E[\|B_n \circ \Gamma_n - B_n\|_S^4]\\
&\lesssim x^{-4} n (E[\|B_n \circ \Gamma_n - B_n\|_S])^4\\
&\lesssim x^{-4} n J_n
\end{aligned}$$

Note that J_n is a function of δ_n and decide δ_n later to minimize the upper bound. To deal with the second term, apply Markov's inequality, theorem 2.14.5 in VW(1996), and CNS(2020) equation (S.314).

$$\begin{aligned}
&P(n^{1/4}\|M_n \circ \Gamma_n - M_n\|_S > x/3) \lesssim x^{-4} n E[\|M_n \circ \Gamma_n - M_n\|_S^4]\\
&\lesssim x^{-4} n\left(\{E[\|M_n \circ \Gamma_n - M_n\|_S]\}^4 + \frac{1}{n}\right) = x^{-4}\left(1 + n\{E[\|M_n \circ \Gamma_n - M_n\|_S]\}^4\right)\\
&\lesssim x^{-4}\left(1 + n J_n^4\left(1 + \frac{J_n}{\delta_n^2\sqrt{n}}\right)^4\right)
\end{aligned}$$

Combine all the bounds at the same time.

$$P(n^{1/4}\|M_n - B_n \circ L\|_\infty > x) \lesssim x^{-4}\left(1 + nJ_n^4\left(1 + \frac{J_n}{\delta_n^2\sqrt{n}}\right)^4\right)$$

Note that $J_n \approx \delta_n\sqrt{\log(1/\delta_n)}$. We need

$$nJ_n^4 \approx n\delta_n^4 \log^2(1/\delta_n) \lesssim O(1)$$

and also

$$\frac{J_n}{\delta_n^2\sqrt{n}} \approx \frac{\sqrt{\log(1/\delta_n)}}{\delta_n\sqrt{n}} \lesssim O(1)$$

Hence, choose $\delta_n = n^{-1/4-\alpha}$ for some small $\alpha > 0$.

Lemma A.2.7 *Under assumption 2.2.1, for all $a \in \mathbb{R}$ and $x \geq K(\frac{\log n}{n^{1/12}})^{1/2}$,*

$$P(n^{1/3}|U_n(a) - f^{-1}(a)| > x) \lesssim \frac{n^{-1/3}}{x^8} + e^{-x^3}$$

In particular, for all $a \in \mathbb{R}$, we have that

$$U_n(a) - f^{-1}(a) = O_p(n^{-1/3})$$

This lemma is a small modification of the lemma 6.4 in Durot et al (2012) supplement. Fix $a \in \mathbb{R}$ and take $y_n = xn^{-1/3}$. By the definition of the inverse process $U_n(a)$, $|U_n(a) - f^{-1}(a)| > y_n$ is true only if there exists u such that $|u - f^{-1}(a)| > y_n$ and $\Lambda_n(u) - au \leq \Lambda(g(a)) - a(g(a))$. Hence,

$$\begin{aligned}
&P(|U_n(a) - f^{-1}(a)| > y_n) \\
&\quad\leq P\left(\inf_{|u-f^{-1}(a)|>y_n}\{\Lambda_n(u) - au\} \leq \{\Lambda_n(f^{-1}(a)) - a(f^{-1}(a))\}\right) \\
&\quad= P\left(\sup_{|u-f^{-1}(a)|>y_n}\{\Lambda_n(f^{-1}(a)) - \Lambda_n(u)\} - a(f^{-1}(a) - u) \geq 0\right) \\
&\quad= P\left(\sup_{|u-f^{-1}(a)|>y_n}\{P_{1n} + P_{2n} + P_{3n} + P_{4n}\} \geq 0\right)
\end{aligned}$$

where

$$P_{1n} = -\{\Lambda_n(u) - \Lambda(u) - n^{-1/2}B_n \circ L(u)\}$$

$$P_{2n} = \Lambda_n(f^{-1}(a)) - \Lambda(g(a)) - n^{-1/2}B_n \circ L(f^{-1}(a))$$

$$P_{3n} = \Lambda(f^{-1}(a)) - \Lambda(u) - a(f^{-1}(a) - u)$$

$$P_{4n} = n^{-1/2}(B_n \circ L(f^{-1}(a)) - B_n \circ L(u))$$

Note that for some $c > 0$ by application of Taylor's expansion $\Lambda(f^{-1}(a)) - \Lambda(u) \approx f(f^{-1}(a))(f^{-1}(a) - u) + \frac{f'(\tilde{s})}{2}(f^{-1}(a) - u)^2$ where $\tilde{s}$ is located between $f^{-1}(a)$ and u. Due to the monotonicity of $\Lambda(s)$, (i) if $f^{-1}(a) > u$, then $\Lambda(f^{-1}(a)) - \Lambda(u) \geq f(f^{-1}(a))(f^{-1}(a) - u) - c(f^{-1}(a) - u)^2$. (ii) Otherwise, $\Lambda(u) - \Lambda(f^{-1}(a)) \leq f(f^{-1}(a))(u - f^{-1}(a)) + c(f^{-1}(a) - u)^2$. Either of cases, we get

$$P_{3n} \leq -c(f^{-1}(a) - u)^2$$

Then, we have that

$$P\left(\sup_{|u-f^{-1}(a)|>y_n} \{P_{1n} + P_{2n} + P_{4n} - c(f^{-1}(a) - u)^2\} \geq 0\right)$$

$$\lesssim P\left(\sup_{s\in[0,1]} |\Lambda_n(s) - \Lambda(s) - n^{-1/2}B_n \circ L(s)| > c_1 y_n^2\right) + P\left(\sup_{|u-f^{-1}(a)|>y_n} \{P_{4n} - c_2(f^{-1}(a) - u)^2\} \geq 0\right)$$

for some c_1, c_2 such that $c_1 + c_2 = c$. Note that $y_n^2 = x^2 n^{-2/3}$. So, there exist some $K > 0$ such that satisfies $K \log n < n^{3/4} y_n^2 = n^{1/12} x^2$. To apply lemma A.2.6, we need $x \geq K\left(\frac{\log n}{n^{1/12}}\right)^{1/2}$. Then we can bound P_{1n}, P_{2n}.

$$P\left(n^{3/4} \sup_{s\in[0,1]} |\Lambda_n(s) - \Lambda(s) - n^{-1/2}B_n \circ L(s)| > c_1 n^{3/4} y_n^2\right) \lesssim n^{-3} y_n^{-8} = n^{-1/3} x^{-8}$$

The remainder is P_{4n}.

$$\begin{aligned} &P\left(\sup_{|u-f^{-1}(a)|>y_n} \{P_{4n} - c_2(f^{-1}(a) - u)^2\} \geq 0\right) \\ &\quad = P\left(\sup_{|u-f^{-1}(a)|>y_n} \{n^{-1/2}(B_n \circ L(f^{-1}(a)) - B_n \circ L(u) - c_2(f^{-1}(a) - u)^2\} \geq 0\right) \end{aligned}$$

By Durot et al (2012) supplement lemma 6.4 page 8,9, and Revuz, Yor (1999) (Continuous martingales and Brownian motion [3rd ed.]) page 55 Proposition (1.8), it is bounded by $c_3 e^{-n y_n^3} = c_3 e^{-x^3}$ where c_3 is a positive constant.

To argue $U_n(a) - f^{-1}(a) = O_p(n^{-1/3})$, it suffices to show that for any $\epsilon > 0$, there exists $M < \infty$ such that satisfies

$$\lim_{n\to\infty} P\bigg(n^{1/3}|U_n(a) - f^{-1}(a)| > M\bigg) < \epsilon$$

Choose M such that satisfies $M > K\left(\frac{\log n}{n^{1/12}}\right)^{1/2}$ for some $K > 0$. Then we can apply the previous inequality for large n. Hence,

$$\lim_{n\to\infty} P\bigg(n^{1/3}|U_n(a) - f^{-1}(a)| > M\bigg) \lesssim \lim_{n\to\infty}\left(\frac{n^{-1/3}}{M^8} + e^{-M^3}\right) = e^{-M^3}$$

Then we can always choose M large enough to bound the probability less than ϵ.

Lemma A.2.8 *Under assumption 2.2.1, for any $a, b \in \mathbb{R}$ with $b - a \geq n^{-1/3}$, and for $y_n \geq K_1\left(\frac{\log n}{n^{1/12}}\right)^{1/2}$ for some $K_1 > 0$. Then we have*

$$P\bigg(\sup_{c\in[a,b]} n^{1/3}|U_n(c) - f^{-1}(c)| > y_n\bigg) \lesssim (b-a)\Big(n^{-1/3}y_n^{-8} + e^{-y_n^{-3}}\Big)$$

This lemma is a modification of Durot et al (2012) supplement lemma 6.5. Since we have $f^{-1}(a) = 1$ for all $a \geq f(1)$, $f^{-1}(a) = 0$ for all $a \leq f(0)$, and $U_n(a)$ is increasing by definition, we have

$$\sup_{a\leq f(0)}\Big|U_n(a) - f^{-1}(a)\Big| = \Big|U_n(f(0)) - f^{-1}(f(0))\Big|$$

and

$$\sup_{a\geq f(1)}\Big|U_n(a) - f^{-1}(a)\Big| = \Big|U_n(f(1)) - f^{-1}(f(1))\Big|$$

Hence, without loss of generality, we can assume $[a, b] \in [f(0), f(1)]$. Decompose the interval $[a, b]$ into K intervals $[c_k, c_{k+1}]$ where

$$c_k = a + \frac{k(b-a)}{K}, \qquad \text{for } k = 0, 1, \cdots, K$$

where $K = [n^{1/3}(b-a) + 1]$. Then we have

$$\sup_{c\in[a,b]}\Big|U_n(c) - f^{-1}(c)\Big| = \max_{0\leq k\leq K-1}\sup_{c\in[c_k,c_{k+1}]}\Big|U_n(c) - f^{-1}(c)\Big|$$

Note that f^{-1} is Lipschitz continuous and has bounded derivative. Then we have

$$\max_{0\le k\le K-1}\sup_{c\in[c_k,c_{k+1}]}\Big|f^{-1}(c_k)-f^{-1}(c)\Big| \le \|f^{-1\prime}\|_\infty \max_{0\le k\le K-1}|c_{k+1}-c_k| \le \|f^{-1\prime}\|_\infty \frac{(b-a)}{K}$$

By triangular inequality, we have

$$\max_{0\le k\le K-1}\sup_{c\in[c_k,c_{k+1}]}\Big|U_n(c)-f^{-1}(c)\Big| \le \max_{0\le k\le K-1}\sup_{c\in[c_k,c_{k+1}]}\Big|U_n(c)-f^{-1}(c_k)\Big| + \|f^{-1\prime}\|_\infty \frac{(b-a)}{K}$$

Now, by monotonicity of $U_n(a)$ and triangular inequality, we have

$$\begin{aligned}\sup_{c\in[c_k,c_{k+1}]}\Big|U_n(c)-f^{-1}(c_k)\Big| &= \max\left\{\Big|U_n(c_k)-f^{-1}(c_k)\Big|, \Big|U_n(c_{k+1})-f^{-1}(c_k))\Big|\right\}\\ &\lesssim \max_{0\le k\le K}\Big|U_n(c_k)-f^{-1}(c_k)\Big| + \max_{0\le k\le K-1}\Big|f^{-1}(c_{k+1})-f^{-1}(c_k)\Big|\end{aligned}$$

Then combine all the inequalities above to get

$$\begin{aligned}\sup_{c\in[a,b]} n^{1/3}\Big|U_n(c)-f^{-1}(c)\Big| &\lesssim \max_{0\le k\le K} n^{1/3}\Big|U_n(c_k)-f^{-1}(c_k)\Big| + \frac{n^{1/3}(b-a)}{K}\\ &\lesssim \max_{0\le k\le K} n^{1/3}\Big|U_n(c_k)-f^{-1}(c_k)\Big| + \tilde{c}\end{aligned}$$

where $\tilde{c}$ is a large positive constant. Now, for $y_n > K_1\left(\frac{\log n}{n^{1/12}}\right)^{1/2}$, we can apply the previous lemma.

$$\begin{aligned}&P\left(\sup_{c\in[a,b]} n^{1/3}|U_n(c)-f^{-1}(c)| > y_n\right)\\ &\le P\left(\max_{0\le k\le K} n^{1/3}|U_n(c_k)-f^{-1}(c_k)| > y_n-\tilde{c}\right)\\ &\le \sum_{k=0}^{K} P\left(n^{1/3}|U_n(c_k)-f^{-1}(c_k)| > y_n-\tilde{c}\right)\\ &\lesssim (K+1)\left(\frac{n^{-1/3}}{y_n^8}+e^{-y_n^3}\right)\\ &\lesssim n^{1/3}(b-a)\left(\frac{n^{-1/3}}{y_n^8}+e^{-y_n^3}\right)\end{aligned}$$

Lemma A.2.9 *Define* S_n

$$S_n = n^{1/3}\sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} A(a)|U_n(a)-f^{-1}(a)|$$

where

$$A(a) = \left[\frac{f'(f^{-1}(a))^2}{4L'(f^{-1}(a))}\right]^{1/3}$$

Let $0 \leq u < v \leq 1$ fixed, and let α_n, β_n be sequences such that $\alpha_n, \beta_n \to 0$ and $0 \leq u + \alpha_n < v - \beta_n \leq 1$ for n sufficiently large. Under assumption 2.2.1,

$$S_n = \sup_{a \in [f(u+\alpha_n), f(v-\beta_n)]} \frac{A(a)}{L'(f^{-1}(a))}|V_n(a)| + O_p(n^{-1/3}(\log n)^{-2/3})$$

where

$$V_n(a) = n^{1/3}\{L(U_n(a)) - L(f^{-1}(a))\}$$

This lemma is the simplification of Durot et al (2012) supplement lemma 6.6 under some modification. For simplicity, define $s_n = u + \alpha_n$, $t_n = v - \beta_n$, and $J_n = \{a : f(s_n) \leq a \leq f(t_n)\}$. Note that L is twice differentiable, L',L'' are bounded. For all $\mathbb{R}$, there exists θ_a between $g(a)$ and $U_n(a)$ such that

$$\begin{aligned} &L(U_n(a)) - L(f^{-1}(a)) \\ &\quad = L'(f^{-1}(a))(U_n(a) - f^{-1}(a)) + \frac{L''(\theta_a)}{2}(U_n(a) - f^{-1}(a))^2 \end{aligned}$$

From the previous lemma, we will derive

$$\sup_{a \in \mathbb{R}} |U_n(a) - f^{-1}(a)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$

It suffices to show that for all $\epsilon > 0$, there exists $M < \infty$ such that satisfies

$$\lim_{n\to\infty} P\left(\sup_{a\in\mathbb{R}} \left(\frac{n}{\log n}\right)^{1/3} \Big|U_n(a) - f^{-1}(a)\Big| > M\right) < \epsilon$$

Apply previous lemma under assuming $M(\log n)^{1/3} > K\left(\frac{\log n}{n^{1/12}}\right)^{1/2}$ for some $K > 0$. Note that we can bound the support of a as $[f(0), f(1)]$. Then we have

$$\begin{aligned} &\lim_{n\to\infty} P\left(\sup_{a\in\mathbb{R}} \left(\frac{n}{\log n}\right)^{1/3} \Big|U_n(a) - f^{-1}(a)\Big| > M\right) \\ &\qquad \lesssim \lim_{n\to\infty} n^{1/3}\|f\|_\infty\left(\frac{n^{-1/3}}{M^8(\log n)^{8/3}} + e^{-M^3 \log n}\right) \end{aligned}$$

Then we can choose $M \geq K_1 \left(\frac{1}{3}\right)^{1/3}$ for some $K_1 > 0$ to make the probability be less than ϵ. By using the fact, we can derive the first order approximation of $L(\cdot)$ function centered on $f^{-1}(a)$.

$$\sup_{a\in\mathbb{R}} |L(U_n(a)) - L(f^{-1}(a)) - L'(f^{-1}(a))(U_n(a) - f^{-1}(a))| = O_p(n^{-2/3}(\log n)^{2/3})$$

Note that $\inf_{s\in[0,1]} L'(s) > 0$ and $A(\cdot)$ is bounded. Then

$$S_n = n^{1/3} \sup_{a\in[f(s_n),f(t_n)]} A(a)\frac{|L(U_n(a)) - L(f^{-1}(a))|}{L'(f^{-1}(a))} + O_p(n^{-1/3}(\log n)^{-2/3})$$

Lemma A.2.10 *Let $V_n(a)$ defined as in the previous lemma. Then under assumption 2.2.1, $V_n(a)$ has a different representation*

$$V_n(a) = -\underset{s\in I_n(a)}{argmax}\{-W_{f^{-1}(a)}(s) - D_n(a,s) - R_n(a,s)\}$$

where

$$I_n(a) = [n^{1/3}(L(0) - L(f^{-1}(a))), n^{1/3}(L(1) - L(f^{-1}(a)))]$$

$$W_{f^{-1}(a)}(s) = n^{1/6}\{B_n(L(f^{-1}(a)) + sn^{-1/3}) - B_n(L(f^{-1}(a)))\}$$

$$D_n(a,s) = n^{2/3}\{(\Lambda \circ L^{-1} - aL^{-1})(L(f^{-1}(a)) + sn^{-1/3}) - (\Lambda(f^{-1}(a)) - af^{-1}(a))\}$$

$$R_n(a,s) = n^{2/3}\{(\Lambda_n \circ L^{-1} - \Lambda \circ L^{-1} - n^{-1/2}B_n)(L(f^{-1}(a)) + sn^{-1/3}) - (\Lambda_n \circ L^{-1} - \Lambda \circ L^{-1} - n^{-1/2}B_n)(L(f^{-1}(a)))$$

V_n is defined as

$$V_n(a) = n^{1/3}\{L(U_n(a)) - L(f^{-1}(a))\}$$

By using the definition of $U_n(a)$

$$U_n(a) = \underset{s\in[0,1]}{argmin}\{\Lambda_n(s) - as\}$$

$L(\cdot)$ is monotonic increasing, so we can rewrite $L(U_n(a))$ as

$$L(U_n(a)) = _{s\in[L(0),L(1)]}\{\Lambda_n(L^{-1}(s)) - aL^{-1}(s)\}$$

Then,

$$V_n(a) = \underset{s\in I_n(a)}{argmin}\{\Lambda_n(L^{-1}(L(f^{-1}(a)) + sn^{-1/3})) - aL^{-1}(L(f^{-1}(a)) + sn^{-1/3})\}$$

The location of the minimum of the process is invariant under addition of constants or multiplication by $n^{2/3}$. Hence, for all $a \in \mathbb{R}$,

$$\begin{aligned} V_n(a) &= \underset{s\in I_n(a)}{argmin}\, n^{2/3}\{\Lambda_n(L^{-1}(L(g(a)) + sn^{-1/3})) - aL^{-1}(L(f^{-1}(a)) + sn^{-1/3})\} \\ &= \underset{s\in I_n(a)}{argmin}\{W_{f^{-1}(a)} + D_n(a,s) + R_n(a,s)\} \end{aligned}$$

Lemma A.2.11 *Let assumption 2.2.1 hold. Define* $\tilde{V}_n(a)$,

$$\tilde{V}_n(a) = -\underset{s\in I_n(a):|s|\le\log n}{argmax}\{-W_{f^{-1}(a)}(s) - D_n(a,s) - R_n(a,s)\}$$

Let $0 \le u < v \le 1$ *fixed, and let* α_n, β_n *be sequences such that* $\alpha_n, \beta_n \to 0$ *and* $0 \le u+\alpha_n < v-\beta_n \le 1$ *for* n *sufficiently large. Then for any* $b(a) \in \mathbb{R}$ *that satisfies* $|a-b(a)| \le n^{-1/3}(\log n)^2$,

$$\begin{aligned} &\sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(a)}{L'(f^{-1}(a))}|V_n(a)| \\ &= \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(a)}{L'(f^{-1}(a))}|\tilde{V}_n(a)| + o_p((\log n)^{-2/3}) \end{aligned}$$

This is lemma 6.7 in Durot et al (2012) supplement with taking $a^{\xi} = a$.

Lemma A.2.12 *Let assumption 2.2.1 hold. Let* $\epsilon_n = 1/\log n$ *and define*

$$\tilde{V}_n^{+}(a,b) = -\underset{s\in I_n(a):|s|\le\log n}{argmax}\left\{-W_{f^{-1}(a)}(s) - \left(\frac{f'(f^{-1}(b))}{2(L'(f^{-1}(b)))^2} - 2\epsilon_n\right)s^2\right\}$$

$$\tilde{V}_n^{-}(a,b) = -\underset{s\in I_n(a):|s|\le\log n}{argmax}\left\{-W_{f^{-1}(a)}(s) - \left(\frac{f'(f^{-1}(b))}{2(L'(f^{-1}(b)))^2} + 2\epsilon_n\right)s^2\right\}$$

Then

$$S_n \le \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(b(a))}{L'(f^{-1}(b(a)))}|\tilde{V}_n^{+}(a,b(a))| + o_p((\log n)^{-2/3})$$

and

$$S_n \ge \sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]} \frac{A(b(a))}{L'(f^{-1}(b(a)))}|\tilde{V}_n^{-}(a,b(a))| + o_p((\log n)^{-2/3})$$

for any $b(a) \in \mathbb{R}$ *that satisfies* $|a-b(a)| \le n^{-1/3}(\log n)^2$.

This lemma is similar to Durot et al (2012) lemma 6.8. From the previous lemma,

$$\sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]}\frac{A(a)}{L'(f^{-1}(a))}|V_n(a)|$$
$$=\sup_{a\in[f(u+\alpha_n),f(v-\beta_n)]}\frac{A(a)}{L'(f^{-1}(a))}|\tilde{V}_n(a)|+o_p((\log n)^{-2/3})$$

Note that $\frac{A(a)}{L'(f^{-1}(a))}$ is non zero constant. Hence, it suffices to show that with probability one,

$$|V_n(a)|\leq|\tilde{V}_n(a)|+\gamma_n$$

where $\gamma_n=o_p((\log n)^{-2/3})$. The first step is to show that $D_n(a,s)$ can be approximated by $\frac{f'(f^{-1}(a))}{2L'(f^{-1}(a))^2}s^2$. Let $H_a=\Lambda\circ L^{-1}-aL^{-1}$. Then, we can derive H_a' and H_a''.

$$H_a'(s)=\frac{f(L^{-1}(s))}{L'(L^{-1}(s))}-\frac{a}{L'(L^{-1}(s))}$$
$$H_a''(s)=\frac{f(L^{-1}(s))}{[L'(L^{-1}(s))]^2}-\frac{f(L^{-1}(s))L''(L^{-1}(s))}{[L'(L^{-1}(s))]^3}+\frac{aL''(L^{-1}(s))}{[L'(L^{-1}(s))]^3}$$

By applying Taylor expansion centered with $L(f^{-1}(a))$, for all $a\in[f(0),f(1)]$ and $s\in I_n(a)$,

$$D_n(a,s)=\frac{f'(f^{-1}(a))}{2[L'(f^{-1}(a))]^2}s^2+\frac{1}{2}\{H_a''(\theta_{a,s})-H_a''(L(f^{-1}(a)))\}s^2$$

for some $\theta_{a,s}$ such that $|\theta_{a,s}-L(f^{-1}(a))|\leq n^{-1/3}|s|$. Note that $W_{f^{-1}(a)}(s)$ is Brownian motion which is symmetric with respect to zero. So $W_{f^{-1}(a)}(s)$ and $-W_{f^{-1}(a)}(s)$ has the same distribution. For all $b\in\mathbb{R}$ that satisfies $|a-b|\leq n^{-1/3}(\log n)^2$,

$$|\tilde{V}_n(a)|=\left|-\underset{s\in I_n(a):|s|\leq\log n}{argmax}\{-W_{f^{-1}(a)}(s)-D_n(a,s)-R_n(a,s)\}\right|$$
$$=\left|\underset{s\in I_n(a):|s|\leq\log n}{argmax}\{W_{f^{-1}(a)}(s)-D_n(a,s)-R_n(a,s)\}\right|$$
$$=\left|\underset{s\in I_n(a):|s|\leq\log n}{argmax}\left\{W_{f^{-1}(a)}(s)-\left(\frac{f'(g(b))}{2[L'(g(b))]^2}-2\epsilon_n\right)s^2+h_n(a,b,s)\right\}\right|$$

where

$$h_n(a,b,s)=h_{1n}(a,b,s)+h_{2n}(a,s)$$
$$h_{1n}(a,b,s)=-D_n(a,s)+\frac{f'(f^{-1}(b))}{2[L'(f^{-1}(b))]^2}s^2-\epsilon_n s^2$$
$$h_{2n}(a,s)=-R_n(a,s)-\epsilon_n s^2$$

By lemma 4.1 in Durot et al (2012), it suffices to show that with probability one, for all $|t| < |s| \leq \log n$,

$$h_{1n}(a, b, s) < h_{1n}(a, b, t)$$

$$h_{2n}(a, s) < h_{2n}(a, t)$$

for all $(a, b) \in \mathcal{E}_n$ where
$\mathcal{E}_n = \left\{(a, b) : a \in [f(u + \alpha_n), f(v - \beta_n)], |a - b| \leq n^{-1/3}(\log n)^2\right\}$. By some algebra,

$$\begin{aligned} h_{1n}(a, b, s) &- h_{1n}(a, b, t) \\ &= D_n(a, t) - D_n(a, s) + \epsilon_n(t^2 - s^2) \\ &= \frac{1}{2}[H_a''(\theta_{a,s}) - H_a''(L(f^{-1}(a)))](t^2 - s^2) + \epsilon_n(t^2 - s^2) \end{aligned}$$

Under the assumptions on (f, f'), H_a'' is Lipschitz continuous. Hence,

$$\begin{aligned} P(h_{1n}(a, b, s) &- h_{1n}(a, b, t) < 0) \\ &= P\left(\frac{1}{2}[H_a''(\theta_{a,s}) - H_a''(L(f^{-1}(a)))](t^2 - s^2) + \epsilon_n(t^2 - s^2) < 0\right) \\ &= P\left(O_p(n^{-1/3}\log n) < \epsilon_n\right) \to 1 \end{aligned}$$

Similarly,

$$\begin{aligned} h_{2n}(a, s) &- h_{2n}(a, t) \\ &= R_n(a, t) - R_n(a, s) + \epsilon_n(t^2 - s^2) \\ &= O_p\left(\frac{\log n}{n}\right) + \epsilon_n(t^2 - s^2) \end{aligned}$$

The last equality comes from lemma A.2.6. Then,

$$\begin{aligned} P(h_{2n}(a, s) &- h_{2n}(a, t) < 0) \\ &= P\left(O_p\left(\frac{\log n}{n}\right) < \epsilon_n(s^2 - t^2)\right) \\ &= P\left(O_p\left(\frac{\log n}{n}\right) < \epsilon_n|s - t||s + t|)\right) \\ &\geq P\left(O_p\left(\frac{\log n}{n}\right) < \epsilon_n\gamma_n O_p(\log n)\right) \to 1 \end{aligned}$$

By Durot et al (2012) lemma 4.1, $[f(u+\alpha_n), f(v-\beta_n)] \in \mathbb{R}$. There exsts $\gamma_n = \frac{1}{\log n}$ such that $h(a,b,s) < h(a,b,t)$ for all s,t such that $|s| > |t| + \gamma_n$. Then,

$$\begin{aligned}
|\tilde{V}_n(a)| &= \left| \underset{s \in I_n(a):|s|\le \log n}{argmax} \left\{ W_{f^{-1}(a)}(s) - \left(\frac{f'(f^{-1}(b))}{2[L'(f^{-1}(b))]^2} - 2\epsilon_n \right) s^2 + h_n(a,b,s) \right\} \right| \\
&\le \left| \underset{s \in I_n(a):|s|\le \log n}{argmax} \left\{ W_{f^{-1}(a)}(s) - \left(\frac{f'(f^{-1}(b))}{2[L'(f^{-1}(b))]^2} - 2\epsilon_n \right) s^2 \right\} \right| + \gamma_n \\
&= |\tilde{V}_+(a,b)| + o_p((\log n)^{-2/3})
\end{aligned}$$

Similarly, to show $|\tilde{V}_n(a)| + \gamma_n \ge |\tilde{V}_n^-(a,s)|$ with probability 1, take

$$h_{1n}(a,b,s) = -D_n(a,s) + \frac{f'(f^{-1}(b))}{2[L'(f^{-1}(b))]^2} s^2 + \epsilon_n s^2$$

$$h_{2n}(a,s) = -R_n(a,s) + \epsilon_n s^2$$

Then we can easily show that $h(a,b,s) > h(a,b,t)$ for all s,t such that $|s| + \gamma_n < |t|$ with probability 1. Then apply Durot et al (2012) lemma 4.1 again.

Lemma A.2.13 *Let assumption 2.2.1 hold. Define S_n as above with $0 \le u < v \le 1$, and α_n, β_n as above. Let $\epsilon_n = \frac{1}{\log n}$ and let $\zeta(c)$ be a random variable such that for all $c \in \mathbb{R}$, $\zeta(c) = \underset{t\in\mathbb{R}}{argmax}\{B_n(t+c) - t^2\}$* Let $(\zeta_i(\cdot))_{i\in\mathbb{N}}$ be a sequence of independent processes, all distributed like $\zeta(\cdot)$. Define S_B, $S_B^{(1)}$, $S_A^{(2)}$ as

$$S_B \overset{d}{=} \max_{1\le i\le K_n} \sup_{c\in[0,\Delta_{in}]} |\zeta_i(c)|$$

$$S_B^{(1)} \overset{d}{=} \max_{1\le i\le K_n} \sup_{c\in[0,\Delta_{in}]} |\zeta_i^{(1)}(c)|$$

$$S_A^{(2)} \overset{d}{=} \max_{2\le i\le K_n} \sup_{c\in[0,\delta_{in}]} |\zeta_i^{(2)}(c)|$$

where

(a) $(\zeta_j^{(1)}), (\zeta_j^{(2)})$ are the copies of $(\zeta_i)_{i\in\mathbb{N}}$,

(b) $0 \le \delta_{in} \le C \log n$ for some $C > 0$,

(c) $K_n = \left\{ \frac{f(v-\beta_n) - f(u+\alpha_n)}{l_n + L_n} \right\} - 1$,

(d) $l_n = \frac{2\|f'\|_\infty}{\inf_{s\in[0,1]} L'(s)} n^{-1/3} \log n$, $L_n = 2n^{-1/3}(\log n)^2$,

(e)

$$A_i = [f(u+\alpha_n) + (i-1)(l_n + L_n), f(u+\alpha_n) + il_n + (i-1)L_n],$$

$$B_i = [f(u+\alpha_n) + il_n + (i-1)L_n, f(u+\alpha_n) + i(l_n + L_n)],$$

$$R_n = [f(u+\alpha_n) + K_n(l_n + L_n), f(v-\beta_n)],$$

(f) b_i is the midpoint of the interval B_i for all i, and

(g) $\Delta_{in} = (1+o(1))(\log n)^2 \left\{ \frac{L'(g(b_i))f'(g(b_i))}{2} \right\}^{-1/3}$.

Then

$$S_B \leq \frac{S_n}{1+O(\epsilon_n)} \leq \max\{S_B^{(1)}, S_A^{(2)}\} + o_p((\log n)^{-2/3})$$

This can be proved by applying lemma 6.9 in Durot et al (2012) supplement.

Lemma A.2.14 *Let assumption 2.2.1 hold. Define the ordered jump point of* $\hat{f}$ *as* $(\tau_1 < \cdots < \tau_{N_\tau} - 1)$ *with setting* $\tau_0 = 0$ *and* $\tau_{N_\tau} = 1$*. Then*

$$\max_{1 \leq i \leq N_\tau} |\tau_i - \tau_{i-1}| = O_p\left(\left(\frac{\log n}{n} \right)^{1/3} \right)$$

This lemma is similar to Durot et al (2012) lemma 5.1. Note that $\hat{f}$ and U_n are non-decreasing and right-continuous step functions. Also, the maximal height of the jumps of U_n is the maximal length of the flat parts of $\hat{f}$. Hence,

$$\max_{1 \leq i \leq N_\tau} |\tau_i - \tau_{i-1}| = \sup_{a \in \mathbb{R}} \left| \lim_{b \uparrow a} U_n(b) - U_n(a) \right|$$

By triangular inequality, continuity of f^{-1}, and lemma 6.8,

$$\begin{aligned} \max_{1 \leq i \leq N_\tau} |\tau_i - \tau_{i-1}| &= \sup_{a \in \mathbb{R}} \left\{ \left| \lim_{b \uparrow a} U_n(b) - f^{-1}(a) \right| + \left| U_n(a) - f^{-1}(a) \right| \right\} \\ &\leq 2 \sup_{a \in \mathbb{R}} |U_n(a) - f^{-1}(a)| = O_p\left(\left(\frac{\log n}{n} \right)^{1/3} \right) \end{aligned}$$

Lemma A.2.15 *Let assumption 2.2.1 hold. Then there exist* $C > 0$ *such that*

$$E[|\hat{f}_n(s) - f(s)|] \leq Cn^{-1/3}$$

for all $s \in [n^{-1/3}, 1 - n^{-1/3}]$*, and*

$$E[|\hat{f}_n(s) - f(s)|] \leq C[n(\min\{s, 1-s\})]^{-1/2}$$

for all $s \in (0, n^{-1/3}] \cup [1 - n^{-1/3}, 1)$.

This lemma is a modification of the theorem 1 in Durot (2007). In the first step, we need to show lemma 2 in Durot (2007) which is

$$P\left(\left|U_n(a) - g(a)\right| \geq x\right) \lesssim \frac{1}{nx^3}$$

where $x > 0$ satisfies a certain restriction. As in the proof of lemma 6.7, we have

$$\begin{aligned} &P(|U_n(a) - f^{-1}(a)| > x) \\ &\quad = P\left(\sup_{|u-f^{-1}(a)|>x} \{P_{1n} + P_{2n} + P_{3n} + P_{4n}\} \geq 0\right) \end{aligned}$$

where

$$P_{1n} = -\{\Lambda_n(u) - \Lambda(u) - n^{-1/2} B_n \circ L(u)\}$$

$$P_{2n} = \Lambda_n(f^{-1}(a)) - \Lambda(f^{-1}(a)) - n^{-1/2} B_n \circ L(f^{-1}(a))$$

$$P_{3n} = \Lambda(f^{-1}(a)) - \Lambda(u) - a(f^{-1}(a) - u)$$

$$P_{4n} = n^{-1/2}(B_n \circ L(f^{-1}(a)) - B_n \circ L(u))$$

Then we can rewrite it as below.

$$\begin{aligned} &P(|U_n(a) - f^{-1}(a)| > x) \\ &\quad \lesssim P\left(\sup_{s\in[0,1]} |\Lambda_n(s) - \Lambda(s) - n^{-1/2} B_n \circ L(s)| > c_1 x^2\right) \\ &\quad + P\left(\sup_{|u-f^{-1}(a)|>y_n} \{P_{4n} - c_2(f^{-1}(a) - u)^2\} \geq 0\right) \end{aligned}$$

for some $c_1, c_2 > 0$. The first probability term can be controlled by using the coupling inequality in lemma A.2.6. Hence suppose that there exist $K > 0$ such that satisfies $x^2 n^{3/4} > K \log n$. Then we have

$$\begin{aligned} &P\left(\sup_{s\in[0,1]} |\Lambda_n(s) - \Lambda(s) - n^{-1/2} B_n \circ L(s)| > c_1 x^2\right) \\ &\quad = P\left(n^{1/4} \|M_n - B_n \circ L\|_\infty > c_1 x^2 n^{3/4}\right) \\ &\quad \lesssim \frac{1}{x^8 n^3} \leq \frac{1}{nx^3} \end{aligned}$$

To control the latter term,

$$
\begin{aligned}
&P\left(\sup_{|u-f^{-1}(a)|>x}\{P_{4n}-c_2(f^{-1}(a)-u)^2\}\geq 0\right)\\
&\qquad = P\left(\sup_{|u-f^{-1}(a)|>x} n^{-1/2}\left\{B_n\circ L(f^{-1}(a))-B_n\circ L(u)\right\}-c_2(f^{-1}(a)-u)^2\geq 0\right)\\
&\qquad \leq P\left(\sup_{|u-f^{-1}(a)|>x}\left\{B_n\circ L(f^{-1}(a))-B_n\circ L(u)\right\}\geq c_2x^2n^{1/2}\right)\\
&\qquad \lesssim \sum_{k\geq 0}P\left(\sup_{|u-f^{-1}(a)|\in[x2^{k-1},x2^k]} B_n\circ(L(f^{-1}(a))-L(u))\geq c_3(x2^{k-1})^2n^{1/2}\right)\\
&\qquad \lesssim \sum_{k\geq 0}P\left(\sup_{|u-f^{-1}(a)|\leq x2^k} B_n(f^{-1}(a)-u)\geq c_4(x2^{k-1})^2n^{1/2}\right)\\
&\qquad \lesssim \sum_{k\geq 0}P\left(\sup_{|u-f^{-1}(a)|\leq x2^k} \{B_n(f^{-1}(a)-u)\}^2\geq c_5(x^4 2^{4k})n\right)\\
&\qquad \lesssim \sum_{k\geq 0}\frac{E\left[\{B_n(x2^k)\}^2\right]}{x^4n2^{4k}}=\sum_{k\geq 0}\frac{x2^k}{x^4n2^{4k}}\lesssim\frac{1}{nx^3}
\end{aligned}
$$

On the 4th line, we use the property that $B_n(u-s)\overset{d}{=}B_n(u)-B_n(s)$. The next inequality holds since $L(\cdot)$ is differentiable and bounded for all $s\in(0,1)$. The last inequality can be derived by Doob's sub-martingale inequality. In result, for all $xn^{3/4}>K\log n$, we have

$$P(|U_n(a)-f^{-1}(a)|>x)\lesssim\frac{1}{nx^3}$$

On the second step, we need to derive an alternative condition for lemma 3 in Durot (2007) such that for all $a\notin[f(0),f(1)]$, and under some restriction on $x>0$,

$$P(|U_n(a)-f^{-1}(a)|>x)\lesssim\frac{1}{nx(f(f^{-1}(a)-a))^2}$$

To show this inequality, we borrow the similar idea on the proof of lemma 6.7 again.

$$
\begin{aligned}
&P(|U_n(a)-f^{-1}(a)|>x)\\
&\quad = P\left(\sup_{|u-f^{-1}(a)|>x}\{P_{1n}+P_{2n}+P_{3n}+P_{4n}\}\geq 0\right)
\end{aligned}
$$

Note that $f^{-1}(a)=0$ if $a<f(0)$ and $f^{-1}(a)=1$ if $a>f(1)$. Since $f'(s)$ is not well defined on $s\notin(0,1)$, we cannot use the second order approximation of P_{3n}. Instead, for all $a\notin$

$[f(0), f(1)]$, we can still use the inequality such as $\Lambda(f^{-1}(a)) - \Lambda(u) \leq f(f^{-1}(a))(f^{-1}(a) - u))$. Then with the similar derivation as above, we have

$$
\begin{aligned}
&P\biggl(\sup_{|u-f^{-1}(a)|>x} \{P_{1n} + P_{2n} + P_{3n} + P_{4n}\} \geq 0\biggr) \\
&\quad \lesssim P\biggl(n^{1/4}\|M_n - B_n \circ L\|_\infty > c_6\{f(f^{-1}(a)) - a\}^2 x n^{3/4}\biggr) \\
&\quad + P\biggl(\sup_{|u-f^{-1}(a)|>x} \{B_n(f^{-1}(a) - u)\}^2 > c_7\{f(f^{-1}(a)) - a\}^2 x^2 n\biggr) \\
&\quad \lesssim \frac{1}{\{f(f^{-1}(a)) - a\}^8 x^4 n^3}
\end{aligned}
$$

Lemma A.2.16 *Let assumption 2.2.1 hold. Let the jump points be* $\tau_1 < \tau_2 < \cdots < \tau_{N_\tau}$ *such that satisfies* $\gamma_i = \hat{f}_n(\tau_i)$ *for all* $i = 1, \cdots, N_\tau - 1$. *Let* $i_1(s)$ *and* $i_2(t)$ *be*

$$i_1(s) = \min\{i \in \{0, 1, \cdots, N_\tau\} \quad \textit{such that} \quad \tau_i \geq s\}$$

$$i_2(t) = \max\{i \in \{0, 1, \cdots, N_\tau\} \quad \textit{such that} \quad \tau_i < 1 - s\}$$

Then

$$\tau_i = s + O_p(n^{-1/3}), for\, i = i_1(s) - 1, i_1(s), i_1(s) + 1$$

and

$$\tau_i = 1 - t + O_p(n^{-1/3}), for\, i = i_2(t) - 1, i_2(t), i_2(t) + 1$$

It can be proved by Durot et al (2012) supplement lemma 6.11 under the condition, $\hat{f}(s) = f(s) + O_p(n^{-1/3})$.

Lemma A.2.17 *Suppose that* $0 < s < 1 - t < 1$ *where* $n^{1/3}s \to \infty$, $n^{1/3}t \to \infty$, *and* $n^{1/3}(1 - s - t) \to \infty$. *Then*

$$P(s \leq \tau_{i_1(s)} \leq \tau_{i_2(t)} < 1 - t) \to 1$$

It can be proved by Durot et al (2012) supplement lemma 6.12 under $\hat{f}(s) = f(s) + O_p(n^{-1/3})$ and previous lemma.

Lemma A.2.18 *Let assumption 2.2.1 hold. If all the assumptions on the lemma A.2.16 and A.2.17 hold, then*

$$P(\gamma_i < f(1) \text{ for all } i \le i_2(t)) \to 1$$
$$P(\gamma_i > f(0) \text{ for all } i \ge i_1(s)) \to 1$$

This lemma is similar to Durot et al (2012) supplement lemma 6.13.

$$\begin{aligned}
&P(\gamma_i < f(1) \text{ for all } i \le i_2(t)) \\
&\le P(\hat{f}(\tau_{i_2(t)}) < f(1)) = P(\hat{f}(\tau_{i_2(t)}) - f(\tau_{i_2}(t)) < f(1) - f(\tau_{i_2}(t))) \\
&\lesssim P(O_p(n^{-1/3} < f(1) - f(1-t) + f(1-t) - f(\tau_{i_2(t)})) \\
&\le P(O_p(n^{-1/3}) < \|f'\|_\infty t) \to 1
\end{aligned}$$

Similarly, we can show $P(\gamma_i > f(0) \text{ for all } i \ge i_1(s)) \to 1$.

Lemma A.2.19 *Let assumption 2.2.1 and assumptions in lemma A.2.16 and A.2.17 hold. Then we can show that*

$$\sup_{u \in (s, 1-t]} B(u)|\hat{f}(u) - f(u)| = \max\left\{ \sup_{u \in (\tau_{i_1(s)}, \tau_{i_2(t)}]} B(u)|\hat{f}(u) - f(u)|, O_p(n^{-1/3}) \right\}$$

where $B(u) = [4f'(u)L'(u)]^{-1/3}$.

This is similar to Durot et al (2012) supplement lemma 6.14. Note that $\hat{f}$ is constant on the interval $(s, \tau_{i_1(s)})$ and $(\tau_{i_2(t)}, 1-t)$ by definition. Note that $B(u)$ is bounded for all $u \in [0,1]$. Then by lemma A.2.16 and A.2.17,

$$\begin{aligned}
&\sup_{u \in \tau_{i_2(t)}, 1-t]} B(u)|\hat{f}(u) - f(u)| \\
&\qquad \le \|B\|_\infty |\hat{f}(1-t) - f(1-t)| + \|B\|_\infty |\hat{f}(1-t) - f(\tau_{i_2(t)})| \\
&\qquad \lesssim \|B\|_\infty \{|\hat{f}(1-t) - f(1-t)| + |f(1-t) - f(\tau_{i_2(t)})|\} \\
&\qquad \lesssim \|B\|_\infty \{|\hat{f}(1-t) - f(1-t)| + \|f'\|_\infty |1 - t - \tau_{i_2(t)}|\} = O_p(n^{-1/3})
\end{aligned}$$

Similarly, we can show that

$$\sup_{u \in (s, \tau_{i_1(s)}]} B(u)|\hat{f}(u) - f(u)| \lesssim O_p(n^{-1/3})$$

Lemma A.2.20 *Let assumption 2.2.1 hold. Let* $\alpha_n \geq K_1 n^{-1/3}(\log n)^{-2/3}$ *and* $\beta_n \geq K_2 n^{-1/3}(\log n)^{-2/3}$ *for some* $K_1, K_2 > 0$ *that do not depend on* n*. Then*

$$\sup_{s\in(\alpha_n,1-\beta_n]} |\hat{f}(s) - f(s)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$

This lemma is Durot et al (2012) theorem 2.1. For the simplicity, let $\tau_{i_1} = \tau_{i_1(s_n)}$ and $\tau_{i_2} = \tau_{i_2(t_n)}$ where $\alpha_n = s_n$ and $\beta_n = t_n$. Then we have

$$\begin{aligned}&\sup_{u\in(s_n,1-t_n]} |\hat{f}(u) - f(u)| \\ &\leq \max_{i=i_1,\cdots,i_2} \sup_{u\in[\tau_{i-1},\tau_i)} |\hat{f}(u) - f(u)| + \sup_{u\in(\tau_{i_2},1-t_n]} |\hat{f}(u) - f(u)|\end{aligned}$$

Note that $\hat{f}$ is right-continuous function and constant on the interval $[\tau_{i-1}, \tau_i)]$ for all $i = 1, \cdots, N_\tau - 1$. By the triangular inequality,

$$\begin{aligned}\sup_{u\in[\tau_{i-1},\tau_i)} |\hat{f}(u) - f(u)| &= \sup_{u\in[\tau_{i-1},\tau_i)} |\hat{f}(\tau_{i-1}) - f(u)| \\ &\leq |\hat{f}(\tau_{i-1}) - f(\tau_{i-1})| + \|f'\|_\infty |\tau_i - \tau_{i-1}|\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\sup_{u\in[\tau_{i_2},1-t_n]} |\hat{f}(u) - f(u)| \\ &\leq |\hat{f}(1-t_n) - f(1-t_n)| + \|f'\|_\infty |\tau_{i_2} - \tau_{i_2-1}|\end{aligned}$$

Note that $1 - t_n \in [n^{-1/3}, 1 - n^{-1/3}]$ if n is large enough. Then apply lemma A.2.14 and A.2.15.

$$\begin{aligned}&\sup_{u\in(s_n,1-t_n]} |\hat{f}(u) - f(u)| \\ &\leq \max_{i=i_1,\cdots,i_2} |\hat{f}(\tau_i) - f(\tau_i)| + O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)\end{aligned}$$

Now we define $E_n = \{s_n \leq \tau_{i_1} \leq \tau_{i_2} < 1 - t_n\}$. Then by lemma A.2.16 and A.2.17, $P(E_n) \to 1$ and $\gamma_i = f \circ f^{-1}(\gamma_i)$ for all $i = i_1, \cdots, i_2$. In result,

$$\sup_{u\in(s_n,1-t_n]}|\hat{f}(u)-f(u)| \leq \max_{i=i_1,\cdots,i_2}|\hat{f}(\tau_i)-f(\tau_i)| + O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$
$$= \max_{i=i_1,\cdots,i_2}|\gamma_i - f(U_n(\gamma_i))| + O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$
$$\leq \max_{i=i_1,\cdots,i_2}\|f'\|_\infty|f^{-1}(\gamma_i)-U_n(\gamma_i)| + O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$
$$\leq \|f'\|_\infty \sup_{a\in\mathbb{R}}\left|U_n(a)-f^{-1}(a)\right| + O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right) = O_p\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$

α_n and β_n can be more relaxed by Durot et al (2012) p.1599. Detailed proofs are skipped.

Lemma A.2.21 *Let assumption 2.2.1 and all the assumptions on the lemma A.2.16 hold. Then*

$$\sup_{u\in(s,1-t]} B(u)|\hat{f}(u)-f(u)|$$
$$= \sup_{a\in[f(s),f(1-s)]} A(a)\left|U_n(a)-f^{-1}(a)\right| + O_p\left(\left(\frac{\log n}{n}\right)^{2/3}\right)$$

where $A(a) = \frac{f'(f^{-1}(a))^{2/3}}{\{4L'(f^{-1}(a))\}^{1/3}}$ *and* $B(u) = [4f'(u)L'(u)]^{-1/3}$.

This lemma is similar to Durot et al (2012) lemma 5.2. In Durot et al (2012) on p.1600∼1604, change the argument from decreasing function to increasing function.

APPENDIX B

Implementation of Salanié and Wolak (2019)

Most of the setup will be the similar to Dube, Fox, Su (2012) and Salanié and Wolak (2019). Let T, J, K, and n denote the total number of markets, the number of products, the number of observed product characteristics (other than price), and the number of consumers in each market, respectively. We set $T = 50$, $J = 5$, $K = 3$, and $n = 1000$. The conditional indirect utility of consumer i in market t from purchasing product j is defined as

$$u_{ijt} = \beta_{i1} + X'_{jt}\beta_i^x - \beta_i^p p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where the utility of the "outside" good, $u_{i0t} = \epsilon_{i0t}$. The random coefficients, β_{ik}'s, are drawn independently from the normal distribution with mean $\bar{\beta}_k$ and variance σ_k^2 where $E[\beta_{ik}] = \bar{\beta}_k$ and $\text{Var}(\beta_{ik}) = \sigma_k^2$. β_i^x indicates $(\beta_{i2}, \beta_{i3}, \beta_{i4})'$ and $\beta_i^p = \beta_{i5}$. We set the true parameters $\bar{\beta} = (\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4, \bar{\beta}_5)' = (-1, 1.5, 1.5, 0.5, -1)'$, and $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2) = (0.5, 0.5, 0.5, 0.5, 0.2)$. The x's are observed product characteristics generated from the multivariate normal distribution as follows:

$$\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \end{pmatrix} \sim N\left(\begin{bmatrix} 2j^2/25 - 1 \\ 2j^2/25 - 1 \\ 2j^2/25 - 1 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix} \right)$$

The ξ_{jt} is the j product specific characteristic in market t, and normally distributed with mean 0 for all j and t. In order to avoid the weak IV problem complicating the analysis, we assume the price to be exogenous. Hence, the price is generated by

$$p_{jt} = |e_{jt} + 1.1(x_{1j} + x_{2j} + x_{3j})|, \qquad \forall j = 1, \cdots, 5 \quad \forall t = 1, \cdots, 50$$

where e_{jt} is standard normal distribution random variable for all j and t independently. To make the notation simpler, define $X_t = (X'_{1t}, \cdots, X'_{Jt})'$, $\xi_t = (\xi_{1t}, \cdots, \xi_{Jt})'$, $p_t = (p_{1t}, \cdots, p_{Jt})'$, and $\theta = (\bar{\beta}_1'^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2)'$. The individual's unobserved characteristics, ϵ_{ijt}'s are assumed to have Type-I extreme value distribution, so we can compute the probability that consumer i purchases good j in market t by using logit model.

$$s_{ijt}\left(X_t, p_t, \xi_t \middle| \beta_i\right) = \frac{\exp(\beta_{i1} + X'_{jt}\beta_i^x - \beta_i^p p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta_{i1} + X'_{kt}\beta_i^x - \beta_i^p p_{kt} + \xi_{kt})}.$$

Then we can generate the observed market share of good j in market t is calculated as the average of $s_{ijt}\left(X_t, p_t, \xi_t \middle| \beta_i\right)$ over $i = 1, \ldots, n$. Given the exogeneity assumption, the natural IV is $(1, x_{1j}, x_{2j}, x_{3j}, p_j)$, which is adopted in our implementation of the MPEC. he number of Monte Carlo simulation is $M = 1,000$. For each simulation, we choose the initial value as the true parameter which is infeasible in practice.

To derive Salanié and Wolak's estimator, we follow section 7.2 in Salanié and Wolak (2019). Salanié and Wolak's (2019) approximate model is the linear model with the dependent variable $y_{jt} = \log\left(S_{jt}/\, S_{0t}\right)$ and explanatory variable $1, x_1, x_2, x_3, p$ along with the artificial regressors described in their Theorem 2. We estimated the linear model is estimated by 2SLS with IV $1, x_1, x_2, x_3, p$ as well as their second degree polynomials including all the interactions. As in Salanié and Wolak (2019, Section 7.2), if any coefficient for the last four variables is negative, we set that coefficient to 0 and rerun the regression without that variable. We iterate this process until all the coefficients are positive, or all four variables are excluded from the regression.

As for the misspecified model, all other things are the similar to DGP 1 case except that we choose the top five products out of $J = 25$, and work with them. Compute the observed market shares for each market and average market shares for each good j. We pick top 5 products whose average market shares are bigger than other goods. Then we estimate the MPEC estimator and the Salanié and Wolak estimator with these 5 products only. All the other Monte Carlo setting is the same as the full specification model case.

APPENDIX C

Tables Related to Chapter 3

Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.0059007	0.00076029	-0.00039608	-0.00353805
Mean Bias	0.02053342	0.00339363	-0.00135234	-0.00299598
RMSE	0.56017176	0.5433295	0.52646188	0.5163012
MAE	0.41672348	0.3943496	0.37919647	0.37666595
Interquartile Range	0.64512991	0.59407214	0.56953993	0.56707552
Geary's Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	3.38695638	3.50646605	3.59341872	3.56975641
Mean Bias	2.62125886	1.94946088	-3.64213812	5.97050425
RMSE	634.6141964	1200.487573	658.4733105	168.9616274
MAE	32.47634986	35.83359323	27.91168193	19.16115362
Interquartile Range	3.84986346	0.59407214	3.6049031	3.46211157

Table C.1: Erros in Variables. x^* is Lognormal mean 2 variance 1, ϵ is Conditionally Normal, v is Normal

Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.56593994	-0.32586635	-0.15533879	-0.05701099
Mean Bias	-0.55534234	-0.30880789	-0.13823101	-0.03739943
RMSE	0.56832234	0.33987435	0.20169069	0.16036092
MAE	0.55575617	0.31623643	0.17250103	0.12582738
Interquartile Range	0.14863846	0.17173319	0.17789519	0.18670953
Geary's Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.51325992	0.47657559	0.45248973	0.45854324
Mean Bias	0.96081907	0.57474091	0.54343496	0.53674692
RMSE	14.22912204	1.05156461	0.7511413	0.73044273
MAE	1.56116272	0.63568127	0.5721476	0.56093292
Interquartile Range	0.91526922	0.6805154	0.59647641	0.58267553

Table C.2: Erros in Variables. x^* is Lognormal mean 2 variance 1, ϵ is Conditionally exponential, v is Exponential

Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.51077675	-0.2173441	-0.08218083	-0.03775122
Mean Bias	-0.50168868	-0.20167067	-0.06169185	-0.01378808
RMSE	0.51710169	0.240458	0.14555036	0.13459034
MAE	0.50338678	0.21671581	0.11602989	0.09744076
Interquartile Range	0.14992975	0.14661187	0.14310532	0.14298898
Geary's Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.30337322	0.27056077	0.2594931	0.25547228
Mean Bias	0.3387876	0.45793172	0.43525852	0.42273372
RMSE	29.80462094	0.90963741	0.79557016	0.775105
MAE	1.18010251	0.51131818	0.47637143	0.45934885
Interquartile Range	0.7932881	0.58790664	0.56083893	0.53654966

Table C.3: Erros in Variables. x^* is Lognormal mean 2 variance 1, ϵ is Conditionally Lognormal, v is Lognormal

OLS y on x^2				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.66774573	-0.48254914	-0.30719512	-0.14486135
Mean Bias	-0.66602503	-0.4813456	-0.30640648	-0.14497245
RMSE	0.66705349	0.48413019	0.31285023	0.16142035
MAE	0.66602503	0.4813456	0.30640648	0.14606273
Interquartile Range	0.04911076	0.07011052	0.08445887	0.09708419
IV Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.44674715	-0.27532724	-0.15235236	-0.06399066
Mean Bias	-0.44534586	-0.27391852	-0.15184726	-0.06384898
RMSE	0.44712235	0.27853358	0.16292023	0.09010106
MAE	0.44534586	0.27391852	0.15221225	0.07427426
Interquartile Range	0.05335457	0.06826231	0.08054825	0.08666268
Amemiya Corrected Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.0797564	0.38038877	0.60020237	0.76977432
Mean Bias	0.10225365	0.41417161	0.6471922	0.82038181
RMSE	0.18834066	0.46565072	0.69856512	0.87049724
MAE	0.13578402	0.41420406	0.6471922	0.82038181
Interquartile Range	0.05335457	0.06826231	0.08054825	0.08666268
Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.00590192	0.00266413	0.00138948	0.00269225
Mean Bias	0.03431635	0.01488819	0.00711491	0.00358145
RMSE	0.25789363	0.20279282	0.18335521	0.17018096
MAE	0.19747258	0.15946337	0.14494053	0.13409058
Interquartile Range	0.32520042	0.26757543	0.24271422	0.21998333

Table C.4: Nonlinear EIV. x^* is Normal, ϵ is Conditionally Normal, v is Normal

OLS y on x^2				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.7479597	-0.51671439	-0.30719512	-0.14486135
Mean Bias	-0.75088258	-0.51788976	-0.30640648	-0.14497245
RMSE	0.75264114	0.52022934	0.31285023	0.16142035
MAE	0.75088258	0.51788976	0.30640648	0.14606273
Interquartile Range	0.06769558	0.06425112	0.08445887	0.09708419
IV Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.46252604	-0.29254301	-0.15235236	-0.06399066
Mean Bias	-0.46220701	-0.29171937	-0.15184726	-0.06384898
RMSE	0.4633994	0.29431769	0.16292023	0.09010106
MAE	0.46220701	0.29171937	0.15221225	0.07427426
Interquartile Range	0.04468101	0.05287052	0.08054825	0.08666268
Amemiya Corrected Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.45512922	0.17273369	0.00796434	0.11893447
Mean Bias	0.71202971	0.24338289	0.02083869	0.14667907
RMSE	1.15305214	0.40524683	0.08423576	0.21657233
MAE	0.71565705	0.25029957	0.05239403	0.14795398
Interquartile Range	0.64206953	0.23798145	0.07605631	0.11470033
Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.02580142	0.00446472	0.00138948	-0.00570922
Mean Bias	0.05457943	0.01888276	0.00711491	0.00203709
RMSE	0.26569576	0.16542408	0.18335521	0.0978058
MAE	0.19252873	0.12696627	0.14494053	0.07518445
Interquartile Range	0.29883803	0.20609076	0.24271422	0.12201324

Table C.5: Nonlinear EIV. x^* is Normal, ϵ is Conditionally Lognormal, v is Lognormal

OLS y on x^2				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.50126498	-0.30591465	-0.17012242	-0.07224074
Mean Bias	-0.48704139	-0.29517029	-0.16204472	-0.06799991
RMSE	0.4960967	0.30817462	0.17993787	0.09761337
MAE	0.4871255	0.29586717	0.16584607	0.08226621
Interquartile Range	0.11230046	0.10872575	0.09645099	0.08716481
IV Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.29425488	-0.15240263	-0.07496064	-0.02965799
Mean Bias	-0.28132842	-0.14279608	-0.06862583	-0.0266858
RMSE	0.29827652	0.16676118	0.10220636	0.07307715
MAE	0.28331316	0.15050095	0.0862621	0.0576126
Interquartile Range	0.12251059	0.10624908	0.09464238	0.0838979
Amemiya Corrected Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.01963609	0.01556076	0.02881465	0.03461751
Mean Bias	0.00452011	0.03279451	0.03953529	0.04089205
RMSE	0.17019799	0.13422275	0.10782767	0.09137732
MAE	0.12593831	0.09848054	0.08020462	0.06844704
Interquartile Range	0.19791128	0.15746291	0.12283809	0.09818892
Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.00627275	0.00149063	0.0018266	0.00007712
Mean Bias	0.0200356	0.00931974	0.005892	0.00142527
RMSE	0.1617006	0.12325474	0.1007321	0.08515778
MAE	0.12472022	0.09613137	0.07848716	0.06645545
Interquartile Range	0.20401244	0.15804816	0.12945255	0.10849898

Table C.6: Nonlinear EIV. x^* is LogNormal, ϵ is Conditionally Normal, v is Normal

OLS y on x^2				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.59741813	-0.32887623	-0.16956284	-0.06993126
Mean Bias	-0.58931121	-0.32551008	-0.16815382	-0.06907677
RMSE	0.59938662	0.33774672	0.18025816	0.0812769
MAE	0.58931121	0.32563379	0.16862681	0.07178086
Interquartile Range	0.13536983	0.11664459	0.08585688	0.05563245
IV Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.325791	-0.17131089	-0.08654402	-0.03501108
Mean Bias	-0.31688135	-0.16765803	-0.08408385	-0.03341395
RMSE	0.33104973	0.18415136	0.10143134	0.05185354
MAE	0.317191	0.16915982	0.0884671	0.04293441
Interquartile Range	0.12402853	0.10171224	0.07510554	0.05135329
Amemiya Corrected Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.44166482	0.10102046	0.02560862	0.00934375
Mean Bias	0.6264693	0.13706725	0.0355928	0.01459266
RMSE	0.99985043	0.25442999	0.09452867	0.05599176
MAE	0.63511139	0.16062314	0.06501722	0.03909061
Interquartile Range	0.53705289	0.18796216	0.09450349	0.06000326
Small Sigma Estimator				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.02257894	0.01225509	0.00364651	-0.00145249
Mean Bias	0.02686493	0.01267295	0.00558811	0.00074829
RMSE	0.16952677	0.11063062	0.07862495	0.05327558
MAE	0.13079945	0.08678314	0.06130857	0.04144279
Interquartile Range	0.20816054	0.14445049	0.10021287	0.0685468

Table C.7: Nonlinear EIV. x^* is Lognormal, ϵ is Conditionally LogNormal, v is Lognormal

x is Normal				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.00598048	-0.00325592	-0.00142989	-0.00056369
Mean Bias	-0.0011145	-0.00100048	-0.00048711	-0.00003059
RMSE	0.09330586	0.06930314	0.05021091	0.0322496
MAE	0.07344119	0.05462343	0.03982862	0.02553086
Interquartile Range	0.12110774	0.09087639	0.06701781	0.04272559
x is Lognormal				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.00392041	0.00216836	0.00098878	0.00053627
Mean Bias	0.04764268	0.0288944	0.01391341	0.0051318
RMSE	0.19121108	0.1367126	0.08752681	0.04754973
MAE	0.11629754	0.08151423	0.05314492	0.03131034
Interquartile Range	0.14182684	0.10054595	0.06919837	0.04374255

Table C.8: Panel Cubic.

x is Normal				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.03802318	-0.0122674	-0.00632409	-0.00182573
Mean Bias	-0.02235625	-0.00509132	-0.0020255	-0.00065373
RMSE	0.17058624	0.12167038	0.08819034	0.05727959
MAE	0.13370623	0.09524048	0.06933336	0.04487821
Interquartile Range	0.20968719	0.15489932	0.11502123	0.07317141
x is Lognormal				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.00286273	0.00091688	0.00054648	0.00094852
Mean Bias	0.12625453	0.08859159	0.05779883	0.02998001
RMSE	0.38082646	0.30058958	0.22424949	0.14524774
MAE	0.24158203	0.18230822	0.13122422	0.08154466
Interquartile Range	0.32552011	0.22521886	0.15593321	0.09779762

Table C.9: Panel Quartic.

x is Normal				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.18984658	-0.12064102	-0.0311165	0.02066977
Mean Bias	-0.17300443	-0.06300058	0.02795941	0.07959617
RMSE	0.25523411	0.20548907	0.19472567	0.2172741
MAE	0.21064286	0.17781417	0.15649797	0.15947687
Interquartile Range	0.28904859	0.29102961	0.28185638	0.27989457
x is Lognormal				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.20029394	-0.10971856	-0.02597325	0.02136421
Mean Bias	-0.13853082	-0.01496322	0.07628208	0.12980679
RMSE	0.28087786	0.2641336	0.28003166	0.31070715
MAE	0.24011425	0.22275249	0.21236031	0.21872283
Interquartile Range	0.35211116	0.37324621	0.38003962	0.39127706

Table C.10: Panel Probit.

CMLE				
Var(v)	1	0.5	0.25	0.1
Median Bias	0.00205078	0.00097656	0.00439453	0.00039063
Mean Bias	0.00708029	0.0069249	0.00792036	0.0060149
RMSE	0.1050914	0.1030499	0.09938424	0.09885032
MAE	0.08317732	0.081374	0.07893228	0.07833572
Interquartile Range	0.13857422	0.13603516	0.13193359	0.13085938
Small Sigma (NLS)				
Var(v)	1	0.5	0.25	0.1
Median Bias	-0.39885633	-0.17940823	-0.07612222	-0.04940441
Mean Bias	-0.4376699	-0.15068319	-0.00298548	0.02350013
RMSE	0.50020954	0.32101035	0.26181267	0.25082368
MAE	0.44354135	0.26746249	0.21537681	0.20054547
Interquartile Range	0.38306654	0.31581093	0.36869457	0.35162067

Table C.11: Panel Logit. x is normal.

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	0.00320982	0.00410643	0.00247197	0.00213128	-0.00061482
Mean Bias	0.00327289	-0.00118299	-0.00117452	0.00127605	-0.00087011
RMSE	0.14336127	0.09266291	0.09309032	0.10218564	0.03942856
MAE	0.11513557	0.07463456	0.07483739	0.08134004	0.0311196
Interquartile Range	0.1980887	0.12878413	0.12274048	0.13911703	0.05244731
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.21836539	-0.1281617	-0.12630201	-0.1683262	0.21368087
Mean Bias	-0.23191272	-0.12836203	-0.12660798	-0.16775433	0.20947628
RMSE	0.32712378	0.16400237	0.1643049	0.2049105	0.22464712
MAE	0.25495421	0.13827713	0.13947236	0.17815839	0.21108042
Interquartile Range	0.25767255	0.14241261	0.13681291	0.14911396	0.10724794

Table C.12: BLP. $Var(\xi) = 1$. Correct Specification (5 Products)

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-1.2015194	-0.13770422	-0.14096711	-0.19971165	0.16033575
Mean Bias	-1.20012674	-0.14508589	-0.14431966	-0.20060055	0.16181497
RMSE	1.22015924	0.17614525	0.17495241	0.23159645	0.18514548
MAE	1.20012674	0.14989246	0.15055837	0.20420278	0.16474448
Interquartile Range	0.30264835	0.13859443	0.13810483	0.15528841	0.11987642
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.44482452	-0.21671778	-0.21203814	-0.24366902	0.15598143
Mean Bias	-0.41651366	-0.20362023	-0.20116452	-0.23244942	0.15325438
RMSE	0.52005419	0.24395978	0.24288293	0.29208016	0.17649116
MAE	0.4613044	0.21787159	0.21377749	0.25344871	0.15766659
Interquartile Range	0.27495204	0.1607882	0.1844804	0.22560219	0.1187552

Table C.13: BLP. $Var(\xi) = 1$. Misspecification (5 out of 25 Products). Average Total Share = 0.56252245

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.20506335	-0.13574494	-0.13280326	-0.17548911	0.21744681
Mean Bias	-0.22207401	-0.13443309	-0.13334403	-0.17361629	0.21372628
RMSE	0.28941384	0.15433174	0.15435799	0.19504274	0.22261266
MAE	0.23058396	0.13662437	0.1372282	0.17612091	0.21393411
Interquartile Range	0.19939628	0.1072708	0.09816325	0.11216283	0.08611314
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.20506335	-0.13574494	-0.13280326	-0.17548911	0.21744681
Mean Bias	-0.22207401	-0.13443309	-0.13334403	-0.17361629	0.21372628
RMSE	0.28941384	0.15433174	0.15435799	0.19504274	0.22261266
MAE	0.23058396	0.13662437	0.1372282	0.17612091	0.21393411
Interquartile Range	0.19939628	0.1072708	0.09816325	0.11216283	0.08611314

Table C.14: BLP. $Var(\xi) = 0.5$. Correct Specification (5 Products)

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-1.23094782	-0.14669893	-0.14965234	-0.20158297	0.16797169
Mean Bias	-1.23077349	-0.15106099	-0.15081505	-0.20507029	0.16743647
RMSE	1.24274868	0.16930013	0.16870399	0.22411429	0.18241124
MAE	1.23077349	0.15209501	0.1523521	0.20581631	0.16815145
Interquartile Range	0.23799566	0.10837596	0.10326225	0.12065542	0.09706861
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.42262488	-0.2286124	-0.22754145	-0.24767152	0.16565774
Mean Bias	-0.40002089	-0.21624104	-0.21402544	-0.23290066	0.16172903
RMSE	0.48724347	0.24674092	0.24560652	0.28058986	0.18098031
MAE	0.43819176	0.22521259	0.22107006	0.24673006	0.16482162
Interquartile Range	0.24193754	0.14331607	0.16029932	0.20552968	0.1105574

Table C.15: BLP. $Var(\xi) = 0.5$. Misspecification (5 out of 25 Products). Average Total Share = 0.55897623

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	0.0010128	0.00129972	0.0007823	0.00067123	-0.00017931
Mean Bias	0.00101821	-0.00037422	-0.00037142	0.0004009	-0.00027269
RMSE	0.04525624	0.02929754	0.02943317	0.0323096	0.01245354
MAE	0.03634602	0.02359739	0.02366179	0.02571804	0.00982886
Interquartile Range	0.06252715	0.04072018	0.03877752	0.04397514	0.0165419
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.18918418	-0.14056245	-0.14363304	-0.18067884	0.21981331
Mean Bias	-0.20160012	-0.14053435	-0.14034675	-0.18013461	0.21772261
RMSE	0.23471013	0.14779609	0.14782882	0.18829451	0.22148601
MAE	0.20190523	0.14053846	0.14046197	0.18021244	0.21772261
Interquartile Range	0.12622517	0.06541273	0.06067378	0.0717111	0.05474974

Table C.16: BLP. $Var(\xi) = 0.1$. Correct Specification (5 Products)

MPEC					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-1.25329744	-0.15554562	-0.15820311	-0.20841095	0.17332504
Mean Bias	-1.25608526	-0.15584194	-0.15616071	-0.20925829	0.17138933
RMSE	1.26166717	0.16385952	0.16405966	0.21868453	0.17949418
MAE	1.25608526	0.15584194	0.15622153	0.20925829	0.17141752
Interquartile Range	0.15844844	0.070389	0.06735977	0.08348667	0.06864748
Salanie Wolak					
Estimator	Intercept	β_{x1}	β_{x1}	β_{x1}	β_p
Median Bias	-0.40733109	-0.24075116	-0.23855063	-0.24854034	0.177201
Mean Bias	-0.39695824	-0.22646207	-0.22504928	-0.23349533	0.16885883
RMSE	0.44863975	0.24919278	0.2485182	0.27001805	0.18434562
MAE	0.41856364	0.23046657	0.2289629	0.24078397	0.17038986
Interquartile Range	0.190428	0.13431571	0.14206482	0.1819861	0.1096493

Table C.17: BLP. $Var(\xi) = 0.1$. Misspecification (5 out of 25 Products). Average Total Share = 0.55611005