Title
Essays on Econometrics

Author
Shi, Ruoyao

Publication Date
2017

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Econometrics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Ruoyao Shi

2017

ABSTRACT OF THE DISSERTATION

Essays on Econometrics

by

Ruoyao Shi

Doctor of Philosophy in Economics

University of California, Los Angeles, 2017

Professor Jinyong Hahn, Co-Chair

Professor Zhipeng Liao, Co-Chair

This dissertation studies econometric questions in the context of three different methods that are frequently used by empirical economists.

Chapter 1 provides a short introduction to the contexts, questions, methods and results studied in Chapter 2 to Chapter 4.

Chapter 2 studies a nonparametric hedonic equilibrium model in which certain product characteristics are unobserved. Unlike most previously studied hedonic models, both the observed and unobserved agent heterogeneities enter the structural functions nonparametrically. Prices are endogenously determined in equilibrium. Using both within-market and cross-market price variation, I show that all the structural functions of the model are nonparametrically identified up to normalization. In particular, the unobserved product quality function is identified if the relative prices of the agent characteristics differ in at least two

markets. Following the constructive identification strategy, I provide easy-to-implement series minimum distance estimators of the structural functions and derive their uniform rates of convergence. To illustrate the estimation procedure, I estimate the unobserved efficiency of American full-time workers as a function of age and unobserved ability.

Chapter 3 studies the averaging GMM estimator that combines a conservative GMM estimator based on valid moment conditions and an aggressive GMM estimator based on both valid and possibly misspecified moment conditions, where the weight is the sample analog of an infeasible optimal weight. We establish asymptotic theory on uniform approximation of the upper and lower bounds of the finite-sample risk difference between two estimators, which is used to show that the averaging estimator uniformly dominates the conservative estimator by reducing the risk under any degree of misspecification. Extending seminal results on the James-Stein estimator, the uniform dominance is established in non-Gaussian semiparametric nonlinear models. The simulation results support our theoretical findings.

Chapter 4 examines properties of permutation tests in the context of synthetic control. Permutation tests are frequently used method of inference for synthetic control when the number of potential control units is small. We show that the size of permutation tests may be distorted. Several alternative methods are discussed.

The dissertation of Ruoyao Shi is approved.

Moshe Buchinsky

Rosa Liliana Matzkin

Peter E. Rossi

Jinyong Hahn, Committee Co-Chair

Zhipeng Liao, Committee Co-Chair

University of California, Los Angeles

2017

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply indebted to Jinyong Hahn, Zhipeng Liao and Rosa Liliana Matzkin for guidance and help. I am grateful to Moshe Buchinsky for constant encouragement and help. I also thank Jaap Abbring, Youssef Benzarti, Leah Boustan, Denis Chetverikov, Yingying Dong, Edward Kung, Adriana Lleras-Muney, Elena Manresa, Maurizio Mazzocco, Rodrigo Pinto, Peter E. Rossi, Shuyang Sheng and Till Von Wachter for helpful discussions. Comments from the seminar participants of the 2015 California Econometrics Conference, 2016 Bristol Econometrics Study Group, 2016 Econometric Society Asian Meeting, 2016 Midwest Econometrics Group Annual Meeting, UCLA econometrics proseminars, UCLA applied micro proseminars, Boston University, Syracuse University, Australian National University, University of Sydney, University of New South Wales, University of California Riverside helped shape the paper as well.

Chapter 3 is a version of Cheng, Liao, and Shi (2016). It received a revise and resubmit request from *Quantitative Economics.* The authors thank Donald Andrews, Denis Chetverikov, Patrik Guggenberger, Jinyong Hahn, Jia Li, Rosa Liliana Matzkin, Hyunsik Roger Moon, Ulrich Mueller, Joris Pinkse, Frank Schorfheide, Shuyang Sheng, and seminar participants at Brown University, Duke University, University of Pennsylvania, Pennsylvania State University, University of California Los Angeles, Yale University, and 2014 New York Area Econometrics Colloquium for helpful comments.

Chapter 4 is a version of Hahn and Shi (2016). It is in preparation for publication. Helpful comments by Alberto Abadie, Bruno Ferman and Guido Imbens are greatly appreciated.

Finally, I want to thank my husband, Cheng Chou for his encouragement, patience and humor.

RUOYAO SHI

Education

University of California, Los Angeles Los Angeles, CA, USA	C.Phil., December 2013 M.A., December 2012 Economics
Peking University Beijing, China	M.A., July 2011 Economics
	B.S., July 2008 Statistics
	B.B.A., July 2008 Management
Hong Kong University of Science and Technology Hong Kong, China	Exchange Student, February-June 2006 Business and Economics

Fellowships and Awards

Dissertation Year Fellowship, UCLA, 2016
Honor Pass in econometrics comprehansive exam, UCLA, 2012
Pauley Fellowship, UCLA, 2011
Distinguished Student Award, Peking University, 2011
Kwang-Hwa Graduate Research Award, Peking University, 2009
Graduate Fellowships, Peking University, 2008- 2010
First Class Undergraduate Scholarships, Peking University, 2005-2007
Dean's List, Hong Kong University of Science and Technology, 2006

Journal Referee

Empirical Economics, 2016-2017
Quantitative Economics, 2016

Conferences Organized

1^{st} China Center for Economic Research (CCER) Conference, Beijing, China, June 2010

Research Assistance

For Prof. Zhipeng Liao

• "Cross-Validation for High-Dimensional and Nonparametric Models in Econometrics", 2016-2017, with Denis Chetverikov

For Prof. Rosa Liliana Matzkin

• "Estimation of Nonparametric Models with Simultaneity", 2015, *Econometrica*, Vol. 83, No. 1, p. 1-66

For Prof. Zhipeng Liao

• "On Standard Inference for GMM with Seeming Local Identification Failure", 2016, with Ji Hyung Lee, R&R at *Econometric Theory*

Conferences and Invited Seminars

2017 Boston University (USA), Syracuse University (USA),
Australian National University (Australia),
University of Sydney (Australia),
University of New South Wales (Australia),
University of California Riverside (USA)

2016 Annual Meeting of Midwest Econometrics Group (USA),
Econometric Society Asian Meeting (Japan),
Econometric Study Group Bristol Conference (UK)

2015 California Econometrics Conference (USA)

2010 Annual Conference of China Economic Association (China)

Teaching Assistance

Ph.D.:

• Econometrics III, UCLA, Spring 2013-2015

MBA:

• Game Theory, Peking University, Spring 2010
• Corporate Finance, Peking University, Fall 2010

Undergraduate:

• Principle of Economics, UCLA, Fall 2012 & Winter 2013
• Statistics for Economists, UCLA, Fall 2013, Winter 2014, Fall 2014 & Winter 2015
• Microeconomic Theory, UCLA, Fall 2015
• Introduction to Econometrics, UCLA, Summer 2015

Chapter 1

Introduction

This dissertation studies econometric questions in the context of three different methods that are frequently used by empirical economists.

Chapter 2 provides a new structural method that accounts for potential equilibrium effects in the construction of counterfactual outcome distributions when the composition of the agents in the market is substantially altered by policies or social changes. One application of such counterfatual distributions is in earning inequality. Suppose the college education expanded such that the number of college graduates in the labor force surged, what would be the counterfactual earning distribution for workers of all education levels? This question is very important because the same policy might affect different workers differently. If more college educated workers entered the labor force, then the market price of college education would be likely to drop, and the prices of other education levels would be likely to rise (at least relatively). As a result, the labor supply and demand choice of everyone, including both the workers and the firms, would be changing. So it is essential to allow the prices and agents' choices to adjust in an equilibrium framework. Currently available methods in the literature, however, essentially exclude equilibrium effects (e.g., DiNardo, Fortin, and Lemieux, 1996). On the contrary, my approach is to use a two-sided, competitive equilibrium model with heterogenous workers and firms. Each worker chooses how much time to work, given the market prices and her own efficiency. A big challenge here is that worker efficiency is usually not observed by researchers. A unique feature or my approach is that I allow worker efficiency to be a completely nonparametric function of workers' observable attributes and unobservable ability. I show that this unobserved worker efficiency function can be nonparametrically identified up to normalization, along with other structural functions of the model. This identification analysis permits counterfactual predictions. The main idea of identification is that I use both within-market and cross-market variation in observed equilibrium earnings and working time to recover the unobserved worker efficiency. In addition to identification analysis, I provide an easy-to-implement, least square type estimation procedure that echoes each step of the identification strategy, and derive the rates of convergence of the estimators. Moreover, I also apply the procedure to the data from American Time Use Survey (ATUS) to estimate the efficiency function of full-time American workers. This paper is the first to provide a structural method for constructing counterfactual earning distributions (or other outcome distributions) that accounts for both equilibrium effects and unobserved worker efficiency (or in general, unobserved product quality) in a nonparametric setting.

Chapter 3 was motivated by a trade-off often faced by empirical researchers, that some invalid instrumental variables may contain useful information that helps with the estimation of the parameters of interest. Conservative researchers may choose to use only the valid instrumental variables. But we provide an averaging GMM estimator that allows researchers to deliberately use the information contained in the potentially invalid instrumental variables to improve the point estimation. This averaging GMM estimator also applies to other generalized method of moments models. The proposed averaging GMM estimator is a weighted average of the conservative estimator that uses only the valid moment conditions and the aggressive estimator that uses both valid and potentially misspecified moment conditions, with the weight being the sample analog of an infeasible optimal weight. By characterizing the asymptotic behaviors of the bounds of the finite-sample risk difference between the two estimators, we show that the averaging GMM estimator has smaller risk than the conservative estimator under any degree of misspecification.

Chapter 4 investigates the inference methods for synthetic control. Synthetic control is a popular method for treatment effect estimation in comparative studies and the inference is often done using permutation tests. We examin the size properties of the permutation tests for a synthetic control method and find that the size of the permutation tests could be distorted. And we discuss two alternative inference methods that correct the size distortion under a factor model framework.

Chapter 2

Identification and Estimation of Nonparametric Hedonic Equilibrium Model with Unobserved Quality

2.1 Introduction

Counterfactual distributions are indispensable components for the evaluation of distributional effects of large-scale policy interventions or social changes; they can also be used to measure the values of public good or natural resources. For example, labor economists might be interested in constructing the counterfactual wage distribution in 1988 had there been no de-unionization or decline in real minimum wage during the 1979-1988 period to evaluate the effect of labor market institutions on inequality (see DiNardo, Fortin, and Lemieux, 1996 for details). Another application of interest would be to measure heterogenous willingness to pay for clean air as exhibited in housing prices (e.g., Sieg, Smith, Banzhaf, and Walsh, 2004, and Chay and Greenstone, 2005).

Three features should be acknowledged in the counterfactual distributional analysis. First, large-scale interventions usually affect a substantial proportion of the agents (e.g., DiNardo, Fortin, and Lemieux, 1996, and Chernozhukov, Fernández-Val, and Melly, 2013), hence the importance of accounting for the equilibrium effects is of first order (e.g., Sieg, Smith, Banzhaf, and Walsh, 2004). Second, some product characteristics might not be observed by researchers and their importance in price determination is widely recognized (e.g., Berry, Levinsohn, and Pakes, 1995), workers' efficiency being an important example. Third, there is considerable observed and unobserved heterogeneity among the agents. Ignoring any of them (e.g., ignoring changes in return to college education as more college graduates entered the labor force and other factors remained constant) is likely to result in biased counterfactual predictions.

This paper is the first to provide an economic model and an econometric method that admits all these features in a nonparametric setting. In this paper, I study a hedonic equilibrium model with unobserved product quality. I show that the quality function, together with all the other structural functions of the model, can be nonparametrically identified. I also provide easy-to-implement estimators for the structural functions and an algorithm to solve the counterfactual equilibrium. In contrast to widely used distributional decomposition methods, the counterfactuals thus constructed account for equilibrium effects of large-scale interventions.

I incorporate unobserved product quality captured by a *structural quality function* $e(x, a)$ into standard hedonic equilibrium models (e.g., Heckman, Matzkin, and Nesheim, 2010), which have been used to analyze the market equilibria of differentiated products with heterogenous agents. Let z_i^m denote the *effective amount* of the product traded between seller-buyer pair i in market m upon which the payment is determined, and assume that

$$z_i^m \equiv h_i^m \cdot e(x_i^m, a_i^m), \tag{2.1.1}$$

where h_i^m represents observed quantity, vector x_i^m and scalar a_i^m represent the seller's observed and unobserved heterogeneity, respectively. I relax the restriction in standard hedonic equilibrium models that $e(x, a) \equiv 1$ by allowing the functional form of the quality function (and hence, the values of $e(x_i^m, a_i^m)$ and z_i^m) to be unknown to researchers.

I demonstrate how to nonparametrically identify the structural quality function $e(x, a)$, along with the *structural marginal (dis)utility functions* of sellers and buyers.[1] The identification strategy consists of three steps. First, I show that the *reduced form* (equilibrium outcome) payment function $I^m(x,a)$ and quantity function $h^m(x,a)$ are nonparametrically identified within each market m using the method developed in Matzkin (2003). Second, I exploit within- and cross-market variation in the reduced form functions to identify the unobserved quality function up to normalization. Specifically, equation (2.1.1) indicates that quantity and quality are substitutes in determining the payment. As a result, variation in quality is manifested inversely in the variation in quantity among sellers who receive the same payment within the same market. Moreover, since quantity is optimally chosen by sellers, it suffers from an endogeneity problem. The different distributions of observed agent characteristics across markets serve as aggregate supply or demand shifters that induce cross-market variation in the payment functions, which facilitates the full identification of the quality function. The identification requirement boils down to a rank condition on the payment functions, which requires that relative prices of the agent characteristics vary across markets.[2] Finally, the third step utilizes the agents' first-order conditions to identify the marginal utility functions, in the spirit of the second step of Rosen (1974)'s method.[3]

[1]This paper focuses on the supply side, since the identification and estimation of the demand side structural functions is completely symmetric.

[2]To focus on the key identification problem that arises because of unobserved quality, I concentrate on the scalar-valued quality function $e(x,a)$ in the main text. It is, however, easy to extend the argument to a vector-valued quality function $e(x,a)$ captured by a single-index structure as in Epple and Sieg (1999) and Sieg, Smith, Banzhaf, and Walsh (2004). I elaborate this extension in Appendix 2.C. Recent work by Chernozhukov, Galichon, and Henry (2014) and Nesheim (2015) discussed identification of hedonic equilibrium models with vector-valued unobserved agent characteristics, while still assuming that all product characteristics are observed. Extending the model in this paper to account for vector-valued unobserved product quality that is more general than the single-index structure is an interesting topic for future research.

[3]Unlike Rosen (1974), the estimation procedure introduced in Section 2.4.1 does not require explicitly estimating the price schedule functions.

The constructive identification strategy suggests an easy-to-implement series estimation procedure. I derive uniform rates of convergence of the estimators and demonstrate the procedure by estimating the unobserved efficiency of American full-time workers using data from the 2015 American Time Use Survey (ATUS).[4]

The literature on the identification and estimation of hedonic equilibrium models is vast. In his seminal work, Rosen (1974) originated a two-step method, of which the first step obtains the hedonic price function and its derivatives by fitting a parametric regression of prices on product characteristics, and the second step combines the hedonic price function and agents' first-order conditions to recover the preference and production parameters. Ekeland, Heckman, and Nesheim (2004) considered the identification of a nonparametric hedonic equilibrium model with additive marginal utility and marginal production functions using single market data. Heckman, Matzkin, and Nesheim (2010) formalized the argument in Brown and Rosen (1982), Epple (1987) and Kahn and Lang (1988) that, in general, cross-market variation in price functions is necessary to nonparametrically identify the structural functions. They then focused on the sufficient restrictions for the identification using single market data, and generalized Rosen (1974)'s two-step method to a nonparametric setting. This paper builds on the work of Heckman, Matzkin, and Nesheim (2010) and advances the literature in two ways. First, it allows product quality to be unobserved by researchers, which captures a crucial feature of many applications. It extends Heckman, Matzkin, and

[4]I also propose an algorithm to solve for the counterfactual equilibrium of the model in Appendix 2.A. It is based on the equilibrium condition and Chiappori, McCann, and Nesheim (2010)'s insight that hedonic equilibrium models are mathematically equivalent to an optimal transportation problem. A simple simulation experiment indicates that the numerical equilibrium solution is stable with regard to the estimation errors in the structural functions.

Nesheim (2010)'s method by adding one step at the beginning, which nonparametrically identifies the unobserved quality function $e(x, a)$. Second, this paper is the first to present a nonparametric estimation procedure and to provide convergence rates for the structural functions in hedonic equilibrium models using multiple market data.

The counterfactual analysis enabled by this model is closely related to an extensive literature on distributional decomposition methods (elegantly reviewed in Fortin, Lemieux, and Firpo (2011)), which aims to evaluate the distributional effects of policy interventions or historical changes. Several methods have been proposed, including the imputation method (Juhn, Murphy, and Pierce, 1993), the reweighting method (DiNardo, Fortin, and Lemieux, 1996), the quantile regression-based method (Machado and Mata, 2005), the re-centered influence function method (Firpo, Fortin, and Lemieux, 2009), among many others (e.g., Fessler, Kasy, and Lindner, 2013, and Fessler and Kasy, 2016). Moreover, Rothe (2010) and Chernozhukov, Fernández-Val, and Melly (2013) considered inference in the context of distributional decomposition. This literature is based on the "selection on observables" assumption, which excludes general equilibrium effects. On the contrary, this paper establishes an equilibrium model, which allows the prices of agent characteristics (e.g., the returns to college education) to change in response to changes in the distribution of the characteristics in the population (e.g., as more college graduates enter the labor force).

Characteristic-based demand models in industrial organization and marketing permit unobserved product characteristic as well. This immense literature dates back at least to Berry (1994) and Berry, Levinsohn, and Pakes (1995) and includes Rossi, McCulloch, and Allenby (1996), Nevo (2001), Petrin (2002), Berry, Levinsohn, and Pakes (2004), Bajari and Benkard (2005), Berry and Pakes (2007), and many others.[5] The econometric methods used to analyze characteristic-based demand models are reviewed by Ackerberg, Benkard, Berry, and Pakes (2007).[6] Characteristic-based demand models often assume additively separable utility functions and parametric distributions for the random error terms, which facilitates the identification and estimation using market level data. In this paper, however, the utility functions are nonparametrically identified and estimated, and the estimators are of least-square type (and hence easy to implement). Moreover, this paper investigates how individual level data can be used to predict individual level counterfactual outcomes, which permits richer counterfactual analyses.

The rest of this chapter is organized as follows. Section 2.2.1 sets up the model and describes some important properties of the equilibrium; Section 2.2.2 discusses several applications to which my model and method can be applied for counterfactual policy analysis. Section 2.3 explains the nonparametric identification of the structural functions of the model. The key step is the identification of the unobserved quality function; the intuition and formal results of this step are given in Section 2.3.2. Section 2.4.1 describes the series estimators, and Section 2.4.2 derives their uniform rates of convergence. An illustration of the estimation procedure using the 2015 ATUS data is given in Section 2.5. Section 2.6 points out several directions for future research and concludes the paper. The algorithm to solve for the counterfactual equilibrium, a few complementary results and most of the proofs are collected in

[5]The utility functions in Bajari and Benkard (2005) and Berry and Pakes (2007) are closer to those in this paper, where only the characteristics of the products bear utility, but not the products *per se*. The consumers' utility functions in models of Berry, Levinsohn and Pakes (1995) type have independently and identically distributed random error terms, which represent taste for products for reasons besides product characteristics.

[6]The estimation of production functions, dynamic models and other issues are also reviewed.

Appendices.

2.2 Model

The hedonic equilibrium model with unobserved quality studied in this paper extends the model of Heckman, Matzkin, and Nesheim (2010) by allowing some product characteristics to be unobserved by researchers. Section 2.2.1 introduces the model and discusses its properties that facilitate identifying the structural functions and solving for the counterfactual equilibrium of the model.[7] Section 2.2.2 describes two markets (labor and housing markets) in which the model and the econometric method provided in this paper could be applied to analyze the distributional effects of counterfactual interventions.

2.2.1 Model Setup and Properties of Equilibrium

The model analyzed in this paper pertains to competitive markets (indexed by $m \in \mathcal{M}$) of a product (good or service), of which the quantity is observed by researchers but quality is not. Each seller and buyer only trades once, and chooses the *effective amount* z, where $z \in \mathcal{Z}$. I assume that $\mathcal{Z} \subset \mathbb{R}$ is compact. Let $P^m(z)$ be a twice continuously differentiable price schedule function defined on $\mathcal{Z}$. Then the value of $P^m(z)$ is the payment for an effective amount z of the product in market m.

The following is the key assumption of this model, and distinguishes it from other hedonic equilibrium models (e.g., Heckman, Matzkin, and Nesheim, 2010).

[7] Parallel discussion for hedonic equilibrium models without unobserved quality can be found in Heckman, Matzkin, and Nesheim (2010), Ekeland, Heckman, and Nesheim (2004) and Ekeland (2010).

Assumption 2.2.1. *Suppose that the unobserved effective amount z of the product is determined by the unobserved quality e and observed quantity h in a multiplicative way, i.e.,* $z = h \cdot e$.

Assumption 2.2.1 implies that quantity h and quality e are substitutes in production. Existing hedonic equilibrium models (e.g., Heckman, Matzkin, and Nesheim, 2010) assume that $e \equiv 1$, hence z is observed. But this paper allows quality e and z to be unknown to researchers.

Sellers and buyers both observe quality. As a result, there is no principal-agent problem in this model.

Each seller's quality e is exogenously determined by a *quality function* $e(x, a)$, where the $d_x \times 1$ vector x is the seller's observed characteristics, and the scalar a is the seller's unobserved characteristic. Sellers have quasilinear utility $P^m(z) - U(h, x, a)$, where $U(h, x, a)$ is the disutility that a seller with characteristics (x, a) endures by producing the product of quantity $h \in \mathcal{H}$ (the set $\mathcal{H} \subset \mathbb{R}$ is compact).[8] The population of sellers in market m is described by the density $f^m_{x,a}$, which is assumed to be differentiable and strictly positive on the compact sets $\mathcal{X} \times \mathcal{A} \subset \mathbb{R}^{d_x+1}$. Sellers may choose not to trade, then they obtain reservation utility V_0.

Each buyer has a utility function $R(z, y, b)$, where the $d_y \times 1$ vector y is the buyer's observed characteristics and the scalar b is the buyer's unobserved characteristic. The population of buyers in market m is described by the density $f^m_{y,b}$, which is assumed to be

[8] I concentrate on scalar-valued quantity h in the main text. But it is easy to extend the argument to a vector-valued h captured by a single-index structure as those in Epple and Sieg (1999) and Sieg, Smith, Banzhaf, and Walsh (2004). I elaborate the identification of the quality function $e(x, a)$ under this extension in Appendix 2.C.

differentiable and strictly positive on the compact set $\mathcal{Y} \times \mathcal{B} \subset \mathbb{R}^{d_y+1}$. If a buyer chooses not to participate, she gets reservation utility S_0.

For the structural functions $e(x,a)$, $U(h,x,a)$ and $R(z,y,b)$, assume the following assumptions hold.

Assumption 2.2.2. *Suppose that buyers' utility function $R(z,y,b)$, sellers' disutility function $U(h,x,a)$ and quality function $e(x,a)$ are all twice continuously differentiable with respect to all arguments on their respective supports. Also suppose that $e(x,a)$ is bounded below away from zero.*

Assumption 2.2.3. *Suppose that $U_h > 0$, $U_a < 0$, $U_{ha} < 0$ and $U_{hh} > 0$ for all $(h,x,a) \in \mathcal{H} \times \mathcal{X} \times \mathcal{A}$, and suppose that $R_z > 0$, $R_b > 0$, $R_{zb} > 0$ and $R_{zz} < 0$ for all $(z,y,b) \in \mathcal{Z} \times \mathcal{Y} \times \mathcal{B}$.*

Assumption 2.2.4. *Suppose $e_a > 0$, that is, the quality function is strictly increasing in the unobserved characteristic of the seller, for all $(x,a) \in \mathcal{X} \times \mathcal{A}$.*

If reservation utilities V_0 and S_0 are sufficiently small, then sellers and buyers always participate.[9] In addition, similar to the discussion in Heckman, Matzkin, and Nesheim (2010) and Chiappori, McCann, and Nesheim (2010), Assumptions 2.2.2-2.2.4 (Spence-Mirrlees type single-crossing condition) are sufficient for each seller and buyer who participates to have a unique interior optimum.

A seller with characteristics (x,a) in market m chooses $h \in \mathcal{H}$, a quantity supplied, to

[9] Allowing for binding reservation utilities serves as an important topic for future research.

maximize

$$\max_{h \in \mathcal{H}} P^m(h \cdot e(x,a)) - U(h,x,a).$$

Since quality $e(x,a)$ is fixed for seller (x,a), choosing $h \in \mathcal{H}$ is equivalent to choosing $z \in \mathcal{Z}$. Under Assumptions 2.2.2-2.2.4, there exists an effective amount supply function $z^s \equiv s^m(x,a)$ (hence a quantity supply function $h^m(x,a) \equiv s^m(x,a)/e(x,a)$) that satisfies the seller's first-order condition (FOC)

$$P_z^m(s^m(x,a)) \cdot e(x,a) - U_h\left(\frac{s^m(x,a)}{e(x,a)}, x, a\right) = 0. \tag{2.2.1}$$

Applying the Implicit Function Theorem (Hildebrandt and Graves, 1927) to equation (2.2.1) gives rise to

$$\frac{\partial z^s}{\partial a} = \frac{\partial s^m(x,a)}{\partial a} = \frac{eU_{ha} - P_z^m e e_a - U_{hh} h^m e_a}{P_{zz}^m e^2 - U_{hh}}, \tag{2.2.2}$$

where the arguments of the functions on the right-hand side of equation (2.2.2) are suppressed for simplicity. By Assumptions 2.2.2 and 2.2.3 and the FOC in equation (2.2.1), $P_z^m > 0$. Then Assumptions 2.2.2-2.2.4 imply that $s^m(x,a)$ is strictly increasing in a.[10] Then the inverse effective amount supply function $a = (s^m)^{-1}(x,z)$ exists and satisfies

$$\frac{\partial (s^m)^{-1}(x,z)}{\partial z^s} = \frac{P_{zz}^m e^2 - U_{hh}}{eU_{ha} - P_z^m e e_a - U_{hh} h^m e_a}.$$

[10] As discussed later, equations (2.2.6) and (2.2.7) imply that $P_{zz}^m e^2 - U_{hh} < 0$.

The payment received by seller (x, a) in market m is then determined by

$$I^m(x, a) = P^m(s^m(x, a)) = P^m(h^m(x, a) \cdot e(x, a)). \tag{2.2.3}$$

Note that the payment function $I^m(x, a)$ is also strictly increasing in a. But since $h^m(x, a) = s^m(x, a)/e(x, a)$, the quantity function $h^m(x, a)$ is not necessarily monotonic in a.

Similar argument applies to the buyers. Each buyer chooses $z \in \mathcal{Z}$ to maximize

$$\max_{z \in \mathcal{Z}} R(z, y, b) - P^m(z).$$

There exists an effective amount demand function $z^d \equiv d^m(y, b)$ that satisfies the buyers' FOC

$$R_z(d^m(y, b), y, b) - P^m(d^m(y, b)) = 0, \tag{2.2.4}$$

and an inverse effective amount demand function $b = (d^m)^{-1}(y, z)$ that satisfies

$$\frac{\partial (d^m)^{-1}(y, z)}{\partial z^d} = \frac{R_{zb}}{P^m_{zz} - R_{zz}}.$$

Define the range of equilibrium effective amount supplied

$$\begin{aligned} \mathcal{Z}_s \quad = \quad & \{z \in \mathcal{Z}\text{: there exists a market } m \in \mathcal{M} \text{ and some} \\ & (x, a) \in \mathcal{X} \times \mathcal{A} \text{ such that in equilibrium } z = h^m(x, a) \cdot e(x, a)\}, \end{aligned}$$

and the range of equilibrium effective amount demanded

$$\mathcal{Z}_d = \{z \in \mathcal{Z}\colon \text{ there exists a market } m \in \mathcal{M} \text{ and some } (y, b) \in \mathcal{Y} \times \mathcal{B} \text{ such that in equilibrium } z = d^m(y, b)\}.$$

In a unique interior equilibrium, the density of effective amount supplied z^s equals that of effective amount demanded z^d for all $z \in \mathcal{Z}$. Using standard change-of-variables formula, this requires $\mathcal{Z}_s = \mathcal{Z}_d$ and

$$\int_{\mathcal{X}} f^m_{x,a}\left(x, (s^m)^{-1}(x, z)\right) \frac{\partial (s^m)^{-1}(x, z)}{\partial z^s} dx = \int_{\mathcal{Y}} f^m_{y,b}\left(y, (d^m)^{-1}(y, z)\right) \frac{\partial (d^m)^{-1}(y, z)}{\partial z^d} dy, \tag{2.2.5}$$

for $\forall z \in \mathcal{Z}_s \cap \mathcal{Z}_d$.

Figure 2.1 illustrates the market equilibrium. Under the price schedule function P^m, each seller (x, a) (drawn from distribution $f^m_{x,a}$) chooses her optimal effective amount supplied z^s. The distribution of z^s is represented by the green line in the figure. Similarly, each buyer (y, b) (drawn from distribution $f^m_{y,b}$) chooses her optimal effective amount demanded z^d. The distribution of z^d is represented by the blue line in the figure. If the green density equals the blue density for $\forall z \in \mathcal{Z}$, then the market is in equilibrium.

On the contrary, Figure 2.2 illustrates a case where the market is not in equilibrium. For example, sellers who are willing to supply the effective amount z_1 outnumber the buyers who demand z_1, and more buyers than sellers are willing to trade effective amount z_2. This mismatch between supply and demand will drive the price schedule function P^m to adjust to clear the market.

Following Chiappori, McCann, and Nesheim (2010), the equilibrium of this model is defined as follows.

Definition 2.2.1. *(Equilibrium) Let μ^m be a joint density on the space of effective amount z, characteristics (x, a) of sellers and (y, b) of buyers. A pair (μ^m, P^m) is an equilibrium if:*

(i) the marginal of μ^m with respect to (x, a) equals $f^m_{x,a}$, and that with respect to (y, b) equals $f^m_{y,b}$ (market clears); and

(ii) if (z, x, a, y, b) is in the support of μ^m, then $z = s^m(x, a) = d^m(y, b)$ (agents optimize).

By the argument provided in Chiappori, McCann, and Nesheim (2010) (also in Ekeland (2010) and Heckman, Matzkin, and Nesheim (2010)), Assumptions 2.2.2-2.2.4 are sufficient for the equilibrium to exist and to be unique and pure. A pure equilibrium means that each seller matches to a single buyer, and each pair chooses a single effective amount z.

Note that the effective amount supply function $s^m(x, a)$ and demand function $d^m(y, b)$ have a superscript m, since they both depend on the market-specific price schedule function P^m. And price schedule function P^m is itself an equilibrium outcome, which in turn depends on the market primitives $(f^m_{x,a}, f^m_{y,b}, U, e, R)$. To see this more clearly, substitute $\frac{\partial (s^m)^{-1}(x,z)}{\partial z^s}$ and $\frac{\partial (d^m)^{-1}(y,z)}{\partial z^d}$, rearrange equation (2.2.5) and suppress the arguments of the functions, one gets

$$P^m_{zz}(z) = \frac{\int_{\mathcal{Y}} \frac{f^m_{y,b}}{R_{zb}} R_{zz} dy + \int_{\mathcal{X}} \frac{f^m_{x,a}}{-(U_{ha}e - P^m_z e e_a - U_{hh} h^m e_a)} U_{hh} dx}{\int_{\mathcal{Y}} \frac{f^m_{y,b}}{R_{zb}} dy + \int_{\mathcal{X}} \frac{e^2 f^m_{x,a}}{-(U_{ha}e - P^m_z e e_a - U_{hh} h^m e_a)} dx}. \tag{2.2.6}$$

Equation (2.2.6) implies that the curvature of the price schedule function P^m can be regarded as a weighted average of the curvatures of the sellers' disutility and the buyers' utility functions. Assumptions 2.2.2 and 2.2.3 imply that the second-order condition (SOC)[11]

$$R_{zz} \cdot e^2 - U_{hh} < 0 \tag{2.2.7}$$

holds for all $(h, x, a) \in \mathcal{H} \times \mathcal{X} \times \mathcal{A}$ and all $(z, y, b) \in \mathcal{Z} \times \mathcal{Y} \times \mathcal{B}$. Together, equation (2.2.6) and equation (2.2.7) imply $P^m_{zz} e^2 - U_{hh} < 0$.

Since the structural functions (U, e, R) remain invariant across markets, equation (2.2.6) implies that cross-market variation in the price schedule functions P^m is driven by that in the distributions $f^m_{x,a}$ and $f^m_{y,b}$. As a result, cross-market variation in other reduced form (equilibrium outcome) functions, such as s^m, d^m, h^m and I^m, also depends on that in $f^m_{x,a}$ and $f^m_{y,b}$. Throughout this paper, I summarize this dependence using the superscript m.

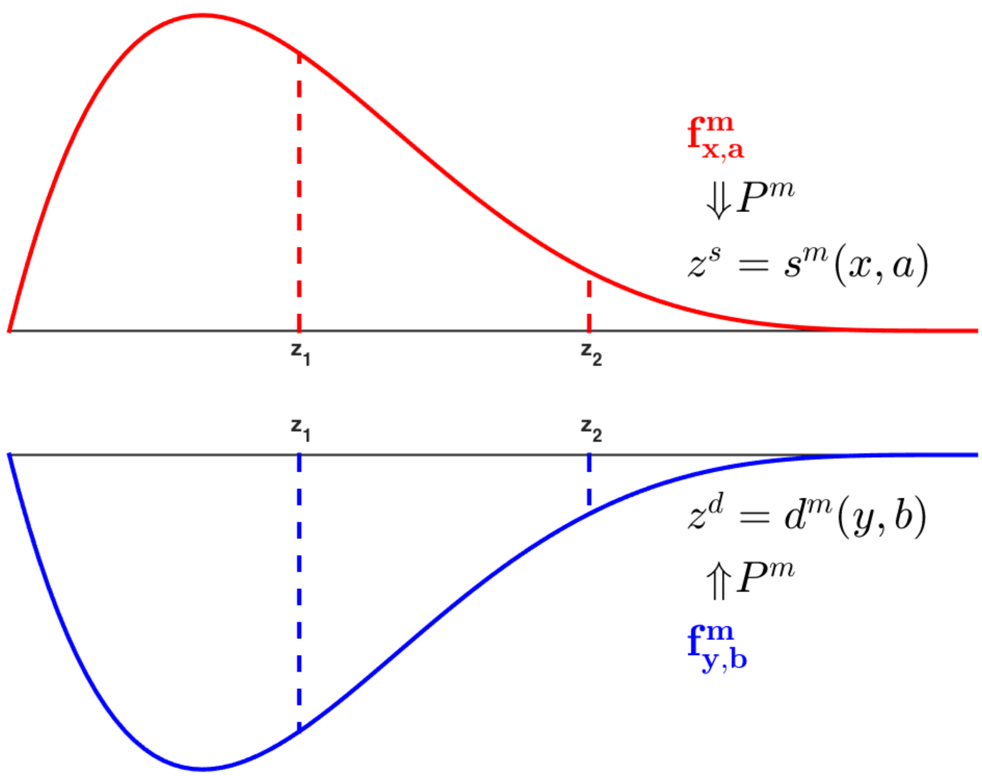
In the same market, all sellers and buyers face the same price schedule function P^m, so sellers with the same characteristics (x, a) always choose the same quantity $h^m(x, a)$ to supply. Without restrictions on sellers' marginal disutility function $U_h(h, x, a)$, its identification using single market data is obstructed by this endogeneity problem. With multiple market data, however, the distributions $f^m_{x,a}$ and $f^m_{y,b}$ serve as aggregate supply or demand shifters (i.e., instruments) that induce variation in P^m (and hence $h^m(x, a)$) while maintaining individual values of (x, a). In practice, multiple markets could be geographically isolated locations, or repeated observations of the same market over time.

Chiappori, McCann, and Nesheim (2010) showed that the classic hedonic equilibrium

[11]SOC of a pair-wise surplus maximization problem.

model is mathematically equivalent to a stable matching problem and to an optimal transportation problem. The same argument applies to the model in this paper as well, since quality is observable to both sellers and buyers. This insight suggests an algorithm for solving for the counterfactual equilibria, which is provided in Appendix 2.A.

Figure 2.1: Equilibrium



The green line illustrates the distribution of the optimal effective labor supply z^s under the price schedule function P^m in market m, as a function of sellers' observed characteristics x and unobserved characteristic a, which follow the distribution $f^m_{x,a}$. Similarly, the blue line illustrates the distribution of the optimal effective labor demand z^d under the same price schedule function P^m in market m, as a function of buyers' observed characteristics y and unobserved characteristic b, which follow the distribution $f^m_{y,b}$. As is shown in this figure, when the distributions of z^s and z^d are the same, the market clears.

2.2.2 Applications

In this section, I discuss two markets to which the model just introduced could be applied to conduct counterfactual distributional analysis. In these examples, the unobserved quality of product plays a key role in determining the payment.

2.2.2.1 Labor Markets

In labor markets, workers are the sellers, and (single-employed) firms are the buyers.[12] Both workers and firms exhibit considerable heterogeneity. Workers differ in observed characteristics x (e.g., age, education and skills) and unobserved characteristic a (e.g., ability). Likewise, employers differ in observed characteristics y (e.g., capital stock) and unobserved characteristic b (e.g., productivity). For a worker with characteristics (x, a), her efficiency is given by the function $e(x, a)$, which is the same across markets and is unknown to researchers.[13] On the other hand, distributions of agent heterogeneity ($f^m_{x,a}$ and $f^m_{y,b}$) could vary among markets, which induce market-specific earnings schedule functions $P^m(z)$. As a result, workers with the same characteristics may choose to work different amount of time $h^m(x, a)$ and make different earnings $I^m(x, a)$ in different markets. Workers' working time and efficiency are substitutes in production, and firms care about how much work is done, but not the working time in itself.[14] Therefore, earnings depend on the effective amount

[12] It is also helpful to think of the buyers as job positions.

[13] Firms know z and $e(x, a)$ by looking at how much work the worker gets done.

[14] Ideally, researchers would want to measure the actual time workers spend in working. The required working time written on the contract deviates from the actual time, since workers could shirk or work over-time.

of labor z via the earnings schedule functions $P^m(z)$, but not on working time $h^m(x, a)$ or efficiency $e(x, a)$ *per se*.

The model in this paper could be used to answer various counterfactual questions that labor economists are interested in. For example, to understand the distributional effects of the changes in labor market institutions during 1979-1988, one may want to construct counterfactual earnings distribution in 1988 had there been no de-unionization since 1979 (e.g., DiNardo, Fortin, and Lemieux, 1996). This corresponds to the equilibrium earnings of a market in which workers' union status (one element of x) had remained what it was in 1979 and other agent characteristics (other variables in (x, a, y, b)) had shifted to their 1988 values.

2.2.2.2 Housing Markets

In housing markets, renters are the buyers, and rental companies (or landlords) are the sellers.[15] Renters' observed characteristics y include income and family structure, and their unobserved characteristic b may be preference over amenities. Rental companies diversify in their characteristics (x, a) as well. For a rental company (x, a), the quality of its apartments is given by $e(x, a)$, which does not depend on which neighborhood m the rental company is in and is unknown to researchers. However, varying composition of renters and rental companies ($f^m_{y,b}$ and $f^m_{x,a}$) across neighborhoods result in neighborhood-specific rental price schedule functions $P^m(z)$, which in turn prompt rental companies with the same characteristics to offer apartments with different sizes $h^m(x, a)$ (e.g., square footage) across neighborhoods.

[15] I focus on housing rental markets, but the same logic applies to housing sale markets.

Rental payments $I^m(x,a)$ depend on the effective amount of housing z via $P^m(z)$, but not directly on the sizes.

A number of interesting counterfactual questions in housing markets could be analyzed using the model in this paper. For example, one may be interested in the distributional effects on housing prices if some public good (e.g., improvement in air quality) were provided.[16] The public good enhances effective amount of housing for all apartments (by all rental companies) in the neighborhood, and it is manifested in increased value of $e(x,a)$ for any given (x,a). Therefore, the counterfactual analysis could be conducted by solving the new equilibrium with a higher quality function $e(x,a)$ estimated using data for neighborhoods with more public good.

2.3 Identification

This section explains identification of the reduced form (equilibrium outcome) functions and the structural functions of the model. The analysis in this section assumes that seller characteristics x, buyer characteristics y, equilibrium payment I and equilibrium quantity h in all markets are observed. The effective amount z, however, is unknown to researchers.

The identification consists of three steps. First, identify the reduced form payment functions $I^m(x,a)$ and the quantity functions $h^m(x,a)$ using single market data. This step employs an existing method (Matzkin, 2003) and facilitates the identification of structural functions. Second, exploit the variation of the payment and quantity functions within and

[16]Harrison and Rubinfeld (1978) and Chay and Greenstone (2005) used housing prices to evaluate willingness-to-pay for clean air. Another example is predicting the effects of cleaning up a hazardous waste site on the distribution of housing prices (Stock, 1991).

across markets to identify the quality function $e(x, a)$. This is the key step, and I will provide both graphical illustration of intuition and general results. The key identification condition requires that the relative returns to sellers' characteristics differ across markets. Finally, combine the functions identified from the first two steps and sellers' FOC to recover sellers' marginal disutility function $U_h(h, x, a)$. To overcome the endogeneity problem of h, this final step requires multiple market data as well. Section 2.3.1, 2.3.2 and 2.3.3 elaborate these steps, respectively.

This section focuses on the quality function $e(x, a)$ and sellers' marginal disutility function $U_h(h, x, a)$. The identification of buyers' marginal utility function $R_z(z, y, b)$ can be achieved via the same method as that used for $U_h(h, x, a)$, and is briefly discussed in Section 2.3.4. Although $f^m_{x,a}$ and $f^m_{y,b}$ are also primitives of the model and serve as aggregate supply or demand shifters that generate cross-market variation in equilibria, their identification is straightforward. The convergence rate results in Section 2.4.2 account for the fact that they need to be estimated.

2.3.1 Identification of Payment Functions $I^m(x, a)$ and Quantity Functions $h^m(x, a)$ Using Single Market Data

In each market m, there is a payment function $I^m(x, a)$ and a quantity function $h^m(x, a)$ in equilibrium. This section uses the method developed by Matzkin (2003) to identify these reduced form functions using data from their own markets.

Assumption 2.3.1. *Suppose that $x \perp\!\!\!\perp a$ and $y \perp\!\!\!\perp b$ within each market $m \in \mathcal{M}$.*[17]

[17] Like Heckman, Matzkin, and Nesheim (2010), because a enters the quality function and sellers' marginal

Assumption 2.3.2. *Suppose that the sellers' unobserved characteristic a follows the uniform distribution $U[0,1]$ in all markets.*

Assumption 2.3.2 may seem restrictive at first glance. But an equivalent interpretation is that a is the quantile of the seller's unobserved characteristic. Based on this interpretation, Assumption 2.3.2 requires that the sellers' unobserved characteristic has the same distribution (probably unknown) across all markets.[18] In Appendix 2.B, I relax this requirement to allow for a finite number of types of markets: markets of the same type have the same distribution of a, yet markets of different types have different distributions of a. The method discussed in the main text can be applied to each type without modification, as long as the type of each market is known and each type has multiple markets.[19] Assumption 2.3.2 is also a normalization that facilitates identification of nonseparable functions like $I^m(x,a)$ (see Matzkin, 2003 for details).[20] But this normalization does not affect counterfactuals.

Lemma 2.3.1. *Under Assumptions 2.2.1-2.3.2, the payment function $I^m(x,a)$ is strictly increasing in the seller's unobserved characteristic a, and $I^m(x,a)$ is nonparametrically identified within each market m.*

Proof. By the payment equation (2.2.3), $I^m(x,a)$ is strictly increasing in a if P^m is strictly increasing in z and s^m is strictly increasing in a. Given the sellers' FOC in equation (2.2.1),

disutility function nonparametrically, this independence assumption is much weaker than it would be if a entered additively.

[18]To see this clearly, suppose that F_a is the distribution function of a, and suppose $\tilde{U}(h,x,a)$ and $\tilde{e}(x,a)$ are the "real" supply side structural functions. Then, based on the quantile interpretation, the supply side structural functions identified in this paper are compounds of F_a and the "real" structural functions. That is, $U(h,x,a) = \tilde{U}\left(h,x,F_a^{-1}(a)\right)$ and $e(x,a) = \tilde{e}\left(x,F_a^{-1}(a)\right)$. Therefore, Assumption 7 implicitly requires that F_a is invariant across markets.

[19]One example of such market level heterogeneity might be large cities v.s. small cities.

[20]One could normalize the distribution of a to any other distributions.

Assumptions 2.2.2 and 2.2.3 guarantee that $P_z^m > 0$. On the other hand, by the SOC in equation (2.2.7) and the equilibrium condition in equation (2.2.6), we have $P_{zz}^m \cdot e^2 - U_{hh} < 0$. Then the expression of $\partial s^m(x,a)/\partial a$ in equation (2.2.2) is positive under Assumptions 2.2.3 and 2.2.4 and the setup of the model. This proves the first statement of the lemma.

Given the strict monotonicity of $I^m(x,a)$ and Assumption 2.3.2, the identification of $I^m(x,a)$ follows the same argument as in Matzkin (2003) (Specification I). In particular, by monotonicity, Assumptions 2.3.1 and 2.3.2, we have

$$F_{I^m|x^m=x}\left(I^m(x,a)\right) = F_{a^m}(a) = a.$$

Then

$$I^m(x,a) = F^{-1}_{I^m|x^m=x}(a),$$

where $F^{-1}_{I^m|x^m=x}$ is the inverse function of the conditional distribution function $F_{I^m|x^m=x}$ with respect to I^m. □

Corollary 2.3.2. *Under the conditions for Lemma 2.3.1, the partial derivatives of the payment function $I^m_{x_j}(x,a)$ ($j = 1,\dots,d_x$) and $I^m_a(x,a)$ are nonparametrically identified within each market m.*

Once one identifies the payment function I^m, she can invert it with respect to a to obtain $a = (I^m)^{-1}(x,I)$. Now that a is known, it is easy to identify the quantity function $h^m(x,a)$. Unlike $I^m(x,a)$, monotonicity is not necessary for identification of $h^m(x,a)$.

Lemma 2.3.3. *Under the conditions for Lemma 2.3.1, the quantity function $h^m(x,a)$ is* *nonparametrically identified within each market m. Moreover, its partial derivatives $h^m_{x_j}(x,a)$ ($j = 1,\ldots,d_x$) and $h^m_a(x,a)$ are nonparametrically identified within each market m as well.*

Note that the functional forms of $I^m(x,a)$ and $h^m(x,a)$ vary from market to market due to the cross-market variation in f^m_x and f^m_y, and they are identified within each market. Their variation within and across markets reveals enough information to identify the quality function $e(x,a)$.

2.3.2 Identification of Quality Function $e(x,a)$ Using Multiple Market Data

This section explains how to use within- and cross-market variation in the reduced form functions to identify the structural quality function $e(x,a)$. Section 2.3.2.1 illustrates the intuition for scalar-valued x. The intuition applies to vector-valued x as well. Section 2.3.2.2 gives general results.

Since quality e and effective amount z are both unobserved, one can always re-scale the price schedule function to make two quality functions observationally equivalent. So we need the following normalization.

Assumption 2.3.3. *Suppose that for a known fixed vector $(\bar{x},\bar{a}) \in \mathcal{X} \times \mathcal{A}$, we have $e(\bar{x},\bar{a}) = 1$.*

The vector $(\bar{x},\bar{a})$ corresponds to a normalization seller, and the quality of other sellers will be expressed as ratio relative to her.

2.3.2.1 Intuition

This section illustrates the intuition for identifying the unobserved quality function $e(x, a)$ for scalar-valued x. The interpretation of the key identification condition is that relative returns to sellers' characteristics differ across markets.

Recall the payment equation (2.2.3),

$$I^m(x, a) = P^m(s^m(x, a)) = P^m(h^m(x, a) \cdot e(x, a)).$$

Since all sellers in the same market face the same price schedule function $P^m(z)$, those sellers who receive the same payment must have sold the same effective amount z of the product. In other words, if $I_i^m = I_j^m$ for two sellers i and j in the same market m, then

$$h^m(x_i^m, a_i^m) \cdot e(x_i^m, a_i^m) = h^m(x_j^m, a_j^m) \cdot e(x_j^m, a_j^m),$$

which implies

$$\frac{e(x_i^m, a_i^m)}{e(x_j^m, a_j^m)} = \frac{h^m(x_j^m, a_j^m)}{h^m(x_i^m, a_i^m)}. \tag{2.3.1}$$

That is, the quality ratio between sellers who receive the same payment in the same market equals the inverse ratio of their quantities.

This is illustrated by Figure 2.3. The solid green line in Step 1 of Figure 2.3 represents the iso-payment curve in Market 1 that contains the normalization seller $(\bar{x}, \bar{a})$. By equation (2.3.1), the quality of any seller (x_1, a_1) on the same iso-payment curve can be identified as

$$e(x_1, a_1) = \frac{h^1(\bar{x}, \bar{a})}{h^1(x_1, a_1)}.$$

The same argument applies to other iso-payment curves in Market 1, which are represented by dashed green lines in Step 1. For example, for sellers $(\tilde{x}, \tilde{a})$ and (x_2, a_2) on another iso-payment curve, we get

$$\frac{e(x_2, a_2)}{e(\tilde{x}, \tilde{a})} = \frac{h^1(\tilde{x}, \tilde{a})}{h^1(x_2, a_2)}. \tag{2.3.2}$$

Since iso-payment curves in the same market are disjoint, neither $e(x_2, a_2)$ nor $e(\tilde{x}, \tilde{a})$ could be identified relative to the normalization seller $(\bar{x}, \bar{a})$. The dashed green lines in Step 1 indicate that the quality of the sellers on those iso-payment curves are not identified yet. This is the most one can get from variation of reduced form functions in one market.

With data from another market, however, it is possible to connect the disjoint iso-payment curves. Suppose that in Market 2, there is an iso-payment curve that contains both $(\bar{x}, \bar{a})$ and (x_2, a_2), then

$$e(x_2, a_2) = \frac{h^2(\bar{x}, \bar{a})}{h^2(x_2, a_2)}. \tag{2.3.3}$$

Combining equation (2.3.2) and equation (2.3.3), we now can identify the quality for seller $(\tilde{x}, \tilde{a})$ as

$$e(\tilde{x}, \tilde{a}) = \frac{h^1(x_2, a_2)}{h^1(\tilde{x}, \tilde{a})} \cdot \frac{h^2(\bar{x}, \bar{a})}{h^2(x_2, a_2)}.$$

Once $e(\tilde{x}, \tilde{a})$ is identified, so is the quality of other sellers on the same iso-payment curve.

In Step 2 of Figure 2.3, the iso-payment curve in Market 2 is represented by the solid blue line. It connects the Market 1 iso-payment curve that contains $(\bar{x}, \bar{a})$ with the one that contains $(\tilde{x}, \tilde{a})$, and thus helps determine the quality level of the latter. In Step 3 of Figure 2.3, the latter becomes solid green as the quality of those sellers are identified. Step 4 shows that by applying this idea recursively to the iso-payment curves from the two markets that cross with each other, one will be able to identify the quality of all sellers with characteristics in the support of their distribution.

As suggested by Figure 2.3, the key identification condition is that for any seller characteristics (x, a), one could find two markets that have iso-payment curves with different slopes. Otherwise, all the iso-payment curves are disjoint, and it is impossible to connect a seller (x, a) with the normalization seller $(\bar{x}, \bar{a})$ if they do not belong to the same iso-payment curve.

Note that the slope of an iso-payment curve can be expressed in terms of the partial derivatives of the payment function, then the identification condition is

$$\frac{I_x^1(x,a)}{I_a^1(x,a)} \neq \frac{I_x^2(x,a)}{I_a^2(x,a)},$$

for $\forall (x, a) \in \mathcal{X} \times \mathcal{A}$, scalar-valued x and two markets. This condition is also equivalent to that the matrix

$$\begin{pmatrix} I_a^1(x,a) & -I_x^1(x,a) \\ I_a^2(x,a) & -I_x^2(x,a) \end{pmatrix}$$

has full column rank.

This key condition is easy to understand. Partial derivatives of the payment functions represent the equilibrium market returns to respective seller characteristics. For example,

$I_x^m(x,a)$ could represent labor market return to education, and $I_a^m(x,a)$ to ability. Then the identification condition requires that the relative equilibrium returns to education and to ability differ in at least two markets. This in turn requires that cross-market variation in f_x^m and f_y^m is sufficiently rich to induce such cross-market variation in equilibria.

2.3.2.2 General Results

It is not hard to generalize the intuition explained in Section 2.3.2.1 to vector-valued x. This section formalizes this intuition and gives general results on the identification of the unobserved quality function $e(x,a)$.

When x is vector-valued ($d_x > 1$), the key identification condition is still that relative market returns to seller characteristics differ in at least two markets. Without loss of generality, one could measure returns as relative to that to the unobserved characteristic a. Suppose $I_a^m(x,a) \neq 0$ and $I_a^{m'}(x,a) \neq 0$ for markets m and m'. Then it is required that

$$\left(\frac{I_{x_1}^m(x,a)}{I_a^m(x,a)}, \ldots, \frac{I_{x_{d_x}}^m(x,a)}{I_a^m(x,a)}, 1\right) \neq \left(\frac{I_{x_1}^{m'}(x,a)}{I_a^{m'}(x,a)}, \ldots, \frac{I_{x_{d_x}}^{m'}(x,a)}{I_a^{m'}(x,a)}, 1\right). \tag{2.3.4}$$

These are just the gradient vectors of the payment functions $I^m(x,a)$ and $I^{m'}(x,a)$.

Cross-market variation in equilibria is crucial for identifying the quality function. The following assumption requires that neither sellers nor buyers move across markets on a large scale. Otherwise, the distributions f_x^m and f_y^m will tend to equalize across markets, which diminishes the cross-market variation.

Assumption 2.3.4. *Suppose that the sellers and buyers do not move across markets.*

In order to state the formal identification condition and the theorem, I need some notation. Let $\nabla_x I^m(x,a)$ denote the $d_x \times 1$ vector of the derivatives of $I^m(x,a)$ with respect to $(x_1,\ldots,x_{d_x})'$, let $\nabla_x h^m(x,a)$ denote those of $h^m(x,a)$ and let $\nabla_x e(x,a)$ denote those of $e(x,a)$. For any integer d, let $\mathbb{I}_d$ denote a $d \times d$ identity matrix.

Assumption 2.3.5. *Suppose that there exist M markets such that the $(Md_x) \times (d_x+1)$ matrix $B(x,a)$ defined as*

$$B(x,a) \equiv \begin{pmatrix} \mathbb{I}_{d_x} \otimes I_a^1(x,a) & -\nabla_x I^1(x,a) \\ \vdots & \vdots \\ \mathbb{I}_{d_x} \otimes I_a^M(x,a) & -\nabla_x I^M(x,a) \end{pmatrix}$$

has full column rank for all $(x,a) \in \mathcal{X} \times \mathcal{A}$.[21]

It only takes some basic algebra to see that if equation (2.3.4) holds for all $(x,a) \in \mathcal{X} \times \mathcal{A}$, then Assumption 2.3.5 is satisfied. Moreover, if Assumption 2.3.5 holds, there could be more than two markets satisfying equation (2.3.4).

Define the $(Md_x) \times 1$ vector $A(x,a)$ as

$$A(x,a) \equiv \begin{pmatrix} \left[h_a^1(x,a)\nabla_x I^1(x,a) - I_a^1(x,a)\nabla_x h^1(x,a)\right]/h^1(x,a) \\ \vdots \\ \left[h_a^M(x,a)\nabla_x I^M(x,a) - I_a^M(x,a)\nabla_x h^M(x,a)\right]/h^M(x,a) \end{pmatrix}.$$

[21] Note that a necessary condition for $B(x,a)$ to have full column rank is that there are d_x+1 linearly independent rows in $B(x,a)$. Therefore we need at least two markets. But when data from more markets is available, and multiple combinations of rows satisfy the requirement, we get over-identification.

And define $d_x + 1$ real-valued functions $g_1(x,a), \ldots, g_{d_x+1}(x,a)$ as

$$(g_1(x,a), \ldots, g_{d_x+1}(x,a))' \equiv [B(x,a)'B(x,a)]^{-}[B(x,a)'A(x,a)],$$

where the superscript "$-$" indicates the generalized inverse of a matrix.

Theorem 2.3.4. *Suppose that Assumptions 2.3.3-2.3.5 and the conditions for Lemma 2.3.1 are satisfied. The quality function is then nonparametrically identified on $\mathcal{X} \times \mathcal{A}$ as*

$$e(x,a) = \exp\left(\sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} g_j(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a}) ds_j + \int_{\bar{a}}^{a} g_{d_x+1}(x,t)dt \right) \quad (2.3.5)$$

where $\bar{x}_j$ $(j = 1, \ldots, d_x)$ and $\bar{a}$ are coordinates of the normalization vector $(\bar{x}, \bar{a})$.

Proof. Suppose that all the functions involved are continuously differentiable. Then, taking the partial derivatives of the payment equation (2.2.3) yields

$$\begin{aligned} \nabla_x I^m(x,a) &= P_z^m\left(h^m(x,a) \cdot e(x,a)\right) \cdot \left[\nabla_x h^m(x,a)e(x,a) + h^m(x,a)\nabla_x e(x,a)\right], \\ I_a^m(x,a) &= P_z^m\left(h^m(x,a) \cdot e(x,a)\right) \cdot \left[h_a^m(x,a)e(x,a) + h^m(x,a)e_a(x,a)\right]. \end{aligned} \quad (2.3.6)$$

Provided that $I_a^m(x,a) \neq 0$ and $h^m(x,a) \neq 0$, one may take the ratios of the first d_x equations to the last equation. One then obtains d_x equations of the same form:

$$\frac{\nabla_x I^m(x,a)}{I_a^m(x,a)} = \frac{\nabla_x h^m(x,a)e(x,a) + h^m(x,a)\nabla_x e(x,a)}{h_a^m(x,a)e(x,a) + h^m(x,a)e_a(x,a)} = \frac{\frac{\nabla_x h^m(x,a)}{h^m(x,a)} + \frac{\nabla_x e(x,a)}{e(x,a)}}{\frac{h_a^m(x,a)}{h^m(x,a)} + \frac{e_a(x,a)}{e(x,a)}}$$

$$\Longrightarrow \quad I_a^m(x,a)\frac{\nabla_x e(x,a)}{e(x,a)} - \nabla_x I^m(x,a)\frac{e_a(x,a)}{e(x,a)}$$
$$= \left[h_a^m(x,a)\nabla_x I^m(x,a) - I_a^m(x,a)\nabla_x h^m(x,a)\right]/h^m(x,a), \tag{2.3.7}$$

for all $m \in \mathcal{M}$ and all $(x,a) \in \mathcal{X} \times \mathcal{A}$. Stack equation (2.3.7) for all markets, one gets a system of equations

$$B(x,a) \cdot \left(\frac{\nabla_x e(x,a)'}{e(x,a)}, \quad \frac{e_a(x,a)}{e(x,a)}\right)' = A(x,a), \tag{2.3.8}$$

for all $(x,a) \in \mathcal{X} \times \mathcal{A}$. Suppose that Assumption 2.3.5 is satisfied. Then, there is a unique solution of $e_a(x,a)/e(x,a)$ and $e_{x_j}(x,a)/e(x,a)$ $(j = 1,\ldots,d_x)$ for all $(x,a) \in \mathcal{X} \times \mathcal{A}$.[22] Define a system of differential equations in an unknown function $\epsilon(x,a)$ as follows

$$\left(\frac{\nabla_x \epsilon(x,a)'}{\epsilon(x,a)}, \quad \frac{\epsilon_a(x,a)}{\epsilon(x,a)}\right)' = [B(x,a)'B(x,a)]^{-1}[B(x,a)'A(x,a)], \tag{2.3.9}$$

which depends only on the identified reduced form functions $I^m(x,a)$, $h^m(x,a)$ and their derivatives. Then the identification of the quality function $e(x,a)$ amounts to a unique solution to the differential equations in (2.3.9).

First fix $(x_2,\ldots,x_{d_x},a) = (\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})$, and only consider the first equation in (2.3.9). Note that

$$\frac{e_{x_1}(x_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})}{e(x_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})} = \frac{d\log\left(e(x_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})\right)}{dx_1} = g_1(x_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a}).$$

[22] In fact, it also requires that the vector $A(x,a)$ lies in the space spanned by the column vectors of $B(x,a)$ for all $(x,a) \in \mathcal{X} \times \mathcal{A}$, but it is implied by Assumption 2.2.1.

Then,

$$\begin{aligned}\log\left(e(x_1, \bar{x}_2, \ldots, \bar{x}_{d_x}, \bar{a})\right) &= \int_{\bar{x}_1}^{x_1} g_1(s_1, \bar{x}_2, \ldots, \bar{x}_{d_x}, \bar{a}) ds_1 + \log\left(e(\bar{x}, \bar{a})\right) \\ &= \int_{\bar{x}_1}^{x_1} g_1(s_1, \bar{x}_2, \ldots, \bar{x}_{d_x}, \bar{a}) ds_1, \qquad (2.3.10)\end{aligned}$$

for all $x_1 \in \mathcal{X}_1$, where the second equality holds by Assumption 2.3.3. Then, consider the second equation in (2.3.9). Similarly, for any given $x_1 \in \mathcal{X}_1$ and fixed $(x_3, \ldots, x_{d_x}, a) = (\bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a})$, we have

$$\frac{e_{x_2}(x_1, x_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a})}{e(x_1, x_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a})} = \frac{d \log\left(e(x_1, x_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a})\right)}{dx_2} = g_2(x_1, x_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a}),$$

which implies

$$\begin{aligned}&\log\left(e(x_1, x_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a})\right) \\ = &\int_{\bar{x}_2}^{x_2} g_2(x_1, s_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a}) ds_2 + \log\left(e(x_1, \bar{x}_2, \ldots, \bar{x}_{d_x}, \bar{a})\right) \\ = &\int_{\bar{x}_2}^{x_2} g_2(x_1, s_2, \bar{x}_3, \ldots, \bar{x}_{d_x}, \bar{a}) ds_2 + \int_{\bar{x}_1}^{x_1} g_1(s_1, \bar{x}_2, \ldots, \bar{x}_{d_x}, \bar{a}) ds_1,\end{aligned}$$

for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. Continue to integrate over $(x_3, \ldots, x_{d_x}, a)$ once at a time in this manner, one will eventually obtain the solution to the initial value problem in equation (2.3.9) and $e(\bar{x}, \bar{a}) = 1$ as

$$e(x, a) = \exp\left(\sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} g_j(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a}) ds_j + \int_{\bar{a}}^{a} g_{d_x+1}(x, t) dt\right).$$

Moreover, this solution is unique by the first fundamental theorem of calculus. This completes the proof of the theorem. $\square$

Define the range of equilibrium effective amount supplied in market m as

$$\begin{aligned}\mathcal{Z}_s^m &= \{z \in \mathcal{Z}\text{: there exists some } (x,a) \in \mathcal{X} \times \mathcal{A} \text{ in market} \\ &\qquad m \in \mathcal{M} \text{ such that in equilibrium } z = h^m(x,a) \cdot e(x,a)\}.\end{aligned}$$

Corollary 2.3.5. *Under the conditions for Theorem 2.3.4, the unobserved effective amount* $z = h^m(x,a) \cdot e(x,a)$ *is identified.*

Corollary 2.3.6. *Under the conditions for Theorem 2.3.4, the price schedule function* $P^m(z)$ *for market* $m \in \mathcal{M}$ *is nonparametrically identified on* $\mathcal{Z}_s^m$*.*

Proof. Assumption 2.2.1, earnings equation (2.2.3), Lemmas 2.3.1 and 2.3.3, and Theorem 2.3.4 together imply the result. $\square$

2.3.3 Identification of Sellers' Marginal Disutility Function $U_h(h,x,a)$ Using Multiple Market Data

The next important result is the identification of the marginal disutility function U_h. Before stating the theorem, define the equilibrium support for sellers' marginal disutility function as:

$$\mathcal{HXA} = \{(h,x,a) \in \mathcal{H} \times \mathcal{X} \times \mathcal{A}\text{: there exists a market } m \in \mathcal{M} \text{ and}$$

some $(x, a) \in \mathcal{X} \times \mathcal{A}$ such that in equilibrium $h = h^m(x, a)\}$.

If $|\mathcal{M}| = 1$, then $\mathcal{HXA}$ is degenerate since h is endogenous. As discussed in Section 2.2.1, different distributions f_x^m and f_y^m serve as aggregate supply or demand shifters (i.e., instruments) that induce variation in P^m (and hence $h^m(x, a)$) while maintaining individual values of (x, a). The richer the variation in f_x^m and f_b^m, the larger the set $\mathcal{HXA}$ will be.

Theorem 2.3.7. *Under the conditions for Theorem 2.3.4, the sellers' marginal disutility function $U_h(h, x, a)$ is nonparametrically identified on $\mathcal{HXA}$.*

Proof. The result follows from Theorem 2.3.4, Corollary 2.3.6, and the sellers' FOC

$$P_z^m(h^m(x, a) \cdot e(x, a)) \cdot e(x, a) = U_h(h^m(x, a), x, a)$$

in each market $m \in \mathcal{M}$. □

2.3.4 Identification of Buyers' Marginal Utility Function $R_z(z, y, b)$ Using Multiple Market Data

Identifying buyers' marginal utility function $R_z(z, y, b)$ and the effective amount demand function $d^m(y, b)$ makes little difference from Heckman, Matzkin, and Nesheim (2010)'s method. The only tweak stems from the fact that z is not directly observed. Once one recovers z from the supply side, Heckman, Matzkin, and Nesheim (2010)'s method can be applied without modification. The relevant definition, assumption and results are given below.

Define the equilibrium support for buyers' marginal utility function $R_z(z, y, b)$ as:

$$\begin{aligned}\mathcal{ZYB} &= \{(z, y, b) \in \mathcal{Z} \times \mathcal{Y} \times \mathcal{B}\text{: there exists a market } m \in \mathcal{M} \\ &\quad \text{and some } (x, a) \in \mathcal{X} \times \mathcal{A} \text{ such that } z = d^m(y, b) \\ &\quad \text{and } z = h^m(x, a) \cdot e(x, a) \text{ in equilibrium}\}.\end{aligned}$$

Assumption 2.3.6. *Suppose that the buyers' unobserved characteristic b follows the uniform distribution $U[0, 1]$ in all markets.*

Lemma 2.3.8. ***(Heckman, Matzkin, and Nesheim 2010 Theorem 4.1)*** *Under Assumption 2.3.6 and the conditions for Theorem 2.3.4, the buyers' marginal utility function $R_z(z, y, b)$ is nonparametrically identified on $\mathcal{ZYB}$.*[23]

2.4 Estimation

This section provides an estimation procedure for the structural functions. Section 2.4.1 describes the estimation procedure step by step, and in Section 2.4.2 I derive the uniform rates of convergence for the estimators.

2.4.1 Series Estimation of Structural Functions

The estimators introduced in this section are premised on the following data structure. Suppose that linked seller-buyer data for M independent markets are available. Within

[23] In labor markets, if the firms' revenue is observed by researchers, then the function $R(z, y, b)$ is also nonparametrically identified under the conditions of Lemma 3.

each market m, suppose that there are N^m seller-buyer pairs, and each pair is indexed by i. Researchers observe which seller is matched with which buyer. For each pair i ($i = 1, \ldots, N^m$ and $m = 1, \ldots, M$), researchers observe $(I_i^m, x_i^m, h_i^m, y_i^m)$.[24]

Assumption 2.4.1. *Suppose* $\{(I_i^m, x_i^m, h_i^m, y_i^m)\}_{i=1}^{N^m}$ *are i.i.d. for* $m = 1, \ldots, M$.

Assumption 2.4.2. *For notational simplicity, suppose that the sample sizes from all the markets are equal, i.e.,* $N^1 = N^2 = \cdots = N^M = N$.

In the rest of this chapter, I maintain Assumptions 2.2.1-2.4.2. Assumption 2.4.2 is not essential for deriving the convergence rates, but relaxing it will complicate the notation and will not provide any new insights. In principle, even though the sample sizes from all the markets are the same, one still could use market-specific numbers of series basis functions $k_{Q,N}^m$, $k_{I,N}^m$ and $k_{h,N}^m$ to estimate $\hat{a}^m$, $\hat{I}^m(x, a)$ and $\hat{h}^m(x, a)$ respectively within each market. To keep the notation simple, however, I assume that one uses the same tuning parameters for all markets for the rest of the chapter, i.e., $k_{Q,N}^m = k_{Q,N}$, $k_{I,N}^m = k_{I,N}$ and $k_{h,N}^m = k_{h,N}$. All the convergence rate results in Section 2.4.2 hold if one relaxes this assumption.[25]

For any vector v, let $\|v\| \equiv (v'v)^{1/2}$ denote its Euclidean norm; for any matrix A, let $\|A\| \equiv [\text{trace}(A'A)]^{1/2}$ denote its Euclidean norm.

The estimation of the structural functions (U_h, e, R_z) follows the steps suggested by the identification strategy. I start with the within market estimation of two reduced form functions, namely, the payment function $I^m(x, a)$ and the quantity function $h^m(x, a)$, as well as their partial derivative functions for each market. Then in light of the proof of Theorem

[24]In labor markets, it is possible that the employers' revenue R_i^m is also observed in the data.

[25]With minor changes in notation to accommodate market-specific tuning parameters.

2.3.4, the quality function $e(x, a)$ can be estimated by first solving an estimated version of the equations (2.3.8) and then integrating over x and a. Finally, sellers' marginal disutility function $U_h(h, x, a)$ can be estimated by a series minimum distance (MD) estimator using the sellers' FOCs.

Following the identification steps in Section 2.3, this section describes the steps for estimating $e(x, a)$ and $U_h(h, x, a)$ in details. The steps for the buyers' marginal utility function $R_z(z, y, b)$ are similar and will be briefly summarized at the end.

2.4.1.1 Estimation of Payment Functions $I^m(x, a)$ and Quantity Functions $h^m(x, a)$ Using Single Market Data

Let me first clarify some notation used in this section: $I^m(x, a)$ and $h^m(x, a)$ indicate the reduced form functions; I^m (or h^m, x^m, or a^m) is a random variable, denoting the payment received by (or the quantity supplied by, the observed characteristics of, or the unobserved characteristic of) a randomly chosen seller from market m; and I_i^m (or h_i^m, x_i^m, or a_i^m) represents the observed payment (or the observed quantity, the observed characteristics, or the unobserved characteristic) value of a specific seller i in market m.

In Section 2.2.1, I showed that the payment function $I^m(x, a)$ is strictly increasing in a under Assumptions 2.2.1-2.2.4. Recall that a^m is the conditional quantile of the payment I^m given observed characteristics x^m of the seller in market m. That is

$$F_{I^m|x^m=x}(I^m(x, a)) = F_{a^m}(a) = a.$$

Use a series of basis functions $\Lambda_{k_{Q,N}}(x) \equiv (\lambda_1(x), \ldots, \lambda_{k_{Q,N}}(x))'$ to approximate the indicator function $\mathbb{I}(I^m \leq I_i^m)$, where $k_{Q,N}$ is the number of basis functions. Then one can estimate a_i^m, the conditional quantile of I^m given x^m by

$$\begin{aligned}
\hat{a}_i^m &\equiv \hat{F}_{I^m|x^m=x_i^m}(I_i^m) \\
&\equiv \Lambda_{k_{Q,N}}(x_i^m)' \left(\sum_{j=1}^{N} \Lambda_{k_{Q,N}}(x_j^m) \Lambda_{k_{Q,N}}(x_j^m)' \right)^{-} \left(\sum_{j=1}^{N} \Lambda_{k_{Q,N}}(x_j^m) \mathbb{I}(I_j^m \leq I_i^m) \right) \quad (2.4.1)
\end{aligned}$$

Note that the tuning parameter $k_{Q,N}$ might depend on the sample size N. Here, $\hat{a}_i^m$ serves as a generated regressor when we estimate functions $I^m(x,a)$ and $h^m(x,a)$.

Use a series of basis functions $\Phi_{k_{I,N}}(x,a) \equiv (\phi_1(x,a), \ldots, \phi_{k_{I,N}}(x,a))'$ to approximate the unknown payment function $I^m(x,a)$, where $k_{I,N}$ is the number of basis functions. Then, the estimated series coefficients for the payment function $I^m(x,a)$ are the solution to the following least square problem

$$\hat{\xi}_{I,k_{I,N}}^m \equiv \arg \min_{\xi \in \mathbb{R}^{k_{I,N}}} \sum_{i=1}^{N} \left(I_i^m - \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m)' \xi \right)^2 .$$

Therefore, the estimated payment function is

$$\hat{I}^m(x,a) \equiv \Phi_{k_{I,N}}(x,a)' \hat{\xi}_{I,k_{I,N}}^m . \quad (2.4.2)$$

Note that there is an explicit solution for $\hat{\xi}_{I,k_{I,N}}^m$,

$$\hat{\xi}_{I,k_{I,N}}^m = \left(\sum_{i=1}^{N} \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m) \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m)' \right)^{-} \left(\sum_{i=1}^{N} \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m) I_i^m \right) . \quad (2.4.3)$$

Because $\Phi_{k_{I,N}}(x,a)$ is a series of known functions, their first-order derivatives are also known. Therefore, the series estimator of the partial derivatives of $I^m(x,a)$ can be obtained immediately

$$\hat{I}^m_{x_j}(x,a) \equiv \left(\frac{\partial\phi_1(x,a)}{\partial x_j},\ldots,\frac{\partial\phi_{k_{I,N}}(x,a)}{\partial x_j}\right)\hat{\xi}^m_{I,k_{I,N}}, \tag{2.4.4}$$

for $j = 1,\ldots,d_x$, and

$$\hat{I}^m_a(x,a) \equiv \left(\frac{\partial\phi_1(x,a)}{\partial a},\ldots,\frac{\partial\phi_{k_{I,N}}(x,a)}{\partial a}\right)\hat{\xi}^m_{I,k_{I,N}}. \tag{2.4.5}$$

Similarly, use the series of basis functions $\Phi_{k_{h,N}}(x,a) \equiv (\phi_1(x,a),\ldots,\phi_{k_{h,N}}(x,a))'$ to approximate the unknown quantity function $h^m(x,a)$.

Then the estimated series coefficients for the quantity function $h^m(x,a)$ is

$$\hat{\xi}^m_{h,k_{h,N}} \equiv \left(\sum_{i=1}^N \Phi_{k_{h,N}}(x^m_i,\hat{a}^m_i)\Phi_{k_{h,N}}(x^m_i,\hat{a}^m_i)'\right)^- \left(\sum_{i=1}^N \Phi_{k_{h,N}}(x^m_i,\hat{a}^m_i)h^m_i\right). \tag{2.4.6}$$

Therefore, the estimated quantity function and its first-order derivatives are

$$\hat{h}^m(x,a) \equiv \Phi_{k_{h,N}}(x,a)'\hat{\xi}^m_{h,k_{h,N}}, \tag{2.4.7}$$

$$\hat{h}^m_a(x,a) \equiv \left(\frac{\partial\phi_1(x,a)}{\partial a},\ldots,\frac{\partial\phi_{k_{h,N}}(x,a)}{\partial a}\right)\hat{\xi}^m_{h,k_{h,N}}, \tag{2.4.8}$$

and

$$\hat{h}^m_{x_j}(x,a) \equiv \left(\frac{\partial\phi_1(x,a)}{\partial x_j},\ldots,\frac{\partial\phi_{k_{h,N}}(x,a)}{\partial x_j}\right)\hat{\xi}^m_{h,k_{h,N}}. \tag{2.4.9}$$

for $j = 1, \ldots, d_x$.

2.4.1.2 Estimation of Quality Function $e(x, a)$ Using Multiple Market Data

Just like the identification strategy, estimating the quality function $e(x, a)$ starts with the system of equations (2.3.8). Replace $I^m(x, a)$, $h^m(x, a)$ and their derivatives in equation (2.3.8) (i.e., in the expressions of $B(x, a)$ and $A(x, a)$) with their counterparts estimated in Section 2.4.1.1. Use the series of basis functions $\Phi_{k_{x_j,MN}}(x, a) = (\phi_1(x, a), \ldots, \phi_{k_{x_j,MN}}(x, a))'$ to approximate $e_{x_j}(x, a)/e(x, a)$ and $\Phi_{k_{a,MN}}(x, a) = (\phi_1(x, a), \ldots, \phi_{k_{a,MN}}(x, a))'$ to approximate $e_a(x, a)/e(x, a)$. Let the series coefficients be $\beta_{x_j,k_{x_j,MN}}$ $(j = 1, \ldots, d_x)$ and $\beta_{a,k_{a,MN}}$, respectively. And let $\beta_{MN} \equiv (\beta'_{x_1,k_{x_1,MN}}, \ldots, \beta'_{x_{d_x},k_{x_{d_x},MN}}, \beta'_{a,k_{a,MN}})'$. Then, for each seller i and each market m, one obtains an estimated version of the equations (2.3.8) as follows:

$$\hat{B}^m(x_i^m, \hat{a}_i^m) \cdot \begin{pmatrix} \Phi_{k_{x_1,MN}}(x_i^m, \hat{a}_i^m)'\hat{\beta}_{x_1,k_{x_1,MN}} \\ \vdots \\ \Phi_{k_{x_{d_x},MN}}(x_i^m, \hat{a}_i^m)'\hat{\beta}_{x_{d_x},k_{x_{d_x},MN}} \\ \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'\hat{\beta}_{a,k_{a,MN}} \end{pmatrix} = \hat{A}^m(x_i^m, \hat{a}_i^m),$$

where the $d_x \times (d_x + 1)$ matrix $\hat{B}^m(x_i^m, \hat{a}_i^m)$ is

$$\hat{B}^m(x_i^m, \hat{a}_i^m) \equiv \Big(\mathbb{I}_{d_x} \otimes \hat{I}_a^m(x_i^m, \hat{a}_i^m)\ ,\ \ -\nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m)\Big),$$

and the $d_x \times 1$ vector $\hat{A}^m(x_i^m, \hat{a}_i^m)$ is

$$\hat{A}^m(x_i^m, \hat{a}_i^m) \equiv \left[\hat{h}_a^m(x_i^m, \hat{a}_i^m)\nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m) - \hat{I}_a^m(x_i^m, \hat{a}_i^m)\nabla_x \hat{h}^m(x_i^m, \hat{a}_i^m)\right] \Big/ \hat{h}^m(x_i^m, \hat{a}_i^m).$$

Therefore, the estimated series coefficients are the solutions to the following least square problem

$$\hat{\beta}_{MN} \equiv \arg\min_{\beta} \sum_{m=1}^{M} \sum_{i=1}^{N} LS(x_i^m, \hat{a}_i^m; \beta),$$

where

$$LS(x_i^m, \hat{a}_i^m; \beta) \equiv \left\| \hat{B}^m(x_i^m, \hat{a}_i^m) \cdot \begin{pmatrix} \Phi_{k_{x_1,MN}}(x_i^m, \hat{a}_i^m)'\beta_{x_1} \\ \vdots \\ \Phi_{k_{x_{d_x},MN}}(x_i^m, \hat{a}_i^m)'\beta_{x_{d_x}} \\ \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'\beta_a \end{pmatrix} - \hat{A}^m(x_i^m, \hat{a}_i^m) \right\|^2 .$$

There is an explicit expression for $\hat{\beta}_{MN}$ as follows:

$$\hat{\beta}_{MN} = \hat{S}_{\Phi\Phi}^{-} \hat{S}_{\Phi A},$$

where

$$\hat{S}_{\Phi\Phi} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m), \tag{2.4.10}$$

$$\hat{S}_{\Phi A} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \hat{A}^m(x_i^m, \hat{a}_i^m). \tag{2.4.11}$$

In equations (2.4.10) and (2.4.11),

$$\hat{S}_{\Phi}(x_i^m, \hat{a}_i^m) \equiv \left(\hat{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) \ , \ \hat{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) \right),$$

where

$$\hat{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) \equiv \begin{pmatrix} \Phi_{k_{x_1,MN}}(x_i^m, \hat{a}_i^m)' & & 0 \\ & \ddots & \\ 0 & & \Phi_{k_{x_{d_x}},MN}(x_i^m, \hat{a}_i^m)' \end{pmatrix} \otimes \hat{I}_a^m(x_i^m, \hat{a}_i^m),$$

$$\hat{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) \equiv -\nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m) \otimes \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'.$$

Then the estimated ratios of the quality function are

$$\begin{cases} \widehat{\frac{e_{x_1}(x,a)}{e(x,a)}} & \equiv \hat{g}_1(x,a) = \Phi_{k_{x_1,MN}}(x,a)' \hat{\beta}_{x_1,k_{x_1,MN}}, \\ \vdots & \vdots \\ \widehat{\frac{e_{x_{d_x}}(x,a)}{e(x,a)}} & \equiv \hat{g}_{d_x}(x,a) = \Phi_{k_{x_{d_x}},MN}(x,a)' \hat{\beta}_{x_{d_x},k_{x_{d_x}},MN}, \\ \widehat{\frac{e_a(x,a)}{e(x,a)}} & \equiv \hat{g}_{d_x+1}(x,a) = \Phi_{k_{a,MN}}(x,a)' \hat{\beta}_{a,k_{a,MN}}. \end{cases} \tag{2.4.12}$$

By replacing the relevant ratios of the quality function in equation (2.3.5) with their estimators given in equation (2.4.12), one obtains the estimator of the quality function

$$\begin{aligned}\hat{e}(x,a) &= \exp\left(\sum_{j=1}^{d_x}\int_{\bar{x}_j}^{x_j}\hat{g}_j(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})ds_j + \int_{\bar{a}}^{a}\hat{g}_{d_x+1}(x,t)dt\right)\\ &= \exp\left(\sum_{j=1}^{d_x}\int_{\bar{x}_j}^{x_j}\left[\Phi_{k_{x_j,MN}}(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})'\hat{\beta}_{x_j,k_{x_j,MN}}\right]ds_j\right.\\ &\quad \left.+\int_{\bar{a}}^{a}\left[\Phi_{k_{a,MN}}(x,t)'\hat{\beta}_{a,k_{a,MN}}\right]dt\right). \qquad (2.4.13)\end{aligned}$$

2.4.1.3 Estimation of Sellers' Marginal Disutility Function $\hat{U}_h(h,x,a)$ Using Multiple Market Data

Estimation of the sellers' marginal disutility function starts from the partial derivatives of the payment equation (2.3.6). Combined with the sellers' FOC in equation (2.2.1), they imply that for $\forall(x,a)\in\mathcal{X}\times\mathcal{A}$,

$$\begin{cases}\nabla_x I^m(x,a) &= \left[\nabla_x h^m(x,a) + h^m(x,a)\frac{\nabla_x e(x,a)}{e(x,a)}\right]\cdot U_h(h^m(x,a),x,a),\\ I_a^m(x,a) &= \left[h_a^m(x,a) + h^m(x,a)\frac{e_a(x,a)}{e(x,a)}\right]\cdot U_h(h^m(x,a),x,a).\end{cases} \qquad (2.4.14)$$

Now, use a series of basis functions $\Psi_{k_{U,MN}}(h,x,a) \equiv (\psi_1(h,x,a),\ldots,\psi_{k_{U,MN}}(h,x,a))'$ to approximate the unknown marginal disutility function. Then, one wants to choose the series coefficients $\hat{\gamma}_{k_{U,MN}}$ to minimize the sum of the squared distances between the left-hand sides and the right-hand sides of the equations (2.4.14). Specifically, define

$$G_{x,MN}(h_i^m, x_i^m, \hat{a}_i^m;\gamma)$$

$$\equiv \left[\nabla_x \hat{h}^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{\nabla_x \widehat{e(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma - \nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m),$$

and

$$\begin{aligned} & G_{a,MN}(h_i^m, x_i^m, \hat{a}_i^m; \gamma) \\ \equiv & \left[\hat{h}_a^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{e_a\widehat{(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma - \hat{I}_a^m(x_i^m, \hat{a}_i^m). \end{aligned}$$

And the minimum distance (MD) estimator of the series coefficients are defined as

$$\hat{\gamma}_{k_{U,MN}} \equiv \arg \min_{\gamma \in \mathbb{R}^{k_{U,MN}}} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\| \begin{pmatrix} G_{x,MN}(h_i^m, x_i^m, \hat{a}_i^m; \gamma) \\ G_{a,MN}(h_i^m, x_i^m, \hat{a}_i^m; \gamma) \end{pmatrix} \right\|^2.$$

The estimator $\hat{\gamma}_{k_{U,MN}}$ has a closed-form expression given by

$$\hat{\gamma}_{k_{U,MN}} = \hat{S}_{\Psi\Psi}^{-} \hat{S}_{\Psi I}$$

where

$$\hat{S}_{\Psi\Psi} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)' \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m), \tag{2.4.15}$$

$$\hat{S}_{\Psi I} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)' \hat{S}_{I}(h_i^m, x_i^m, \hat{a}_i^m). \tag{2.4.16}$$

In equations (2.4.15) and (2.4.16), the $(d_x+1)\times k_{U,MN}$ matrix $\hat{S}_\Psi(h_i^m, x_i^m, \hat{a}_i^m)$ is

$$\hat{S}_\Psi(h_i^m, x_i^m, \hat{a}_i^m) \equiv \begin{pmatrix} \left[\nabla_x \hat{h}^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{\widehat{\nabla_x e(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)}\right] \otimes \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \\ \left[\hat{h}_a^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{\widehat{e_a(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \end{pmatrix}$$

and the $(d_x+1)\times 1$ vector $\hat{S}_I(h_i^m, x_i^m, \hat{a}_i^m) \equiv (\nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m)', \hat{I}_a^m(x_i^m, \hat{a}_i^m))'$. As a result, the estimated sellers' marginal disutility function is

$$\hat{U}_h(h, x, a) \equiv \Psi_{k_{U,MN}}(h, x, a)'\hat{\gamma}_{k_{U,MN}}.$$

The steps described in Sections 2.4.1.1-2.4.1.3 complete the estimation of the supply side structural functions (e, U_h).

2.4.1.4 Estimation of Buyers' Marginal Utility Function $\hat{R}_z(z, y, b)$ Using Multiple Market Data

The buyers' marginal utility function R_z can be estimated by similar steps. First, within each market m, estimate the conditional quantile b_i^m of the payment I_i^m using a formula similar to equation (2.4.1). The unobserved effective amounts can also be estimated as $\hat{z}_i^m \equiv h_i^m \cdot \hat{e}(x_i^m, \hat{a}_i^m)$, since researchers observe which seller is matched with which buyer. Second, estimate the reduced form payment function $I^m(y, b)$ and effective amount demand function $d^m(y, b)$ using the generated regressor $\hat{b}_i^m$ and generated dependent variable $\hat{z}_i^m$ from the single market m. Third, taking the partial derivatives of the payment equation for the buyers yields

$$\begin{aligned}\nabla_y I^m(y,b) &= P_z^m(d^m(y,b)) \cdot \nabla_y d^m(y,b), \\ I_a^m(y,b) &= P_z^m(d^m(y,b)) \cdot d_a^m(y,b).\end{aligned}$$

Combine these equations with the buyers' FOC in equation (2.2.4), and use a series of basis functions $\Theta_{k_N}(z,y,b) \equiv (\theta_1(z,y,b),\ldots,\theta_{k_N}(z,y,b))$ to approximate the unknown buyers' marginal utility function $R_z(z,y,b)$. Then, the function can be estimated by an MD estimator similar to that in Section 2.4.1.3. Moreover, if the buyers' utility values R_i^m are observed,[26] then the second and third steps are not necessary. The series estimation of R and its derivative functions boils down to a linear regression of R_i^m on $\Theta_{k_N}(\hat{z}_i^m, y_i^m, \hat{b}_i^m)$ using multiple market data.

2.4.2 Uniform Rates of Convergence of Structural Function Estimators

In this section and Appendix 2.D, C denotes a sufficiently large, generic positive constant, and c denotes a sufficiently small, generic positive constant, both of which may take different values in different uses.

[26] For example, firm revenue in labor markets.

2.4.2.1 Unobserved Heterogeneity Estimators $\hat{a}_i^m$

This subsection derives the convergence rates of the within market series estimators of the conditional quantile a_i^m given in equation (2.4.1).

Assumption 2.4.3. *Suppose that $F_{I^m|x^m}(I|x) \equiv F_{I^m|x^m=x}(I)$ is continuously differentiable of order $d_1 > d_x$ on the support with derivatives uniformly bounded in I and x.*

Define

$$\nu_{a,N} \equiv \left(\frac{k_{Q,N}}{N} + k_{Q,N}^{1-2d_1/d_x} \right)^{1/2}.$$

And I will assume that $k_{Q,N}/N \to 0$ and $k_{Q,N}$.

Theorem 2.4.1. *Suppose that Assumption 2.4.3 is satisfied. Then,*

$$\sum_{i=1}^{N} |\hat{a}_i^m - a_i^m|^2/N = \mathcal{O}\left(\nu_{a,N}^2\right).$$

2.4.2.2 Payment Function Estimators $\hat{I}^m(x,a)$ and Quantity Function Estimators $\hat{h}^m(x,a)$

This subsection derives the convergence rates of the within market series estimators of the reduced form payment functions $I^m(x,a)$ and quantity functions $h^m(x,a)$ and their first-order derivatives.

Assumption 2.4.4. *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Cartesian products of closed intervals.*

Assumption 2.4.5. *Suppose that* $\Phi_k(x,a) = \Phi_{1,k_1}(x_1) \otimes \cdots \otimes \Phi_{d_x,k_{d_x}}(x_{d_x}) \otimes \Phi_{a,k_a}(a)$*. This implies that* $k = k_a \cdot \prod_{j=1}^{d_x} k_j$*.*

In Assumption 2.4.5, if k denotes the number of series basis functions used to approximate an unknown function of (x,a) (or of (h,x,a)), then let k_h, k_j and k_a denote the numbers of series basis functions used to approximate the h component, x_j component and a component in the Cartesian space, respectively.

Let $\zeta_0(k) \equiv k$, $\zeta_a(k) \equiv k_a^2 k$, and $\zeta_j(k) \equiv k_j^2 k$.

Assumption 2.4.6. *Suppose that for all* $m \in \mathcal{M}$*,* $I^m(x,a)$ *and* $h^m(x,a)$ *are continuously differentiable of order* $d \geq 2$ *on the support.*[27]

For a function $l(x,a) : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$, define the norm $|l|_\delta$ as $|l|_\delta \equiv \max_{|\mu| \leq \delta} \sup_{(x,a) \in \mathcal{X} \times \mathcal{A}}$ $|\partial^\mu l(x,a)/\partial x_1^{\mu_1} \cdots \partial x_{d_x}^{\mu_{d_x}} \partial a^{\mu_a}|$, with $\mu_1 + \cdots + \mu_{d_x} + \mu_a = \mu$ ($\mu_1, \ldots, \mu_{d_x}, \mu_a$ are integers).

One implication of Assumptions 2.3.1, 2.3.2 and 2.4.6 is that there exist some positive constants B_I and B_{hu} such that for all $m \in \mathcal{M}$, $|I^m|_2 \leq B_I$, and $|h^m|_2 \leq B_{hu}$.

Suppose that the following assumption about the approximation error by the basis functions holds.

Assumption 2.4.7. *Suppose that for a positive integer* $\delta_I \geq 1$*, there exist a constant* $\alpha_I > 0$ *and pseudo-true series coefficients* $\xi^m_{0,I,k_I} \in \mathbb{R}^{k_I}$ *such that* $|I^m - \Phi'_{k_I} \xi^m_{0,I,k_I}|_{\delta_I} \leq C k_I^{-\alpha_I}$ *for all positive integers* k_I*. Suppose as well that for a positive integer* $\delta_h \geq 1$*, there exist a constant* $\alpha_h > 0$ *and pseudo-true series coefficients* $\xi^m_{0,h,k_h} \in \mathbb{R}^{k_h}$ *such that* $|h^m - \Phi'_{k_h} \xi^m_{0,h,k_h}|_{\delta_h} \leq C k_h^{-\alpha_h}$ *for all positive integers* k_h*.*[28]

[27] Without loss of generality, here I assume that d is the same across all markets $m \in \mathcal{M}$.

[28] Without loss of generality, here I assume that α_I and α_h are the same across all markets $m \in \mathcal{M}$.

Let $l^m(x,a)$ denote either the payment function $I^m(x,a)$ or the quantity function $h^m(x,a)$. Let $\hat{l}^m(x,a)$ denote the series estimator of $l^m(x,a)$ defined in equation (2.4.2) or equation (2.4.7), and let $\hat{l}^m_{x_j}(x,a)$ $(j=1,\ldots,d_x)$ and $\hat{l}^m_a(x,a)$ denote the series estimators of the first-order derivatives of $l^m(x,a)$ defined in equation (2.4.4), equation (2.4.5), equation (2.4.8) or equation (2.4.9).

Define

$$\begin{aligned}
\nu_{l,N} &\equiv \zeta_0(k_{l,N})\left(\nu_{a,N}+k_{l,N}^{-\alpha_l}\right),\\
\nu_{l_j,N} &\equiv \zeta_j(k_{l,N})\left(\nu_{a,N}+k_{l,N}^{-\alpha_l}\right),\\
\nu_{l_a,N} &\equiv \zeta_a(k_{l,N})\left(\nu_{a,N}+k_{l,N}^{-\alpha_l}\right).
\end{aligned}$$

And I will assume that $\nu_{l,N}\to 0$, $\nu_{l_j,N}\to 0$ and $\nu_{l_a,N}\to 0$ as $N\to\infty$ for the rest of the chapter. Moreover, note that $\nu_{l,N}=\mathcal{O}(\nu_{l_j,N})$, and $\nu_{l,N}=\mathcal{O}(\nu_{l_a,N})$.

Theorem 2.4.2. *Suppose that Assumptions 2.4.4-2.4.7 and the conditions of Theorem 2.4.1 are satisfied. Suppose as well that the numbers of series basis functions used to approximate each component in $k_{l,N}$ all increase to infinity with N, and $\sqrt{k_{l,N}}\nu_{a,N}\zeta_a(k_{l,N})\to 0$ as $N\to\infty$. Then*

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}|\hat{l}^m(x,a)-l^m(x,a)| \;=\; \mathcal{O}_p\left(\nu_{l,N}\right).$$

Theorem 2.4.3. *Suppose that the conditions for Theorem 2.4.2 are satisfied . Then*

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left|\hat{l}^m_{x_j}(x,a) - l^m_{x_j}(x,a)\right| = \mathcal{O}_p\left(\nu_{l_j,N}\right),$$

and

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left|\hat{l}^m_{a}(x,a) - l^m_{a}(x,a)\right| = \mathcal{O}_p\left(\nu_{l_a,N}\right).$$

Since $\hat{a}^m_i$ is used as a generated regressor,[29] the convergence rates of the reduced form functions and their derivatives depend on the estimation errors of $\hat{a}^m_i$ as well as on the series approximation errors of the functions themselves.

2.4.2.3 Quality Function Estimator $\hat{e}(x,a)$

This subsection derives the convergence rates of the cross-market series estimators of the quality function $e(x,a)$ and its first-order derivative ratios.

Assumption 2.4.8. *Suppose that for a positive integer $\delta_e \geq 0$, there exist a constant $\alpha_e > 0$ and pseudo-true series coefficients $\beta_{0,x_j,k_{x_j}} \in \mathbb{R}^{k_{x_j}}$ (for $j = 1,\dots,d_x$) and $\beta_{0,a,k_a} \in \mathbb{R}^{k_a}$ such that $|e_{x_j}/e - \Phi'_{k_{x_j}}\beta_{0,x_j,k_{x_j}}|_{\delta_e} \leq Ck_{x_j}^{-\alpha_e}$ and $|e_a/e - \Phi'_{k_a}\beta_{0,a,k_a}|_{\delta_e} \leq Ck_a^{-\alpha_e}$ for all positive integers k_{x_j} ($j = 1,\dots,d_x$) and k_a.*

Define

$$S_{\Phi\Phi} \equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N} S_\Phi(x^m_i,a^m_i)'S_\Phi(x^m_i,a^m_i),$$

[29]Recall equations (2.4.3) and (2.4.6).

where

$$S_\Phi(x_i^m, a_i^m) \equiv (S_{\Phi,1}(x_i^m, a_i^m)\ ,\ S_{\Phi,2}(x_i^m, a_i^m)),$$

$$S_{\Phi,1}(x_i^m, a_i^m) \equiv \begin{pmatrix} I_a^m(x_i^m, a_i^m)\Phi_{k_{x_1,MN}}(x_i^m, a_i^m)' & & 0 \\ & \ddots & \\ 0 & & I_a^m(x_i^m, a_i^m)\Phi_{k_{x_{d_x},MN}}(x_i^m, a_i^m)' \end{pmatrix},$$

and

$$S_{\Phi,2}(x_i^m, a_i^m) \equiv -\nabla_x I^m(x_i^m, a_i^m) \otimes \Phi_{k_{a,MN}}(x_i^m, a_i^m)'.$$

Assumption 2.4.9. *Suppose that there exist some positive constants B_{eu} and B_{el} such that the quality function $e(x, a)$ satisfies $|e|_2 \leq B_{eu}$ and $|e|_0 \geq B_{el}$.*

Assumption 2.4.10. *Suppose:*

(i) $\lambda_{\min}(\mathbb{E}(S_{\Phi\Phi})) \geq c > 0$;

(ii) There exists some positive constant B_{hl} such that for all $m \in \mathcal{M}$, $|h^m|_0 \geq B_{hl}$.

For $j = 1, \ldots, d_x$, define

$$\nu_{e_j,M,N} \equiv \zeta_0(k_{x_j,MN}) \left[\sum_{j=1}^{d_x} \nu_{h_j,N} + \nu_{h_a,N} + \sum_{j=1}^{d_x} \nu_{I_j,N} + \nu_{I_a,N} + k_{a,MN}^{-\alpha_e} + \sum_{j=1}^{d_x} k_{x_j,MN}^{-\alpha_e} \right],$$

and

$$\nu_{e_a,M,N} \equiv \zeta_0(k_{a,MN}) \left[\sum_{j=1}^{d_x} \nu_{h_j,N} + \nu_{h_a,N} + \sum_{j=1}^{d_x} \nu_{I_j,N} + \nu_{I_a,N} + k_{a,MN}^{-\alpha_e} + \sum_{j=1}^{d_x} k_{x_j,MN}^{-\alpha_e} \right].$$

And I will assume that $\nu_{e_j,M,N} \to 0$ $(j = 1, \ldots, d_x)$ and $\nu_{e_a,M,N} \to 0$ as $N \to \infty$ for the rest of the chapter.

Lemma 2.4.4. *Suppose that Assumptions 2.4.8-2.4.10 and the conditions of Theorems 2.4.1-2.4.3 are satisfied. Suppose as well that the numbers of series basis functions* $k_{x_j,MN} \to \infty$, $\nu_a^2(\sigma_N)\left(\zeta_a^2(k_{x_j,MN}) + \zeta_a^2(k_{a,MN})\right) \to 0$, $\nu_{a,N}(\zeta_a(k_{x_j,MN})+\zeta_a(k_{a,MN}))(\zeta_0(k_{x_j,MN})+\zeta_0(k_{a,MN})) \to 0$, $[\nu_{l_a,N}\zeta_0(k_{x_j,MN}) + \zeta_0(k_{a,MN})\nu_{l_j,N}](\zeta_0(k_{x_j,MN}) + \zeta_0(k_{a,MN})) \to 0$ *for* $j = 1, \ldots, d_x$, $k_{a,MN} \to \infty$, *and* $[\zeta_0^2\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}\right) + \zeta_0^2(k_{a,MN})](\max_{j=1,\ldots,d_x} k_{x_j,MN} + k_{a,MN})/(MN) \to 0$ *as* $N \to \infty$. *Then*

$$\left\|\hat{\beta}_{MN} - \beta_{0,MN}\right\| = \mathcal{O}_p\left(\sum_{j=1}^{d_x}\nu_{h_j,N} + \nu_{h_a,N} + \sum_{j=1}^{d_x}\nu_{I_j,N} + \nu_{I_a,N} + k_{a,MN}^{-\alpha_e} + \sum_{j=1}^{d_x} k_{x_j,MN}^{-\alpha_e}\right).$$

Theorem 2.4.5. *Suppose that the conditions for Lemma 2.4.4 are satisfied. Then, for* $j = 1, \ldots, d_x$

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}\left|e_{x_j}\widehat{(x,a)/e}(x,a) - e_{x_j}(x,a)/e(x,a)\right| = \mathcal{O}_p\left(\nu_{e_j,M,N}\right),$$

and

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}\left|e_a\widehat{(x,a)/e}(x,a) - e_a(x,a)/e(x,a)\right| = \mathcal{O}_p\left(\nu_{e_a,M,N}\right).$$

Theorem 2.4.6. *Suppose that the conditions for Theorem 2.4.5 are satisfied. Then*

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\hat{e}(x,a)-e(x,a)| \;=\; \mathcal{O}_p\left(\sum_{j=1}^{d_x}\nu_{e_j,M,N}+\nu_{e_a,M,N}\right).$$

The convergence rates of the quality function and its derivative ratios depend on the estimation errors of the reduced form functions and the series approximation errors of the quality function itself. Note that the estimation errors in $\hat{a}_i^m$ affect the estimation errors of $e_{x_j}(x, a)/e(x,a)$, $e_a(x,a)/e(x,a)$ and $e(x,a)$ only through $\hat{I}^m$, $\hat{h}^m$ and their partial derivatives.

2.4.2.4 Sellers' Marginal Disutility Function Estimator $\hat{U}_h(h,x,a)$

This subsection derives the convergence rate of the cross-market series estimator of the sellers' marginal disutility function $U_h(h,x,a)$.

For a function $l(h,x,a)$: $\mathcal{H}\times\mathcal{X}\times\mathcal{A}\to\mathbb{R}$, define the norm $|l|_\delta$ as $|l|_\delta \equiv \max_{|\mu|\le\delta} \sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}} |\partial^\mu l(x,a)/\partial h^{\mu_h}\partial x_1^{\mu_1}\cdots\partial x_{d_x}^{\mu_{d_x}}\partial a^{\mu_a}|$, with $\mu_h+\mu_1+\cdots+\mu_{d_x}+\mu_a=\mu$ (μ_h, $\mu_1,\ldots,\mu_{d_x},\mu_a$ are integers).

Assumption 2.4.11. *Suppose that for a positive integer $\delta_U\ge 0$, there exist a constant $\alpha_U>0$ and pseudo-true series coefficients $\gamma_{0,k_U}\in\mathbb{R}^{k_U}$ such that $|U_h-\Psi'_{k_U}\gamma_{0,k_U}|_{\delta_U}\le Ck_U^{-\alpha_U}$ for all positive integers k_U.*

Assumption 2.4.12. *Suppose that there exists some positive constant B_U such that $|U_h|_1\le B_U$.*

Assumption 2.4.13. *Suppose that $\Psi_k(h,x,a)=\Psi_{h,k_h}(h)\otimes\Psi_{1,k_1}(x_1)\otimes\cdots\otimes\Psi_{d_x,k_{d_x}}(x_{d_x})\otimes\Psi_{a,k_a}(a)$. This implies that $k=k_h\cdot k_a\cdot\prod_{j=1}^{d_x}k_j$.*

Assumption 2.4.14. *Suppose that $\mathcal{H}$ is a compact set and the cross-market variation in f_x^m and f_y^m is rich enough that the equilibrium cross-market joint density of (h, x, a) is bounded away from zero.*

Define

$$S_{\Psi\Psi} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} S_{\Psi}(h_i^m, x_i^m, a_i^m)' S_{\Psi}(h_i^m, x_i^m, a_i^m),$$

where

$$S_{\Psi}(h_i^m, x_i^m, a_i^m) \equiv \left(\begin{array}{c} \left[\nabla_x h^m(x_i^m, a_i^m) + h^m(x_i^m, a_i^m) \frac{\nabla_x e(x_i^m, a_i^m)}{e(x_i^m, a_i^m)}\right] \otimes \Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)' \\ \left[h_a^m(x_i^m, a_i^m) + h^m(x_i^m, a_i^m) \frac{e_a(x_i^m, a_i^m)}{e(x_i^m, a_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)' \end{array} \right).$$

Assumption 2.4.15. *Suppose that $\lambda_{\min}\left(\mathbb{E}(S_{\Psi\Psi})\right) \geq c > 0$.*

Lemma 2.4.7. *Suppose that Assumptions 2.4.11-2.4.15 and the conditions for Theorem 2.4.5 are satisfied. Suppose as well that $k_{U,MN} \to \infty$, $\sqrt{k_{U,MN}}\nu_{a,N}\zeta_a(k_{U,MN}) \to 0$, $k_{U,MN}$ $\nu_{e_j,N} \to 0$ and $k_{U,MN}\nu_{e_a,N} \to 0$ as $N \to \infty$. Then*

$$\left\|\hat{\gamma}_{k_{U,MN}} - \gamma_{0,k_{U,MN}}\right\| = \mathcal{O}_p\left(\nu_{e_a,M,N} + \sum_{j=1}^{d_x} \nu_{e_j,M,N} + k_{U,MN}^{-\alpha_U}\right).$$

In addition, define

$$\nu_{U_h,M,N} \equiv \zeta_0(k_{U,MN}) \left[\nu_{e_a,M,N} + \sum_{j=1}^{d_x} \nu_{e_j,M,N} + k_{U,MN}^{-\alpha_U}\right].$$

And I will assume that $\nu_{U_h,M,N} \to 0$ as $N \to \infty$ for the rest of the chapter.

Theorem 2.4.8. *Suppose that the conditions of Lemma 2.4.7 are satisfied. Then*

$$\sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}} \left|\hat{U}_h(h,x,a) - U_h(h,x,a)\right| \quad = \quad \mathcal{O}_p\left(\nu_{U_h,M,N}\right).$$

The convergence rate of the sellers' marginal disutility function depends on the estimation errors of the quality function and on the series approximation error of the sellers' marginal disutility function itself. Note that the estimation errors of $\hat{a}_i^m$ and the reduced form functions and their derivatives directly affect the convergence rate of the sellers' marginal disutility function, but they are dominated by the estimation errors of the quality function and its derivatives.

2.5 Empirical Illustration in Labor Markets

In this section, I apply the estimation procedure provided in Sections 2.4.1 to estimate the efficiency (quality) function e in labor markets. Section 2.5.1 introduces the data set, and Section 2.5.2 estimates the workers' unobserved efficiency function.

2.5.1 Data: the 2015 American Time Use Survey

The data set I use is the American Time Use Survey (ATUS, see Hofferth, Flood, and Sobek, 2013 for details). The ATUS randomly chooses one individual from a subsample of the households that are completing their participation in the Current Population Survey (CPS) and asks them to recall their time spent, minute by minute, on various activities within a randomly picked 24-hour period in the past. The ATUS classifies activities into 17 major categories and many more sub-categories, and provides a quite precise measure of the time that workers actually spent in working.[30]

I consider the 2015 ATUS respondents,[31] and focus on full-time workers in the three largest cities: New York, Los Angeles and Chicago[32]. After dropping observations on Saturdays and Sundays and making some other minor adjustments, I end up with a sample of 92 workers in New York, 74 workers in Los Angeles, and 55 workers in Chicago.

I use the time spent in the "working" sub-category of the ATUS as the measure of working time h_i^m, the weekly earnings in the CPS as the measure of earnings I_i^m, and the age reported in the CPS as the observed characteristic x_i^m of the workers.[33]

Figure 2.4 shows the scatter plots of working time per day and weekly earnings of each worker in the three cities. Within- and cross-market variation appears prominent: (i) both working time and earnings vary substantially within all the markets; (ii) for the same working time, earnings in New York tend to be higher than those in Los Angeles, which in turn, tend to be higher than those in Chicago. In fact, the median of the earnings-to-working-time ratio is 2.47 for the workers in New York, 2.03 in Los Angeles, and 1.62 in Chicago. Such

[30] Major categories include working and work-related activities, household activities, education, traveling and others. For working and work-related activities, it further breaks down to working, looking for a job, eating and drinking on the job (e.g., lunch breaks), security procedures, and so on. I use the time spent in the working sub-category as the measure of working time.

[31] The data were obtained via ATUS-X Extract Builder: Sandra L. Hofferth, Sarah M. Flood, and Matthew Sobek. 2013. American Time Use Survey Data Extract System: Version 2.4 [Machine-readable database]. Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota.

[32] To be precise, the three largest metro areas: New York-Newark-Bridgeport (NY-NH-CT-PA), Los Angeles-Long Beach-Riverside (CA), and Chicago-Naperville-Michigan City (IL-IN-WI).

[33] Individuals in the ATUS can be linked to their observations in the CPS to obtain rich demographic information. In this illustration, I use age as the only observed characteristic for simplicity. The application to more observed variables poses no theoretical problem, but it may take more computing time.

within- and cross-market variation is crucial for the identification of the unobserved efficiency function.

2.5.2 Estimation of Unobserved Efficiency Function

With the observed data (I_i^m, h_i^m, x_i^m) from the three cities, one is able to estimate the efficiency function $e(x, a)$.

As discussed in Sections 2.2.1 and 2.3, distributions f_x^m of workers' observed characteristic x_i^m (age) serve as aggregate instruments that induce cross-market variation in the earnings functions. Figure 2.5 plots the kernel estimated densities of the workers' age distributions in the three cities. It shows that in the 2015 ATUS sample, full-time workers in Chicago are slight younger than in the other two cities. The age distributions in Los Angeles and Chicago are slightly more dispersed than that in New York.

Such variation in the distributions f_x^m appears to be sufficient to generate adequate variation in the earnings functions. Figure 2.6 draws representative iso-earnings curves for the three cities on the support $\mathcal{X} \times \mathcal{A} = [25, 65] \times [0.05, 0.95]$. Recall that Assumption 2.3.5 for identifying the efficiency function requires that the iso-earnings curves from at least two cities have different slopes. For each value of (x, a) on the support, this is the case, except in the very small region with $a > 0.9$ and $x \in [35, 55]$. This suggests that Assumption 2.3.5 is satisfied. Moreover, using estimated derivatives of the earnings functions $\hat{I}_x^m(x, a)$ and $\hat{I}_a^m(x, a)$, $m = 1, \ldots, M$, I compute $\hat{B}(x, a)$, the estimate of the coefficient matrix $B(x, a)$ defined in Assumption 2.3.5 for a grid of (x, a) values on the support $\mathcal{X} \times \mathcal{A}$. The determinants of $\hat{B}(x, a)'\hat{B}(x, a)$ for all these (x, a) values are bounded well away from zero. This indicates that the matrix $B(x, a)$ has full column rank. As a result, I am convinced that the key identification condition for the efficiency function $e(x, a)$ is satisfied.

The normalization worker I choose is $(\bar{x}, \bar{a}) \equiv (25, 0)$. I used the tensor product of quadratic polynomials of x and a to approximate $e_x(x, a)/e(x, a)$ and $e_a(x, a)/e(x, a)$.[34] With the two estimated ratio functions, one could obtain the estimates of the efficiency function defined as in equation (2.4.13). Figure 2.7 plots the estimated efficiency function $\hat{e}(x, a)$ on the support $\mathcal{X} \times \mathcal{A}$.

Figure 2.7 presents a prominent and interesting pattern of the efficiency function. For workers with the same level of unobserved characteristic a ("ability"), efficiency first increases with age, and then decreases. For workers of the same age, efficiency increases with a. At age 25, workers with the highest ability do not exhibit much higher efficiency than their lower ability peers. As they mature, however, their efficiency could be much higher than their peers with the lowest ability.[35]

2.6 Conclusion and Extensions

In this paper, I study the identification and estimation of a nonparametric hedonic equilibrium model with unobserved quality. I explain how to use within- and cross-market variation in equilibrium prices and quantities to identify and estimate the structural functions of the

[34] That is, I approximate the two ratio functions using $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_4 a + \beta_5 zx + \beta_6 ax^2 + \beta_7 a^2 + \beta_8 a^2 x + \beta_9 a^2 x^2$. There is no obvious rule for how one should determine the order of the polynomials for the efficiency function or for the other structural functions in this model. This may serve as a topic for further research.

[35] Since I only control for age and neglect the dynamic perspective of the workers, one should be cautious when interpreting this estimate. But this issue will be investigated in future research, and an in-depth empirical analysis is beyond the scope of this section.

model. Using the estimated structural functions and the equilibrium-solving algorithm suggested in this paper, researchers could solve the counterfactual equilibrium to analyze the distributional effects of policy interventions. In contrast to other widely used methods, the counterfactuals thus constructed account for unobserved quality and equilibrium effects of policy interventions in a nonparametric setting. Yet several directions of extension are worth more research.

First, asymptotic distribution results are necessary for conducting inference on the structural functions and the counterfactuals. In addition, providing an easy-to-implement, data-driven method to determine the tuning parameters for each step of the estimation procedure is relevant to empirical work.

Second, in this paper I assume that agents' unobserved heterogeneity is scalar-valued, which might restrict its applicability (e.g., Roy model is excluded). Chernozhukov, Galichon, and Henry (2014) considered the identification of hedonic equilibrium models with multidimensional unobserved heterogeneity among agents. It might be an interesting research topic to see whether one can extend their method to models with unobserved product characteristic. Another related possible extension is to allow for multidimensional unobserved product characteristics that is more general than the single-index model discussed in Appendix 2.C. Multidimensional quality could be important for a variety of empirical questions.[36]

Third, the results in this paper are based on the assumption that agents' observed and unobserved heterogeneities are independent of each other in each market. While this could be a very restrictive assumption for scenarios in which agents select their own observed

[36]For example, Halket, Nesheim and Oswald (2015) found that the English Housing Survey data rejects unidimensional unobserved housing quality assumption.

characteristics, it might be possible to relax it by controlling on some additional variables.

Fourth, it is necessary to re-examine the identification results under alternative data structures. For example, what can be identified if a positive proportion of workers choose not to work at all?[37] Another example is that quantity only has discrete support in the data (e.g., full-time v.s. part-time work, number of bedrooms in a house). Moreover, assuming that one seller is matched with one buyer might not capture certain decisions they make (e.g., firm size) or over-simplify the production process (e.g., no complementarity among workers).[38]

Finally, the current static model might give biased estimates and counterfactuals if in fact agents optimize over a longer horizon.[39] Investigating identification of a dynamic model will be an important topic for future research.

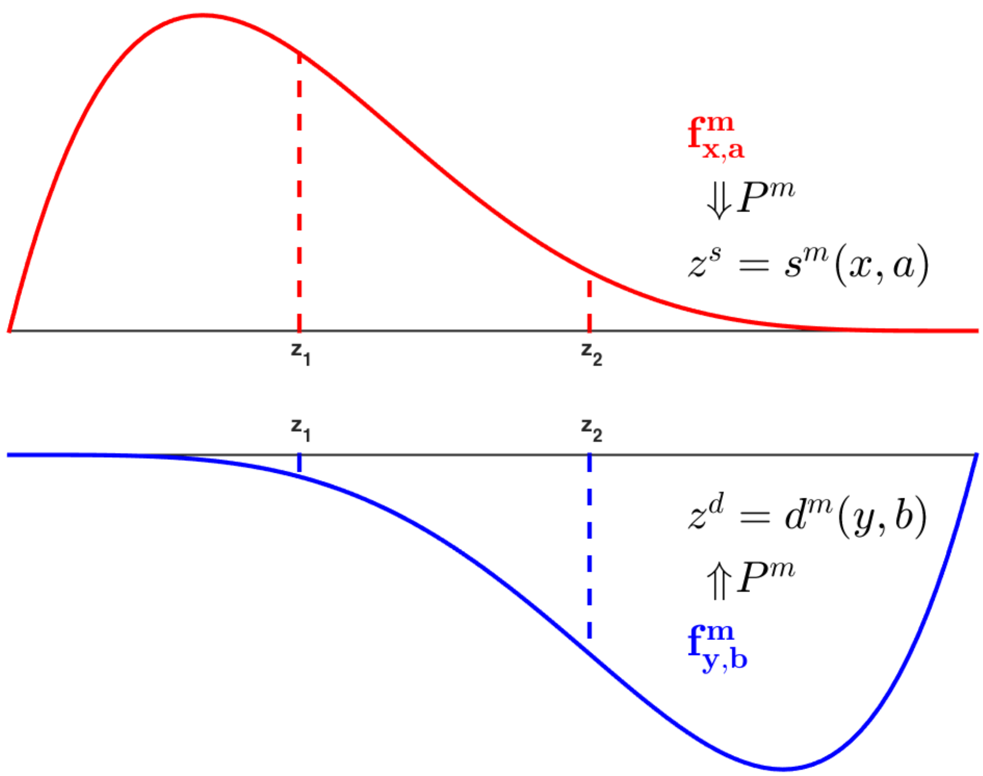
---

[37]Chiappori, McCann, and Nesheim (2010) showed the existence of equilibrium if agents had potentially binding outside options.

[38]Proper frameworks to analyze (non-)identification of these complications remain a question.

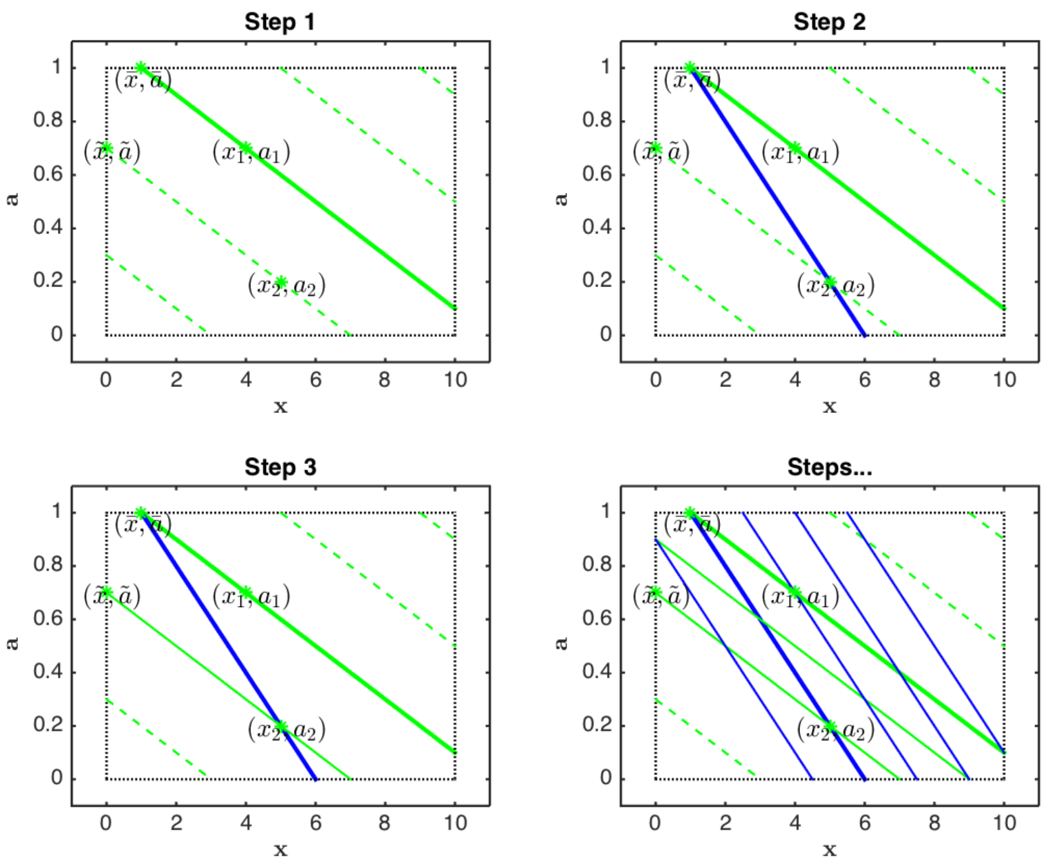
[39]For example, efficiency might be under-estimated for young workers and over-estimated for experienced workers, if young workers choose to work extra time to enhance human capital.

Figure 2.2: Off Equilibrium



The green line illustrates the distribution of the optimal effective labor supply z^s under the price schedule function P^m in market m, as a function of sellers' observed characteristics x and unobserved characteristic a, which follow the distribution $f^m_{x,a}$. Similarly, the blue line illustrates the distribution of the optimal effective labor demand z^d under the same price schedule function P^m in market m, as a function of buyers' observed characteristics y and unobserved characteristic b, which follow the distribution $f^m_{y,b}$. As is shown in this figure, when the distributions of z^s and z^d are different (for example, density of the effective labor supply is larger than that of the demand at z_1, and is the opposite at z_2), the market is off equilibrium and the price schedule function P^m will adjust.

Figure 2.3: Identification of $e(x, a)$ in Two Markets



Green lines (solid and dashed) illustrate the disjoint iso-payment curves in Market 1 and blue lines (solid and dashed) illustrate the disjoint iso-payment curves in Market 2. The quality $e(\bar{x}, \bar{a})$ is normalized to be one. In each market, the relative qualities for sellers on the same iso-payment curves can be identified, but not for those on different iso-payment curves. For example, $e(x_1, a_1)/e(\bar{x}, \bar{a})$ and $e(x_2, a_2)/e(\tilde{x}, \tilde{a})$ are identified from Market 1 (illustrated in Step 1), but not $e(x_2, a_2)/e(\bar{x}, \bar{a})$. From Market 2, however, $e(x_2, a_2)/e(\bar{x}, \bar{a})$ can be identified (illustrated in Step 2). As a result, $e(x_2, a_2)/e(\bar{x}, \bar{a})$ can be identified using the data from both markets (illustrated in Step 3). This idea could be applied repeatedly to identify the quality function $e(x, a)$ (illustrated in the last panel). The identification requires a rank condition on the derivatives of the payment functions $I^m(x, a)$ across markets. As is shown in the figure, this condition can be understood as requiring that the slopes of the iso-payment curves across markets are different.

Figure 2.4: Scatter Plots of Weekly Earnings and Working Time in the Three Cities

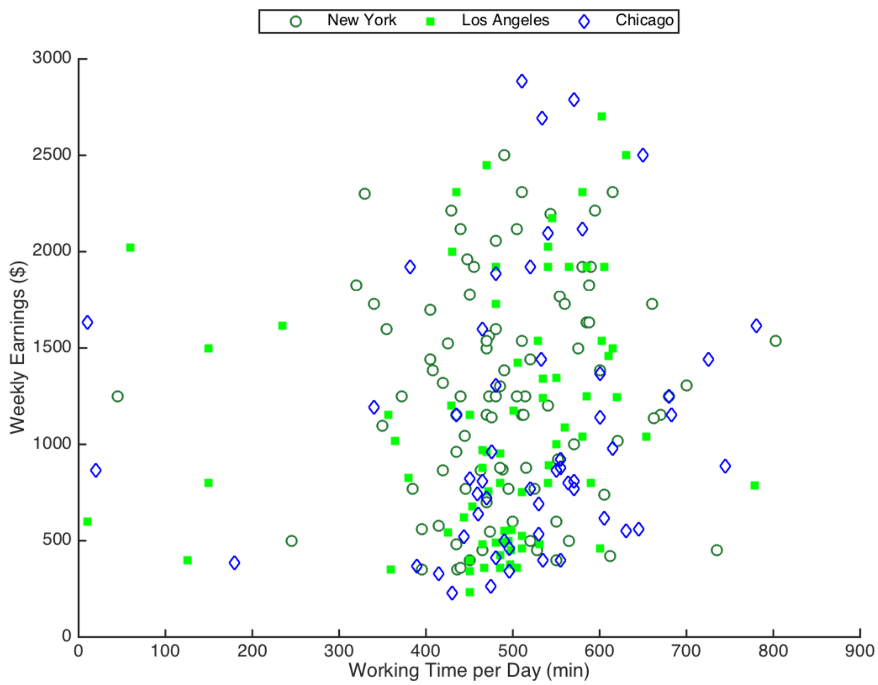
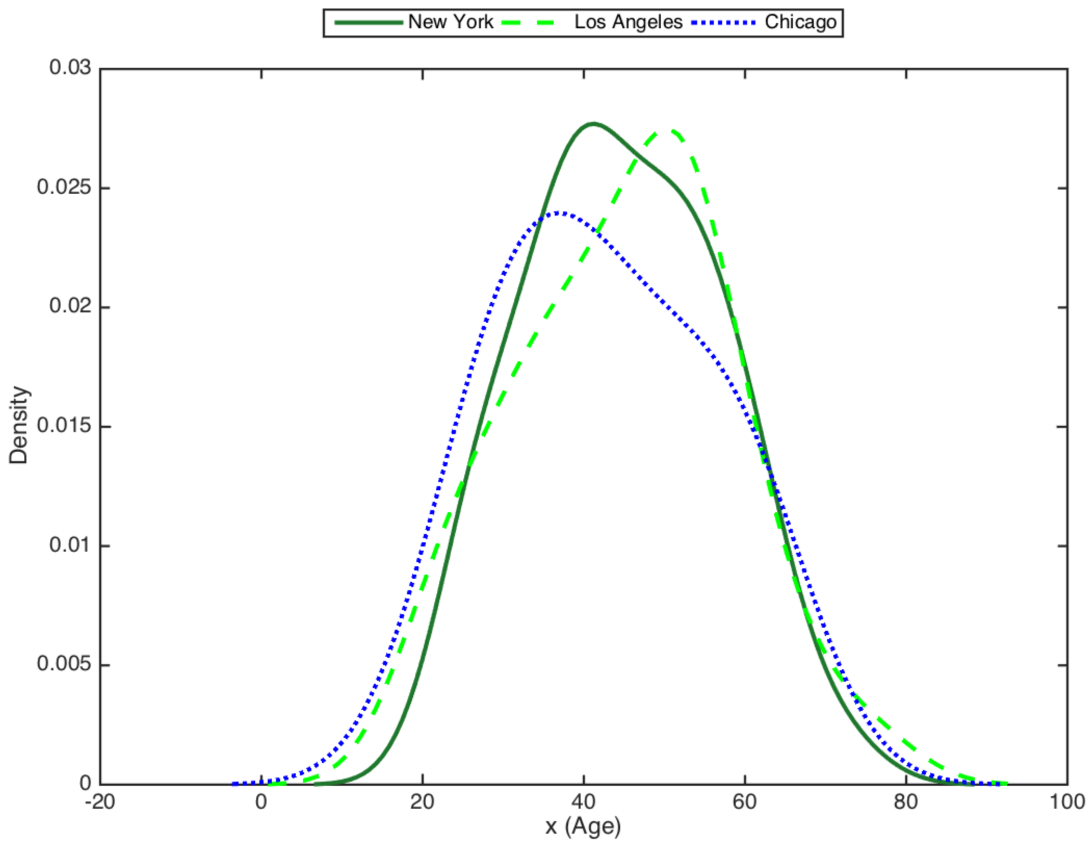
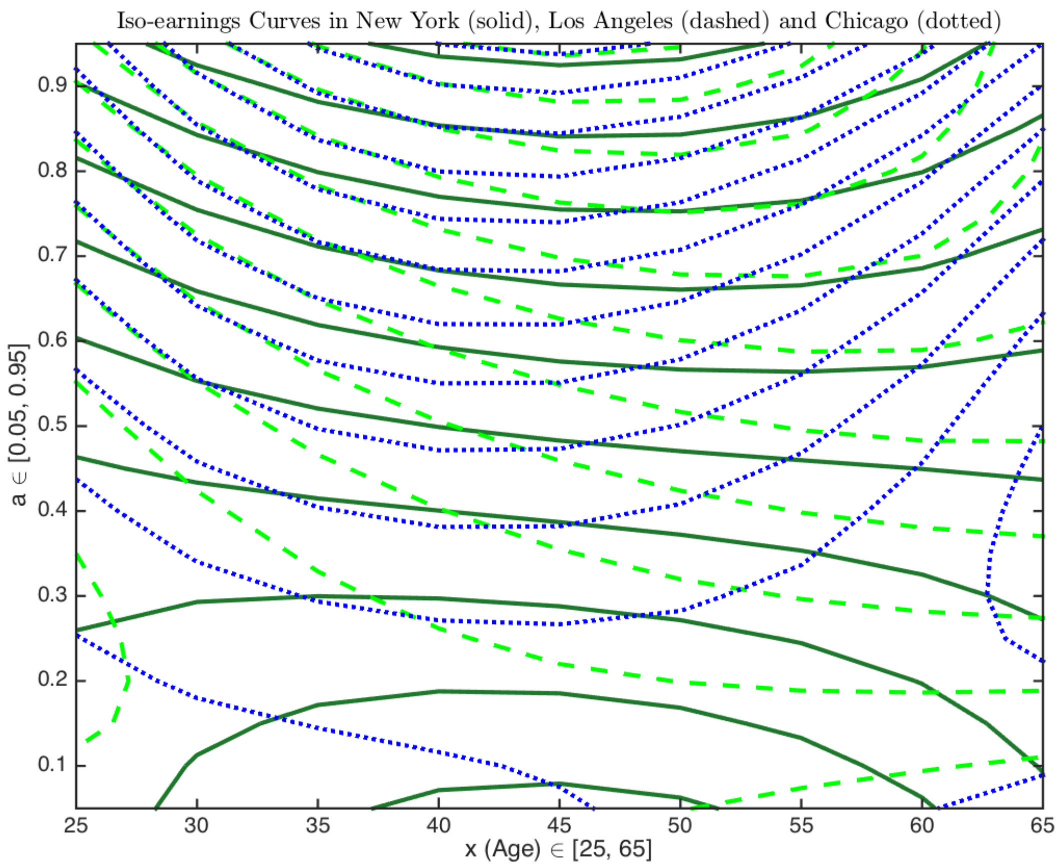


Figure 2.5: Distributions of Age in the Three Cities



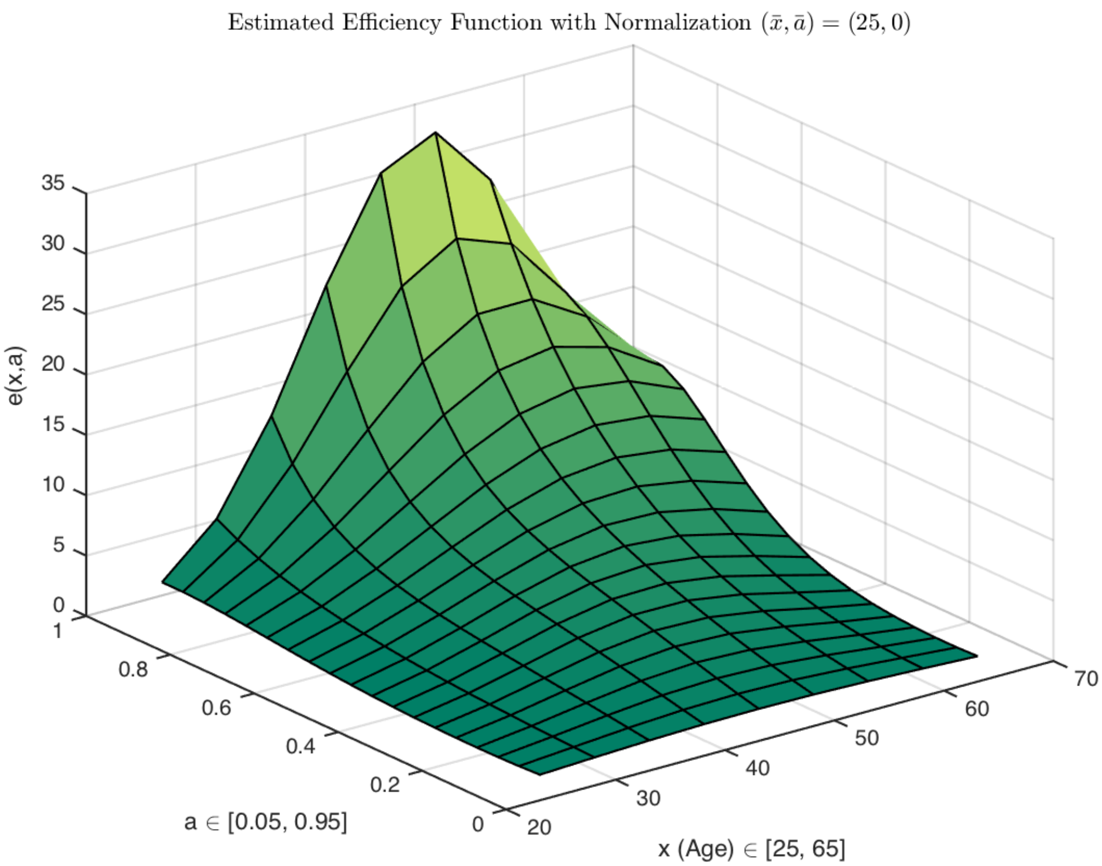
For ease of illustration, age is used as the single observed characteristic (x) of the workers. This figure shows that there is decent cross-market variation in the distributions of x, which drives (partially) the cross-market variation in the equilibrium payment functions.

Figure 2.6: Iso-Earnings Curves in the Three Cities



This figure shows representative iso-earnings curves for the three cities, on the support of age (x) and "ability" (a). For majority of the support, the iso-earnings curves from at least two markets cross. This suggests that the identification condition for the efficiency function $e(x, a)$ is satisfied.

Figure 2.7: Estimated Efficiency Function $e(x, a)$



Estimated worker efficiency function increases with "ability" (a), and is hump-shaped with age (x).

Appendix

2.A Solving for Counterfactual Equilibrium

This section suggests an algorithm to numerically solve for counterfactual equilibrium, as there is in general no closed-form solution to hedonic equilibrium models. This algorithm can be applied to analyze the distributional effects of a wide range of interventions. Take labor markets as example, an expansion of higher education may change education level from x_i^m to $\tilde{x}_i^m$ for a large number of workers; new investment projects may increase firms' capital stock from y_i^m to $\tilde{y}_i^m$; or advances in total factor productivity may give rise to a new revenue function $\tilde{R}$ instead of R. Establishing counterfactual distributions of worker earnings $\tilde{I}_i^m$ (or labor supply $\tilde{h}_i^m$) constitutes a vital part of welfare analysis of interventions like these and helps understanding sources of earnings inequality (or other questions concerning labor supply $\tilde{h}_i^m$).

To fix idea, suppose we are under the first intervention, namely, x_i^m is replaced by $\tilde{x}_i^m$ (for all $i = 1, \dots, N^m$, $m = 1, \dots, M$). And suppose that the estimates of the other market primitives, i.e. the structural functions $(\hat{U}_h, \hat{e}, \hat{R})$, the unobserved worker characteristic $\hat{a}_i^m$, and the firm characteristics $(y_i^m, \hat{b}_i^m)$, have been obtained using the estimation procedure provided in Section 2.4.1 and will remain constant under the intervention.[40] The equilibrium is solved for each market m separately.

2.A.1 Algorithm to Solve for Counterfactual Equilibrium

The algorithm consists of two steps:

1. Obtain the general solution to the ODE in equation (2.2.6) that characterizes the equilibrium.

2. Determine the initial value condition of the ODE in equation (2.2.6) by solving the optimal transportation problem, which is mathematically equivalent to the hedonic equilibrium model.

To implement the first step, let us take a closer look at the equilibrium condition in equation (2.2.6). It is a first-order ODE in the first-order derivative of the earnings schedule function P_z^m. Given any P_z^m function and any value $z \in \mathcal{Z}$, the right-hand side of equation (2.2.6) can be approximated numerically. In particular, the first term of the numerator can be approximated as

$$\sum_{i=1}^{N^m} \frac{\hat{R}_{zz}\left(z, y_i^m, b^*(z, y_i^m)\right)}{\hat{R}_{zb}\left(z, y_i^m, b^*(z, y_i^m)\right)},$$

where $\hat{R}_{zz}$ and $\hat{R}_{zb}$ are the second-order derivatives associated with the estimated revenue function $\hat{R}$, and $b^*(z, y)$ is the inverse effective labor demand function that satisfies firms'

[40]Under multiple concurrent interventions, researchers may also replace y_i^m and/or one or more estimated structural functions with their "tilde" counterparts. The algorithm described in this section still applies.

FOC under the given P_z^m function. That is,

$$P_z^m(z) = \hat{R}_z\left(z, y, b^*(z, y)\right).$$

Similarly, the first term of the denominator can be approximated as

$$\sum_{i=1}^{N^m} \frac{1}{\hat{R}_{zb}\left(z, y_i^m, b^*(z, y_i^m)\right)}.$$

On the other hand, define the inverse effective labor supply function $a^*(z, x)$ that satisfies workers' FOC under the given P_z^m function

$$P_z^m(z) = \hat{U}_h\left(\frac{z}{\hat{e}\left(x, a^*(z, x)\right)}, x, a^*(z, x)\right)/\hat{e}\left(x, a^*(z, x)\right).$$

Thus, the second term of the numerator can be numerically approximated as

$$\sum_{i=1}^{N^m} \frac{-\hat{U}_{hh}\left(\frac{z}{\hat{e}(x, a^*(z,x))}, x, a^*(z, x)\right)}{T_1(z, x) - T_2(z, x) - T_3(z, x)},$$

and the second term of the denominator can be approximated as

$$\sum_{i=1}^{N^m} \frac{-\left[\hat{e}\left(x, a^*(z, x)\right)\right]^2}{T_1(z, x) - T_2(z, x) - T_3(z, x)},$$

where

$$T_1(z, x) \quad \equiv \quad \hat{U}_{ha}\left(\frac{z}{\hat{e}\left(x, a^*(z, x)\right)}, x, a^*(z, x)\right)\hat{e}\left(x, a^*(z, x)\right),$$

$$T_2(z,x) \equiv P_z^m(z)\hat{e}\left(x,a^*(z,x)\right)\hat{e}_a\left(x,a^*(z,x)\right),$$

$$T_3(z,x) \equiv \hat{U}_{hh}\left(\frac{z}{\hat{e}\left(x,a^*(z,x)\right)},x,a^*(z,x)\right)\frac{z}{\hat{e}\left(x,a^*(z,x)\right)}\hat{e}_a\left(x,a^*(z,x)\right),$$

and $\hat{e}_a$, $\hat{U}_{hh}$ and $\hat{U}_{ha}$ are the partial derivatives associated with the estimated efficiency function $\hat{e}$ and marginal disutility function $\hat{U}_h$, respectively.

Therefore, replacing the four components of the right-hand side of the equilibrium condition in equation (2.2.6) with their numerical approximation above, one could numerically solve the ODE for its general solution. That is, for each value $P_{z,0}^m \in \mathbb{R}$, we get a different function P_z^m such that P_z^m satisfies equation (2.2.6) and $P_z^m(z_0) = P_{z,0}^m$ for a fixed value $z_0 \in \mathcal{Z}$.[41]

The next step, therefore, is to determine the value $P_{z,0}^m$. Without loss of generality, one could let $z_0 = 0$. Chiappori, McCann, and Nesheim (2010) showed that the hedonic equilibrium model is mathematically equivalent to an optimal transportation problem. Note that this equivalence holds even when quality is unobserved (by researchers) as long as z is observed by both sellers and buyers. Therefore, the algorithm they proposed to solve the optimal transportation problem can be employed to solve for the equilibrium of my model. With the general solution of the ODE obtained in the first step, one only needs to optimize over a one-dimensional parameter $P_{z,0}^m$ to solve the optimal transportation problem.[42] Other

[41] Matlab provides toolboxes that quickly deliver numerical solutions to first-order ODEs.

[42] In fact, instead of solving the optimal transportation problem directly, they suggested solving the dual problem, a constrained linear programing problem (equation (42) in their paper). They did not have the first step and used a series expansion to approximate the unknown equilibrium price schedule function. Therefore, they needed to optimize over multidimensional series coefficients. Please refer to their paper for details. Depending on the sample size, the shape of the equilibrium price schedule function, and the ranges of the series coefficients, among other factors, my algorithm might be faster or slower than theirs. Further research is needed to investigate the situations to which each algorithm is suited.

equilibrium outcomes, such as $\tilde{I}_i^m$ and $\tilde{h}_i^m$, can be constructed as a result.[43]

2.A.2 Stability of Numerical Equilibrium Solutions

Cautious researchers might be interested in the stability of numerical equilibrium solutions of the model. Two sources of errors might contribute to the difference between the numerical solution and the true counterfactual equilibrium: estimation errors in the estimation of the market primitives and numerical errors in the implementation of the algorithm described in Section 2.A.1. If the mapping from the market primitives to the equilibrium outcomes is not continuous, then the numerical equilibrium solution will be unstable with respect to these errors.

To examine the stability of the numerical equilibrium solutions, I conduct a small-scale simulation experiment. I implement the algorithm in Section 2.A.1 to solve for the equilibrium in a market with 1000 worker-firm pairs.[44] The first panel of Figure 2.8 shows the (kernel estimated) equilibrium densities of effective labor supply z^s and demand z^d when I use the true structural functions. The second panel shows the (kernel estimated) equilibrium densities of z^s and z^d when I perturb the structural functions by them with multiplying normal random variables with mean 1 and standard deviation 0.01.[45] The third and fourth panels show the cases when the standard deviations of the perturbations are 0.05 and 0.1,

[43]With the first-order derivative function P_z^m, one may let $P^m(0) \equiv 0$ to determine the level of the price schedule function P^m.

[44]I assume that x and y follow beta distributions, that is, $x_i \sim \beta(9,1)$ and $y_i \sim \beta(1,9)$. I also assume that $U(h,x,a) = \left[h^2 x^1 + (1-a)^1\right]^1$, $e(x,a) = x^{0.7} a^{0.5}$ and $R(z,y,b) = z^{1/2} y^{1/2} b^{1/2}$.

[45]In the interim steps of the algorithm, every time I need to evaluate a structural function, I compute the true value and multiply it by a new normal random variable.

respectively.

Figure 2.8 has two important implications. First, even though I approximate the integrals in the equilibrium condition in equation (2.2.6) with sample averages and approximate the integrals in the constraints of the optimal transportation problem with quadratures (details in Chiappori, McCann, and Nesheim, 2010), the algorithm in Section 2.A.1 is still able to deliver a very precise numerical equilibrium solution. This is illustrated by the estimated densities of z^s and z^d, which trace each other very closely in the first panel. Second, the mapping from the structural functions to the equilibrium is likely to be continuous; otherwise, small perturbations in the structural functions would result in large changes in the equilibrium quantities or even render the equilibrium non-solvable. However, the last three panels of Figure 2.8 show the contrary. With moderately sized perturbations to the structural functions, I still obtain equilibrium solutions that closely resemble the one obtained using the true structural functions.

2.B Market Level Heterogeneity

In the main text of this paper, I assume that efficiency function $e(x, a)$ takes the same value for all workers with the characteristics (x, a) across markets. This implies that a worker with ath quantile of unobserved characteristic in one market will have the same efficiency as a worker with ath quantile of unobserved characteristic in another market (given that their x's are the same). If the markets (cities, counties, etc. depending on specific applications under investigation) are comparable with each other in terms of the distributions of workers' unobserved characteristics, then this is a plausible assumption. In many applications, however, this may not be true. The distribution of workers' unobserved characteristics in Manhattan, New York may well be different from that in Manhattan, Kansas. My model and all the results still apply if there are finite types of markets. As long as the type of each market is observed (or can be estimated based on some market level observables), then all the results in this paper apply within each market type. One important practical implication is that we may allow large cities to have a different efficiency function from small cities. So long as we have multiple cities of the same type in our sample, then the efficiency functions can be identified and estimated separately. Accommodating this generality formally provides no extra insight, but induces notational complexity.

2.C Multidimensional Quality with Single Index Structure

In this section, I relax the assumption that h^m and e are single-dimensional. Let the constant $L > 1$ denote the dimensional of h^m and e. Let $h^m(x, a) \equiv (h_1^m(x, a), \ldots, h_L^m(x, a))'$ and $e(x, a) \equiv (e_1(x, a), \ldots, e_L(x, a))'$. Assume that the coordinates of h and e enter the price schedule function collectively in a single index. Recall Assumption 2.2.1 and the payment equation (2.2.3) in market m, then we have

$$I^m(x, a) = P^m \left(h_1^m(x, a) \cdot e_1(x, a) + \cdots + h_L^m(x, a) e_L(x, a) \right), \qquad (2.C.1)$$

for all $m \in \mathcal{M}$ and all $(x, a) \in \mathcal{X} \times \mathcal{A}$. Taking the partial derivatives with suppressed arguments gives us:

$$\begin{cases} \nabla_x I^m(x,a) & = P_z^m \cdot \left[\nabla_x h_1^m e_1 + h_1^m \nabla_x e_1 + \cdots + \nabla_x h_L^m e_L + h_L^m \nabla_x e_L \right], \\ \frac{\partial I^m}{\partial a}(x,a) & = P_z^m \cdot \left[\frac{\partial h_1^m}{\partial a} e_1 + h_1^m \frac{\partial e_1}{\partial a} + \cdots + \frac{\partial h_L^m}{\partial a} e_L + h_L^m \frac{\partial e_L}{\partial a} \right]. \end{cases}$$

Provided that $\partial I^m(x,a)/\partial a \neq 0$, we may take the ratio of the first equation to the last equation:

$$\frac{\frac{\partial I^m}{\partial x_1}}{\frac{\partial I^m}{\partial a}} = \frac{\frac{\partial h_1^m}{\partial x_1} + h_1^m \frac{\partial e_1}{\partial x_1}/e_1 + \cdots + \frac{\partial h_L^m}{\partial x_1}\frac{e_L}{e_1} + h_L^m \frac{\partial e_L}{\partial x_1}/e_1}{\frac{\partial h_1^m}{\partial a} + h_1^m \frac{\partial e_1}{\partial a}/e_1 + \cdots + \frac{\partial h_L^m}{\partial a}\frac{e_L}{e_1} + h_L^m \frac{\partial e_L}{\partial a}/e_1},$$

which implies

$$\begin{aligned} & \frac{\partial I^m}{\partial a} h_1^m \frac{\partial e_1}{\partial x_1}/e_1 - \frac{\partial I^m}{\partial x_1} h_1^m \frac{\partial e_1}{\partial a}/e_1 + \cdots + \frac{\partial I^m}{\partial a} h_L^m \frac{\partial e_L}{\partial x_1}/e_1 - \frac{\partial I^m}{\partial x_1} h_L^m \frac{\partial e_L}{\partial a}/e_1 \\ & + \left(\frac{\partial I^m}{\partial a}\frac{\partial h_2^m}{\partial x_1} - \frac{\partial I^m}{\partial x_1}\frac{\partial h_2^m}{\partial a} \right) \frac{e_2}{e_1} + \cdots + \left(\frac{\partial I^m}{\partial a}\frac{\partial h_L^m}{\partial x_1} - \frac{\partial I^m}{\partial x_1}\frac{\partial h_L^m}{\partial a} \right) \frac{e_L}{e_1} \\ = \; & \frac{\partial I^m}{\partial x_1}\frac{\partial h_1^m}{\partial a} - \frac{\partial I^m}{\partial a}\frac{\partial h_1^m}{\partial x_1}. \end{aligned} \tag{2.C.2}$$

Taking the ratio of the second equation to the last equation:

$$\frac{\frac{\partial I^m}{\partial x_2}}{\frac{\partial I^m}{\partial a}} = \frac{\frac{\partial h_1^m}{\partial x_2} + h_1^m \frac{\partial e_1}{\partial x_2}/e_1 + \cdots + \frac{\partial h_L^m}{\partial x_2}\frac{e_L}{e_1} + h_L^m \frac{\partial e_L}{\partial x_2}/e_1}{\frac{\partial h_1^m}{\partial a} + h_1^m \frac{\partial e_1}{\partial a}/e_1 + \cdots + \frac{\partial h_L^m}{\partial a}\frac{e_L}{e_1} + h_L^m \frac{\partial e_L}{\partial a}/e_1},$$

which implies

$$\frac{\partial I^m}{\partial a} h_1^m \frac{\partial e_1}{\partial x_2}/e_1 - \frac{\partial I^m}{\partial x_2} h_1^m \frac{\partial e_1}{\partial a}/e_1 + \cdots + \frac{\partial I^m}{\partial a} h_L^m \frac{\partial e_L}{\partial x_2}/e_1 - \frac{\partial I^m}{\partial x_2} h_L^m \frac{\partial e_L}{\partial a}/e_1$$

$$
\begin{aligned}
&+\left(\frac{\partial I^m}{\partial a}\frac{\partial h_2^m}{\partial x_2}-\frac{\partial I^m}{\partial x_2}\frac{\partial h_2^m}{\partial a}\right)\frac{e_2}{e_1}+\cdots+\left(\frac{\partial I^m}{\partial a}\frac{\partial h_L^m}{\partial x_2}-\frac{\partial I^m}{\partial x_2}\frac{\partial h_L^m}{\partial a}\right)\frac{e_L}{e_1}\\
=\;&\frac{\partial I^m}{\partial x_2}\frac{\partial h_1^m}{\partial a}-\frac{\partial I^m}{\partial a}\frac{\partial h_1^m}{\partial x_2}. \qquad (2.C.3)
\end{aligned}
$$

By the same token, we could get another $(d_x - 2)$ equations like (2.C.2) and (2.C.3). After some rearrangement, we get

$$
B^m(x,a)\left(\frac{\nabla e_1(x,a)'}{e_1(x,a)},\ldots,\frac{\nabla e_L(x,a)'}{e_1(x,a)},\frac{e_2(x,a)}{e_1(x,a)},\ldots,\frac{e_L(x,a)}{e_1(x,a)}\right)' \;=\; A^m(x,a). \quad (2.C.4)
$$

In the above equation, for $l = 1, \ldots, L$, the $(d_x + 1) \times 1$ vector $\nabla e_l(x,a)$ is defined as

$$
\nabla e_l(x,a) \equiv \left(\nabla_x e_l(x,a)', \frac{\partial e_l(x,a)}{\partial a}\right)';
$$

the $d_x \times 1$ vector $A^m(x,a)$ is defined as

$$
A^m(x,a) \equiv \frac{\partial h_1^m(x,a)}{\partial a}\nabla_x I^m(x,a) - \frac{\partial I^m(x,a)}{\partial a}\nabla_x h_1^m(x,a);
$$

and the $d_x \times (d_x L + 2L - 1)$ matrix $B^m(x,a)$ is defined as

$$
B^m(x,a) \equiv \left(B_1^m(x,a), \ldots, B_L^m(x,a), B_{L+1}^m(x,a)\right),
$$

in which for $l = 1, \ldots, L$, the $d_x \times (d_x L + L)$ matrix $B_l^m(x,a)$ is

$$
B_l^m(x,a) \equiv \left(\frac{\partial I^m(x,a)}{\partial a} h_l^m(x,a)\mathbb{I}_{d_x}, -h_l^m(x,a)\nabla_x I^m(x,a)\right),
$$

and the $d_x \times (L-1)$ matrix $B^m_{L+1}(x,a)$ is

$$B^m_{L+1}(x,a) \equiv \left(B^m_{L+1,2}, \dots, B^m_{L+1,L}\right),$$

where for $l' = 2, \dots, L$

$$B^m_{L+1,l'} \equiv \frac{\partial I^m(x,a)}{\partial a} \nabla_x h^m_{l'}(x,a) - \frac{\partial h^m_{l'}(x,a)}{\partial a} \nabla_x I^m(x,a)$$

If we stack the equations like (2.C.4) for all markets, we get a system of Md_x equations with $d_x L + 2L - 1$ unknowns for all $(x,a) \in \mathcal{X} \times \mathcal{A}$,

$$\tilde{B}(x,a) \left(\nabla e_1(x,a)', \dots, \nabla e_L(x,a)', \frac{e_2(x,a)}{e_1(x,a)}, \dots, \frac{e_L(x,a)}{e_1(x,a)}\right)' = \tilde{A}(x,a), \quad (2.C.5)$$

where

$$\tilde{B}(x,a) \equiv \left(B^1(x,a)', \dots, B^M(x,a)' \right)',$$

and

$$\tilde{A}(x,a) \equiv \left(A^1(x,a)', \dots, A^M(x,a)' \right)'.$$

Therefore, there exists a unique solution of $(\nabla e_1(x,a)', \dots, \nabla e_L(x,a)', \frac{e_2(x,a)}{e_1(x,a)}, \dots, \frac{e_L(x,a)}{e_1(x,a)})'$ if the matrix $\tilde{B}(x,a)$ has full column rank. A necessary condition for this is that $M \geq L + (2L-1)/d_x$. The full-column-rank condition here has a similar gradient interpretation as in Section 2.3.2.2, but I will not fully elaborate it.

By normalize $e_1(\bar{x}, \bar{a}) = 1$, and solving the ordinary differential equations for each $e_l(x,a)$

$(l = 1, \dots, L)$ with the steps described in the proof of Theorem 2.3.4, one can recover all the quality functions $e_l(x, a)$ $(l = 1, \dots, L)$.

Finally, note that with large M, one might get over-identification as well.

2.D Proofs of the Theorems in Section 2.4.2

2.D.1 Proof of the Theorem in Section 2.4.2.1

This section provides the proof of Theorem 2.4.1. But some notation is needed first. Let $\Lambda_i^m \equiv \Lambda_{k_{Q,N}}(x_i^m)$, $\omega_{ij} \equiv \mathbb{I}(I_j^m \leq I_i^m) - F_{I^m|x^m}(I_i^m|x_j^m)$ $(i, j = 1, \dots, N)$ and $\hat{W}^m \equiv \sum_{i=1}^N \Lambda_i^m \Lambda_i^{m\prime}/N$.

Lemma 2.D.1. *For $x^m \equiv (x_1^m, \dots, x_N^m)$ and $k_{Q,N} \times 1$ vectors of functions $b_i(x^m)$ $(i = 1, \dots, N)$, if $\sum_{i=1}^N b_i(x^m)'\hat{W}^m b_i(x^m)/N = \mathcal{O}_p(r_N)$, then*

$$\sum_{i=1}^N \left[b_i(x^m)' \sum_{j=1}^N \Lambda_j^m \omega_{ij}/\sqrt{N} \right]^2 /N = \mathcal{O}_p(r_N).$$

Proof. This lemma is the same as Lemma S.1 in Imbens and Newey (2009), only with the notation adapted to that in this paper. □

Lemma 2.D.2. *Suppose that Assumption 2.4.3 is satisfied, then there exists C such that for each I there is $\rho(I)$ with $\sup_{x \in \mathcal{X}} |F_{I^m|x^m}(I|x) - \Lambda_k(x)'\rho(I)| \leq Ck^{-d_1/d_x}$.*

Proof. This lemma is the same as Lemma S.2 in Imbens and Newey (2009),[46] only with the notation adapted to that in this paper. □

[46]It is a reiteration of Theorem 8 (p. 90) in Lorentz (1986).

Proof of Theorem 2.4.1

This theorem is the same as Lemma 11 in Imbens and Newey (2009), only with the notation adapted to that in this paper.

2.D.2 Proofs of the Theorems in Section 2.4.2.2

In the rest of this subsection, I will suppress the superscript m for functions and variables for notational simplicity. The results in Section 2.4.1.1 and the proofs in this subsection hold regardless of the market index m.

Recall that $l^m(x, a)$ denotes either the payment function $I^m(x, a)$ or the quantity function $h^m(x, a)$ in a market m. Let $l \equiv (l^m(x_1, a_1), \ldots, l^m(x_N, a_N))'$, $\tilde{l} \equiv (l^m(x_1, \hat{a}_1), \ldots, l^m(x_N, \hat{a}_N))'$, $\Phi_i \equiv \Phi_{k_{l,N}}(x_i, a_i)$, $\tilde{\Phi}_i \equiv \Phi_{k_{l,N}}(x_i, \hat{a}_i)$, $\Phi \equiv (\Phi_1, \ldots, \Phi_N)'$, $\tilde{\Phi} \equiv (\tilde{\Phi}_1, \ldots, \tilde{\Phi}_N)'$, $Q \equiv \mathbb{E}(\Phi_i \Phi_i')$, $\bar{Q} \equiv \Phi'\Phi/N$, and $\tilde{Q} \equiv \tilde{\Phi}'\tilde{\Phi}/N$. Without loss of generality, we can set $Q = \mathbb{I}_{k_{l,N}}$, the $k_{l,N} \times k_{l,N}$ identity matrix, as in Newey (1997). Note that the estimated series coefficients in equation (2.4.3) and equation (2.4.6) can be written with this notation as $\hat{\xi}_{l,k_{l,N}} \equiv \tilde{Q}^{-}\tilde{\Phi}'l/N$. Finally, let $\tilde{\xi}_{l,k_{l,N}} \equiv \tilde{Q}^{-}\tilde{\Phi}'\tilde{l}/N$.

Recall that the estimated series coefficients $\hat{\xi}_{l,k_{l,N}}$ take least square forms. So the proof in this subsection proceeds in three steps: (i) to show that the "denominator" of the estimated series coefficients converges in probability to a constant matrix; (ii) to find out the rate at which the "numerator" converges to its probability limit, hence the estimated series coefficients converge to the pseudo-true series coefficients at the same rate; (iii) to obtain the convergence rates for $\hat{l}(x, a)$ and its derivatives using the results in step (ii), the compact support assumption, and the assumptions on the approximation errors by the series basis functions. In what follows, Lemma 2.D.4 presents step (i), Lemma 2.D.5 presents step (ii), and step (iii) is given by Theorems 2.4.2 and 2.4.3.

Lemma 2.D.3. *Suppose that Assumptions 2.4.4 and 2.4.5 are satisfied. Then,* $\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \|\Phi_k(x,a)\| \leq C\zeta_0(k)$, $\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \|\partial\Phi_k(x,a)/\partial x_j\| \leq C\zeta_j(k)$ *and* $\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \|\partial\Phi_k(x,a)/\partial a\| \leq C\zeta_a(k)$.

Proof. Under the maintained Assumptions 2.3.1 and 2.3.2, the joint density of (x, a) is bounded away from zero. Combine this with Assumption 2.4.5, then the results follow from equations (3.13)-(3.16) in Andrews (1991). □

Lemma 2.D.4. *Suppose that the conditions of Theorem 2.4.1 and Lemma 2.D.3 are satisfied. Suppose as well that the numbers of series basis functions used to approximate each component in* $k_{l,N}$ *all increase to infinity with* N, *and* $\sqrt{k_{l,N}}\nu_{a,N}\zeta_a(k_{l,N}) \to 0$. *Then, the following results hold:*

(i) $\|\tilde{\Phi} - \Phi\|^2/N = \mathcal{O}_p\left(\nu_{a,N}^2\zeta_a^2(k_{l,N})\right)$;

(ii) $\|\bar{Q} - Q\| = \mathcal{O}_p\left(\zeta_0(k_{l,N})\sqrt{k_{l,N}/N}\right)$;

(iii) $\|\tilde{Q} - \bar{Q}\| = \mathcal{O}_p\left(\nu_{a,N}^2\zeta_a^2(k_{l,N}) + \sqrt{k_{l,N}}\nu_{a,N}\zeta_a(k_{l,N})\right)$;

(iv) $\lambda_{\min}(\tilde{Q}) \geq c > 0$, $\lambda_{\min}(\bar{Q}) \geq c > 0$ *with probability approaching 1, where* $\lambda_{\min}$ *denotes the minimum eigenvalues of a symmetric matrix.*

Proof. For (i), consider a mean value expansion for $i \in \{1, \dots N\}$,

$$\tilde{\Phi}_i = \Phi_i + \frac{\partial\Phi_{k_{l,N}}}{\partial a}(x_i, \tilde{a}_i) \cdot (\hat{a}_i - a_i),$$

where $\tilde{a}_i$ lies between $\hat{a}_i$ and a_i. Since $\hat{a}_i$ and a_i are in $[0,1]$, so is $\tilde{a}_i$. By Lemma 2.D.3, $\|\partial\Phi_{k_{l,N}}(x_i,\tilde{a}_i)/\partial a\| \le C\zeta_a(k_{l,N})$. Then by Cauchy-Schwarz inequality, $\|\tilde{\Phi}_i - \Phi_i\| \le C\zeta_a(k_{l,N})$ $|\hat{a}_i - a_i|$. Together with Theorem 2.4.1, this implies

$$\|\tilde{\Phi} - \Phi\|^2/N \;=\; \sum_{i=1}^{N} \|\tilde{\Phi}_i - \Phi_i\|^2/N = \mathcal{O}_p\left(\nu_{a,N}^2\zeta_a^2(k_{l,N})\right).$$

So (i) holds.

For (ii), let $\mathbb{I}_{jl}$ denote the (j,l)-element of an identity matrix. Note that $\mathbb{E}(\phi_j(x,a)$ $\phi_l(x,a)) = \mathbb{I}_{jl}$, then

$$\begin{aligned}
\mathbb{E}\left[\|\bar{Q} - Q\|^2\right] &= \mathbb{E}\left[\left\|N^{-1}\sum_{i=1}^{n}\Phi_i\Phi_i' - Q\right\|^2\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{k_{l,N}}\sum_{l=1}^{k_{l,N}}\left(N^{-1}\sum_{i=1}^{n}\phi_j(x_i,a_i)\phi_l(x_i,a_i) - \mathbb{I}_{jl}\right)^2\right] \\
&\le N^{-1}\mathbb{E}\left(\sum_{j=1}^{k_{l,N}}\phi_j^2(x_i,a_i)\sum_{l=1}^{k_{l,N}}\phi_l^2(x_i,a_i)\right) \\
&\le N^{-1}\zeta_0^2(k_{l,N})\mathrm{tr}(\mathbb{I}_{k_{l,N}}) \\
&= \zeta_0^2(k_{l,N})k_{l,N}/N.
\end{aligned}$$

So (ii) follows by the Markov's inequality.

For (iii), by the triangular inequality and the Cauchy-Schwarz inequality,

$$\|\tilde{Q} - \bar{Q}\| \;\le\; \sum_{i=1}^{N}\|\tilde{\Phi}_i\tilde{\Phi}_i' - \Phi_i\Phi_i'\|/N$$

$$\leq \sum_{i=1}^{N} \|\tilde{\Phi}_i - \Phi_i\|^2/N + 2\left(\sum_{i=1}^{N} \|\tilde{\Phi}_i - \Phi_i\|^2/N\right)^{1/2} \left(\sum_{i=1}^{N} \|\Phi_i\|^2/N\right)^{1/2}.$$

Moreover, by the Markov's inequality

$$\sum_{i=1}^{N} \|\Phi_i\|^2/N = \mathcal{O}_p\left(\mathbb{E}\left(\|\Phi_i\|^2\right)\right) = \mathcal{O}_p\left(\mathrm{tr}(Q)\right) = \mathcal{O}_p\left(\mathrm{tr}(\mathbb{I}_{k_{l,N}})\right) = \mathcal{O}_p\left(k_{l,N}\right). \quad (2.D.1)$$

So the result follows from (i).

For (iv), by the definition of $\zeta_0(k_{l,N})$ and $\zeta_a(k_{l,N})$, and the fact that $\nu_{a,N}$ converges to zero slower than $N^{-1/2}$, we have that $\sqrt{k_{l,N}}\nu_{a,N}\zeta_a(k_{l,N}) \to 0$ implies $\zeta_0(k_{l,N})\sqrt{k_{l,N}/N} \to 0$ and $\nu_{a,N}^2\zeta_a^2(k_{l,N}) \to 0$. Therefore by (ii) and (iii), we have that $\|\bar{Q}-Q\| \xrightarrow{p.} 0$ and $\|\tilde{Q}-\bar{Q}\| \xrightarrow{p.} 0$. By the same argument following equation (A.1) in Newey (1997), $|\lambda_{\min}(\bar{Q}) - \lambda_{\min}(Q)|$ and $|\lambda_{\min}(\tilde{Q}) - \lambda_{\min}(\bar{Q})|$ are bounded by $\|\bar{Q} - Q\|$ and $\|\tilde{Q} - \bar{Q}\|$, respectively. Since $Q \equiv I_{k_{l,N}}$, $\lambda_{\min}(\bar{Q}) \xrightarrow{p.} 1$ and $\lambda_{\min}(\tilde{Q}) \xrightarrow{p.} 1$. So the result follows. □

Lemma 2.D.5. *Suppose that Assumptions 2.4.6 and 2.4.7, and the conditions of Theorem 2.4.1 and Lemma 2.D.4 are satisfied. Then, the following results hold:*

(i) $\|\hat{\xi}_{l,k_{l,N}} - \tilde{\xi}_{l,k_{l,N}}\| = \mathcal{O}_p\left(\nu_{a,N}\right)$;

(ii) $\|\hat{\xi}_{l,k_{l,N}} - \xi_{l,0,k_{l,N}}\| = \mathcal{O}_p\left(k_{l,N}^{-\alpha_l}\right)$.

Proof. For (i), consider a mean value expansion for $i \in \{1, \dots N\}$,

$$l(x_i, \hat{a}_i) = l(x_i, a_i) + \frac{\partial l}{\partial a}(x_i, \tilde{a}_i) \cdot (\hat{a}_i - a_i),$$

where $\tilde{a}_i$ lies between $\hat{a}_i$ and a_i and might take a different value from that in the proof of Lemma 2.D.4. Since $\hat{a}_i$ and a_i are in $[0,1]$, so is $\tilde{a}_i$. Together with Assumptions 2.4.4 and 2.4.6, this implies that $|\partial l(x_i,\tilde{a}_i)/\partial a| \le C$. Moreover, by Lemma 2.D.4, we have that $\lambda_{\min}(\tilde{Q}) \ge c$ with probability 1, so

$$\begin{aligned}
\left\|\tilde{Q}^{1/2}(\hat{\xi}_{l,k_{l,N}} - \tilde{\xi}_{l,k_{l,N}})\right\|^2 &= (l-\tilde{l})'\tilde{\Phi}\tilde{Q}^{-}\tilde{\Phi}'(l-\tilde{l})/N^2 \\
&\le C\|\tilde{l}-l\|^2/N \\
&\le C\sum_{i=1}^{N}|\hat{a}_i - a_i|^2/N,
\end{aligned}$$

Then (i) holds by Theorem 2.4.1 and Lemma 2.D.4 (iv).

Similarly, for (ii), by the definition of $\tilde{\xi}_{l,k_{l,N}}$,

$$\begin{aligned}
\left\|\tilde{Q}^{1/2}(\tilde{\xi}_{l,k_{l,N}} - \xi_{l,0,k_{l,N}})\right\|^2 &= \left\|\tilde{Q}^{1/2}(\tilde{\xi}_{l,k_{l,N}} - \tilde{Q}^{-}\tilde{\Phi}'\tilde{\Phi}\xi_{l,0,k_{l,N}}/N)\right\|^2 \\
&= (\tilde{l}-\tilde{\Phi}\xi_{l,0,k_{l,N}})'\tilde{\Phi}\tilde{Q}^{-}\tilde{\Phi}'(\tilde{l}-\tilde{\Phi}\xi_{l,0,k_{l,N}})/N^2 \\
&\le C\|\tilde{l}-\tilde{\Phi}\xi_{l,0,k_{l,N}}\|^2/N \\
&\le C\left(\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}|l(x,a)-\Phi_{k_{l,N}}(x,a)'\xi_{l,0,k_{l,N}}|^2\right) \\
&= \mathcal{O}_p\left(k_{l,N}^{-2\alpha_l}\right),
\end{aligned}$$

where the last equality holds by Assumption 2.4.7. Therefore the result holds by Lemma 2.D.4 (iv). □

Proof of Theorem 2.4.2

Proof. By the definition of $\zeta_0(k_{l,N})$ and $\zeta_a(k_{l,N})$, the condition $k_{l,N}^{3/2}k_{a,l,N}^2\nu_{a,N} \to 0$ implies that $\sqrt{k_{l,N}}\zeta_a(k_{l,N})\nu_{a,N} \to 0$.

By the triangular inequality,

$$
\begin{aligned}
& \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\hat{l}(x,a) - l(x,a)| \\
\leq\ & \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\Phi_{k_{l,N}}(x,a)'(\hat{\xi}_{l,k_{l,N}} - \xi_{l,0,k_{l,N}})| + \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\Phi_{k_{l,N}}(x,a)'\xi_{l,0,k_{l,N}} - l(x,a)| \\
=\ & \mathcal{O}_p\left(\zeta_0(k_{l,N})\left(\nu_{a,N} + k_{l,N}^{-\alpha_l}\right)\right) + \mathcal{O}_p\left(k_{l,N}^{-\alpha_l}\right) \\
=\ & \mathcal{O}_p\left(\zeta_0(k_{l,N})\left(\nu_{a,N} + k_{l,N}^{-\alpha_l}\right)\right),
\end{aligned}
$$

The first equality holds by the Cauchy-Schwarz inequality, Assumption 2.4.7, and Lemmas 2.D.3 and 2.D.5. The second equality holds since $\zeta_0(k_{l,N}) \to \infty$ as $N \to \infty$. This completes the proof. □

Proof of Theorem 2.4.3

Proof. For $j = 1, \ldots, d_x$, by the triangular inequality,

$$
\begin{aligned}
& \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left|\hat{l}_{x_j}^m(x,a) - l_{x_j}^m(x,a)\right| \\
\leq\ & \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left|\frac{\partial}{\partial x_j}\Phi_{k_{l,N}}(x,a)'(\hat{\xi}_{l,k_{l,N}} - \xi_{l,0,k_{l,N}})\right| \\
& + \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left|l_{x_j}^m(x,a) - \frac{\partial}{\partial x_j}\Phi_{k_{l,N}}(x,a)'\xi_{l,0,k_{l,N}}\right| \\
=\ & \mathcal{O}_p\left(\zeta_j(k_{l,N})\left(\nu_{a,N} + k_{l,N}^{-\alpha_l}\right)\right) + \mathcal{O}_p\left(k_{l,N}^{-\alpha_l}\right)
\end{aligned}
$$

$$= \mathcal{O}_p\left(\zeta_j(k_{l,N})\left(\nu_{a,N} + k_{l,N}^{-\alpha_l}\right)\right).$$

The first equality holds by the Cauchy-Schwarz inequality, Assumption 2.4.7, and Lemmas 2.D.3 and 2.D.5. The second equality holds by that $\zeta_j(k_{l,N}) \to \infty$ as $N \to \infty$. This completes the proof of the first statement. The proof of the second statement follows the same argument. □

2.D.3 Proofs of the Theorems in Section 2.4.2.3

This subsection proceeds with the same steps as in Appendix 2.D.2. In the rest of the proof, I will spell out the superscripts of the market index m.

Define

$$\bar{S}_{\Phi\Phi} \equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\bar{S}_{\Phi}(x_i^m, \hat{a}_i^m)'\bar{S}_{\Phi}(x_i^m, \hat{a}_i^m).$$

In this equation,

$$\bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) \equiv \left(\bar{S}_{\Phi,1}(x_i^m, \hat{a}_i^m)\ ,\ \bar{S}_{\Phi,2}(x_i^m, \hat{a}_i^m)\right),$$

where

$$\bar{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) \equiv \begin{pmatrix} \Phi_{k_{x_1},MN}(x_i^m, \hat{a}_i^m)' & & 0 \\ & \ddots & \\ 0 & & \Phi_{k_{x_{d_x}},MN}(x_i^m, \hat{a}_i^m)' \end{pmatrix} \otimes I_a^m(x_i^m, \hat{a}_i^m),$$

and

$$\bar{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) \equiv -\nabla_x I^m(x_i^m, \hat{a}_i^m) \otimes \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'.$$

Lemma 2.D.6. *Suppose that Assumption 2.4.6, and the conditions of Theorems 2.4.1 and 2.4.3 and Lemma 2.D.3 are satisfied. Suppose as well that the numbers of series basis functions used to approximate each component in $k_{x_j,MN}$ ($j = 1, \ldots, d_x$) and $k_{a,MN}$ all increase to infinity with N, $\nu_{a,N}(\zeta_a(k_{x_j,MN}) + \zeta_a(k_{a,MN})) \to 0$, $k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN} \to 0$, $(\nu_{I_a,N} \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}) \to 0$. Then*

(i)

$$\begin{aligned}
&\|\bar{S}_{\Phi\Phi} - S_{\Phi\Phi}\| \\
=\ & \mathcal{O}_p\Bigg(\nu_{a,N}^2 \sum_{j=1}^{d_x} \left(\zeta_a^2(k_{x_j,MN}) + \zeta_a^2(k_{a,MN})\right) \\
&+\nu_{a,N}\left[\sum_{j=1}^{d_x} \left(\zeta_a(k_{x_j,MN}) + \zeta_a(k_{a,MN})\right)\right]\left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right)^{1/2}\Bigg);
\end{aligned}$$

(ii)

$$\begin{aligned}
&\|\hat{S}_{\Phi\Phi} - \bar{S}_{\Phi\Phi}\| \\
=\ & \mathcal{O}_p\left(\left(\nu_{I_a,N}^2 \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2\right)^{1/2}\left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right)^{1/2}\right);
\end{aligned}$$

(iii) $\lambda_{\min}(\hat{S}_{\Phi\Phi}) \geq c$, $\lambda_{\min}(\bar{S}_{\Phi\Phi}) \geq c$ and $\lambda_{\min}(S_{\Phi\Phi}) \geq c$ with probability approaching 1, where $\lambda_{\min}$ denotes the minimum eigenvalue of a symmetric matrix.

Proof. To prove (i), some preliminary results are needed. For $j = 1, \ldots, d_x$, consider the mean value expansion

$$
\begin{aligned}
& I_a^m(x_i^m,\hat{a}_i^m)\Phi_{k_{x_j,MN}}(x_i^m,\hat{a}_i^m)' - I_a^m(x_i^m,a_i^m)\Phi_{k_{x_j,MN}}(x_i^m,a_i^m)' \\
= & I_{aa}^m(x_i^m,\tilde{a}_i^m)\Phi_{k_{x_j,MN}}(x_i^m,\tilde{a}_i^m)'(\hat{a}_i^m - a_i^m) \\
& + I_a^m(x_i^m,\tilde{a}_i^m)\frac{\partial}{\partial a}\Phi_{k_{x_j,MN}}(x_i^m,\tilde{a}_i^m)'(\hat{a}_i^m - a_i^m),
\end{aligned}
$$

where $\tilde{a}_i^m$ is between a_i^m and $\hat{a}_i^m$, so it must be in $[0,1]$. By Lemma 2.D.3, Assumption 2.4.6, the triangular inequality, and the Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
& \|I_a^m(x_i^m,\hat{a}_i^m)\Phi_{k_{x_j,MN}}(x_i^m,\hat{a}_i^m)' - I_a^m(x_i^m,a_i^m)\Phi_{k_{x_j,MN}}(x_i^m,a_i^m)'\|^2 \\
\leq & C\left(\zeta_0^2(k_{x_j,MN}) + \zeta_a^2(k_{x_j,MN})\right)|\hat{a}_i^m - a_i^m|^2 \\
\leq & C\zeta_a^2(k_{x_j,MN})|\hat{a}_i^m - a_i^m|^2. \qquad (2.D.2)
\end{aligned}
$$

By the same token, we have that for $j = 1, \dots, d_x$,

$$
\begin{aligned}
& \|I_{x_j}^m(x_i^m,\hat{a}_i^m)\Phi_{k_{a,MN}}(x_i^m,\hat{a}_i^m)' - I_{x_j}^m(x_i^m,a_i^m)\Phi_{k_{a,MN}}(x_i^m,a_i^m)'\|^2 \\
\leq & C\left(\zeta_0^2(k_{a,MN}) + \zeta_a^2(k_{a,MN})\right)|\hat{a}_i^m - a_i^m|^2 \\
\leq & C\zeta_a^2(k_{a,MN})|\hat{a}_i^m - a_i^m|^2. \qquad (2.D.3)
\end{aligned}
$$

Equation (2.D.2) implies that

$$
\left\|\bar{S}_{\Phi,1}(x_i^m,\hat{a}_i^m) - S_{\Phi,1}(x_i^m,a_i^m)\right\|^2
$$

$$
\begin{aligned}
&= \sum_{j=1}^{d_x} \|I_a^m(x_i^m, \hat{a}_i^m)\Phi_{k_{x_j,MN}}(x_i^m, \hat{a}_i^m)' - I_a^m(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\|^2 \\
&\leq C \quad |\hat{a}_i^m - a_i^m|^2 \sum_{j=1}^{d_x} \zeta_a^2(k_{x_j,MN}).
\end{aligned}
$$

And equation (2.D.3) implies that

$$
\begin{aligned}
& \left\|\bar{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) - S_{\Phi,2}(x_i^m, a_i^m)\right\|^2 \\
= & \sum_{j=1}^{d_x} \|I_{x_j}^m(x_i^m, \hat{a}_i^m)\Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)' - I_{x_j}^m(x_i^m, a_i^m)\Phi_{k_{a,MN}}(x_i^m, a_i^m)'\|^2 \\
\leq & \; C d_x |\hat{a}_i^m - a_i^m|^2 \zeta_a^2(k_{a,MN}).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
& \left\|\bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) - S_{\Phi}(x_i^m, a_i^m)\right\|^2 \\
= & \left\|\bar{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) - S_{\Phi,1}(x_i^m, a_i^m)\right\|^2 + \left\|\bar{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) - S_{\Phi,2}(x_i^m, a_i^m)\right\|^2 \\
= & \; C|\hat{a}_i^m - a_i^m|^2 \sum_{j=1}^{d_x} \left(\zeta_a^2(k_{x_j,MN}) + \zeta_a^2(k_{a,MN})\right). \qquad (2.D.4)
\end{aligned}
$$

On the other hand, by Lemma 2.D.3, Assumption 2.4.6 and the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\|S_{\Phi,1}(x_i^m, a_i^m)\|^2 &= \sum_{j=1}^{d_x} \left\|I_a^m(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2 \\
&\leq \sum_{j=1}^{d_x} |I_a^m(x_i^m, a_i^m)|^2 \cdot \left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\right\|^2
\end{aligned}
$$

$$= \mathcal{O}_p\left(\sum_{j=1}^{d_x} k_{x_j,MN}\right).$$

In this expression, the inequality holds by the Cauchy-Schwarz inequality. The second equality holds because I set the basis functions to be orthonormal without loss of generality, and hence for $j = 1, \ldots, d_x$,

$$\mathbb{E}\left(\left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\right\|^2\right) = \operatorname{tr}(\mathbb{I}_{k_{x_j,MN}}) = k_{x_j,MN}.$$

Then by the Markov's inequality,

$$\left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(k_{x_j,MN}\right). \tag{2.D.5}$$

By similar argument, we also have

$$\left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(k_{a,MN}\right), \tag{2.D.6}$$

which implies that

$$\|S_{\Phi,2}(x_i^m, a_i^m)\|^2 = \mathcal{O}_p\left(k_{a,MN}\right).$$

As a result,

$$\begin{aligned} \|S_\Phi(x_i^m, a_i^m)\|^2 &= \|S_{\Phi,1}(x_i^m, a_i^m)\|^2 + \|S_{\Phi,2}(x_i^m, a_i^m)\|^2 \\ &= \mathcal{O}_p\left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right). \end{aligned} \tag{2.D.7}$$

Now consider (i),

$$
\begin{aligned}
& \|\bar{S}_{\Phi\Phi} - S_{\Phi\Phi}\| \\
= \; & (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\| \bar{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) - S_{\Phi}(x_i^m, a_i^m)' S_{\Phi}(x_i^m, a_i^m) \right\| \\
\leq \; & (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\| \bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) - S_{\Phi}(x_i^m, a_i^m) \right\|^2 \\
& +2(MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left(\left\| \bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) - S_{\Phi}(x_i^m, a_i^m) \right\|^2 \right)^{1/2} \left(\| S_{\Phi}(x_i^m, a_i^m) \|^2 \right)^{1/2},
\end{aligned}
$$

where the inequality holds by the triangular inequality and the Cauchy-Schwarz inequality. Combine this result with Theorem 2.4.1, equation (2.D.4) and equation (2.D.7), we get

$$
\begin{aligned}
& \|\bar{S}_{\Phi\Phi} - S_{\Phi\Phi}\| \\
= \; & \mathcal{O}_p \left(\nu_{a,N}^2 \sum_{j=1}^{d_x} \left(\zeta_a^2(k_{x_j,MN}) + \zeta_a^2(k_{a,MN}) \right) \right. \\
& \left. + \nu_{a,N} \left[\sum_{j=1}^{d_x} \left(\zeta_a(k_{x_j,MN}) + \zeta_a(k_{a,MN}) \right) \right] \left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN} \right)^{1/2} \right).
\end{aligned}
$$

So (i) holds.

To prove (ii), some preliminary results are necessary. Note that the Cauchy-Schwarz inequality, Theorem 2.4.3, equation (2.D.5) and equation (2.D.6) imply that

$$
\begin{aligned}
& \left\| \left(\hat{I}_{x_j}^m(x_i^m, \hat{a}_i^m) - I_{x_j}^m(x_i^m, \hat{a}_i^m) \right) \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m) \right\|'^2 \\
\leq \; & \left| \hat{I}_{x_j}^m(x_i^m, \hat{a}_i^m) - I_{x_j}^m(x_i^m, \hat{a}_i^m) \right|^2 \cdot \left\| \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m) \right\|^2 \\
= \; & \mathcal{O}_p \left(\nu_{I_j,N}^2 k_{a,MN} \right),
\end{aligned}
$$

for $j = 1, \dots, d_x$, and

$$\begin{aligned} & \left\| \left(\hat{I}_a^m(x_i^m, \hat{a}_i^m) - I_a^m(x_i^m, \hat{a}_i^m) \right) \Phi_{k_{x_j,MN}}(x_i^m, \hat{a}_i^m)' \right\|^2 \\ \leq & \left| \hat{I}_a^m(x_i^m, \hat{a}_i^m) - I_a^m(x_i^m, \hat{a}_i^m) \right|^2 \cdot \left\| \Phi_{k_{x_j,MN}}(x_i^m, \hat{a}_i^m) \right\|^2 \\ = & \mathcal{O}_p \left(\nu_{I_a}^2(\sigma_N, k_{I,N}) k_{x_j,MN} \right). \end{aligned}$$

They further imply that

$$\begin{aligned} & \left\| \hat{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) - \bar{S}_{\Phi,1}(x_i^m, \hat{a}_i^m) \right\|^2 \\ = & \sum_{j=1}^{d_x} \left\| \left(\hat{I}_a^m(x_i^m, \hat{a}_i^m) - I_a^m(x_i^m, \hat{a}_i^m) \right) \Phi_{k_{x_j,MN}}(x_i^m, \hat{a}_i^m)' \right\|^2 \\ = & \mathcal{O}_p \left(\nu_{I_a}^2(\sigma_N, k_{I,N}) \sum_{j=1}^{d_x} k_{x_j,MN} \right), \end{aligned}$$

and

$$\begin{aligned} & \left\| \hat{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) - \bar{S}_{\Phi,2}(x_i^m, \hat{a}_i^m) \right\|^2 \\ = & \sum_{j=1}^{d_x} \left\| \left(\hat{I}_{x_j}^m(x_i^m, \hat{a}_i^m) - I_{x_j}^m(x_i^m, \hat{a}_i^m) \right) \Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)' \right\|^2 \\ = & \mathcal{O}_p \left(k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2 \right). \end{aligned}$$

As a result,

$$\left\| \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m) - \bar{S}_{\Phi}(x_i^m, \hat{a}_i^m) \right\|^2$$

$$
\begin{aligned}
&= \left\|\hat{S}_{\Phi,1}(x_i^m,\hat{a}_i^m) - \bar{S}_{\Phi,1}(x_i^m,\hat{a}_i^m)\right\|^2 + \left\|\hat{S}_{\Phi,2}(x_i^m,\hat{a}_i^m) - \bar{S}_{\Phi,2}(x_i^m,\hat{a}_i^m)\right\|^2 \\
&= \mathcal{O}_p\left(\nu_{I_a,N}^2 \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2\right). \qquad (2.D.8)
\end{aligned}
$$

On the other hand, by the fact that $\hat{a}_i^m \in [0,1]$, Lemma 2.D.3, Assumption 2.4.6, the Cauchy-Schwarz inequality, and that the basis functions are orthonormal, we have

$$
\left\|\bar{S}_{\Phi,1}(x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(\sum_{j=1}^{d_x} k_{x_j,MN}\right),
$$

$$
\left\|\bar{S}_{\Phi,2}(x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(k_{a,MN}\right).
$$

As a result

$$
\begin{aligned}
\left\|\bar{S}_{\Phi}(x_i^m, a_i^m)\right\|^2 &= \left\|\bar{S}_{\Phi,1}(x_i^m, a_i^m)\right\|^2 + \left\|\bar{S}_{\Phi,2}(x_i^m, a_i^m)\right\|^2 \\
&= \mathcal{O}_p\left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right). \qquad (2.D.9)
\end{aligned}
$$

Now consider (ii),

$$
\begin{aligned}
&\|\hat{S}_{\Phi\Phi} - \bar{S}_{\Phi\Phi}\| \\
&= (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\|\hat{S}_{\Phi}(x_i^m,\hat{a}_i^m)'\hat{S}_{\Phi}(x_i^m,\hat{a}_i^m) - \bar{S}_{\Phi}(x_i^m,a_i^m)'\bar{S}_{\Phi}(x_i^m,a_i^m)\right\| \\
&\leq (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\|\hat{S}_{\Phi}(x_i^m,\hat{a}_i^m) - \bar{S}_{\Phi}(x_i^m,a_i^m)\right\|^2 \\
&\quad +2(MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left(\left\|\hat{S}_{\Phi}(x_i^m,\hat{a}_i^m) - \bar{S}_{\Phi}(x_i^m,a_i^m)\right\|^2\right)^{1/2} \left(\left\|\bar{S}_{\Phi}(x_i^m,a_i^m)\right\|^2\right)^{1/2},
\end{aligned}
$$

where the inequality holds by the triangular inequality and the Cauchy-Schwarz inequality. Combine this result with equation (2.D.8) and equation (2.D.9), we get

$$
\begin{aligned}
&\|\hat{S}_{\Phi\Phi} - \bar{S}_{\Phi\Phi}\| \\
= \ & \mathcal{O}_p\left(\left(\nu_{I_a,N}^2 \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2\right)\right. \\
&\left. + \left(\nu_{I_a,N}^2 \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2\right)^{1/2} \left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right)^{1/2}\right) \\
= \ & \mathcal{O}_p\left(\left(\nu_{I_a,N}^2 \sum_{j=1}^{d_x} k_{x_j,MN} + k_{a,MN} \sum_{j=1}^{d_x} \nu_{I_j,N}^2\right)^{1/2} \left(k_{a,MN} + \sum_{j=1}^{d_x} k_{x_j,MN}\right)^{1/2}\right).
\end{aligned}
$$

So (ii) holds.

To prove (iii), note that

$$
\begin{aligned}
&\mathbb{E}\left[\|S_{\Phi\Phi} - \mathbb{E}(S_{\Phi\Phi})\|^2\right] \\
\leq \ & \sum_{j=1}^{d_x} \mathbb{E}\left[(MN)^{-1} \left(I_a^m(x_i^m, a_i^m)\right)^4 \left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
&+2\sum_{j=1}^{d_x} \mathbb{E}\left[(MN)^{-1} \left(I_a^m(x_i^m, a_i^m)\right)^2 \left(I_{x_j}^m(x_i^m, a_i^m)\right)^2 \left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
&+\mathbb{E}\left[\left(\sum_{j=1}^{d_x}\left(I_{x_j}^m(x_i^m, a_i^m)\right)^2\right)^2 \left\|\Phi_{k_{a,,N}}(x_i^m, a_i^m)\Phi_{k_{a,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
\leq \ & B_I^4 \sum_{j=1}^{d_x} \mathbb{E}\left[(MN)^{-1} \left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
&+2B_I^4 \sum_{j=1}^{d_x} \mathbb{E}\left[(MN)^{-1} \left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
&+d_x^2 B_I^4 \mathbb{E}\left[\left\|\Phi_{k_{a,,N}}(x_i^m, a_i^m)\Phi_{k_{a,MN}}(x_i^m, a_i^m)'\right\|^2\right],
\end{aligned}
\tag{2.D.10}
$$

where the first inequality holds by the definition of $S_{\Phi\Phi}$, Assumption 2.4.1, and that the second moment of a random variable is no less than its variance; the second inequality holds by Assumption 2.4.6. Recall that I assume the series basis functions are orthonormal (i.e. $Q = \mathbb{I}$), then by Lemma 2.D.3, we have that for $j = 1, \ldots, d_x$,

$$\begin{aligned}
&\mathbb{E}\left[(MN)^{-1}\left\|\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \\
= \;& (MN)^{-1}\mathbb{E}\left[\left(\sum_{k=1}^{k_{x_j,MN}} \phi_k^2(x_i^m, a_i^m)\right)\left(\sum_{l=1}^{k_{x_j,MN}} \phi_l^2(x_i^m, a_i^m)\right)\right] \\
\leq \;& (MN)^{-1}\zeta_0^2(k_{x_j,MN})\mathrm{tr}(\mathbb{I}_{k_{x_j,MN}}) \\
= \;& \zeta_0^2(k_{x_j,MN})k_{x_j,MN}/(MN). \qquad (2.D.11)
\end{aligned}$$

By the same token,

$$\mathbb{E}\left[(MN)^{-1}\left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \leq \zeta_0^2(k_{x_j,MN})k_{a,MN}/(MN), \qquad (2.D.12)$$

$$\mathbb{E}\left[(MN)^{-1}\left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\Phi_{k_{x_j,MN}}(x_i^m, a_i^m)'\right\|^2\right] \leq \zeta_0^2(k_{a,MN})k_{x_j,MN}/(MN), \qquad (2.D.13)$$

and

$$\mathbb{E}\left[(MN)^{-1}\left\|\Phi_{k_{a,MN}}(x_i^m, a_i^m)\Phi_{k_{a,MN}}(x_i^m, a_i^m)'\right\|^2\right] \leq \zeta_0^2(k_{a,MN})k_{a,MN}/(MN). \qquad (2.D.14)$$

Plug the bounds in equations (2.D.11)-(2.D.14) into equation (2.D.10), we get

$$\mathbb{E}\left[\|S_{\Phi\Phi} - \mathbb{E}(S_{\Phi\Phi})\|^2\right]$$

$$\leq \left[\zeta_0^2\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}\right)+\zeta_0^2(k_{a,MN})\right]\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}+k_{a,MN}\right)\Big/(MN).$$

Then by the Markov's inequality,

$$\begin{aligned}&\|S_{\Phi\Phi}-\mathbb{E}(S_{\Phi\Phi})\|\\ =\ &\mathcal{O}_p\left(\sqrt{\left[\zeta_0^2\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}\right)+\zeta_0^2(k_{a,MN})\right]\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}+k_{a,MN}\right)\Big/(MN)}\right).\end{aligned}$$

Since $\nu_{a,N}$ converges to zero at a slower rate than $N^{-1/2}$, $\nu_{a,N}(\zeta_a(k_{x_j,MN})+\zeta_a(k_{a,MN}))\to 0$ and $k_{a,MN}+\sum_{j=1}^{d_x} k_{x_j,MN}\to 0$ imply that $[\zeta_0^2\left(\max_{j=1,\ldots,d_x} k_{x_j,MN}\right)+\zeta_0^2(k_{a,MN})](\max_{j=1,\ldots,d_x}$ $k_{x_j,MN}+k_{a,MN})/(MN)\to 0$. As a result, $\|S_{\Phi\Phi}-\mathbb{E}(S_{\Phi\Phi})\|=o_p(1)$.

Note that $\nu_{a,N}(\zeta_a(k_{x_j,MN})+\zeta_a(k_{a,MN}))\to 0$ implies $\nu_{a,N}^2(\zeta_a^2(k_{x_j,MN})+\zeta_a^2(k_{a,MN}))\to 0$. Then by result (i), we have $\|\bar{S}_{\Phi\Phi}-S_{\Phi\Phi}\|=o_p(1)$. Moreover, by result (ii), and the conditions that the numbers of series basis functions used to approximate each component in $k_{x_j,MN}$ $(j=1,\ldots,d_x)$ and $k_{a,MN}$ all increase to infinity with N, $(\nu_{I_a,N}\sum_{j=1}^{d_x} k_{x_j,MN}+k_{a,MN}\sum_{j=1}^{d_x}$ $\nu_{I_j,N})\to 0$ for $j=1,\ldots,d_x$, $k_{a,MN}+\sum_{j=1}^{d_x} k_{x_j,MN}\to 0$, we have $\|\hat{S}_{\Phi\Phi}-\bar{S}_{\Phi\Phi}\|=o_p(1)$. Then (iii) follows by the same argument for the proof of Lemma 2.D.4(iv). This completes the proof of the lemma. □

Define

$$\begin{aligned}\bar{S}_{\Phi A} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_\Phi(x_i^m,\hat{a}_i^m)'\bar{A}^m(x_i^m,\hat{a}_i^m),\\ \bar{S}_{0,\Phi A} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_\Phi(x_i^m,\hat{a}_i^m)'\bar{A}_0^m(x_i^m,\hat{a}_i^m),\end{aligned}$$

$$\hat{S}_{0,\Phi A} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \hat{A}_0^m(x_i^m, \hat{a}_i^m),$$

where

$$\bar{A}^m(x_i^m, \hat{a}_i^m) \equiv \left[h_a^m(x_i^m, \hat{a}_i^m)\nabla_x I^m(x_i^m, \hat{a}_i^m) - I_a^m(x_i^m, \hat{a}_i^m)\nabla_x h^m(x_i^m, \hat{a}_i^m)\right] / h^m(x_i^m, \hat{a}_i^m),$$

$$\begin{aligned}
\bar{A}_0^m(x_i^m, \hat{a}_i^m) \equiv & \begin{pmatrix} I_a^m(x_i^m, \hat{a}_i^m)\Phi_{k_{x_1,MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,x_1,k_{x_1,MN}} \\ \vdots \\ I_a^m(x_i^m, \hat{a}_i^m)\Phi_{k_{x_{d_x},MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,x_{d_x},k_{x_{d_x},MN}} \end{pmatrix} \\
& - \nabla_x I^m(x_i^m, \hat{a}_i^m) \otimes \left[\Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,a,k_{a,MN}}\right], \\
\hat{A}_0^m(x_i^m, \hat{a}_i^m) \equiv & \begin{pmatrix} \hat{I}_a^m(x_i^m, \hat{a}_i^m)\Phi_{k_{x_1,MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,x_1,k_{x_1,MN}} \\ \vdots \\ \hat{I}_a^m(x_i^m, \hat{a}_i^m)\Phi_{k_{x_{d_x},MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,x_{d_x},k_{x_{d_x},MN}} \end{pmatrix} \\
& - \nabla_x \hat{I}^m(x_i^m, \hat{a}_i^m) \otimes \left[\Phi_{k_{a,MN}}(x_i^m, \hat{a}_i^m)'\beta_{0,a,k_{a,MN}}\right].
\end{aligned}$$

Now we need some intermediate coefficients which help analyze the estimated series coefficients for the quality function. Define

$$\begin{aligned}
\bar{\beta}_{MN} &\equiv \left(\bar{\beta}'_{x_1,k_{x_1,MN}}, \ldots, \bar{\beta}'_{x_{d_x},k_{x_{d_x},MN}}, \bar{\beta}'_{a,k_{a,MN}}\right)' \equiv \hat{S}^{-}_{\Phi\Phi}\bar{S}_{\Phi A}, \\
\bar{\beta}_{0,MN} &\equiv \left(\bar{\beta}'_{0,x_1,k_{x_1,MN}}, \ldots, \bar{\beta}'_{0,x_{d_x},k_{x_{d_x},MN}}, \bar{\beta}'_{0,a,k_{a,MN}}\right)' \equiv \hat{S}^{-}_{\Phi\Phi}\bar{S}_{0,\Phi A}.
\end{aligned}$$

And with some standard algebra, we get

$$\beta_{0,MN} \equiv \left(\beta'_{0,x_1,k_{x_1,MN}}, \ldots, \beta'_{0,x_{d_x},k_{x_{d_x},MN}}, \beta'_{0,a,k_{a,MN}}\right)' \equiv \hat{S}^{-}_{\Phi\Phi}\hat{S}_{0,\Phi A}.$$

Moreover,

$$\hat{\beta}_{MN} \equiv \left(\hat{\beta}'_{x_1,k_{x_1,MN}}, \ldots, \hat{\beta}'_{x_{d_X},k_{x_{d_x},MN}}, \hat{\beta}'_{a,k_{a,MN}}\right)'.$$

Note that if we let $\hat{S}_\Phi$ denote the stack of $\hat{S}_\Phi(x_i^m, \hat{a}_i^m)$ for all $i \in \{1, \ldots, N\}$ and all $m \in \{1, \ldots, M\}$, then $\hat{S}_{\Phi\Phi} = \hat{S}'_\Phi \hat{S}_\Phi/(MN)$. Let $\bar{S}_A$, $\bar{S}_{0,A}$ and $\hat{S}_{0,A}$ denote the similar stacks of $\bar{A}^m(x_i^m, \hat{a}_i^m)$, $\bar{A}_0^m(x_i^m, \hat{a}_i^m)$ and $\hat{A}_0^m(x_i^m, \hat{a}_i^m)$, respectively. Then $\bar{S}_{\Phi A} = \hat{S}'_\Phi \bar{S}_A/(MN)$, $\bar{S}_{0,\Phi A} = \hat{S}'_\Phi \bar{S}_{0,A}/(MN)$ and $\hat{S}_{0,\Phi A} = \hat{S}'_\Phi \hat{S}_{0,A}/(MN)$. Then we have the following lemma.

Lemma 2.D.7. *Suppose that Assumptions 2.4.8-2.4.10, and the conditions of Theorem 2.4.2 and Lemma 2.D.6 are satisfied. Then*

(i) $\left\|\hat{\beta}_{MN} - \bar{\beta}_{MN}\right\| = \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{h_j,N} + \nu_{h_a,N} + \sum_{j=1}^{d_x} \nu_{I_j,N} + \nu_{I_a,N}\right)$;

(ii) $\left\|\bar{\beta}_{MN} - \bar{\beta}_{0,MN}\right\| = \mathcal{O}_p\left(k_{a,MN}^{-\alpha_e} + \sum_{j=1}^{d_x} k_{x_j,MN}^{-\alpha_e}\right)$;

(iii) $\left\|\bar{\beta}_{0,MN} - \beta_{0,MN}\right\| = \mathcal{O}_p\left(\nu_{I_a,N} + \sum_{j=1}^{d_x} \nu_{I_j,N}\right)$.

Proof. For (i), by Theorems 2.4.2 and 2.4.3, and the conditions that $\nu_{l,N} \to 0$, $\nu_{l_j,N} \to 0$ and $\nu_{l_a,N} \to 0$ ($l = I^m$ or $l = h^m$), we have that $|\hat{I}^m - I^m|_1 \xrightarrow{p.} 0$ and $|\hat{h}^m - h^m|_1 \xrightarrow{p.} 0$ for $m = 1, \ldots, M$.

Some notation is necessary before I proceed with the proof. Let $\mathcal{I}^m$ ($m = 1, \ldots, M$) denote a set of functions I: $\mathbb{R}^{d_x+1} \to \mathbb{R}$ such that each function in $\mathcal{I}^m$ is continuously differentiable of order one. Similarly, let $\mathcal{H}^m$ $(m = 1, \ldots, M)$ denote a set of functions h: $\mathbb{R}^{d_x+1} \to \mathbb{R}$ such that each function in $\mathcal{H}^m$ is continuously differentiable of order one.

For any functions $(I, h) \in \mathcal{I}^m \times \mathcal{H}^m$, define d_x functionals $\Gamma^{(j)}_{x,a}(I, h)$ indexed by $(x, a) \in \mathcal{X} \times \mathcal{A}$ and $j \in \{1, \ldots, d_x\}$ as follows:

$$\Gamma^{(j)}_{x,a}(I,h) \equiv I_{x_j}(x,a)\frac{h_a(x,a)}{h(x,a)} - I_a(x,a)\frac{h_{x_j}(x,a)}{h(x,a)}.$$

Note that

$$\hat{A}^m(x_i^m, \hat{a}_i^m) = \left(\Gamma^{(1)}_{x_i^m,\hat{a}_i^m}(\hat{I}^m, \hat{h}^m), \ldots, \Gamma^{(d_x)}_{x_i^m,\hat{a}_i^m}(\hat{I}^m, \hat{h}^m) \right)',$$

and

$$\bar{A}^m(x_i^m, \hat{a}_i^m) = \left(\Gamma^{(1)}_{x_i^m,\hat{a}_i^m}(I^m, h^m), \ldots, \Gamma^{(d_x)}_{x_i^m,\hat{a}_i^m}(I^m, h^m) \right)'.$$

In what follows, I will omit the explicit dependence of $\Gamma^{(j)}$ on (x, a), and the results in this proof hold uniformly for all $(x, a) \in \mathcal{X} \times \mathcal{A}$. In particular, since $\hat{a}_i^m \in [0, 1]$, it must be the case that $(x_i^m, \hat{a}_i^m) \in \mathcal{X} \times \mathcal{A}$. Let $\Delta I^m \equiv \hat{I}^m - I^m$ and $\Delta h^m \equiv \hat{h}^m - h^m$. Then we have

$$\begin{aligned}
& \left\| \hat{A}^m(x_i^m, \hat{a}_i^m) - \bar{A}^m(x_i^m, \hat{a}_i^m) \right\|^2 \\
= & \sum_{j=1}^{d_x} \left| \Gamma^{(j)}(\hat{I}^m, \hat{h}^m) - \Gamma^{(j)}(I^m, h^m) \right|^2 \\
= & \sum_{j=1}^{d_x} \left| D\Gamma^{(j)}(I^m, h^m; \Delta I^m, \Delta h^m) + R\Gamma^{(j)}(I^m, h^m; \Delta I^m, \Delta h^m) \right|^2 \\
\leq & C \left(\sum_{j=1}^{d_x} \left| D\Gamma^{(j)}(I^m, h^m; \Delta I^m, \Delta h^m) \right|^2 \right. \\
& \left. + \sum_{j=1}^{d_x} \left| R\Gamma^{(j)}(I^m, h^m; \Delta I^m, \Delta h^m) \right|^2 \right), \qquad (2.D.15)
\end{aligned}$$

where the inequality holds by the triangular inequality. In equation (2.D.15), the first terms of the summands are linear functionals with

$$
\begin{aligned}
& |D\Gamma^{(j)}(I^m, h^m;\, \Delta I^m, \Delta h^m)| \\
= & \left| \frac{h_a^m}{h^m} \Delta I_{x_j}^m + \frac{I_{x_j}^m}{h^m} \Delta h_a^m - \frac{I_{x_j}^m h_a^m}{(h^m)^2} \Delta h^m \right. \\
& \left. - \frac{h_{x_j}^m}{h^m} \Delta I_a^m - \frac{I_a^m}{h^m} \Delta h_{x_j}^m + \frac{I_a^m h_{x_j}^m}{(h^m)^2} \Delta h^m \right| \\
\leq & \; C \left(|\Delta I^m|_1 + |\Delta h^m|_1 \right),
\end{aligned} \tag{2.D.16}
$$

where the inequality holds by Assumptions 2.4.6 and 2.4.10(ii), and the triangular inequality. And in equation (2.D.15), the second terms of the summands are nonlinear functionals with

$$
\begin{aligned}
& |R\Gamma^{(j)}(I^m, h^m;\, \Delta I^m, \Delta h^m)| \\
= & \left| \frac{1}{(h^m)^2 (h^m + \Delta h^m)} \left[(h^m)^2 \left(\Delta I_{x_j}^m \Delta h_a^m - \Delta I_a^m \Delta h_{x_j}^m \right) + (I_{x_j}^m h_a^m - I_a^m h_{x_j}^m)(\Delta h^m)^2 \right. \right. \\
& \left. \left. - h^m (I_{x_j}^m \Delta h_a^m + h_a^m \Delta I_{x_j}^m - I_a^m \Delta h_{x_j}^m - h_{x_j}^m \Delta I_a^m) \Delta h^m \right] \right| \\
\leq & \; C \left(|\Delta I^m|_1^2 + |\Delta h^m|_1^2 \right),
\end{aligned} \tag{2.D.17}
$$

where the inequality holds by Assumptions 2.4.6 and 2.4.10(ii), the triangular inequality, and the Cauchy-Schwarz inequality.

By the consistency of $\hat{I}^m$, $\hat{h}^m$ and their derivatives, equation (2.D.16) and equation (2.D.17) imply that $|R\Gamma^{(j)}(I^m, h^m;\, \Delta I^m, \Delta h^m)| = o\left(|D\Gamma^{(j)}(I^m, h^m;\, \Delta I^m, \Delta h^m)|\right)$. Then combine equation (2.D.15), equation (2.D.16) and Assumptions 2.4.6 and 2.4.10(ii), we get

$$
\begin{aligned}
& \left\| \hat{A}^m(x_i^m, \hat{a}_i^m) - \bar{A}^m(x_i^m, \hat{a}_i^m) \right\|^2 \\
= \;& \mathcal{O}_p\left(\sum_{j=1}^{d_x} \left| D\Gamma^{(j)}(I^m, h^m; \Delta I^m, \Delta h^m) \right|^2 \right) \\
= \;& \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{h_j,N}^2 + \nu_{h_a,N}^2 + \nu_{h_a,N}^2 + \sum_{j=1}^{d_x} \nu_{I_j,N}^2 + \nu_{I_a,N}^2 \right) \\
= \;& \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{h_j,N}^2 + \nu_{h_a,N}^2 + \sum_{j=1}^{d_x} \nu_{I_j,N}^2 + \nu_{I_a,N}^2 \right).
\end{aligned}
$$

Recall that this result holds for all $i \in \{1, \ldots, N\}$ and all $m \in \{1, \ldots, M\}$. By Lemma 2.D.6(iii), we have that $\lambda_{\min}(\hat{S}_{\Phi\Phi}) \geq c$ with probability approaching 1, then we have

$$
\begin{aligned}
& \left\| \hat{S}_{\Phi\Phi}^{1/2} \left(\hat{\beta}_{MN} - \bar{\beta}_{MN} \right) \right\|^2 \\
= \;& \left(\hat{S}_A - \bar{S}_A \right)' \hat{S}_\Phi \hat{S}_{\Phi\Phi}^{-} \hat{S}_\Phi' \left(\hat{S}_A - \bar{S}_A \right) / (MN)^2 \\
\leq \;& C \left(\hat{S}_A - \bar{S}_A \right)' \left(\hat{S}_A - \bar{S}_A \right) / (MN) \\
= \;& \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{h_j,N}^2 + \nu_{h_a,N}^2 + \sum_{j=1}^{d_x} \nu_{I_j,N}^2 + \nu_{I_a,N}^2 \right).
\end{aligned}
$$

So (i) holds by Lemma 2.D.6(iii).

For (ii), consider

$$
\begin{aligned}
& \left\| \bar{A}^m(x_i^m, \hat{a}_i^m) - \bar{A}_0^m(x_i^m, \hat{a}_i^m) \right\|^2 \\
= \;& \sum_{j=1}^{d_x} \left| I_{x_j}^m(x_i^m, \hat{a}_i^m) \frac{h_a^m(x_i^m, \hat{a}_i^m)}{h^m(x_i^m, \hat{a}_i^m)} - I_a^m(x_i^m, \hat{a}_i^m) \Phi_{k_{x_j,MN}}(x_i^m, \hat{a}_i^m)' \beta_{0,x_j,k_{x_j},MN} \right.
\end{aligned}
$$

$$
-I_a^m(x_i^m,\hat{a}_i^m)\frac{h_{x_j}^m(x_i^m,\hat{a}_i^m)}{h^m(x_i^m,\hat{a}_i^m)}+I_{x_j}^m(x_i^m,\hat{a}_i^m)\Phi_{k_{a,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,a,k_{a,MN}}\Bigg|^2
$$

$$
\begin{aligned}
&= \sum_{j=1}^{d_x}\Bigg|I_a^m(x_i^m,\hat{a}_i^m)\left[\frac{e_{x_j}(x_i^m,\hat{a}_i^m)}{e(x_i^m,\hat{a}_i^m)}-\Phi_{k_{x_j,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,x_j,k_{x_j,MN}}\right] \\
&\quad -I_{x_j}^m(x_i^m,\hat{a}_i^m)\left[\frac{e_a(x_i^m,\hat{a}_i^m)}{e(x_i^m,\hat{a}_i^m)}-\Phi_{k_{a,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,a,k_{a,MN}}\right]\Bigg|^2 \\
&= \mathcal{O}_p\left(\sum_{j=1}^{d_x}k_{x_j,MN}^{-2\alpha_e}+k_{a,MN}^{-2\alpha_e}\right).
\end{aligned}
$$

where the second equality holds by equation (2.3.7); the third equality holds by Assumptions 2.4.6 and 2.4.8, the triangular inequality and the Cauchy-Schwarz inequality. By the same argument as in the proof of (i), we have

$$
\left\|\hat{S}_{\Phi\Phi}^{1/2}\left(\bar{\beta}_{MN}-\bar{\beta}_{0,MN}\right)\right\|^2 = \mathcal{O}_p\left(k_{a,MN}^{-2\alpha_e}+\sum_{j=1}^{d_x}k_{x_j,MN}^{-2\alpha_e}\right).
$$

So (ii) holds by Lemma 2.D.6(iii).

For (iii), note that

$$
\begin{aligned}
&\left\|\bar{A}_0^m(x_i^m,\hat{a}_i^m)-\hat{A}_0^m(x_i^m,\hat{a}_i^m)\right\|^2 \\
&= \sum_{j=1}^{d_x}\Bigg|\left(I_a^m(x_i^m,\hat{a}_i^m)-\hat{I}_a^m(x_i^m,\hat{a}_i^m)\right)\Phi_{k_{x_j,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,x_j,k_{x_j,MN}} \\
&\quad -\left(I_{x_j}^m(x_i^m,\hat{a}_i^m)-\hat{I}_{x_j}^m(x_i^m,\hat{a}_i^m)\right)\Phi_{k_{a,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,a,k_{a,MN}}\Bigg|^2 \\
&= \sum_{j=1}^{d_x}\Bigg|\left(I_a^m(x_i^m,\hat{a}_i^m)-\hat{I}_a^m(x_i^m,\hat{a}_i^m)\right)\left[\Phi_{k_{x_j,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,x_j,k_{x_j,MN}}-\frac{e_{x_j}(x_i^m,\hat{a}_i^m)}{e(x_i^m,\hat{a}_i^m)}\right] \\
&\quad +\left(I_a^m(x_i^m,\hat{a}_i^m)-\hat{I}_a^m(x_i^m,\hat{a}_i^m)\right)\frac{e_{x_j}(x_i^m,\hat{a}_i^m)}{e(x_i^m,\hat{a}_i^m)} \\
&\quad -\left(I_{x_j}^m(x_i^m,\hat{a}_i^m)-\hat{I}_{x_j}^m(x_i^m,\hat{a}_i^m)\right)\left[\Phi_{k_{a,MN}}(x_i^m,\hat{a}_i^m)'\beta_{0,a,k_{a,MN}}-\frac{e_a(x_i^m,\hat{a}_i^m)}{e(x_i^m,\hat{a}_i^m)}\right]
\end{aligned}
$$

$$
\begin{aligned}
& - \left(I^m_{x_j}(x^m_i, \hat{a}^m_i) - \hat{I}^m_{x_j}(x^m_i, \hat{a}^m_i)\right) \frac{e_a(x^m_i, \hat{a}^m_i)}{e(x^m_i, \hat{a}^m_i)}\Bigg|^2 \\
= \;& \mathcal{O}_p\left(\nu^2_{I_a,N}\sum_{j=1}^{d_x} k^{-2\alpha_e}_{x_j,MN} + \nu^2_{I_a,N} + k^{-2\alpha_e}_{a,MN}\sum_{j=1}^{d_x}\nu^2_{I_j,N} + \sum_{j=1}^{d_x}\nu^2_{I_j,N}\right) \\
= \;& \mathcal{O}_p\left(\nu^2_{I_a,N} + \sum_{j=1}^{d_x}\nu^2_{I_j,N}\right),
\end{aligned}
$$

where the third equality holds by Theorem 2.4.3, Assumptions 2.4.8 and 2.4.9, the triangular inequality and the Cauchy-Schwarz inequality; the fourth equality holds by that $k_{a,MN} \to \infty$ and $k_{x_j,MN} \to \infty$. By the same argument as in the proof of (i), we have

$$
\left\|\hat{S}^{1/2}_{\Phi\Phi}\left(\bar{\beta}_{0,MN} - \beta_{0,MN}\right)\right\|^2 = \mathcal{O}_p\left(\nu^2_{I_a,N} + \sum_{j=1}^{d_x}\nu^2_{I_j,N}\right).
$$

So (iii) holds by Lemma 2.D.6(iii). This completes the proof of the lemma. □

Proof of Lemma 2.4.4

Proof. By the triangular inequality, we have

$$
\begin{aligned}
& \left\|\left(\hat{\beta}_{MN} - \beta_{0,MN}\right)\right\| \\
\leq \;& \left\|\left(\hat{\beta}_{MN} - \bar{\beta}_{MN}\right)\right\| + \left\|\left(\bar{\beta}_{MN} - \bar{\beta}_{0,MN}\right)\right\| + \left\|\left(\bar{\beta}_{0,MN} - \beta_{0,MN}\right)\right\|.
\end{aligned}
$$

So the result follows by Lemma 2.D.7. □

Proof of Theorem 2.4.5

Proof. By the triangular inequality, for $j = 1, \ldots, d_x$,

$$\begin{aligned}
& \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left| \frac{\widehat{e_{x_j}(x,a)}}{e(x,a)} - \frac{e_{x_j}(x,a)}{e(x,a)} \right| \\
\leq & \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left| \Phi_{k_{x_j},MN}(x,a)' \left(\hat{\beta}_{k_{x_j},MN} - \beta_{0,x_j,k_{x_j},MN} \right) \right| \\
& + \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left| \Phi_{k_{x_j},MN}(x,a)'\beta_{0,x_j,k_{x_j},MN} - \frac{e_{x_j}(x,a)}{e(x,a)} \right|.
\end{aligned}$$

Then the result follows by Lemma 2.4.4, 2.D.3, Assumption 2.4.8, the triangular inequality and the Cauchy-Schwarz inequality.

The uniform convergence rate of $e_a(\widehat{x,)/e(}x,a)$ holds by the same argument. $\square$

Proof of Theorem 2.4.6

Proof. By Assumption 2.4.4, Theorem 2.4.5 and the conditions that for $j = 1, \ldots, d_x$, $\nu_{e_j,M,N} \to 0$ and $\nu_{e_a,M,N} \to 0$, then we have

$$\begin{aligned}
& \int_{\bar{x}_j}^{x_j} \frac{e_{x_j}(x_1, \ldots, \widehat{x_{j-1}, s_j}, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})}{e(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})} ds_j - \int_{\bar{x}_j}^{x_j} \frac{e_{x_j}(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})}{e(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})} ds_j \\
= & \mathcal{O}_p\left(\left| \frac{e_{x_j}(x_1, \ldots, \widehat{x_{j-1}, s_j}, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})}{e(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})} - \frac{e_{x_j}(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})}{e(x_1, \ldots, x_{j-1}, s_j, \bar{x}_{j+1}, \ldots, \bar{x}_{d_x}, \bar{a})} \right| \right) \\
= & \mathcal{O}_p\left(\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left| \frac{\widehat{e_{x_j}(x,a)}}{e(x,a)} - \frac{e_{x_j}(x,a)}{e(x,a)} \right| \right) \\
= & o_p(1), \qquad (2.D.18)
\end{aligned}$$

and

$$
\begin{aligned}
& \int_{\bar{a}}^{a} \frac{\widehat{e_a(x,t)}}{e(x,t)} dt - \int_{\bar{a}}^{a} \frac{e_a(x,t)}{e(x,t)} dt \\
= \;& \mathcal{O}_p \left(\left| \frac{\widehat{e_a(x,t)}}{e(x,t)} - \frac{e_a(x,t)}{e(x,t)} \right| \right) \\
= \;& \mathcal{O}_p \left(\sup_{(x,a) \in \mathcal{X} \times \mathcal{A}} \left| \frac{\widehat{e_a(x,a)}}{e(x,a)} - \frac{e_a(x,a)}{e(x,a)} \right| \right) \\
= \;& o_p(1). \qquad (2.D.19)
\end{aligned}
$$

Let $\mathcal{J}$ denote a set of functions $J\colon \mathbb{R}^{d_x+1} \to \mathbb{R}$. Define a family of functionals $\Xi_{x,a}(J)$ indexed by $(x,a) \in \mathcal{X} \times \mathcal{A}$ as follows

$$
\Xi_{x,a}(J) \equiv \exp(J(x,a)).
$$

Let

$$
\begin{aligned}
\hat{J}(x,a) &\equiv \sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} \frac{\widehat{e_{x_j}(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})}}{e(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})} ds_j + \int_{\bar{a}}^{a} \frac{\widehat{e_a(x,t)}}{e(x,t)} dt, \\
J(x,a) &\equiv \sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} g_j(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a}) ds_j + \int_{\bar{a}}^{a} g_{d_x+1}(x,t) dt.
\end{aligned}
$$

Then it is easy to see that $\hat{e}(x,a) = \Xi_{x,a}\left(\widehat{J}\right)$ and $e(x,a) = \Xi_{x,a}(J)$. In what follows, I will omit the explicit dependence of Ξ on (x,a), and the results in this proof hold uniformly for all $(x,a) \in \mathcal{X} \times \mathcal{A}$. Let $\Delta J \equiv \hat{J} - J$, then we have

$$
\begin{aligned}
\widehat{e} - e &= \Xi(J + \Delta J) - \Xi(J) \\
&= D\Xi(J; \Delta J) + R\Xi(J; \Delta J).
\end{aligned}
$$

The first term in this decomposition is a linear functional with

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |D\Xi(J;\,\Delta J)| \;\equiv\; \sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\exp(J)\Delta J| \leq C|\Delta J|_0,$$

where the inequality holds since $\mathcal{X}\times\mathcal{A}$ is compact by Assumption 2.4.4. And the second term in the decomposition is a nonlinear functional with

$$\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |R\Xi(J;\,\Delta J)| = o\left(|\Delta J|_0\right).$$

Then by the triangular inequality,

$$\begin{aligned}\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}} |\hat{e}(x,a)-e(x,a)| \;&\leq\; C|\Delta J|_0 + o\left(|\Delta J|_0\right)\\ &=\; \mathcal{O}_p\Bigg(C\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}\left|\int_{\bar{a}}^{a}\left(\widehat{\frac{e_a(x,t)}{e(x,t)}}\right)dt-\int_{\bar{a}}^{a}\frac{e_a(x,t)}{e(x,t)}dt\right|\\ &\quad +C\sup_{(x,a)\in\mathcal{X}\times\mathcal{A}}\left|\int_{\bar{x}}^{x}\left(\widehat{\frac{e_x(s,\bar{a})}{e(s,\bar{a})}}\right)ds-\int_{\bar{x}}^{x}\frac{e_x(s,\bar{a})}{e(s,\bar{a})}ds\right|\Bigg).\end{aligned}$$

And the result follows by equation (2.D.18), equation (2.D.19) and Theorem 2.4.5. □

2.D.4 Proofs of the Theorems in Section 2.4.2.4

This subsection proceeds with the same steps as in Appendix 2.D.2.

Lemma 2.D.8. *Suppose that Assumption 2.4.13 and the conditions for Theorem 2.4.1 are satisfied. Then,* $\sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}}\|\Psi_k(h,x,a)\| \leq C\zeta_0(k)$, $\sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}}\|\partial\Psi_k(h,x,a)/\partial h\| \leq C\zeta_h(k)$, $\sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}}\|\partial\Psi_k(h,x,a)/\partial x_j\| \leq C\zeta_j(k)$ *and* $\sup_{(h,x,a)\in\mathcal{H}\times\mathcal{X}\times\mathcal{A}}\|\partial\Psi_k(h,x,a)/$

$\partial a\| \leq C\zeta_a(k)$.

Proof. This lemma holds by the same argument as Lemma 2.D.3. □

Define

$$\bar{S}_{\Psi\Psi} \equiv (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \bar{S}'_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m),$$

where

$$\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) \equiv \begin{pmatrix} \left[\nabla_x h^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{\nabla_x e(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right] \otimes \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \\ \left[h_a^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_a(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \end{pmatrix}.$$

Lemma 2.D.9. *Suppose that Assumptions 2.4.12, 2.4.13 and 2.4.15, and the conditions for Theorem 2.4.5 are satisfied. Suppose as well that the numbers of series basis functions used to approximate each component in* $k_{U,MN}$ *all increase to infinity with* N, $\sqrt{k_{U,MN}}\nu_a(\sigma_N)\zeta_a(k_{U,MN}) \to 0$, $k_{U,MN}\nu_{e_j,M,N} \to 0$ *for* $j = 1, \ldots, d_x$, *and* $k_{U,MN}\nu_{e_a,M,N} \to 0$. *Then*

(i) $\|\bar{S}_{\Psi\Psi} - S_{\Psi\Psi}\| = \mathcal{O}_p\left(\nu_{a,N}^2\zeta_a^2(k_{U,MN}) + \sqrt{k_{U,MN}}\nu_a(\sigma_N)\zeta_a(k_{U,MN})\right)$;

(ii) $\|\hat{S}_{\Psi\Psi} - \bar{S}_{\Psi\Psi}\| = \mathcal{O}_p\left(k_{U,MN}\left[\sum_{j=1}^{d_x}\nu_{e_j,M,N} + \nu_{e_a,M,N}\right]\right)$;

(iii) $\lambda_{\min}(\hat{S}_{\Psi\Psi}) \geq c$, $\lambda_{\min}(\bar{S}_{\Psi\Psi}) \geq c$ *and* $\lambda_{\min}(S_{\Psi\Psi}) \geq c$ *with probability approaching* 1, *where* $\lambda_{\min}$ *denotes the minimum eigenvalue of a symmetric matrix.*

Proof. To prove (i), some preliminary results are needed. For $j = 1, \ldots, d_x$, consider the mean value expansion

$$\begin{aligned} &\left[h_{x_j}^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_{x_j}(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \\ &- \left[h_{x_j}^m(x_i^m, a_i^m) + h^m(x_i^m, a_i^m)\frac{e_{x_j}(x_i^m, a_i^m)}{e(x_i^m, a_i^m)}\right] \Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)' \end{aligned}$$

$$
\begin{aligned}
= \; & \Bigg[h^m_{x_j a}(x^m_i, \tilde{a}^m_i) + h^m_a(x^m_i, \tilde{a}^m_i)\frac{e_{x_j}(x^m_i, \tilde{a}^m_i)}{e(x^m_i, \tilde{a}^m_i)} + h^m(x^m_i, \tilde{a}^m_i)\frac{e_{x_j a}(x^m_i, \tilde{a}^m_i)}{e(x^m_i, \tilde{a}^m_i)} \\
& -h^m(x^m_i, \tilde{a}^m_i)\frac{e_{x_j}(x^m_i, \tilde{a}^m_i)e_a(x^m_i, \tilde{a}^m_i)}{[e(x^m_i, \tilde{a}^m_i)]^2}\Bigg] \cdot \Psi_{k_{U,MN}}(h^m_i, x^m_i, \tilde{a}^m_i)'(\hat{a}^m_i - a^m_i) \\
& + \left[h^m_{x_j}(x^m_i, \tilde{a}^m_i) + h^m(x^m_i, \tilde{a}^m_i) \cdot \frac{e_{x_j}(x^m_i, \tilde{a}^m_i)}{e(x^m_i, \tilde{a}^m_i)}\right] \cdot \frac{\partial}{\partial a}\Psi_{k_{U,MN}}(h^m_i, x^m_i, \tilde{a}^m_i)'(\hat{a}^m_i - a^m_i),
\end{aligned}
$$

where $\tilde{a}^m_i$ is between a^m_i and $\hat{a}^m_i$, so it must be in $[0, 1]$. Note that $\tilde{a}^m_i$ might take a different value from the previous uses. By Lemma 2.D.8, Assumptions 2.4.6, 2.4.9 and 2.4.10(ii), the triangular inequality, and the Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
& \Bigg\| \left[h^m_{x_j}(x^m_i, \hat{a}^m_i) + h^m(x^m_i, \hat{a}^m_i)\frac{e_{x_j}(x^m_i, \hat{a}^m_i)}{e(x^m_i, \hat{a}^m_i)}\right] \Psi_{k_{U,MN}}(h^m_i, x^m_i, \hat{a}^m_i)' \\
& - \left[h^m_{x_j}(x^m_i, a^m_i) + h^m(x^m_i, a^m_i)\frac{e_{x_j}(x^m_i, a^m_i)}{e(x^m_i, a^m_i)}\right] \Psi_{k_{U,MN}}(h^m_i, x^m_i, a^m_i)' \Bigg\|^2 \\
\leq \; & C|\hat{a}^m_i - a^m_i|^2 \left(\zeta^2_0(k_{U,MN}) + \zeta^2_a(k_{U,MN})\right) \\
\leq \; & C|\hat{a}^m_i - a^m_i|^2 \zeta^2_a(k_{U,MN}). \qquad (2.D.20)
\end{aligned}
$$

By the same token,

$$
\begin{aligned}
& \Bigg\| \left[h^m_a(x^m_i, \hat{a}^m_i) + h^m(x^m_i, \hat{a}^m_i)\frac{e_a(x^m_i, \hat{a}^m_i)}{e(x^m_i, \hat{a}^m_i)}\right] \Psi_{k_{U,MN}}(h^m_i, x^m_i, \hat{a}^m_i)' \\
& - \left[h^m_a(x^m_i, a^m_i) + h^m(x^m_i, a^m_i)\frac{e_a(x^m_i, a^m_i)}{e(x^m_i, a^m_i)}\right] \Psi_{k_{U,MN}}(h^m_i, x^m_i, a^m_i)' \Bigg\|^2 \\
\leq \; & C|\hat{a}^m_i - a^m_i|^2 \left(\zeta^2_0(k_{U,MN}) + \zeta^2_a(k_{U,MN})\right) \\
\leq \; & C|\hat{a}^m_i - a^m_i|^2 \zeta^2_a(k_{U,MN}). \qquad (2.D.21)
\end{aligned}
$$

Equation (2.D.20) and equation (2.D.21) imply that

$$\begin{aligned} & \left\|\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) - S_{\Psi}(h_i^m, x_i^m, a_i^m)\right\|^2 \\ \leq \;& C|\hat{a}_i^m - a_i^m|^2 \zeta_a^2(k_{U,MN}). \end{aligned} \tag{2.D.22}$$

Without loss of generality, I can set the basis functions to be orthonormal. So

$$\left(\mathbb{E}\left(\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\right\|^2\right)\right) = \operatorname{tr}\left(\mathbb{I}_{k_{U,MN}}\right) = k_{U,MN}.$$

Then by the Markov's inequality, we have

$$\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(k_{U,MN}\right). \tag{2.D.23}$$

This implies, together with Lemma 2.D.8, Assumptions 2.4.6, 2.4.9 and 2.4.10(ii), and the Cauchy-Schwarz inequality, that

$$\left\|S_{\Psi,k_{U,MN},x}(x_i^m, a_i^m)\right\|^2 = \mathcal{O}_p\left(k_{U,MN}\right). \tag{2.D.24}$$

Now consider (i),

$$\begin{aligned} \|\bar{S}_{\Psi\Psi} - S_{\Psi\Psi}\| &= (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\|\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)' \bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\right. \\ & \quad \left. - S_{\Psi}(h_i^m, x_i^m, a_i^m)' S_{\Psi}(h_i^m, x_i^m, a_i^m)\right\| \\ &= (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \left\|\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) - S_{\Psi}(h_i^m, x_i^m, a_i^m)\right\|^2 \end{aligned}$$

$$+2(MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\left(\left\|\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) - S_{\Psi}(h_i^m, x_i^m, a_i^m)\right\|^2\right)^{1/2}$$
$$\left(\left\|S_{\Psi,k_{U,MN}}(x_i^m, a_i^m)\right\|^2\right)^{1/2},$$

where the inequality holds by the triangular inequality and the Cauchy-Schwarz inequality. Combine this result with Theorem 2.4.1, equation (2.D.22) and equation (2.D.24), we get

$$\|\bar{S}_{\Psi\Psi} - S_{\Psi\Psi}\| = \mathcal{O}_p\left(\nu_{a,N}^2\zeta_a^2(k_{U,MN}) + \sqrt{k_{U,MN}}\nu_a(\sigma_N)\zeta_a(k_{U,MN})\right).$$

So (i) holds.

To prove (ii), some preliminary results are necessary. Recall that $\mathcal{H}^m$ ($m = 1, \ldots, M$) denotes a set of functions h: $\mathbb{R}^{d_x+1} \to \mathbb{R}$ such that each function in $\mathcal{H}^m$ is continuously differentiable of order one; and that $\mathcal{J}$ denotes a set of functions J: $\mathbb{R}^{d_x+1} \to \mathbb{R}$.

For any functions $(h, J) \in \mathcal{H}^m \times \mathcal{J}$, define a family of functionals $\Upsilon_{x,a}^{(j)}(h, J)$ indexed by $(x, a) \in \mathcal{X} \times \mathcal{A}$ and $j \in \{1, \ldots, d_x\}$ as follows:

$$\Upsilon_{x,a}^{(j)}(h, J) \equiv h_{x_j}(x, a) + h(x, a)\frac{e_{x_j}(x, a)}{e(x, a)}. \tag{2.D.25}$$

And define another family of functionals $\Upsilon_{x,a}^{(a)}(h, J)$ indexed by $(x, a) \in \mathcal{X} \times \mathcal{A}$ as follows:

$$\Upsilon_{x,a}^{(a)}(h, J) \equiv h_a(x, a) + h(x, a)\frac{e_a(x, a)}{e(x, a)}. \tag{2.D.26}$$

Note that

$$\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) = \left(\Upsilon^{(1)}_{x_i^m,\hat{a}_i^m}\left(\hat{h}^m, \widehat{\frac{e_{x_1}}{e}}\right), \ldots, \Upsilon^{(d_x)}_{x_i^m,\hat{a}_i^m}\left(\hat{h}^m, \widehat{\frac{e_{x_{d_x}}}{e}}\right), \Upsilon^{(a)}_{x_i^m,\hat{a}_i^m}\left(\hat{h}^m, \widehat{\frac{e_a}{e}}\right)\right)' \otimes \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)',$$

and

$$\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) = \left(\Upsilon^{(1)}_{x_i^m,\hat{a}_i^m}\left(h^m, \frac{e_{x_1}}{e}\right), \ldots, \Upsilon^{(d_x)}_{x_i^m,\hat{a}_i^m}\left(h^m, \frac{e_{x_{d_x}}}{e}\right), \Upsilon^{(a)}_{x_i^m,\hat{a}_i^m}\left(h^m, \frac{e_a}{e}\right)\right)' \otimes \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'.$$

In what follows, I will omit the explicit dependence of $\Upsilon^{(j)}$ and $\Upsilon^{(a)}$ on (x, a), and the results in this proof hold uniformly for all $(x, a) \in \mathcal{X} \times \mathcal{A}$. In particular, since $\hat{a}_i^m \in [0, 1]$, it must be the case that $(x_i^m, \hat{a}_i^m) \in \mathcal{X} \times \mathcal{A}$. Let $\Delta h^m \equiv \hat{h}^m - h^m$, let $\Delta\left(\frac{e_{x_j}}{e}\right) \equiv \widehat{\frac{e_{x_j}}{e}} - \frac{e_{x_j}}{e}$ for $j = 1, \ldots, d_x$, and let $\Delta\left(\frac{e_a}{a}\right) \equiv \widehat{\frac{e_a}{e}} - \frac{e_a}{e}$. Then we have

$$\begin{aligned} & \left\|\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m) - \bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\right\|^2 \\ \leq & \left[\sum_{j=1}^{d_x}\left|\Upsilon^{(j)}\left(\hat{h}^m, \widehat{\frac{e_{x_j}}{e}}\right) - \Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}\right)\right|^2 \right. \\ & \left. + \left|\Upsilon^{(a)}\left(\hat{h}^m, \widehat{\frac{e_a}{e}}\right) - \Upsilon^{(a)}\left(h^m, \frac{e_a}{e}\right)\right|^2\right] \cdot \left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)\right\|^2, \end{aligned} \quad (2.D.27)$$

where the inequality holds by the Cauchy-Schwarz inequality. By the same argument as for equation (2.D.23), we have

$$\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)\right\|^2 \quad = \quad \mathcal{O}_p\left(k_{U,MN}\right). \tag{2.D.28}$$

Moreover,

$$\begin{aligned}
& \left|\Upsilon^{(j)}\left(\hat{h}^m, \frac{\widehat{e_{x_j}}}{e}\right) - \Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}\right)\right|^2 \\
= & \left|D\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right) + R\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right)\right|^2 \\
\leq & \; C\left|D\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right)\right|^2 \\
& + C\left|R\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right)\right|^2,
\end{aligned} \tag{2.D.29}$$

where the inequality holds by the triangular inequality. In equation (2.D.29), the first term is a linear functional with

$$\begin{aligned}
\left|D\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right)\right| \quad & = \quad \left|\Delta h_{x_j}^m - h^m \Delta\left(\frac{e_{x_j}}{e}\right) - \Delta h^m \frac{e_{x_j}}{e}\right| \\
& \leq \quad C\left(\left|\Delta h^m\right|_1 + \left|\Delta \frac{e_{x_j}}{e}\right|_0\right),
\end{aligned} \tag{2.D.30}$$

where the inequality holds by Assumptions 2.4.6, 2.4.9 and 2.4.10(ii), and the triangular inequality. And the second term in equation (2.D.29) is a nonlinear functional with

$$\begin{aligned}
\left|R\Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta\left(\frac{e_{x_j}}{e}\right)\right)\right| \quad & = \quad \left|\Delta h^m \Delta\left(\frac{e_{x_j}}{e}\right)\right| \\
& \leq \quad C\left(\left|\Delta h^m\right|_0 \cdot \left|\Delta\left(\frac{e_{x_j}}{e}\right)\right|_0\right),
\end{aligned} \tag{2.D.31}$$

where the inequality holds by Assumptions 2.4.6, 2.4.9 and 2.4.10(ii), the triangular inequality, and the Cauchy-Schwarz inequality.

By the consistency of $\hat{h}^m$, $\hat{h}^m_{x_j}$ and $\widehat{\frac{e_{x_j}}{e}}$, equation (2.D.30) and equation (2.D.31) imply that $|R\Upsilon^{(j)}(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta(e_{x_j}/e))| = o(|D\Upsilon^{(j)}(h^m, \frac{e_{x_j}}{e}; \Delta h^m, \Delta(e_{x_j}/e))|)$. Then combine equation (2.D.29), equation (2.D.30), Assumptions 2.4.6 and 2.4.10, we get

$$\left|\Upsilon^{(j)}\left(\hat{h}^m, \widehat{\frac{e_{x_j}}{e}}\right) - \Upsilon^{(j)}\left(h^m, \frac{e_{x_j}}{e}\right)\right|^2 = \mathcal{O}_p\left(\nu^2_{h_j,N} + \nu^2_{e_j,M,N}\right) = \mathcal{O}_p\left(\nu^2_{e_j,M,N}\right) \quad (2.D.32)$$

for $j = 1, \ldots, d_x$. By the same token, we have

$$\left|\Upsilon^{(a)}\left(\hat{h}^m, \widehat{\frac{e_a}{e}}\right) - \Upsilon^{(a)}\left(h^m, \frac{e_a}{e}\right)\right|^2 = \mathcal{O}_p\left(\nu^2_{e_a,M,N}\right). \quad (2.D.33)$$

Together, equation (2.D.27), equation (2.D.28), equation (2.D.32) and equation (2.D.33) imply that

$$\left\|\hat{S}_\Psi(h^m_i, x^m_i, \hat{a}^m_i) - \bar{S}_\Psi(h^m_i, x^m_i, \hat{a}^m_i)\right\|^2 = \mathcal{O}_p\left(k_{U,MN}\left[\sum_{j=1}^{d_x} \nu^2_{e_j,M,N} + \nu^2_{e_a,M,N}\right]\right) \quad (2.D.34)$$

On the other hand, by equation (2.D.28), Assumptions 2.4.12 and 2.4.15, and the Cauchy-Schwarz inequality, we have

$$\left\|\bar{S}_{\Psi,k_N,x}(x^m_i, a^m_i)\right\|^2 = \mathcal{O}_p\left(k_{U,MN}\right). \quad (2.D.35)$$

Now consider (ii),

$$
\begin{aligned}
\|\hat{S}_{\Psi\Psi} - \bar{S}_{\Psi\Psi}\| &= (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\Big\|\hat{S}_{\Psi}(h_i^m, x_i^m, a_i^m)'\hat{S}_{\Psi}(h_i^m, x_i^m, a_i^m) \\
&\quad -\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)'\bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\big\| \\
&= (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\Big\|\hat{S}_{\Psi}(h_i^m, x_i^m, a_i^m) - \bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\Big\|^2 \\
&\quad +2(MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\left(\Big\|\hat{S}_{\Psi}(h_i^m, x_i^m, a_i^m) - \bar{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)\Big\|^2\right)^{1/2} \\
&\quad \left(\big\|\bar{S}_{\Psi,k_{U,MN}}(x_i^m, a_i^m)\big\|^2\right)^{1/2},
\end{aligned}
$$

where the inequality holds by the triangular inequality and the Cauchy-Schwarz inequality. Combine this result with equation (2.D.34) and equation (2.D.35), we get

$$
\|\hat{S}_{\Psi\Psi} - \bar{S}_{\Psi\Psi}\| = \mathcal{O}_p\left(k_{U,MN}\left[\sum_{j=1}^{d_x}\nu_{e_j,M,N} + \nu_{e_a,M,N}\right]\right).
$$

So (ii) holds.

To prove (iii), note that

$$
\begin{aligned}
\mathbb{E}\left[\|S_{\Psi\Psi} - \mathbb{E}(S_{\Psi\Psi})\|^2\right] &\leq \sum_{j=1}^{d_x}\mathbb{E}\left[(MN)^{-1}\left(h_{x_j}^m(x_i^m, a_i^m) + h^m(x_i^m, a_i^m)\frac{e_{x_j}(x_i^m, a_i^m)}{e(x_i^m, a_i^m)}\right)^2\right. \\
&\quad \left.\cdot\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)'\right\|^2\right] \\
&\quad +\mathbb{E}\left[(MN)^{-1}\left(h_a^m(x_i^m, a_i^m) + h^m(x_i^m, a_i^m)\frac{e_a(x_i^m, a_i^m)}{e(x_i^m, a_i^m)}\right)^2\right. \\
&\quad \left.\cdot\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)'\right\|^2\right] \\
&\leq C\mathbb{E}\left[\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)'\right\|^2\right], \qquad (2.D.36)
\end{aligned}
$$

where the first inequality holds by the definition of $S_{\Psi\Psi}$, Assumption 2.4.1, and that the second moment of a random variable is no less than its variance; the second inequality holds by Assumptions 2.4.6, 2.4.9 and 2.4.10(ii). Recall that I assume the series basis functions are orthonormal, then by Lemma 2.D.8, we have

$$\begin{aligned}
& \mathbb{E}\left[\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)\Psi_{k_{U,MN}}(h_i^m, x_i^m, a_i^m)'\right\|^2\right] \\
= \;& (MN)^{-1}\mathbb{E}\left[\left(\sum_{k=1}^{k_{U,MN}} \psi_k^2(h_i^m, x_i^m, a_i^m)\right)\left(\sum_{l=1}^{k_{U,MN}} \psi_l^2(h_i^m, x_i^m, a_i^m)\right)\right] \\
\leq \;& (MN)^{-1}\zeta_o^2(k_{U,MN})\mathrm{tr}(\mathbb{I}_{k_{U,MN}}) \\
= \;& \zeta_o^2(k_{U,MN})k_{U,MN}/(MN). \qquad (2.D.37)
\end{aligned}$$

Plug the bounds in equation (2.D.37) into equation (2.D.36), then we get

$$\mathbb{E}\left[\|S_{\Psi\Psi} - \mathbb{E}(S_{\Psi\Psi})\|^2\right] \;\leq\; C\zeta_o^2(k_{U,MN})k_{U,MN}/(MN).$$

Then by the Markov's inequality,

$$\|S_{\Psi\Psi} - \mathbb{E}(S_{\Psi\Psi})\| \;=\; \mathcal{O}_p\left(\zeta_o(k_{U,MN})\sqrt{k_{U,MN}/(MN)}\right).$$

Again, since $\nu_a(\sigma_N)$ converges to zero at a slower rate than $N^{-1/2}$, $\sqrt{k_{U,MN}}\nu_a(\sigma_N)\zeta_a(k_{U,MN}) \to 0$ implies $\zeta_o^2(k_{U,MN})k_{U,MN}/(MN) \to 0$. As a result, $\|S_{\Psi\Psi} - \mathbb{E}(S_{\Psi\Psi})\| = o_p(1)$.

Note that $\sqrt{k_{U,MN}}\nu_a(\sigma_N)\zeta_a(k_{U,MN}) \to 0$ implies $\nu_{a,N}^2\zeta_a^2(k_{U,MN}) \to 0$. Then by result (i), we have $\|\hat{S}_{\Psi\Psi} - \bar{S}_{\Psi\Psi}\| = o_p(1)$. Moreover, by result (ii), the conditions that the numbers of series basis functions used to approximate each component in $k_{U,MN}$ all increase to infinity with N, $k_{U,MN}\nu_{e_j,M,N} \to 0$ for $j = 1, \ldots, d_x$, and $k_{U,MN}\nu_{e_a,M,N} \to 0$, we have $\|\bar{S}_{\Psi\Psi} - S_{\Psi\Psi}\| = o_p(1)$. Then (iii) follows by the same argument for the proof of Lemma 2.D.4(iv). This completes the proof of the lemma. $\square$

Define

$$\begin{aligned}
\bar{S}_{\Psi I} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)'\bar{S}_I(h_i^m, x_i^m, \hat{a}_i^m),\\
\bar{S}_{0,\Psi I} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)'\bar{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m),\\
\tilde{S}_{0,\Psi I} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)'\tilde{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m),\\
\hat{S}_{0,\Psi I} &\equiv (MN)^{-1}\sum_{m=1}^{M}\sum_{i=1}^{N}\hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)'\hat{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m),
\end{aligned}$$

where

$$\begin{aligned}
\bar{S}_I(h_i^m, x_i^m, \hat{a}_i^m) &\equiv \begin{pmatrix} \nabla_x I^m(x_i^m, \hat{a}_i^m)', \ I_a^m(x_i^m, \hat{a}_i^m) \end{pmatrix}',\\
\bar{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) &\equiv \begin{pmatrix} \nabla_x h^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{\nabla_x e(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)} \\ h_a^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_a(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)} \end{pmatrix}\\
&\quad \otimes \left[\Psi_{k_{U,MN}}(h^m(x_i^m, \hat{a}_i^m), x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right],\\
\tilde{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) &\equiv \begin{pmatrix} \nabla_x h^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{\nabla_x e(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)} \\ h_a^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_a(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)} \end{pmatrix}\\
&\quad \otimes \left[\Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right],
\end{aligned}$$

$$\hat{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) \equiv \begin{pmatrix} \nabla_x \hat{h}^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{\widehat{\nabla_x e(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)} \\ \hat{h}_a^m(x_i^m, \hat{a}_i^m) + \hat{h}^m(x_i^m, \hat{a}_i^m)\frac{\widehat{e_a(x_i^m, \hat{a}_i^m)}}{e(x_i^m, \hat{a}_i^m)} \end{pmatrix} \otimes \left[\Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right].$$

Now we need some intermediate coefficients which help analyze the estimated series coefficients for the sellers' marginal disutility function. Define

$$\bar{\gamma}_{k_{U,MN}} \equiv \hat{S}_{\Psi\Psi}^{-}\bar{S}_{\Psi I},$$

$$\bar{\gamma}_{0,k_{U,MN}} \equiv \hat{S}_{\Psi\Psi}^{-}\bar{S}_{0,\Psi I},$$

$$\tilde{\gamma}_{0,k_{U,MN}} \equiv \hat{S}_{\Psi\Psi}^{-}\tilde{S}_{0,\Psi I}.$$

And with some standard algebra, we get

$$\gamma_{0,k_{U,MN}} \equiv \hat{S}_{\Psi\Psi}^{-}\hat{S}_{0,\Psi I}.$$

Note that if we let $\hat{S}_\Psi$ denote the stack of $\hat{S}_\Psi(h_i^m, x_i^m, \hat{a}_i^m)$ for all $i \in \{1, \ldots, N\}$ and all $m \in \{1, \ldots, M\}$, then $\hat{S}_{\Psi\Psi} = \hat{S}_\Psi'\hat{S}_\Psi/(MN)$. Let $\bar{S}_I$, $\bar{S}_{0,I}$, $\tilde{S}_{0,I}$ and $\hat{S}_{0,I}$ denote the similar stacks of $\bar{S}_I(h_i^m, x_i^m, \hat{a}_i^m)$, $\bar{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m)$, $\tilde{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m)$ and $\hat{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m)$, respectively. Then $\bar{S}_{\Psi I} = \hat{S}_\Psi'\bar{S}_I/(MN)$, $\bar{S}_{0,\Psi I} = \hat{S}_\Psi'\bar{S}_{0,I}/(MN)$, $\tilde{S}_{0,\Psi I} = \hat{S}_\Psi'\tilde{S}_{0,I}/(MN)$ and $\hat{S}_{0,\Psi I} = \hat{S}_\Psi'\hat{S}_{0,I}/(MN)$. Then we have the following lemma.

Lemma 2.D.10. *Suppose that Assumption 2.4.11 and the conditions of Lemma 2.D.9 are satisfied. Then* *(i)* $\|\hat{\gamma}_{k_{U,MN}} - \bar{\gamma}_{k_{U,MN}}\| = \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{I_j,N} + \nu_{I_a,N}\right)$;

(ii) $\|\bar{\gamma}_{k_{U,MN}} - \bar{\gamma}_{0,k_{U,MN}}\| = \mathcal{O}_p\left(k_{U,MN}^{-\alpha_U}\right)$;

(iii) $\|\bar{\gamma}_{0,k_{U,MN}} - \tilde{\gamma}_{0,k_{U,MN}}\| = \mathcal{O}_p\left(k_{U,MN}^{-\alpha_U} + \nu_a(\sigma_N)\right)$;

(iv) $\|\tilde{\gamma}_{0,k_{U,MN}} - \gamma_{0,k_{U,MN}}\| = \mathcal{O}_p\left(\nu_{e_a,M,N} + \sum_{j=1}^{d_x} \nu_{e_j,M,N}\right)$.

Proof. For (i), consider

$$\begin{aligned}\left\|\hat{S}_{\Psi\Psi}^{1/2}\left(\hat{\gamma}_{k_{U,MN}} - \bar{\gamma}_{k_{U,MN}}\right)\right\|^2 &= \left(\hat{S}_I - \bar{S}_I\right)' \hat{S}_\Psi \hat{S}_{\Psi\Psi}^- \hat{S}_\Psi' \left(\hat{S}_I - \bar{S}_I\right)/(MN)^2 \\ &\leq C\left(\hat{S}_I - \bar{S}_I\right)'\left(\hat{S}_I - \bar{S}_I\right)/(MN) \\ &= \mathcal{O}_p\left(\sum_{j=1}^{d_x} \nu_{I_j,N}^2 + \nu_{I_a,N}^2\right), \qquad (2.D.38)\end{aligned}$$

where the second equality holds by Theorem 2.4.3. So (i) holds by Lemma 2.D.9(iii).

For (ii), note that equation (2.4.14) implies for $j = 1, \ldots, d_x$,

$$I_{x_j}^m(x_i^m, \hat{a}_i^m) = \left[h_{x_j}^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_{x_j}(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right] U_h\left(h^m(x_i^m, \hat{a}_i^m), x_i^m, \hat{a}_i^m\right).$$

Then

$$\begin{aligned}&\left\|\bar{S}_I(h_i^m, x_i^m, \hat{a}_i^m) - \bar{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m)\right\|^2 \\ \leq\ &\left(\sum_{j=1}^{d_x}\left|h_{x_j}^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_{x_j}(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right|^2 \right. \\ &\left. + \left|h_a^m(x_i^m, \hat{a}_i^m) + h^m(x_i^m, \hat{a}_i^m)\frac{e_a(x_i^m, \hat{a}_i^m)}{e(x_i^m, \hat{a}_i^m)}\right|^2\right) \\ &\cdot \left|U_h\left(h^m(x_i^m, \hat{a}_i^m), x_i^m, \hat{a}_i^m\right) - \Psi_{k_{U,MN}}(h^m(x_i^m, \hat{a}_i^m), x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right|^2 \\ \leq\ &Ck_{U,MN}^{-2\alpha_U},\end{aligned}$$

where the first inequality holds by the Cauchy-Schwarz inequality; the second inequality holds by Assumptions 2.4.6, 2.4.9, 2.4.10(ii) and 2.4.11. As a result,

$$
\begin{aligned}
\left\|\hat{S}_{\Psi\Psi}^{1/2}\left(\bar{\gamma}_{k_{U,MN}}-\bar{\gamma}_{0,k_{U,MN}}\right)\right\|^2 &= \left(\bar{S}_I-\bar{S}_{0,I}\right)'\hat{S}_{\Psi}\hat{S}_{\Psi\Psi}^{-}\hat{S}_{\Psi}'\left(\bar{S}_I-\bar{S}_{0,I}\right)/(MN)^2 \\
&\leq C\left(\bar{S}_I-\bar{S}_{0,I}\right)'\left(\bar{S}_I-\bar{S}_{0,I}\right)/(MN) \\
&= \mathcal{O}_p\left(k_{U,MN}^{-2\alpha_U}\right).
\end{aligned}
$$

So (ii) holds by Lemma 2.D.9(iii).

For (iii), consider the mean value expansion

$$
\begin{aligned}
& U_h(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)-U_h(h_i^m,x_i^m,\hat{a}_i^m) \\
=\ & U_h(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)-U_h(h^m(x_i^m,a_i^m),x_i^m,\hat{a}_i^m) \\
=\ & U_{hh}(\tilde{h}_i^m,x_i^m,\hat{a}_i^m)h_a^m(x_i^m,\tilde{a}_i^m)(\hat{a}_i^m-a_i^m). \qquad (2.D.39)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
& \left|\Psi_{k_{U,MN}}(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}-\Psi_{k_{U,MN}}(h_i^m,x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right|^2 \\
\leq\ & \left|U_h(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)-\Psi_{k_{U,MN}}(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right|^2 \\
& +\left|U_h(h_i^m,x_i^m,\hat{a}_i^m)-\Psi_{k_{U,MN}}(h_i^m,x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right|^2 \\
& +\left|U_h(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)-U_h(h_i^m,x_i^m,\hat{a}_i^m)\right|^2 \\
\leq\ & C\left(k_{U,MN}^{-2\alpha_U}+|\hat{a}_i^m-a_i^m|^2\right),
\end{aligned}
$$

where the first inequality holds by the triangular inequality and the Cauchy-Schwarz inequality; the second inequality holds by Assumptions 2.4.6, 2.4.10(ii), 2.4.11, 2.4.12, and equation (2.D.39). This implies that

$$
\begin{aligned}
& \left\| \bar{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) - \tilde{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) \right\|^2 \\
\leq \; & C \left| \Psi_{k_{U,MN}}(h^m(x_i^m, \hat{a}_i^m), x_i^m, \hat{a}_i^m)' \gamma_{0,k_{U,MN}} - \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \gamma_{0,k_{U,MN}} \right|^2 \\
\leq \; & \mathcal{O}_p \left(k_{U,MN}^{-2\alpha_U} + |\hat{a}_i^m - a_i^m|^2 \right).
\end{aligned}
$$

Together with Theorem 2.4.1, this implies that

$$
\begin{aligned}
\left\| \hat{S}_{\Psi\Psi}^{1/2} \left(\bar{\gamma}_{0,k_{U,MN}} - \tilde{\gamma}_{0,k_{U,MN}} \right) \right\|^2 &= \left(\bar{S}_{0,I} - \tilde{S}_{0,I} \right)' \hat{S}_{\Psi} \hat{S}_{\Psi\Psi}^{-} \hat{S}_{\Psi}' \left(\bar{S}_{0,I} - \tilde{S}_{0,I} \right) / (MN)^2 \\
&\leq C \left(\bar{S}_{0,I} - \tilde{S}_{0,I} \right)' \left(\bar{S}_{0,I} - \tilde{S}_{0,I} \right) / (MN) \\
&= \mathcal{O}_p \left(k_{U,MN}^{-2\alpha_U} + \nu_{a,N}^2 \right).
\end{aligned}
$$

So (iii) holds by Lemma 2.D.9(iii).

For (iii), recall the definitions of the functionals in equation (2.D.25) and equation (2.D.26). Together with the Cauchy-Schwarz inequality, they imply that

$$
\begin{aligned}
& \left\| \tilde{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) - \hat{S}_{0,I}(h_i^m, x_i^m, \hat{a}_i^m) \right\|^2 \\
\leq \; & \left[\sum_{j=1}^{d_x} \left| \Upsilon^{(j)} \left(\hat{h}^m, \frac{\widehat{e_{x_j}}}{e} \right) - \Upsilon^{(j)} \left(h^m, \frac{e_{x_j}}{e} \right) \right|^2 \right. \\
& \left. + \left| \Upsilon^{(a)} \left(\hat{h}^m, \frac{\widehat{e_a}}{e} \right) - \Upsilon^{(a)} \left(h^m, \frac{e_a}{e} \right) \right|^2 \right] \\
& \cdot \left\| \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)' \gamma_{0,k_{U,MN}} \right\|^2 . \qquad (2.D.40)
\end{aligned}
$$

In equation (2.D.40),

$$
\begin{aligned}
\left\|\Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right\|^2 &\leq \left\|U_h(h_i^m, x_i^m, \hat{a}_i^m) - \Psi_{k_{U,MN}}(h_i^m, x_i^m, \hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right\|^2 \\
&\quad + \left\|U_h(h_i^m, x_i^m, \hat{a}_i^m)\right\|^2 \\
&\leq Ck_{U,MN}^{-2\alpha_U} + B_U, \qquad (2.D.41)
\end{aligned}
$$

where the first inequality holds by the triangular inequality and the Cauchy-Schwarz inequality; the second inequality holds by Assumptions 2.4.11 and 2.4.12. Together, equation (2.D.32), equation (2.D.33), equation (2.D.40), equation (2.D.41) and Lemma 2.D.9(iii) imply

$$
\begin{aligned}
\left\|\hat{S}_{\Psi\Psi}^{1/2}\left(\tilde{\gamma}_{0,k_{U,MN}} - \gamma_{0,k_{U,MN}}\right)\right\|^2 &= \left(\tilde{S}_{0,I} - \hat{S}_{0,I}\right)'\hat{S}_\Psi \hat{S}_{\Psi\Psi}^{-}\hat{S}_\Psi'\left(\tilde{S}_{0,I} - \hat{S}_{0,I}\right)/(MN)^2 \\
&\leq C\left(\tilde{S}_{0,I} - \hat{S}_{0,I}\right)'\left(\tilde{S}_{0,I} - \hat{S}_{0,I}\right)/(MN) \\
&= \mathcal{O}_p\left(\nu_{e_a,M,N}^2 + \sum_{j=1}^{d_x}\nu_{e_j,M,N}^2\right).
\end{aligned}
$$

So (iv) holds. □

Proof of Lemma 2.4.7

Proof. By the triangular inequality, we have

$$
\begin{aligned}
&\left\|\hat{\gamma}_{k_{U,MN}} - \gamma_{0,k_{U,MN}}\right\| \\
\leq\ &\left\|\hat{\gamma}_{k_{U,MN}} - \bar{\gamma}_{k_{U,MN}}\right\| + \left\|\bar{\gamma}_{k_{U,MN}} - \bar{\gamma}_{0,k_{U,MN}}\right\| + \left\|\tilde{\gamma}_{0,k_{U,MN}} - \gamma_{0,k_{U,MN}}\right\|.
\end{aligned}
$$

So the result follows by Lemma 2.D.10. □

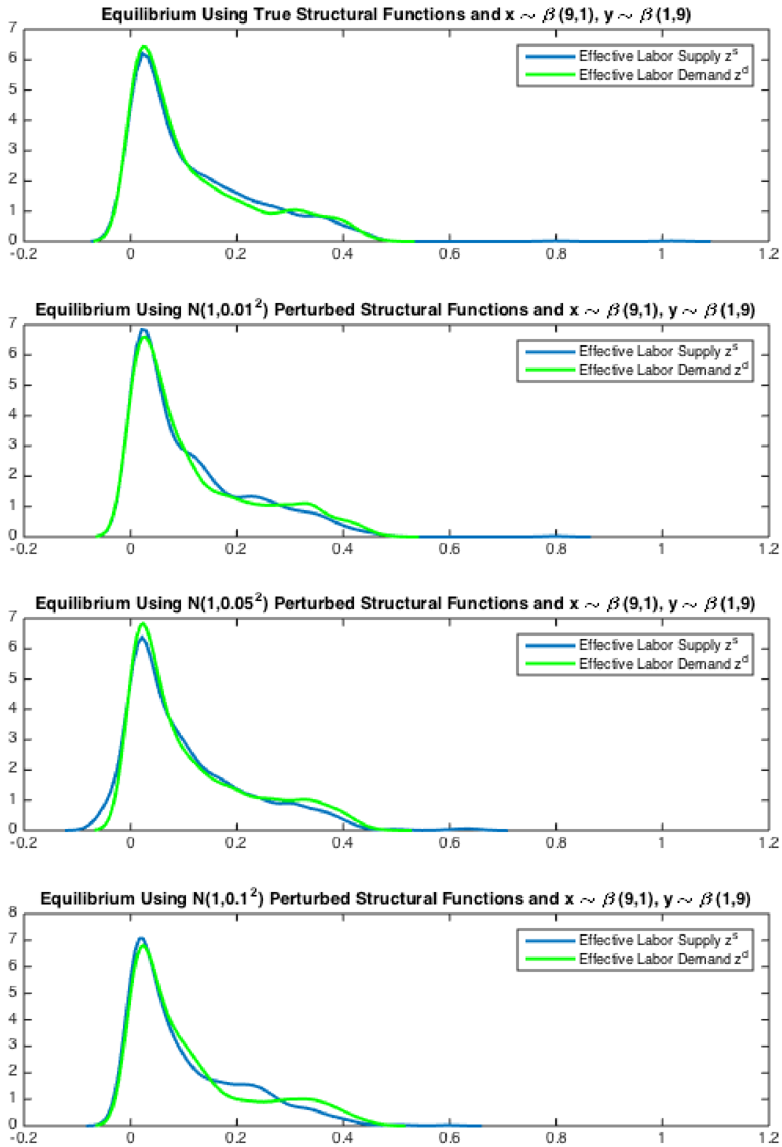
Proof of Theorem 2.4.8

Proof. By the triangular inequality,

$$\begin{aligned}
& \sup_{(h,x,a)} \left|\hat{U}_h(h,x,a) - U_h(h,x,a)\right| \\
\leq \; & \sup_{(h,x,a)} \left|\Psi_{k_{U,MN}}(h,x,a)' \left(\hat{\gamma}_{k_{U,MN}} - \gamma_{0,k_{U,MN}}\right)\right| \\
& + \sup_{(h,x,a)} \left|\Psi_{k_{U,MN}}(h,x,a)'\gamma_{0,k_{U,MN}} - U_h(h,x,a)\right| \\
= \; & \mathcal{O}_p\left(\zeta_0(k_{U,MN})\left[\nu_{e_a,M,N} + \sum_{j=1}^{d_x} \nu_{e_j,M,N} + k_{U,MN}^{-\alpha_U}\right]\right) + \mathcal{O}_p\left(k_{U,MN}^{-\alpha_U}\right) \\
= \; & \mathcal{O}_p\left(\zeta_0(k_{U,MN})\left[\nu_{e_a,M,N} + \sum_{j=1}^{d_x} \nu_{e_j,M,N} + k_{U,MN}^{-\alpha_U}\right]\right),
\end{aligned}$$

where the inequality holds by the triangular inequality; the equality holds by Lemmas 2.4.7 and 2.D.8, Assumption 2.4.11, and the Cauchy-Schwarz inequality. So the result holds. □

Figure 2.8: Numerically Solved Equilibrium Using True and Perturbed Structural Functions



The first panel shows the equilibrium densities of effective labor supply z^s and demand z^d when solving the equilibrium using the true structural functions. The following three panels show the equilibrium when the structural function values are perturbed by multiplying random variables drawn from $\mathcal{N}(1, 0.01^1)$, $\mathcal{N}(1, 0.05^1)$ and $\mathcal{N}(1, 0.1^1)$, respectively. The perturbed equilibria are very close to the true one. This suggests that the equilibrium is a continuous mapping from the structural functions, and that the algorithm approximates the equilibrium well.

Chapter 3

Uniform Asymptotic Risk of Averaging GMM Estimator Robust to Misspecification

3.1 Introduction

We are interested in estimating some finite dimensional parameter $\theta_0 \in \mathbb{R}^{d_\theta}$ which is uniquely identified by the moment restrictions

$$\mathbb{E}_F[g_1(W, \theta_0)] = 0_{r_1 \times 1} \tag{3.1.1}$$

for some known vector functions $g_1(\cdot) : \mathcal{W} \times \Theta \to \mathbb{R}^{r_1}$, where Θ is a compact subset of $\mathbb{R}^{d_\theta}$, W is a random vector with support $\mathcal{W}$ and joint distribution F, and $\mathbb{E}_F[\cdot]$ denotes the expectation operator under F. Suppose we have i.i.d. data $\{W_i\}_{i=1}^n$, where W_i has distribution F for any $i = 1, \ldots, n$.[1] Let $\overline{g}_1(\theta) = n^{-1} \sum_{i=1}^n g_1(W_i, \theta)$. One efficient GMM estimator for θ_0 is

$$\widehat{\theta}_1 = \arg\min_{\theta \in \Theta} \overline{g}_1(\theta)'(\overline{\Omega}_1)^{-1} \overline{g}_1(\theta), \tag{3.1.2}$$

where $\overline{\Omega}_1 = n^{-1} \sum_{i=1}^n g_1(W_i, \widetilde{\theta}_1) g_1(W_i, \widetilde{\theta}_1)' - \overline{g}_1(\widetilde{\theta}_1) \overline{g}_1(\widetilde{\theta}_1)'$ is the efficient weighting matrix with some preliminary consistent estimator $\widetilde{\theta}_1$.[2] In a linear instrumental variable (IV) example, $Y_i = X_i'\theta_0 + U_i$ where the IV $Z_{1,i} \in \mathbb{R}^{r_1}$ satisfies $\mathbb{E}_F[Z_{1,i}U_i] = 0_{r_1 \times 1}$. The moments in (3.1.1) hold with $g_1(W_i, \theta_0) = Z_{1,i}(Y_i - X_i'\theta_0)$ and θ_0 is uniquely identified if $\mathbb{E}_F[Z_{1,i}X_i']$ has full column rank. Under certain regularity conditions, it is well-known that $\widehat{\theta}_1$ is consistent and achieves the lowest asymptotic variance among GMM estimators based on the moments

[1] The main theory of the paper can be easily extended to time series models with dependent data, as long as the preliminary results in Lemma 3.B.1 hold.

[2] For example, $\widetilde{\theta}_1$ could be the GMM estimator similar to $\widehat{\theta}_1$ but with an identity weighting matrix, see (3.B.11) in the Appendices.

in (3.1.1), see Hansen (1982).

If one has additional moments $\mathbb{E}_F[g^*(W_i,\theta_0)] = 0_{r^*\times 1}$ for some known function $g^*(\cdot)$: $\mathcal{W}\times\Theta\to\mathbb{R}^{r^*}$, imposing them together with (3.1.1) can further reduce the asymptotic variance of the GMM estimator. However, if these additional moments are misspecified in the sense that $\mathbb{E}_F[g^*(W_i,\theta_0)] \neq 0_{r^*\times 1}$, imposing $\mathbb{E}_F[g^*(W_i,\theta_0)] = 0_{r^*\times 1}$ may result in inconsistent estimation. The choice of moment conditions is routinely faced by empirical researchers. Take the linear IV model for example. One typically starts with a large number of candidate IVs but only has confidence that a small number of them are valid, denoted by $Z_{1,i}$. The rest of them, denoted by Z_i^*, are valid only under certain economic hypothesis that yet to be tested. In this example, $g^*(W_i,\theta_0) = Z_i^*(Y_i - X_i'\theta_0)$. In contrast to the conservative estimator $\widehat{\theta}_1$, an aggressive estimator $\widehat{\theta}_2$ always imposes $\mathbb{E}_F[g^*(W_i,\theta_0)] = 0_{r^*\times 1}$ regardless of its validity. Let $g_2(W_i,\theta) = (g_1(W_i,\theta)', g^*(W_i,\theta)')'$ and $\overline{g}_2(\theta) = n^{-1}\sum_{i=1}^n g_2(W_i,\theta)$. The aggressive estimator $\widehat{\theta}_2$ takes the form

$$\widehat{\theta}_2 = \underset{\theta\in\Theta}{\arg\min}\,\overline{g}_2(\theta)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta), \tag{3.1.3}$$

where $\overline{\Omega}_2$ is constructed in the same way as $\overline{\Omega}_1$ except that $g_1(W_i,\theta)$ is replaced by $g_2(W_i,\theta)$.[3]

Because imposing $\mathbb{E}_F[g^*(W_i,\theta_0)] = 0_{r^*\times 1}$ is a double-edged sword, a data-dependent decision usually is made to choose between $\widehat{\theta}_1$ and $\widehat{\theta}_2$. To study such a decision and the subsequent estimator, let

$$\delta_F = \mathbb{E}_F[g^*(W_i,\theta_0)] \in \mathbb{R}^{r^*}. \tag{3.1.4}$$

[3]See the first line of equations (3.B.15) for the definition of $\overline{\Omega}_2$.

The pre-testing approach tests the null hypothesis $H_0 : \delta_F = 0_{r^* \times 1}$ and constructs an estimator

$$\widehat{\theta}_{pre} = 1\{T_n > c_\alpha\}\widehat{\theta}_1 + 1\{T_n \leq c_\alpha\}\widehat{\theta}_2 \tag{3.1.5}$$

for some test statistic T_n with the critical value c_α at the significance level α. One popular test is the J-test, see Hansen (1982), and c_α is the $1-\alpha$ quantile of the chi-squared distribution with degree of freedom $r_2 - d_\theta$ where $r_2 = r_1 + r^*$. Because the power of this test against most fixed alternative is 1, $\widehat{\theta}_{pre}$ equals $\widehat{\theta}_1$ with probability 1 asymptotically ($n \to \infty$) for these fixed misspecified model where $\delta_F \neq 0_{r^* \times 1}$. Thus, it seems that $\widehat{\theta}_{pre}$ is immune to moment misspecification. However, we care about the finite-sample mean squared error (MSE) of $\widehat{\theta}_{pre}$ in practice and this standard pointwise asymptotic analysis (δ_F is fixed and $n \to \infty$) provides a poor approximation to the former.[4] To see the comparison between $\widehat{\theta}_{pre}$ and $\widehat{\theta}_1$, the dashed line in Figure 3.1 plots the finite-sample ($n = 250$) MSE of $\widehat{\theta}_{pre}$ while the MSE of $\widehat{\theta}_1$ is normalized to be 1. For some values of δ_F, the MSE of $\widehat{\theta}_{pre}$ is larger than that of $\widehat{\theta}_1$, sometimes by 40%.

The goal of this paper is twofold. First, we propose a data-dependent averaging of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ that takes the form

$$\widehat{\theta}_{eo} = (1 - \widetilde{\omega}_{eo})\widehat{\theta}_1 + \widetilde{\omega}_{eo}\widehat{\theta}_2 \tag{3.1.6}$$

where $\widetilde{\omega}_{eo} \in [0, 1]$ is a data-dependent weight specified in (3.4.7) below. The subscript in $\widetilde{\omega}_{eo}$ is short for empirical optimal because this weight is an empirical analog of an infeasible

[4] The poor approximation of the pointwise asymptotics to the finite sample properties of the pre-test estimator has been noted in Shibata (1986), Pötscher (1991), Kabaila (1995, 2009) and Leeb and Pötscher (2005, 2008), among others.

optimal weight ω_F^* defined in (3.4.3) below. We plot the finite-sample MSE of this averaging estimator as the solid line in Figure 3.1. This averaging estimator is robust to misspecification in the sense that the solid line is below 1 for all values of δ_F, in contrast to the bump in the dashed line that represents the pre-test estimator. Second, we develop a *uniform* asymptotic theory to justify the finite-sample robustness of this averaging estimator. We show that this averaging estimator dominates the conservative estimator uniformly over a large class of models with different degrees of misspecification.[5] The standard asymptotic theory is pointwise and fails to reveal the fragile nature of the pre-test estimator. A stronger uniform notion of robustness is crucial for this model. Furthermore, we quantify the upper and lower bounds of the asymptotic risk differences between the averaging estimator and the conservative estimator.[6]

The rest of the chapter is organized as follows. Section 3.2 discusses the literatures related to our paper. Section 3.3 defines the parameter space over which the uniform result is established and defines uniform dominance. Section 3.4 introduces the averaging weight. Section 3.5 provides an analytical representation of the bounds of the asymptotic risk differences and applies it to show that the averaging GMM estimator uniformly dominates the conservative estimator. Section 3.6 studies the upper bound of the asymptotic risk difference in a local misspecification framework. Section 3.7 investigates the finite sample performance of our averaging estimator using Monte Carlo simulations. Section 3.8 concludes. Proofs and technical arguments are given in the Appendices.

[5]The uniform dominance is established under the truncated weighted loss function which is defined in (3.3.7) below.

[6]The lower and upper bounds of asymptotic risk difference are defined in (3.3.9) below.

Notation. Let $C > 1$ be generic positive constant whose value does not depend on the joint distribution F or the sample size. $a \equiv b$ means a is defined as b. For any real matrix A, we use $||A||$ to denote the Frobenius norm of A. If A is a real square matrix, we use $tr(A)$ denote the trace of A, and $\rho_{\min}(A)$ and $\rho_{\max}(A)$ to denote the smallest and largest eigenvalues of A, respectively. For any positive integers d_1 and d_2, I_{d_1} and $0_{d_1 \times d_2}$ stand for the $d_1 \times d_1$ identity matrix and $d_1 \times d_2$ zero matrix, respectively. Let $\text{vec}(\cdot)$ denotes vectorization of a matrix and $\text{vech}(\cdot)$ denotes the half vectorization of a symmetric matrix. Let $\mathbb{R} = (-\infty, +\infty)$, $\mathbb{R}_+ = [0, +\infty)$, $\mathbb{R}_\infty = \mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{+\infty\}$. For any positive integers d and any set $\mathbb{S}$, $\mathbb{S}^d$ denotes the Cartesian product of d many sets: $\mathbb{S}_1 \times \cdots \times \mathbb{S}_d$ with $\mathbb{S}_j = \mathbb{S}$ for $j = 1, \ldots, d$. For any set $\mathbb{S}$, $int(\mathbb{S})$ denotes the interior of $\mathbb{S}$. We use $\{n\}$ to denote the set of natural numbers and $\{p_n\} = \{p_n : n \geq 1\}$ denote a subsequence of $\{n\}$. For any (possibly random) positive sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, $a_n = O_p(b_n)$ means that $\lim_{c\to\infty} \limsup_n \Pr(a_n/b_n > c) = 0$; $a_n = o_p(b_n)$ means that for all $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(a_n/b_n > \varepsilon) = 0$. Let "$\to_p$" and "$\to_d$" stand for convergence in probability and convergence in distribution, respectively.

3.2 Related Literature

In this section, we discuss some related literature. Our uniform dominance result is related to the Stein's phenomenon (Stein, 1956) in parametric models. The James-Stein (JS) estimator (James and Stein, 1961) is shown to dominate the maximum likelihood estimator in exact normal sampling. Hansen (2016) considers local asymptotic analysis of the JS-type averaging estimator in general parametric models and substantially extends its application in econometrics. The present paper focuses on the uniformity issue and studies the Stein's phenomenon in non-Gaussian semiparametric nonlinear models. The proposed averaging estimator is not a mimic of the JS-type estimator for moment-based models and we find its finite-sample risk compares favorably to that of the latter. The asymptotic results are developed along drifting sequences of data generating processes (DGPs) with different degrees of misspecification. This class of DGPs include the crucial $n^{-1/2}$ local sequences that are considered by Hjort and Claeskens (2003), Liu (2015), Hansen (2007, 2015, 2016), DiTraglia (2016) for various averaging estimators, as well as some more distant sequences. The theoretical results glue all sequences together and show that they are sufficient to provide a uniform approximation of the finite-sample risk differences. The proof uses the techniques developed in Andrews and Guggenberger (2006) and Andrews, Cheng, and Guggenberger (2011) for uniform size control for inference in non-standard problems.

Measured by the MSE, the post-model-selection estimator based on consistent model selection procedure usually does better than the unrestricted estimator in part of the parameter space and worse than the latter in other part of the parameter space. One standard example is the Hodge's estimator, whose scaled maximal MSE diverges to infinity with the growth of the sample size (see, e.g., Lehmann and Casella, 1998). Similar unbounded risk results for other post-model-selection estimators are established in Yang (2005) and Leeb and Pötscher (2008). The post-model-selection estimator has unbounded (scaled) maximal MSE because it is based on a non-smooth transition rule between the restricted and unrestricted estimators and a consistent model selection procedure is employed in the transition rule.

However, the averaging estimator proposed in this paper is based on a smooth combination of the restricted and unrestricted estimators and no model selection procedure is used in the smooth combination. Hence our averaging estimator is essentially different from the post-model-selection estimator and the uniform dominance result established in this paper does not contradict the unbounded risk property of the post-model-selection estimator found in Yang (2005) and Leeb and Pötscher (2008).

The estimator proposed in this paper is a frequentist model averaging (FMA) estimator. FMA estimators have received much attention in recent years. Important works include Buckland, Burnham, and Augustin (1997), Hjort and Claeskens (2003, 2006), Leung and Barron (2006), Claeskens and Carroll (2007), Hansen (2007, 2015), Hansen and Racine (2012), Cheng and Hansen (2015) and Lu and Su (2015), to name only a few. Uniform asymptotic properties are important for frequentist estimators when standard pointwise asymptotic properties fail to capture their finite-sample behaviors. Our paper provides a uniform asymptotic framework to compare different FMA estimators in moment-based models.

Recently, DiTraglia (2016) and Hansen (2015) both consider averaging estimators that combine the ordinary least squares (OLS) estimator and the two-stage-least-squares (2SLS) estimator in linear IV models. In linear IV models with homoskedastic errors, our conservative estimator becomes the 2SLS estimator, and our aggressive estimator using both the IVs and the endogenous variables becomes the OLS estimator[7]. However, the averaging weight is different from those in DiTraglia (2016) and Hansen (2015) and we study it in a different

[7]Consider the linear IV model $Y_i = X_i'\theta_0 + u_i$ with instruments Z_i. The aggressive estimator is equivalent to the OLS estimator because $(X'P_{[X,Z]}X)^{-1}X'P_{[X,Z]}Y = (X'X)^{-1}X'Y$, where $Y = (Y_1, \dots, Y_n)'$, $X = (X_1, \dots, X_n)'$, $Z = (Z_1, \dots, Z_n)'$ and $P_{[X,Z]} = (X, Z)\left[(X, Z)'(X, Z)\right]^{-1}(X, Z)'$ denotes the projection matrix.

asymptotic framework.

There is a large literature studying the validity of GMM moment conditions. Many methods can be applied to detect the validity, including the over-identification tests (see, e.g., Sargan, 1958; Hansen, 1982; and Eichenbaum, Hansen, and Singleton, 1988), the information criteria (see, e.g., Andrews, 1999; Andrews and Lu, 2001; Hong, Preston, and Shum, 2003), and the penalized estimation methods (see, e.g., Liao, 2013 and Cheng and Liao, 2015). Recently, misspecified moments and their consequences are considered by Ashley (2009), Berkowitz, Caner, and Fang (2012), Conley, Hansen, and Rossi (2012), Doko Tchatoka and Dufour (2008, 2014), Guggenberger (2012), Nevo and Rosen (2012), Kolesár, Chetty, Friedman, Glaeser, and Imbens (2015), Small (2007), Kang, Zhang, Cai, and Small (2016) among others. Moon and Schorfheide (2009) explore over-identifying moment inequalities to reduce the MSE. This paper contributes to this literature by providing new uniform results for potentially misspecified semiparametric models.

3.3 Parameter Space and Uniform Dominance

Let $g_{2,j}(w,\theta)$ $(j = 1,\ldots,r_2)$ denote the j-th component function of $g_2(w,\theta)$. We assume that $g_{2,j}(w,\theta)$ for $j = 1,\ldots,r_2$ is twice continuously differentiable with respect to θ for any $w \in \mathcal{W}$. The first and second order derivatives of $g_2(w,\theta)$ with respect to θ are denoted by

$$g_{2,\theta}(w,\theta) \equiv \begin{pmatrix} \frac{\partial g_{2,1}(w,\theta)}{\partial \theta'} \\ \vdots \\ \frac{\partial g_{2,r_2}(w,\theta)}{\partial \theta'} \end{pmatrix} \text{ and } g_{2,\theta\theta}(w,\theta) \equiv \begin{pmatrix} \frac{\partial^2 g_{2,1}(w,\theta)}{\partial \theta \partial \theta'} \\ \vdots \\ \frac{\partial^2 g_{2,r_2}(w,\theta)}{\partial \theta \partial \theta'} \end{pmatrix}$$

respectively.[8] Let $\mathcal{F}$ be a set of distribution functions of W. For $k = 1$ and 2, define the expectation of the moment functions, the Jacobian matrix and the variance-covariance matrix as

$$M_{k,F} \equiv \mathbb{E}_F\left[g_k(W,\theta_F)\right], G_{k,F} \equiv \mathbb{E}_F\left[g_{k,\theta}(W,\theta_F)\right] \text{ and } \Omega_{k,F} \equiv \text{Var}_F\left[g_k(W,\theta_F)\right] \quad (3.3.1)$$

for any $F \in \mathcal{F}$ respectively, where θ_F denotes the solution of the moment restrictions $\mathbb{E}_F\left[g_1(W,\theta)\right] = 0_{r_1\times 1}$. The unique identification of θ_0 by (3.1.1) ensures that $\theta_0 = \theta_F$. The moments above exist by Assumption 3.3.2 below. We consider the risk difference between two estimators uniformly over $F \in \mathcal{F}$ that satisfies Assumptions 3.3.1-3.3.3 below.

Let $Q_F(\theta) \equiv \mathbb{E}_F[g_2(W,\theta)]'\Omega_{2,F}^{-1}\mathbb{E}_F[g_2(W,\theta)]$ for any $\theta \in \Theta$ which denotes the population criterion of the GMM estimation in (3.1.3). For any $\theta \in \Theta$, define $B_\varepsilon^c(\theta) = \{\theta' \in \Theta : ||\theta' - \theta|| \geq \varepsilon\}$.

Assumption 3.3.1. *The following conditions hold for any $F \in \mathcal{F}$:*

(i) $\mathbb{E}_F\left[g_1(W,\theta_F)\right] = 0_{r_1\times 1}$ *for some $\theta_F \in int(\Theta)$;*

(ii) *for any $\varepsilon > 0$ there is $\eta_{1,\varepsilon} > 0$ such that* $\inf_{\theta \in B_\varepsilon^c(\theta_F)} ||\mathbb{E}_F\left[g_1(W,\theta)\right]|| \geq \eta_{1,\varepsilon}$*;*

(iii) *there is $\theta_F^* \in int(\Theta)$ such that for any $\varepsilon > 0$ there is $\eta_{2,\varepsilon} > 0$ with*

$$\inf_{\theta \in B_\varepsilon^c(\theta_F^*)} Q_F(\theta) - Q_F(\theta_F^*) \geq \eta_{2,\varepsilon};$$

(iv) $||G_{2,F}'\Omega_{2,F}^{-1}\delta_{2,F}|| \geq C^{-1}\left\|\delta_{2,F}\right\|$, *where* $\delta_{2,F} = (0_{1\times r_1}, \delta_F')'$.

[8] By definition, $g_{1,\theta}(w,\theta)$ and $g_{1,\theta\theta}(w,\theta)$ are the leading $r_1 \times d_\theta$ and $(r_1 d_\theta) \times d_\theta$ submatrices of $g_{2,\theta}(w,\theta)$ and $g_{2,\theta\theta}(w,\theta)$ respectively.

Assumptions 3.3.1.(i)-(ii) require that the true unknown parameter θ_F is uniquely identified by the moment conditions $\mathbb{E}_F[g_1(W,\theta_F)] = 0_{r_1\times 1}$. Assumption 3.3.1.(iii) implies that a pseudo true value θ_F^* is identified by the unique minimizer of the population GMM criterion $Q_F(\theta)$ under possible misspecification. Assumption 3.3.1.(iv) requires that $\delta_{2,F}$ does not lie in the right null space of the matrix $G'_{2,F}\Omega_{2,F}^{-1}$, which rules out the special case that θ_F may be consistently estimable even with severely misspecified moment conditions.

Assumption 3.3.2. *The following conditions hold for any $F \in \mathcal{F}$:*

(i) $\mathbb{E}_F[\sup_{\theta\in\Theta}(||g_2(W,\theta)||^{2+\gamma} + ||g_{2,\theta}(W,\theta)||^{2+\gamma} + ||g_{2,\theta\theta}(W,\theta)||^{2+\gamma})] \leq C$ *for some* $\gamma > 0$*;*

(ii) $\rho_{\min}(\Omega_{2,F}) \geq C^{-1}$*;*

(iii) $\rho_{\min}(G'_{1,F}G_{1,F}) \geq C^{-1}$.

Assumption 3.3.2.(i) imposes $2+\gamma$ finite moment conditions on the GMM moment functions and their first and second derivatives. Assumptions 3.3.2.(ii) and 3.3.2.(iii) are important sufficient conditions for the local identification of the unknown parameter in GMM with valid moment conditions.

Uniform dominance is of interest only if we allow for different degrees of misspecification in the parameter space. If we only allow for correctly specified models or severely misspecified models, the desired dominance results hold trivially following a pointwise analysis. The next assumption states that the parameter space contains a continuous perturbation from correctly specified models to misspecified models and it forms a closed set. Write

$$v_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta'_F) \tag{3.3.2}$$

for any $F \in \mathcal{F}$. It is clear that v_F includes the Jacobian matrix, the variance-covariance matrix, and the measure of misspecification of the moment conditions $\mathbb{E}_F[g^*(W_i, \theta_F)] = 0_{r^* \times 1}$. Let $\overline{v}_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})')$ for any $F \in \mathcal{F}$.

Assumption 3.3.3. (i) *The set $\Lambda \equiv \{v_F\colon F \in \mathcal{F}\}$ is closed;* (ii) *for any $F \in \mathcal{F}$ with $\delta_F = 0_{r^* \times 1}$, there exists a constant $\varepsilon_F > 0$ such that for any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 \leq ||\widetilde{\delta}|| < \varepsilon_F$, there is $\widetilde{F} \in \mathcal{F}$ with $\delta_{\widetilde{F}} = \widetilde{\delta}$ and $\left\|\overline{v}_F - \overline{v}_{\widetilde{F}}\right\| \leq C||\widetilde{\delta}||^{\kappa}$ for some $\kappa > 0$.*

Assumption 3.3.3.(i) is useful to show that the asymptotic risk of the GMM estimator has a well-defined upper bound which can be represented by v_F for $F \in \mathcal{F}$ and some other nuisance parameter which measures the asymptotic degree of misspecification of the moment conditions. Assumption 3.3.3.(ii) requires that for any $F \in \mathcal{F}$ such that $\mathbb{E}_F[g_2(W, \theta_F)] = 0_{r^* \times 1}$ is valid, there are many DGPs $\widetilde{F} \in \mathcal{F}$ which are close to F. Here the closeness of any two DGPs F and $\widetilde{F}$ are measured by the distance between $\overline{v}_F$ and $\overline{v}_{\widetilde{F}}$. This condition ensures that the aforementioned upper bound of the asymptotic risk is also a lower bound, and hence it is useful to derive the exact expression of the asymptotic risk of GMM estimator.

Example 3.3.1. ***(Linear IV Model)*** *We study a simple linear IV model and provide a set of simple conditions that imply Assumptions 3.3.1, 3.3.2 and 3.3.3. The parameters of interest θ_0 are the coefficients of the endogenous regressors X_i in*

$$Y_i = X_i'\theta_0 + U_i, \text{ where } \mathbb{E}_{F^*}[Z_{1,i}U_i] = 0_{r_1 \times 1} \tag{3.3.3}$$

with some valid instruments $Z_{1,i} \in \mathbb{R}^{r_1}$, where F^* denotes the joint distribution of $(X_i', Z_{1,i}', V_i', U_i)$ and V_i is a $r^* \times 1$ random vector. In addition, we have some potentially misspecified IVs

$Z_i^* \in \mathbb{R}^{r^*}$. *In a reduced form equation obtained by projecting* Z_i^* *on* U_i, *we can write*

$$Z_i^* = U_i\delta_0 + V_i, \text{ where } \mathbb{E}_{F^*}[U_iV_i] = 0_{r^*\times 1}, \tag{3.3.4}$$

where δ_0 *is a* $r^* \times 1$ *real vector which characterizes the degree of misspecification. These additional IVs are valid only if* $\delta_0 = 0_{r^*\times 1}$. *Without loss of generality, we assume* $\mathbb{E}_{F^*}[U_i^2] = 1$ *such that* $\mathbb{E}_F[g^*(W_i,\theta_0)] = \mathbb{E}_{F^*}[Z_i^*U_i] = \delta_0$ *as in (3.1.4), where* F *denotes the joint distribution of* $W_i = (Y_i, Z_{1,i}', Z_i^{*\prime}, X_i')$ *which is determined by* θ_0, δ_0 *and* F^* *through the linear structures in (3.3.3) and (3.3.4).*

Let $\mathcal{F}^*$ *denote a class of distributions containing* F^*, *and let* Θ *and* Δ_δ *denote the parameter spaces of* θ_0 *and* δ_0 *respectively. Let* $\mathcal{F}$ *denote a class of distributions* F *induced by any* $\theta_0 \in \Theta$, *any* $\delta_0 \in \Delta_\delta$ *and any* $F^* \in \mathcal{F}^*$. *The Jacobian matrices are*

$$G_{1,F} = -\mathbb{E}_F[Z_{1,i}X_i'] \text{ and } G_{2,F} = \begin{pmatrix} -\mathbb{E}_F[Z_{1,i}X_i'] \\ -\mathbb{E}_F[Z_i^*X_i'] \end{pmatrix}. \tag{3.3.5}$$

The variance-covariance matrix is

$$\Omega_{2,F} = \mathbb{E}_F[Z_{2,i}Z_{2,i}'(Y_i - X_i'\theta_0)^2] - \mathbb{E}_F[(Y_i - X_i'\theta_0)Z_{2,i}]\mathbb{E}_F[(Y_i - X_i'\theta_0)Z_{2,i}'], \tag{3.3.6}$$

where $Z_{2,i} = (Z_{1,i}', Z_i^{*\prime})'$. *By definition,* $\Omega_{1,F}$ *is the leading* $r_1 \times r_1$ *submatrix of* $\Omega_{2,F}$.

Assumption 3.3.4. *For any* $\theta_0 \in \Theta$, *any* $\delta_0 \in \Delta_\delta$ *and any* $F^* \in \mathcal{F}^*$, *we have:*

(i) $\mathbb{E}_{F^*}[Z_{1,i}U_i] = 0_{r_1\times 1}$, $\mathbb{E}_{F^*}[U_i^2] = 1$ *and* $\mathbb{E}_{F^*}[U_iV_i] = 0_{r^*\times 1}$;

(ii) $\rho_{\min}(G'_{1,F}G_{1,F}) \geq C^{-1}$ *and* $\rho_{\min}(\Omega_{2,F}) \geq C^{-1}$;

(iii) $E_{F^*}[||X_i||^{4+\gamma} + U_i^{4+\gamma}] + E_F[||Z_{2,i}||^{4+\gamma}] \leq C$ *for some* $\gamma > 0$;

(iv) $||G'_{2,F}\Omega_{2,F}^{-1}\delta_{2,0}|| \geq C^{-1} \|\delta_{2,0}\|$, *where* $\delta_{2,0} = (0_{1\times r_1}, \delta'_0)'$;

(v) $\{(\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta'_0) : \theta_0 \in \Theta,\ \delta_0 \in \Delta_\delta$ *and* $F^* \in \mathcal{F}^*\}$ *is closed.*

Let $\theta^*_F = -(G'_{2,F}\Omega_{2,F}^{-1}G_{2,F})^{-1}G'_{2,F}\Omega_{2,F}^{-1}\mathbb{E}_F[Z_{2,i}Y_i]$ which is well defined by Assumptions 3.3.4.(ii)-(iii).

Lemma 3.3.2. *Suppose that* $\{W_i\}_{i=1}^n$ *are i.i.d. and generated by the linear model (3.3.3) and (3.3.4) with:* (i) θ_0, $\theta^*_F \in int(\Theta)$ *and* Θ *is compact;* (ii) $\Delta_\delta = [-C, C]^{r^*}$; *and* (iii) θ_0, δ_0 *and* $\mathcal{F}^*$ *satisfy Assumption* 3.3.4. *Then,* $\mathcal{F}$ *satisfies Assumptions* 3.3.1, 3.3.2 *and* 3.3.3.

For the linear IV model, Lemma 3.3.2 provides simple conditions on θ_0, δ_0 and $\mathcal{F}^*$ on which uniformity results are subsequently established.

Now we get back to the general set up. For a generic estimator $\widehat{\theta}$ of θ, consider a weighted quadratic loss function

$$\ell(\widehat{\theta}, \theta) = n(\widehat{\theta} - \theta)'\Upsilon(\widehat{\theta} - \theta), \tag{3.3.7}$$

where Υ is a $d_\theta \times d_\theta$ pre-determined positive definite matrix. For example, if $\Upsilon = I_{d_\theta}$, $\mathbb{E}_F[\ell(\widehat{\theta})]$ is the MSE of $\widehat{\theta}$. If $\Upsilon = (\Sigma_{1,F} - \Sigma_{2,F})^{-1}$ where $\Sigma_{k,F}$ $(k = 1, 2)$ are defined in (3.4.4), the weighting matrix Υ first rescales $\widehat{\theta}$ by the scale of variance reduction. If $\Upsilon = \mathbb{E}_F[X_iX'_i]$ for regressors X_i, $\mathbb{E}_F[\ell(\widehat{\theta})]$ is the MSE of $X'_i\widehat{\theta}$, an estimator of $X'_i\theta$. For ease of notation, we suppress the dependence of the loss function on θ and write it as $\ell(\widehat{\theta})$ for any estimator $\widehat{\theta}$. Whenever the expectation operator $\mathbb{E}_F[\cdot]$ is taken on the loss function $\ell(\widehat{\theta})$, it is understood that the parameter of interest, i.e. θ in the definition of $\ell(\widehat{\theta}, \theta)$ in (3.3.7), is θ_F which is defined through the moment restrictions (3.1.1).

Below we compare the averaging estimator $\widehat{\theta}_{eo}$ and the conservative estimator $\widehat{\theta}_1$. We are interested in the bounds of the finite sample risk difference

$$
\begin{aligned}
\underline{RD}_n(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \inf_{F \in \mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta}_{eo}) - \ell(\widehat{\theta}_1)], \\
\overline{RD}_n(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \sup_{F \in \mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta}_{eo}) - \ell(\widehat{\theta}_1)]. \qquad (3.3.8)
\end{aligned}
$$

We investigate these finite-sample objects by asymptotic analysis with $n^{1/2}(\widehat{\theta} - \theta_F)$ in $\ell(\widehat{\theta})$ replaced by its asymptotic distribution. To apply the bounded convergence theorem, we approximate $\ell(\widehat{\theta})$ with the trimmed loss $\ell_\zeta(\widehat{\theta}) \equiv \min\{\ell(\widehat{\theta}), \zeta\}$ and consider arbitrarily large trimming ($\zeta \to \infty$). As such, the finite-sample bounds in (3.3.8) are approximated by

$$
\begin{aligned}
Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \lim_{\zeta \to \infty} \liminf_{n \to \infty} \inf_{F \in \mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] \text{ and} \\
Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \lim_{\zeta \to \infty} \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)], \qquad (3.3.9)
\end{aligned}
$$

which are called lower and upper bounds of asymptotic risk difference respectively in this paper. The averaging estimator $\widehat{\theta}_{eo}$ asymptotically uniformly dominates the conservative estimator $\widehat{\theta}_1$ if

$$
Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) < 0 \text{ and } Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) \leq 0. \qquad (3.3.10)
$$

The bounds of the asymptotic risk difference build the uniformity over $F \in \mathcal{F}$ into the definition by taking $\inf_{F \in \mathcal{F}}$ and $\sup_{F \in \mathcal{F}}$ before $\liminf_{n \to \infty}$ and $\limsup_{n \to \infty}$ respectively.

Uniformity is crucial for the asymptotic results to give a good approximation to their finite-sample counterparts. These uniform bounds are different from pointwise results which are either obtained under a fixed DGP or a particular sequence of drifting DGP. The sequence of DGPs $\{F_n\}$ along which the supremum or the infimum are approached often varies with the sample size.[9] Therefore, to determine the bounds of the asymptotic risk difference, one has to derive the asymptotic distributions of these estimators under various sequences $\{F_n\}$. Under $\{F_n\}$, the observations $\{W_{n,i}\}_{i=1}^{n}$ form a triangular array. For notational simplicity, $W_{n,i}$ is abbreviated to W_i.

To study the bounds of asymptotic risk difference, we consider sequences of DGPs $\{F_n\}$ such that δ_{F_n} satisfies

$$\text{(i) } n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*} \text{ or (ii) } ||n^{1/2}\delta_{F_n}|| \to \infty. \tag{3.3.11}$$

Case (i) models mild misspecification, where δ_{F_n} is a $n^{-1/2}$-local deviation from $0_{r^*\times 1}$. Case (ii) includes the severe misspecification where $\|\delta_{F_n}\|$ is bounded away from 0 as well as the intermediate case in which $\delta_{F_n} \to 0$ and $||n^{1/2}\delta_{F_n}|| \to \infty$. To obtain a uniform approximation, all of these sequences are necessary. Once we study the bounds of asymptotic risk difference along each of these sequences, we show that we can glue them together to obtain the bounds of asymptotic risk difference.

[9] In the rest of the chapter, we use $\{F_n\}$ to denote $\{F_n \in \mathcal{F} : n = 1, 2, ...\}$.

3.4 Averaging Weight

We start by deriving the joint asymptotic distribution of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ under different degrees of misspecification. We consider sequences of DGPs $\{F_n\}$ such that (i) $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$ or $||n^{1/2}\delta_{F_n}|| \to \infty$; and (ii) G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} converges to $G_{2,F}$, $\Omega_{2,F}$ and $M_{2,F}$ for some $F \in \mathcal{F}$. [10]

For $k = 1, 2$, define

$$\Gamma_{k,F} = -\left(G'_{k,F}\Omega^{-1}_{k,F}G_{k,F}\right)^{-1} G'_{k,F}\Omega^{-1}_{k,F}. \tag{3.4.1}$$

Let $\mathcal{Z}_{2,F}$ denote a zero mean normal random vector with variance-covariance matrix $\Omega_{2,F}$ and $\mathcal{Z}_{1,F}$ denote its first r_1 components.

Lemma 3.4.1. *Suppose Assumptions 3.3.1 and 3.3.2 hold. Consider any sequence of DGPs $\{F_n\}$ such that $v_{F_n} \to v_F$ for some $F \in \mathcal{F}$, and $n^{1/2}\delta_{F_n} \to d$ for $d \in \mathbb{R}^{r^*}_\infty$.*

(a) *If $d \in \mathbb{R}^{r^*}$, then*

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_d \begin{pmatrix} \xi_{1,F} \\ \xi_{2,F} \end{pmatrix} \equiv \begin{pmatrix} \Gamma_{1,F}\mathcal{Z}_{1,F} \\ \Gamma_{2,F}\left(\mathcal{Z}_{2,F} + d_0\right) \end{pmatrix},$$

where $d_0 = (\mathbf{0}_{1\times r_1}, d')'$.

(b) *If $||d|| = \infty$, then $n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_d \xi_{1,F}$ and $||n^{1/2}(\widehat{\theta}_2 - \theta_{F_n})|| \to_p \infty$.*

[10] The requirement on the convergence of G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} is not restrictive as it seems to be. Lemma 3.B.7 in Appendix 3.B shows that the sequences G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} have subsequences which respectively converge to $G_{2,F}$, $\Omega_{2,F}$ and $M_{2,F}$ for some $F \in \mathcal{F}$. The general result on the lower and upper bounds of the asymptotic risk difference, Lemma 3.C.6 in Appendix 3.C, only requires to consider the subsequence $\{F_{p_n}\}$ such that $G_{2,F_{p_n}}$, $\Omega_{2,F_{p_n}}$ and $M_{2,F_{p_n}}$ are convergent, where $\{p_n\}$ is a subsequence of $\{n\}$. The asymptotic properties of the GMM estimators established in this section under the full sequence of DGPs $\{F_n\}$ holds trivially for its subsequence.

Given the joint asymptotic distribution of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, it is straightforward to study $\widehat{\theta}(\omega) = (1-\omega)\widehat{\theta}_1 + \omega\widehat{\theta}_2$ if ω is deterministic. Following Lemma 3.4.1.(a),

$$n^{1/2}(\widehat{\theta}(\omega) - \theta_{F_n}) \to_d \xi_F(\omega) \equiv (1-\omega)\xi_{1,F} + \omega\xi_{2,F} \tag{3.4.2}$$

for $n^{1/2}\delta_{F_n} \to d$, where $d \in \mathbb{R}^{r^*}$. In Appendix 3.B, a simple calculation shows that $\mathbb{E}[\ell(\xi_F(\omega))]$ is minimized at the infeasible optimal weight

$$\omega_F^* \equiv \frac{\mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}{d_0' \left(\Gamma_{2,F} - \Gamma_{1,F}^*\right)' \Upsilon \left(\Gamma_{2,F} - \Gamma_{1,F}^*\right) d_0 + \mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}, \tag{3.4.3}$$

where Υ is the matrix specified in the loss function,

$$\Sigma_{k,F} \equiv \left(G_{k,F}'\Omega_{k,F}^{-1}G_{k,F}\right)^{-1} \text{ for } k = 1, 2 \text{ and } \Gamma_{1,F}^* \equiv [\Gamma_{1,F}, \mathbf{0}_{d_\theta \times r^*}]. \tag{3.4.4}$$

To gain some intuition, consider the case where $\Upsilon = I_{d_\theta}$ such that the MSE of $\widehat{\theta}(\omega)$ is minimized at ω_F^*. In this case, the infeasible optimal weight ω_F^* yields the ideal bias and variance trade off. However, the bias depends on d, which cannot be consistently estimated. Hence, ω_F^* cannot be consistently estimated. Our solution to this problem is to construct an empirical analog of ω_F^* based on an inconsistent but unbiased estimator of d, such that the resulting averaging estimator reduces the MSE for any value of d.

The empirical analog of ω_F^* is constructed as follows. First, for $k = 1$ and 2, replace $\Sigma_{k,F}$

by its consistent estimator $\widehat{\Sigma}_k \equiv (\widehat{G}_k' \widehat{\Omega}_k^{-1} \widehat{G}_k)^{-1}$,[11] where

$$\widehat{G}_k \equiv n^{-1} \sum_{i=1}^{n} g_{k,\theta}(W_i, \widehat{\theta}_1) \text{ and } \widehat{\Omega}_k \equiv n^{-1} \sum_{i=1}^{n} g_k(W_i, \widehat{\theta}_1) g_k(W_i, \widehat{\theta}_1)' - \overline{g}_k(\widehat{\theta}_1) \overline{g}_k(\widehat{\theta}_1)'. \quad (3.4.5)$$

Note that $\widehat{G}_k$ and $\widehat{\Omega}_k$ are based on the conservative GMM estimator $\widehat{\theta}_1$. Second, replace $(\Gamma_{2,F} - \Gamma_{1,F}^*) d_0$ by its asymptotically unbiased estimator $n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)$ because

$$n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) \rightarrow_d (\Gamma_{2,F} - \Gamma_{1,F}^*) (\mathcal{Z}_{2,F} + d_0), \quad (3.4.6)$$

for $d_0 = (\mathbf{0}_{1 \times r_1}, d')'$ and $d \in \mathbb{R}^{r^*}$ following Lemma 3.4.1(a). Then then empirical optimal weight takes the form

$$\widetilde{\omega}_{eo} \equiv \frac{\operatorname{tr}(\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2))}{n(\widehat{\theta}_2 - \widehat{\theta}_1)' \Upsilon (\widehat{\theta}_2 - \widehat{\theta}_1) + \operatorname{tr}(\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2))}, \quad (3.4.7)$$

and the averaging GMM estimator takes the form

$$\widehat{\theta}_{eo} = (1 - \widetilde{\omega}_{eo}) \widehat{\theta}_1 + \widetilde{\omega}_{eo} \widehat{\theta}_2. \quad (3.4.8)$$

Next we consider the asymptotic distribution of $\widehat{\theta}_{eo}$ under different degrees of misspecification.

Lemma 3.4.2. *Suppose that Assumptions 3.3.1-3.3.3 hold. Consider $\{F_n\}$ such that $v_{F_n} \rightarrow v_F$ for some $F \in \mathcal{F}$, and $n^{1/2} \delta_{F_n} \rightarrow d$ for $d \in \mathbb{R}_{\infty}^{r^*}$.*

[11] The consistency of $\widehat{\Sigma}_k$ is proved in Lemma 3.4.2.

(a) *If* $d \in \mathbb{R}^{r^*}$, *then*

$$\widetilde{\omega}_{eo} \rightarrow_d \overline{\omega}_F \equiv \frac{\mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}{(\mathcal{Z}_{2,F} + d_0)'(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})(\mathcal{Z}_{2,F} + d_0) + \mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}$$

and

$$n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) \rightarrow_d \overline{\xi}_F \equiv (1 - \overline{\omega}_F)\xi_{1,F} + \overline{\omega}_F \xi_{2,F}.$$

(b) *If* $||d|| = \infty$, *then* $\widetilde{\omega}_{eo} \rightarrow_p 0$ *and* $n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) \rightarrow_d \xi_{1,F}$.

To study the bounds of asymptotic risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$, it is important to take into account the data-dependent nature of $\widetilde{\omega}_{eo}$. Unlike $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, the randomness in $\widetilde{\omega}_{eo}$ is non-negligible in the mild misspecification case (a) of Lemma 3.4.2. In consequence, $\widehat{\theta}_{eo}$ does not achieve the same bounds of asymptotic risk difference as the ideal averaging estimator $(1 - \omega^*_F)\widehat{\theta}_1 + \omega^*\widehat{\theta}_2$ does. Nevertheless, below we show that $\widehat{\theta}_{eo}$ is insured against potentially misspecified moments because it uniformly dominates $\widehat{\theta}_1$.

3.5 Bounds of Asymptotic Risk Difference under Misspecification

In this section, we study the bounds of the asymptotic risk difference defined in (3.3.9). Note that the asymptotic distributions of $\widehat{\theta}_1$ and $\widehat{\theta}_{eo}$ in Lemma 3.4.1 and 3.4.2 only depend on d, $G_{2,F}$ and $\Omega_{2,F}$. For notational convenience, define

$$h_{F,d} = (d', \mathrm{vec}(G_{2,F})', \mathrm{vech}(\Omega_{2,F})') \tag{3.5.1}$$

for any $F \in \mathcal{F}$ and any $d \in \mathbb{R}_{\infty}^{r^*}$. For the mild misspecification case, define the parameter space of $h_{F,d}$ as

$$H = \{h_{F,d} : d \in \mathbb{R}^{r^*} \text{ and } F \in \mathcal{F} \text{ with } \delta_F = 0_{r^* \times 1}\} \tag{3.5.2}$$

where δ_F is defined by (3.1.4) for a given F.

Theorem 3.5.1. *Suppose that Assumptions 3.3.1-3.3.3 hold. The bounds of the asymptotic risk difference satisfy*

$$Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \max\left\{\sup_{h \in H} [g(h)], 0\right\},$$
$$Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \min\left\{\inf_{h \in H} [g(h)], 0\right\},$$

where $g(h) \equiv \mathbb{E}[\overline{\xi}_F' \Upsilon \overline{\xi}_F - \xi_{1,F}' \Upsilon \xi_{1,F}]$.

To show that $\widehat{\theta}_{eo}$ uniformly dominates $\widehat{\theta}_1$, Theorem 3.5.1 implies that it is sufficient to show that $\inf_{h \in H} [g(h)] < 0$ and $\sup_{h \in H} [g(h)] \leq 0$. We can investigate $\inf_{h \in H} g(h)$ and $\sup_{h \in H} g(h)$ by simulating $g(h)$. In practice, we replace $G_{2,F}$ and $\Omega_{2,F}$ by their consistent estimators and plot $g(h)$ as a function of d. Even if the uniform dominance condition does not hold, $\min\{\inf_{h \in H} [g(h)], 0\}$ and $\max\{\sup_{h \in H} [g(h)], 0\}$ quantify the most- and least-favorable scenarios for the averaging estimator.

We call $\mathbb{E}[\overline{\xi}_F' \Upsilon \overline{\xi}_F]$ and $\mathbb{E}\left[\xi_{1,F}' \Upsilon \xi_{1,F}\right]$ the asymptotic risks of $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ such that $g(h)$ represents the asymptotic risk difference. For any given h, Figure 3.2 uses simulation to demonstrate that these asymptotic risks provide good approximations to the finite-sample risks $\mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo})]$ and $\mathbb{E}_F[\ell_\zeta(\widehat{\theta}_1)]$. Figure 3.2 normalizes the finite-sample risk of $\widehat{\theta}_1$ to be 1 in all cases and plots the (simulated) asymptotic risks and finite-sample risks of two estimators: one is the averaging estimator $\widehat{\theta}_{eo}$ and one is the pre-test GMM estimator based on the over-identification J-test with significance level 0.01.[12] The asymptotic risk for this pre-test estimator is given by (3.E.2) in the Appendices. It is clear that the finite sample risk and the (simulated) asymptotic risk are fairly close and the averaging GMM estimator uniformly dominates the conservatives estimator while the pre-test estimator does not.

Theorem 3.5.2. *Let* $A_F \equiv \Upsilon(\Sigma_{1,F} - \Sigma_{2,F})$. *Suppose that Assumptions 3.3.1-3.3.3 hold. If* $\text{tr}(A_F) > 0$ *and* $\text{tr}(A_F) \geq 4\rho_{\max}(A_F)$ *for any* $F \in \mathcal{F}$, *we have*

$$Asy\underline{RD}^*(\widehat{\theta}_{eo}, \widehat{\theta}_1) < 0 \text{ and } Asy\overline{RD}^*(\widehat{\theta}_{eo}, \widehat{\theta}_1) = 0.$$

Thus, $\widehat{\theta}_{eo}$ *uniformly dominates* $\widehat{\theta}_1$.

Theorem 3.5.2 indicates that: (i) there exists $\varepsilon_1 < 0$ and some finite integer n_{ε_1} such that the minimum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is less than ε_1 for any n larger than n_{ε_1}; (ii) for any $\varepsilon_2 > 0$, there exists a finite integer n_{ε_2} such that the maximum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is less than ε_2 for any n larger than n_{ε_2}. The pre-test estimator fails to satisfy both properties (i) and (ii) above at the same time. Consider the pre-test estimator based on the J-test for example.[13] Consider three different scenarios: (a) the critical value is

[12] For the finite-sample results, $\delta_{F_n} = c_0 n^{-1/2} 1_{r^* \times 1}$ where c_0 is a scale from 0 to 20. The finite sample risks are calculated using 100,000 simulated samples and the asymptotic risks are simulated by drawing 10,000 normal random vectors with mean zero and variance-covariance $\widehat{\Omega}_2$ in each simulated sample. No truncation is applied to the finite-sample risk.

[13] See Section 3.E in the Appendices for definition and analysis of this estimator.

fixed for any sample size; (b) the critical value diverges to infinity; and (c) the critical value converges to zero. In the pointwise asymptotic framework, the J-test based on the critical values in (a), (b) and (c) leads to inconsistent (but conservative) model selection, consistent model selection and no model selection results respectively. The pre-test estimator based on the J-test violates property (ii) in scenarios (a) and (b), and violates property (i) in scenario (c).

To shed light on the sufficient conditions in Theorem 3.5.2, let us consider a scenario similar to the James-Stein estimator: $\Sigma_{1,F} = \sigma_{1,F}^2 I_{d_\theta}$, $\Sigma_{2,F} = \sigma_{2,F}^2 I_{d_\theta}$, and $\Upsilon = I_{d_\theta}$. In this case, the sufficient conditions become $\sigma_{1,F} > \sigma_{2,F}$ and $d_\theta \geq 4$. The first condition $\text{tr}(A_F) > 0$, which is reduced to $\sigma_{1,F} > \sigma_{2,F}$, requires that the additional moments $\mathbb{E}_F[g^*(W_i, \theta_F)] = 0$ are non-redundant in the sense that they lead to a more efficient estimator of θ_F. The second condition $\text{tr}(A_F) \geq 4\rho_{\max}(A_F)$, which is reduced to $d_\theta \geq 4$, requires that we are interested in the total risk of several parameters rather than that of a single one. In a more general case where $\Sigma_{1,F}$ and $\Sigma_{2,F}$ are not proportional to the identity matrix, the sufficient conditions are reduced to $\Sigma_{1,F} > \Sigma_{2,F}$ and $d_\theta \geq 4$ under the choice $\Upsilon = (\Sigma_{1,F} - \Sigma_{2,F})^{-1}$, which rescales $\widehat{\theta}$ by the variance reduction $\Sigma_{1,F} - \Sigma_{2,F}$. In a simple linear IV model (Example 3.3.1) where Z_i^* is independent of $Z_{1,i}$ and the regression error U_i is homoskedastic conditional on the IVs, $\Sigma_{1,F} > \Sigma_{2,F}$ requires that $\mathbb{E}_{F^*}[Z_i^* X_i']$ and $\mathbb{E}_{F^*}[Z_i^* Z_i^{*\prime}]$ both have full rank.

Note that these conditions are sufficient but not necessary. If these sufficient conditions do not hold, we can still simulate the upper bounds in Theorem 3.5.1 to check the uniform dominance condition. In fact, simulation studies in the next session show that in many cases $\widehat{\theta}_{eo}$ has a smaller finite-sample risk than $\widehat{\theta}_1$ even if these sufficient conditions are violated. Nevertheless, these analytical sufficient conditions can be checked easily before the simulation-based methods are adopted.

3.6 Local Uniform Dominance

In the previous section, we provide a general framework for comparison of two GMM estimators. The moment conditions in this framework can be correctly specified (i.e. $\|\delta_F\| = 0$), locally misspecified (i.e. $\|\delta_F\|$ is close to zero), or severely misspecified ($\|\delta_F\| > C$). One may argue that the uniform dominance result in Theorem 3.5.2 is weak because in finite samples, the maximum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ may be larger than zero for all n although it converges to zero as n goes to infinity. A more favorable and stronger dominance result would be

$$Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) < 0, \tag{3.6.1}$$

which implies that the maximum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is bounded away from zero for all large n. This strong result is impossible if the set of DGPs $\mathcal{F}$ is large such that the DGPs under which the extra moment conditions become severely misspecified are included. On the other hand, (3.6.1) can be established if we only consider the DGPs under which the extra moment condition are correctly specified or locally misspecified.

Assumption 3.6.1. *For each n, $\mathcal{F}_n$ is a set of DGPs such that for any $F \in \mathcal{F}_n$:*

(i) $\mathbb{E}_F[g_1(W, \theta_F)] = 0_{r_1 \times 1}$ *for some $\theta_F \in int(\Theta)$ where Θ is a compact subset of $\mathbb{R}^{d_\theta}$;*

(ii) *for any $\varepsilon > 0$ there is $\eta_{1,\varepsilon} > 0$ such that* $\inf_{\theta \in B^c_\varepsilon(\theta_F)} ||\mathbb{E}_F[g_1(W, \theta)]|| \geq \eta_{1,\varepsilon}$*;*

(iii) $\mathbb{E}_F[g_2(W, \theta_F)] = n^{-1/2} d_F$ *for some $d_F \in \mathbb{R}^{r^*}$;*

(iv) *Assumption 3.3.2 holds for F;*

(v) $\overline{\Lambda} \equiv \{\overline{v}_F$*: $F \in \mathcal{F}_n$ for some $n\}$ is closed;*

(vi) $\|d_F\| \leq D$ *for some fixed constant D.*

Assumption 3.6.1.(i) and (ii) are the same as Assumptions 3.3.1.(i)-(ii), which ensures the unique identification of θ_F. Assumptions 3.3.1.(iii)-(iv) are not needed here as the aggressive GMM estimator $\widehat{\theta}_2$ is $n^{1/2}$-consistent for any $F \in \mathcal{F}_n$ under Assumption 3.6.1.(iii) (and other conditions in this assumption). Assumption 3.3.2 contains some regularity conditions for showing the asymptotic properties of the GMM estimator and it is maintained in Assumption 3.6.1.(iv). Assumption 3.6.1.(v) is a reduced version of Assumption 3.3.3.(i). Assumption 3.6.1.(vi) is an important condition to show the local uniform dominance result.

To introduce the local uniform dominance result, we define

$$H_D = \{h_{F,d} : d \in \mathbb{R}^{r^*} \text{ with } \|d\| \leq D \text{ and } F \in \mathcal{F}_n \text{ for some } n\}. \tag{3.6.2}$$

In the local misspecification framework, the set of DGPs $\mathcal{F}_n$ may change with the sample size n. The upper bound of the finite sample risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ should be defined as

$$\overline{RD}_n(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \sup_{F \in \mathcal{F}_n} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)], \tag{3.6.3}$$

which is approximated by

$$Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \lim_{\zeta \to \infty} \limsup_{n \to \infty} \sup_{F \in \mathcal{F}_n} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]. \tag{3.6.4}$$

To show the strong uniform dominance result, it is sufficient to study the upper bound of the risk difference $Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_{1}.)$

Lemma 3.6.1. *Suppose that Assumption 3.6.1 hold. The bounds of the asymptotic risk difference satisfy*

$$Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_{1}) \leq \sup_{h \in H_D} [g(h)], \tag{3.6.5}$$

where $g(h) = \mathbb{E}[\overline{\xi}_F' \Upsilon \overline{\xi}_F - \xi_{1,F}' \Upsilon \xi_{1,F}]$.

Lemma 3.6.1 provides a upper bound to the maximum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$. The criterion function $g(h)$ in (3.6.5) is the same as its counterpart in Theorem 3.5.1. The upper bound $Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_{1})$in (3.6.5) is not necessarily larger than zero, which is different from the results established in Theorem 3.5.1. To show the local uniform dominance result in (3.6.1), it is sufficient to show that $\sup_{h \in H_D}[g(h)]$ is bounded away from zero which is proved in the following Theorem.

Theorem 3.6.2. *Suppose that Assumption 3.6.1 hold. If* $\text{tr}(A_F) > 0$ *and* $\text{tr}(A_F) \geq 4\rho_{\max}(A_F)$ *for any* $F \in \mathcal{F}$*, we have* $\sup_{h \in H_D}[g(h)] < 0$ *for any finite constant* D*.*

Combining the results in Lemma 3.6.1 and Theorem 3.6.2, we immediately obtain (3.6.1). The sufficient conditions to ensure that the upper bound $\sup_{h \in H_D}[g(h)]$ is bounded away from zero are the same as those in Theorem 3.5.2.

3.7 Simulation Studies

In this section, we investigate the finite sample performance of our averaging GMM estimator in linear IV models. In addition to the empirical optimal weight $\widetilde{\omega}_{eo}$, we consider two other averaging estimators based on the JS type of weights. The first one is based on the positive part of the JS weight[14]:

$$\omega_{P,JS} = 1 - \left(1 - \frac{\operatorname{tr}(\widehat{A}) - 2\rho_{\max}(\widehat{A})}{n(\widehat{\theta}_2 - \widehat{\theta}_1)' H(\widehat{\theta}_2 - \widehat{\theta}_1)}\right)_+ \tag{3.7.1}$$

where $(x)_+ = \max\{0, x\}$ and $\widehat{A}$ is the estimator of A using $\widehat{\Sigma}_k$ for $k = 1, 2$. The second one uses the restricted JS weight

$$\omega_{R,JS} = (\omega_{P,JS})_+ . \tag{3.7.2}$$

By construction, $\omega_{P,JS} \leq 1$ and $0 \leq \omega_{R,JS} \leq 1$. We compare the finite-sample MSEs of these three averaging estimators, the conservative GMM estimator $\widehat{\theta}_1$, and the pre-test GMM estimator based on the J-test. The finite-sample MSE of the conservative GMM estimator is normalized to be 1.

In Theorem 3.5.2, the sufficient condition for the uniform dominance is $\operatorname{tr}(A) \geq 4\rho_{\max}(A)$. When this condition is not satisfied, however, it is still possible that our averaging GMM estimator has a smaller risk than the conservative GMM estimator. Therefore we consider two models in simulation studies. In the first model, $\operatorname{tr}(A) \geq 4\rho_{\max}(A)$ and hence the

[14]This formula is a GMM analog of the generalized JS type shrinkage estimator in Hansen (2016) for parametric models. The shrinkage scalar τ is set to $\operatorname{tr}(\widehat{A}) - 2\lambda_{\max}(\operatorname{tr}(\widehat{A}))$ in a fashion similar to the original JS estimator.

sufficient condition in Theorem 3.5.2 is satisfied. In the second model $2\rho_{\max}(A) < \mathrm{tr}(A) < 4\rho_{\max}(A)$, which means that the sufficient condition in Theorem 3.5.2 does not hold. In each model, we consider four sample sizes, $n = 50, 100, 250, 500$, and use 100,000 simulation repetitions.

3.7.1 Simulation in Model 1

Our first simulation model is

$$Y_i = \sum_{j=1}^{6} \theta_j X_{j,i} + \epsilon_i, \tag{3.7.3}$$

where $X_{j,i}$ are generated by

$$X_{j,i} = \beta_j (Z_{j,i} + Z_{j+6,i}) + Z_{j+12,i} + u_{j,i} \text{ for } j = 1, ..., 6. \tag{3.7.4}$$

We draw i.i.d. random vectors $(Z_{1,i}, ..., Z_{18,i}, u_{1,i}, ..., u_{6,i}, \epsilon_i)'$ from normal distributions with mean zero and variance-covariance matrix $\mathrm{diag}(I_{18\times18}, \Sigma_{7\times7})$, where

$$\Sigma_{7\times7} = \begin{pmatrix} I_{6\times6} & 0.25 \times \mathbf{1}_{6\times1} \\ 0.25 \times \mathbf{1}_{1\times6} & 1 \end{pmatrix}. \tag{3.7.5}$$

We set $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times6}$ and $(\beta_1, ..., \beta_6) = 0.5 \times \mathbf{1}_{1\times6}$. The observed data are $W_i = (Y_i, X_{1,i}, ..., X_{6,i}, Z_{1,i}, ..., Z_{12,i}, \tilde{Z}_{13,i}, ..., \tilde{Z}_{18,i})'$, where

$$\tilde{Z}_{j,i} = Z_{j,i} + n^{-1/2} d_j \epsilon_i, \text{ for } j = 13, ..., 18. \tag{3.7.6}$$

In the main regression equation (3.7.3), all regressors are endogenous because $\mathbb{E}(X_{j,i}\epsilon_i) = 0.25$ for $j = 1, ..., 6$. The instruments $(Z_{1,i}, ..., Z_{12,i})'$ are valid and $(\tilde{Z}_{13,i}, ..., \tilde{Z}_{18,i})'$ are misspecified because $\mathbb{E}(\tilde{Z}_{j,i}\epsilon_i) = n^{-1/2}d_j$ for $j = 13, ..., 18$. In the simulation studies, we consider $d = (d_{13}, ..., d_{18}) = c_0 \times \mathbf{1}_{1\times 6}$ where c_0 is a scalar that takes values on the grid points between 0 and 20 with the grid length 0.1.

Figure 3.3 presents the MSEs of various estimators of the parameters in (3.7.3). Our findings in model 1 are summarized as follows. First, the averaging GMM estimator $\widehat{\theta}_{eo}$ has smaller MSE than $\widehat{\theta}_1$ uniformly over d in all sample sizes considered, which is predicted by our theory because the key sufficient condition is satisfied in this model. Second, the pre-test GMM estimator does not dominate the conservative GMM estimator $\widehat{\theta}_1$. When the location parameter c_0 is close to zero, the pre-test GMM estimator has relative MSE as low as 0.4. However, its relative MSE is above 1 when c_0 is around 5. Third, among the three averaging estimators, the one based on $\widetilde{\omega}_{eo}$ has the smallest MSE. Fourth, the positive JS averaging estimator has relative MSE above 1 when the sample size is small, e.g., $n = 50$ and $n = 100$, and it has relative MSE below 1 when the sample size becomes slightly large. Fifth, it is interesting to see that as the sample size grows, the finite sample MSEs of the positive and restricted JS averaging estimators converge to that of the averaging estimator based on $\widetilde{\omega}_{eo}$.

3.7.2 Simulation in Model 2

The second model is

$$Y_i = \sum_{j=1}^{6} \theta_j X_{j,i} + \epsilon_i, \tag{3.7.7}$$

where $X_{1,i}$, $X_{2,i}$ and $X_{3,i}$ are exogenous variables generated by

$$X_{1,i} = 3^{-1/2}(Z_{1,i}+Z_{2,i}+Z_{4,i}),\ X_{2,i} = 3^{-1/2}(Z_{2,i}+Z_{3,i}+Z_{6,i}),\ X_{3,i} = 3^{-1/2}(Z_{3,i}+Z_{1,i}+Z_{8,i}), \tag{3.7.8}$$

and $X_{j,i}$ $(j = 4, 5, 6)$ are generated by

$$X_{j,i} = \beta_j(Z_{j,i} + Z_{j+3,i}) + Z_{j+6,i} + u_{j,i} \text{ for } j = 4, 5, 6. \tag{3.7.9}$$

We draw i.i.d. random vectors $(Z_{1,i}, ..., Z_{12,i}, u_{4,i}, ..., u_{6,i}, \epsilon_i)'$ from normal distributions with mean zero and variance-covariance matrix $\text{diag}(I_{12\times12}, \Sigma_{4\times4})$, where

$$\Sigma_{4\times4} = \begin{pmatrix} I_{3\times3} & 0.25 \times \mathbf{1}_{3\times1} \\ 0.25 \times \mathbf{1}_{1\times3} & 1 \end{pmatrix}. \tag{3.7.10}$$

The observed data are $W_i = (Y_i, X_{1,i}, ..., X_{6,i}, Z_{4,i}, ..., Z_{9,i}, \tilde{Z}_{4,i}, ..., \tilde{Z}_{6,i})'$, where

$$\tilde{Z}_{j,i} = Z_{j+6,i} + n^{-1/2} d_j \epsilon_i \text{ for } j = 4, 5, 6. \tag{3.7.11}$$

We set $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times6}$ and $(\beta_4, ..., \beta_6) = 0.5 \times \mathbf{1}_{1\times3}$. In this model, $X_{j,i}$ $(j = 4, 5, 6)$ are endogenous regressors, $(Z_{4,i}, ..., Z_{9,i})'$ are valid IVs, and $(\tilde{Z}_{4,i}, ..., \tilde{Z}_{6,i})'$ are misspecified IVs. In the simulation, we consider $d = (d_4, ..., d_6) = c_0 \times \mathbf{1}_{1\times3}$ where c_0 is a scalar that takes values on the grid points between 0 and 20 with grid length 0.1.

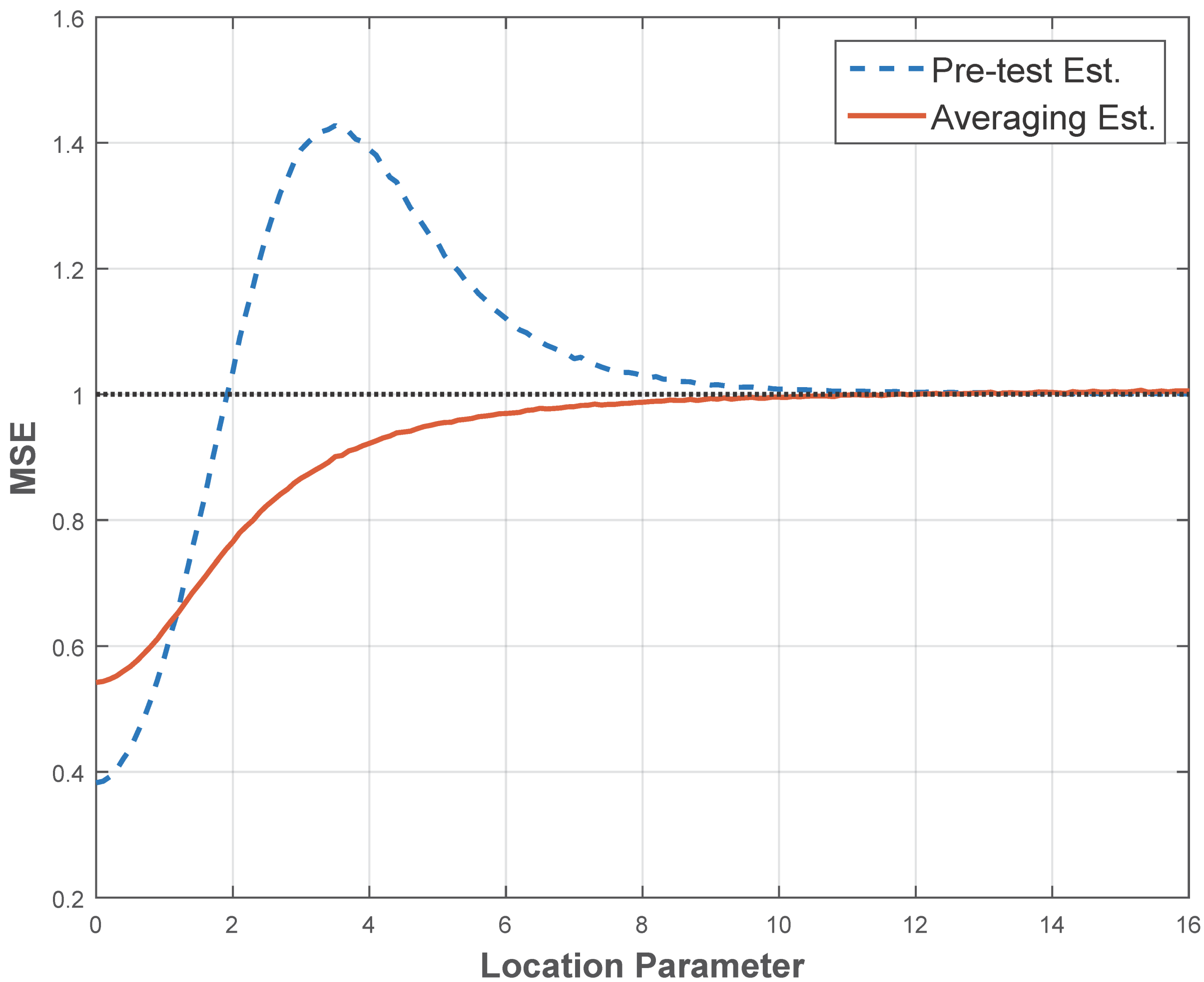
Our findings in Figure 3.4 are summarized as follows. First, even though the sufficient condition in Theorem 3.5.1(c) is not satisfied, the averaging estimator based on $\widetilde{\omega}_{eo}$ has a smaller MSE than $\widehat{\theta}_1$ uniformly over d in all sample sizes considered. Moreover, its MSE is much smaller than that of the other two averaging estimators. Second, the properties of the pre-test estimator are similar to those in model 1. That is, it does not dominate the conservative estimator. Third, the averaging estimator using $\omega_{P,JS}$ has very large and unstable MSE when the sample size is small (e.g., $n = 50$ and 100). When the sample size is 50, its MSE is above 1.5 and hence does not show up in the first panel of Figure 3. When the sample size becomes slightly large, (e.g., $n = 250$ and 500), the positive JS averaging estimator has larger MSE than $\widehat{\theta}_1$ when the location parameter c_0 is close to zero. Fourth, the averaging estimator using $\omega_{R,JS}$ has almost identical MSE as $\widehat{\theta}_1$ when the sample size is small (e.g., $n = 50$ and 100), and its relative MSE becomes slightly lower than 1 when c_0 is close to zero and the sample size is large (e.g., $n = 250$ and 500).

3.8 Conclusion

This paper studies the averaging GMM estimator that combines the conservative estimator and the aggressive estimator with a data-dependent weight. The averaging weight is the sample analog of an optimal non-random weight. We provide a sufficient class of drifting DGPs under which the pointwise asymptotic results combine to yield uniform approximations to the finite-sample risk difference between two estimators. Using this asymptotic approximation, we show that the proposed averaging GMM estimator uniformly dominates the conservative GMM estimator.

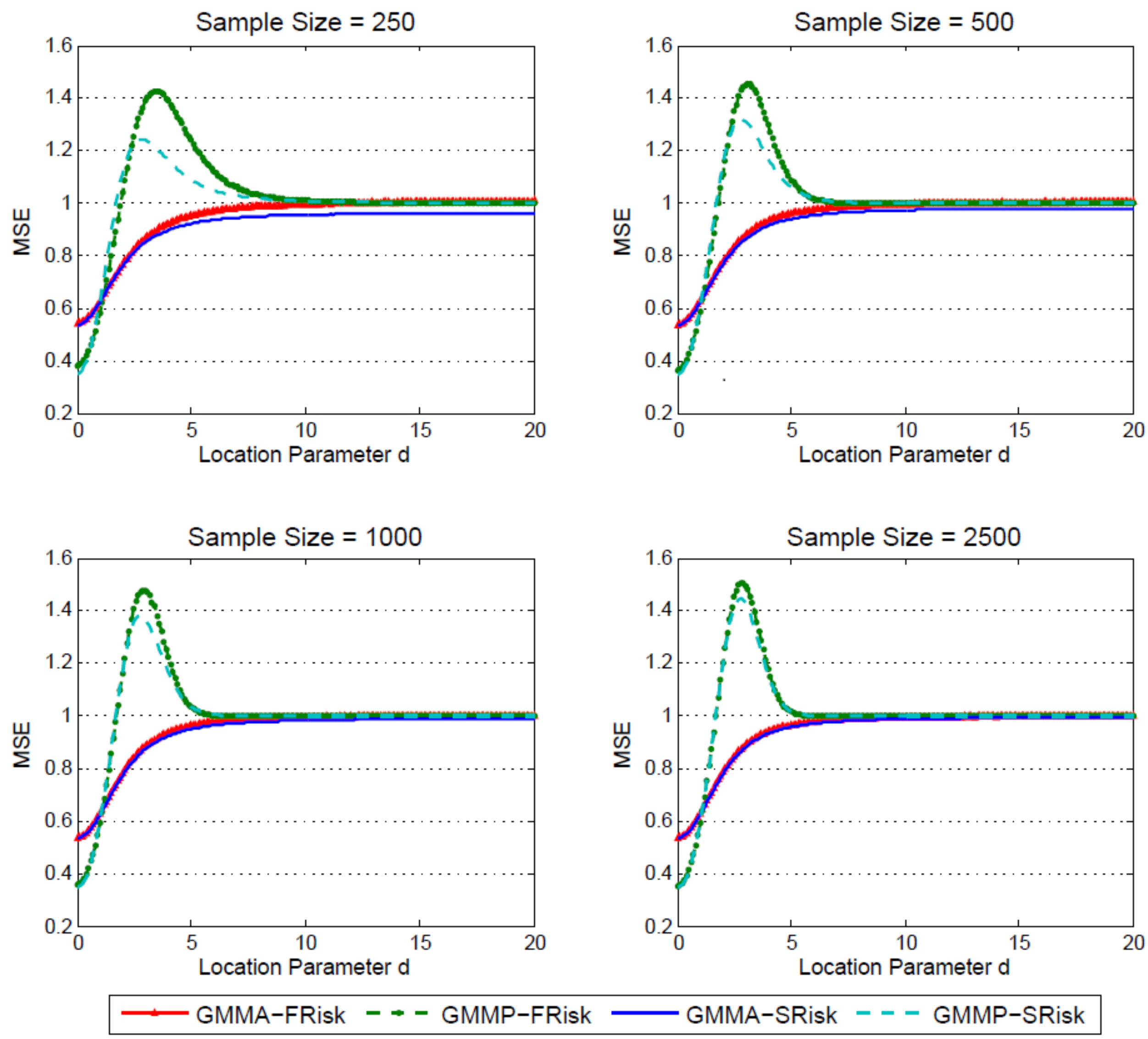
Inference based on the averaging estimator is an interesting and challenging problem. In addition to the uniform validity, a desirable confidence set should have smaller volume than that obtained from the conservative moments alone. We leave the inference issue to future investigation.

Figure 3.1: Finite Sample (n=250) MSEs of the Pre-test and the Averaging GMM Estimators



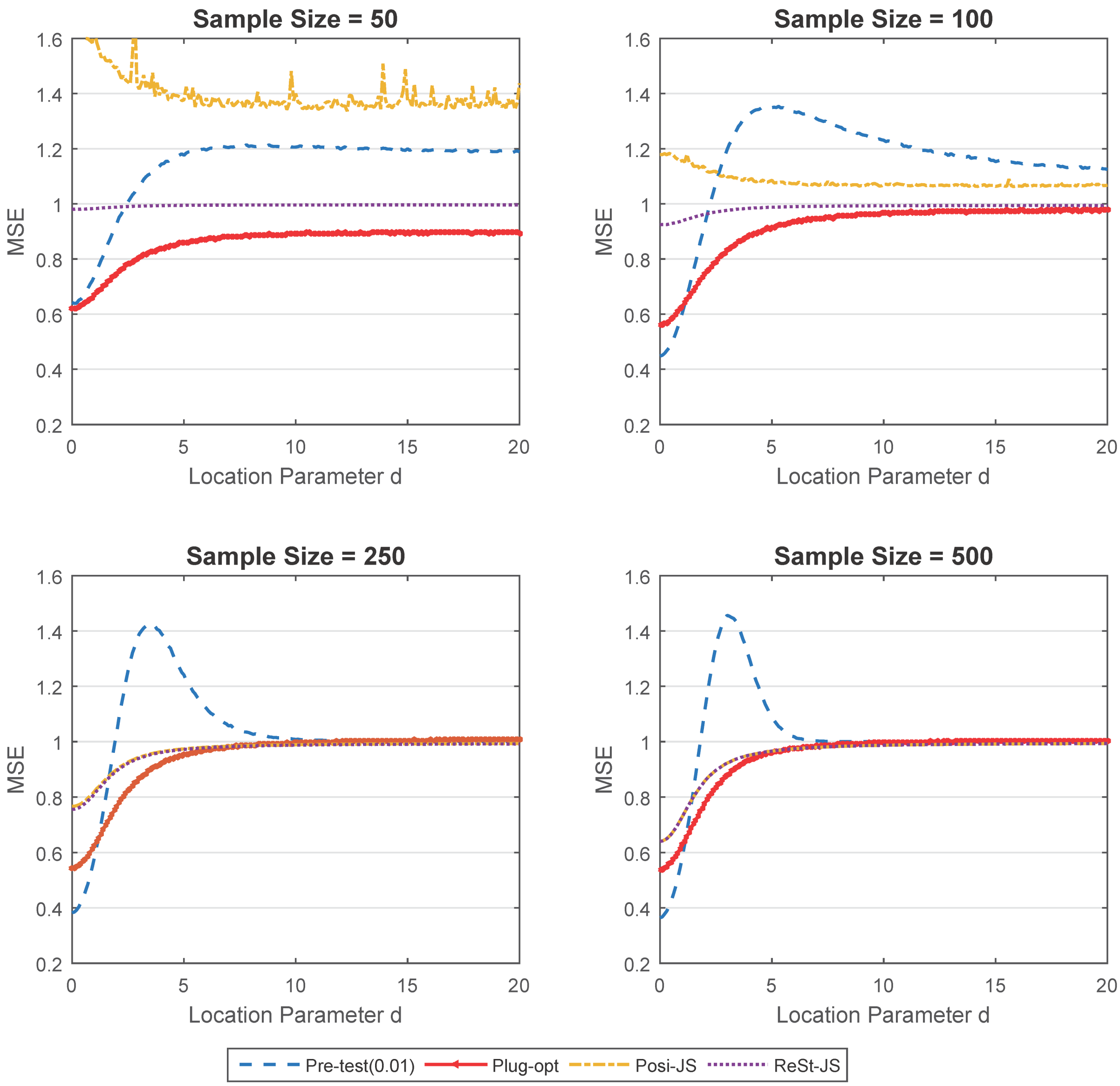
Note: The Pre-test estimator is based on the J-test with a norminal size 0.01. The location parameter in the horizontal axis is δ_F multiplied by $n^{-1/2}$. Details of the simulation design for Figure 3.1 is provided in Subsection 3.7.1.

Figure 3.2: The Finite Sample Risk and the (Simulated) Asymptotic Risk



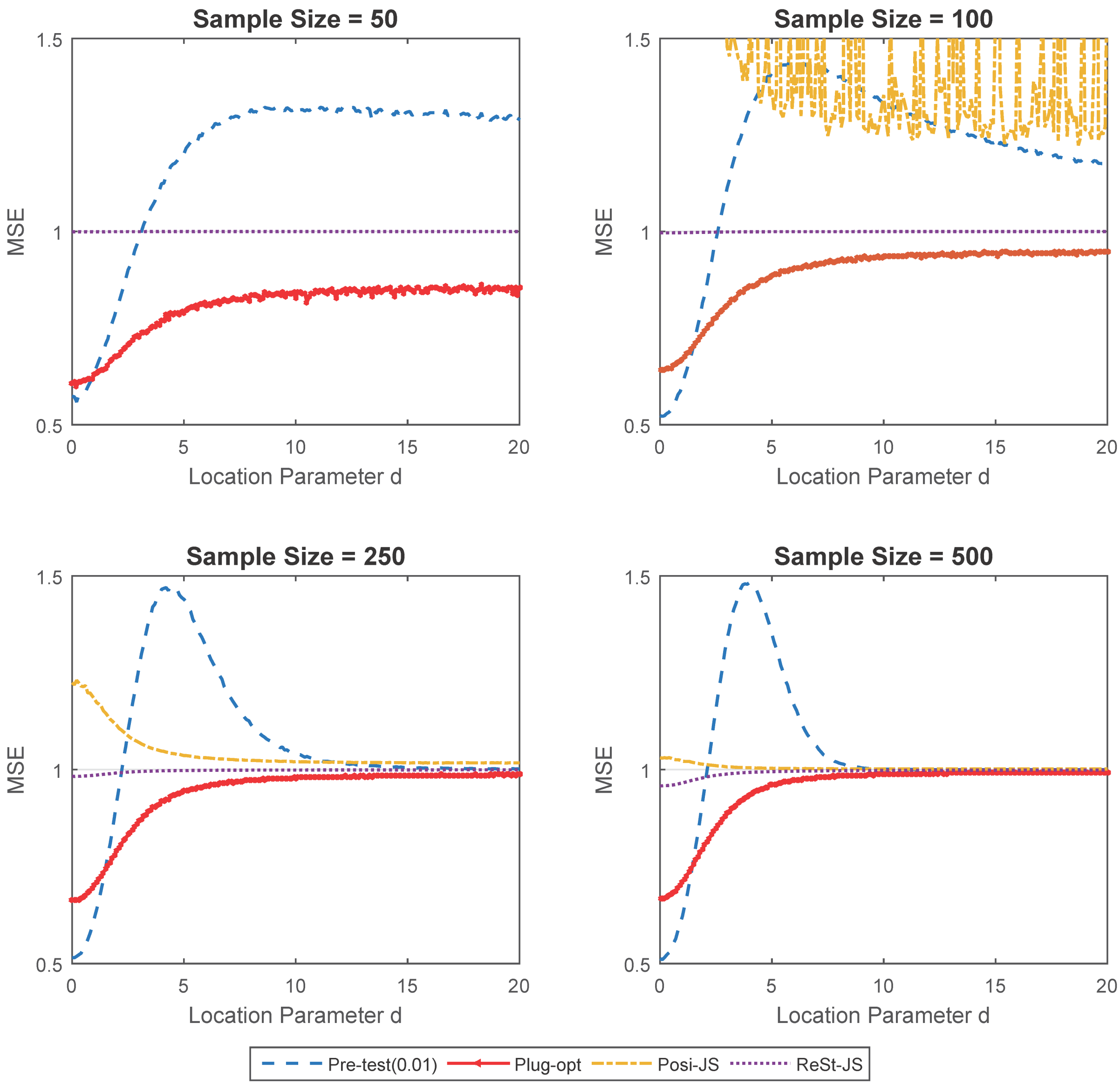
Note: The finite-sample and (simulated) asymptotic risk of the averaging GMM estimator are represented by "GMMA-FRisk" and "GMMA-SRisk", respectively. The finite-sample and (simulated) asymptotic risk of the pre-test GMM estimator are represented by "GMMP-FRisk" and "GMMP-SRisk", respectively.

Figure 3.3: Finite Sample MSEs of the Averaging Estimators in Model 1



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Plug-opt" refers to the averaging GMM estimator based on the empirical optimal weight ; "Posi-JS" and "ReSt-JS" refer to the averaging estimators based on the positive part of the JS weight and the restricted JS weight, respectively.

Figure 3.4: Finite Sample Risks of the Averaging Estimators in Model 2



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Plug-opt" refers to the averaging GMM estimator based on the empirical optimal weight ; "Posi-JS" and "ReSt-JS" refer to the averaging estimators based on the positive part of the JS weight and the restricted JS weight, respectively.

Appendix

3.A Proof of the Results in Section 3.3

Proof of Lemma 3.3.2. By Assumption *3.3.4.(i)* and the definition of $G_{1,F}$,

$$\mathbb{E}_F\left[g_1(W_i,\theta)\right] = \mathbb{E}_{F^*}[Z_{1,i}(U_i - X_i'(\theta-\theta_0))] = G_{1,F}(\theta-\theta_0), \tag{3.A.1}$$

which together with Assumption *3.3.4.(ii)* implies that $\theta_F = \theta_0$ and $\mathbb{E}_F\left[g_1(W_i,\theta_F)\right] = 0_{r_1\times 1}$. Also $\theta_F \in int(\Theta)$ holds by $\theta_F = \theta_0$ and Condition (i) of the Lemma. This verifies Assumption 3.3.1.(i). For any $\theta \in \Theta$ with $||\theta-\theta_F|| \geq \varepsilon$,

$$\left\|\mathbb{E}_F\left[g_1(W_i,\theta)\right]\right\| = \left\|G_{1,F}\left(\theta_F-\theta\right)\right\| \geq \left(\rho_{\min}(G_{1,F}'G_{1,F})\right)^{1/2}\left\|\theta_F-\theta\right\| \geq C^{-1/2}\varepsilon \tag{3.A.2}$$

where the first inequality is by the Cauchy-Schwarz inequality, the second inequality is by Assumption *3.3.4.(ii)*. This verifies Assumption 3.3.1.(ii). By the definitions of $G_{2,F}$ and $\Omega_{2,F}$, and Assumption *3.3.4.(iii)*,

$$\left\|G_{2,F}'G_{2,F}\right\| \leq C \text{ and } \left\|\Omega_{2,F}\right\| \leq C. \tag{3.A.3}$$

As $G_{2,F}' = \left(G_{1,F}', G_{r^*,F}'\right)$ where $G_{r^*,F} = -\delta_0 \mathbb{E}_{F^*}[U_i X_i'] - \mathbb{E}_{F^*}[V_i X_i']$, we have

$$G_{2,F}' G_{2,F} = G_{1,F}' G_{1,F} + G_{r^*,F}' G_{r^*,F}, \tag{3.A.4}$$

which implies that

$$\rho_{\min}(G_{2,F}' G_{2,F}) \geq \rho_{\min}(G_{1,F}' G_{1,F}). \tag{3.A.5}$$

To show Assumption 3.3.1.(iii), we write

$$\begin{aligned} Q_F(\theta) &= \mathbb{E}_F[Z_{2,i}(Y_i - X_i'\theta)]' \Omega_{2,F}^{-1} \mathbb{E}_F[Z_{2,i}(Y_i - X_i'\theta)] \\ &= \theta' G_{2,F}' \Omega_{2,F}^{-1} G_{2,F} \theta + 2\theta' G_{2,F}' \Omega_{2,F}^{-1} B_F + B_F' \Omega_{2,F}^{-1} B_F, \end{aligned} \tag{3.A.6}$$

where $B_F = \mathbb{E}_F\left[Z_{2,i} Y_i\right]$. It is clear that $Q_F(\theta)$ is minimized at θ_F^* as $G_{2,F}' \Omega_{2,F}^{-1} G_{2,F}$ is non-singular by (3.A.5) and Assumption 3.3.4.(ii). For any $\theta \in \Theta$ with $||\theta - \theta_F^*|| \geq \varepsilon$,

$$\begin{aligned} Q_F(\theta) - Q_F(\theta_F^*) &= (\theta - \theta_F^*)' G_{2,F}' \Omega_{2,F}^{-1} G_{2,F} (\theta - \theta_F^*) \\ &\geq \rho_{\min}(G_{2,F}' \Omega_{2,F}^{-1} G_{2,F}) \left\|\theta - \theta_F^*\right\|^2 \\ &\geq C^{-2} \left\|\theta - \theta_F^*\right\|^2 \geq C^{-2} \varepsilon^2, \end{aligned} \tag{3.A.7}$$

where the second inequality is by Assumption 3.3.4.(ii), (3.A.5) and (3.A.3). Assumption 3.3.1.(iv) is imposed in Assumption 3.3.4.(iv).

To verify Assumption 3.3.2, note that $g_2(W_i, \theta) = Z_{2,i}(U_i - X_i'(\theta - \theta_0))$, $g_{2,\theta}(W_i, \theta) = -Z_{2,i} X_i'$ and $g_{2,\theta\theta}(W_i, \theta) = 0_{(r_2 d_\theta) \times d_\theta}$. Assumption 3.3.2.(i) follows by Assumption 3.3.4.(iii) and the assumption that Θ is bounded. Assumptions 3.3.2(ii)-(iii) follow from Assumption 3.3.4.(ii).

We next verify 3.3.3. By definition,

$$v_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta_0) \,. \tag{3.A.8}$$

Hence Assumption 3.3.3.(i) holds by Assumption 3.3.4.(v). Consider any $F \in \mathcal{F}$ induced by δ_0, θ_0 and F^* with $\delta_F = 0_{r^* \times 1}$. Let ε_F in Assumption 3.3.3.(ii) to be $\varepsilon_F = C$ where C is specified in Condition (ii) of the lemma. For any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 \leq ||\widetilde{\delta}|| < \varepsilon_F$, we have $\widetilde{\delta} \in [-C, C]^{r^*}$. Let $\widetilde{F}$ be the distribution induced by $\widetilde{\delta}$, θ_0 and F^*. By definition,

$$G_{2,\widetilde{F}} = \begin{pmatrix} -\mathbb{E}_{F^*}[Z_{1,i}X_i'] \\ \\ -\widetilde{\delta}\mathbb{E}_{F^*}[U_iX_i'] - \mathbb{E}_{F^*}[V_iX_i'] \end{pmatrix} \text{ and } G_{2,F} = \begin{pmatrix} -\mathbb{E}_{F^*}[Z_{1,i}X_i'] \\ \\ -\mathbb{E}_{F^*}[V_iX_i'] \end{pmatrix} \tag{3.A.9}$$

which together with the Cauchy-Schwarz inequality, the Hölder inequality and Assumption 3.3.4.(iii) implies that

$$||G_{2,\widetilde{F}} - G_{2,F}|| = ||\widetilde{\delta}\mathbb{E}_{F^*}[U_iX_i']|| \leq C||\widetilde{\delta}||. \tag{3.A.10}$$

By definition $\theta_{\widetilde{F}} = \theta_0 = \theta_F$. Hence

$$\mathbb{E}_{\widetilde{F}}[Z_{1,i}Z_{1,i}'(Y_i - X_i'\theta_{\widetilde{F}})^2] = \mathbb{E}_{F^*}[Z_{1,i}Z_{1,i}'U_i^2] = \mathbb{E}_F[Z_{1,i}Z_{1,i}'(Y_i - X_i'\theta_F)^2]. \tag{3.A.11}$$

Under $\widetilde{F}$,

$$\mathbb{E}_{\widetilde{F}}[Z_{1,i}Z'_{r^*,i}(Y_i - X_i'\theta_{\widetilde{F}})^2] = \mathbb{E}_{F^*}[Z_{1,i}(U_i\widetilde{\delta} + V_i)'U_i^2]$$
$$= \mathbb{E}_{F^*}[U_i^3 Z_{1,i}]\widetilde{\delta}' + \mathbb{E}_{F^*}[U_i^2 Z_{1,i}V_i'], \tag{3.A.12}$$

and

$$\mathbb{E}_{\widetilde{F}}[Z_{r^*,i}Z'_{r^*,i}(Y_i - X_i'\theta_{\widetilde{F}})^2]$$
$$= \mathbb{E}_{F^*}[(U_i\widetilde{\delta} + V_i)(U_i\widetilde{\delta} + V_i)'U_i^2]$$
$$= \mathbb{E}_{F^*}[U_i^4]\widetilde{\delta}\widetilde{\delta}' + \widetilde{\delta}\mathbb{E}_{F^*}[U_i^3 V_i'] + \mathbb{E}_{F^*}[U_i^3 V_i]\widetilde{\delta}' + \mathbb{E}_{F^*}[U_i^2 V_i V_i']. \tag{3.A.13}$$

While under F,

$$\mathbb{E}_{F}[Z_{1,i}Z'_{r^*,i}(Y_i - X_i'\theta_F)^2] = \mathbb{E}_{F^*}[U_i^2 Z_{1,i}V_i'] \text{ and}$$
$$\mathbb{E}_{F}[Z_{r^*,i}Z'_{r^*,i}(Y_i - X_i'\theta_F)^2] = \mathbb{E}_{F^*}[U_i^2 V_i V_i']. \tag{3.A.14}$$

Collecting the results in (3.A.11), (3.A.12), (3.A.13) and (3.A.14), we get

$$\left\|\mathbb{E}_{\widetilde{F}}[Z_{2,i}Z'_{2,i}(Y_i - X_i'\theta_{\widetilde{F}})^2] - \mathbb{E}_{F}[Z_{2,i}Z'_{2,i}(Y_i - X_i'\theta_F)^2]\right\|$$
$$\leq \left\|\mathbb{E}_{F^*}[U_i^3 Z_{1,i}]\widetilde{\delta}'\right\| + \left\|\mathbb{E}_{F^*}[U_i^4]\widetilde{\delta}\widetilde{\delta}'\right\| + \left\|\widetilde{\delta}\mathbb{E}_{F^*}[U_i^3 V_i']\right\| + \left\|\mathbb{E}_{F^*}[U_i^3 V_i]\widetilde{\delta}'\right\|$$
$$\leq C(||\widetilde{\delta}||^2 + ||\widetilde{\delta}||) \leq C||\widetilde{\delta}||, \tag{3.A.15}$$

where the first inequality is by the triangle inequality, the second inequality is by the Cauchy-

Schwarz inequality, the Hölder inequality and Assumption 3.3.4.(iii), and the last inequality is by $||\widetilde{\delta}|| \leq C$. Next note that

$$\mathbb{E}_{\widetilde{F}}[Z_{2,i}(Y_i - X_i'\theta_{\widetilde{F}})] = \begin{pmatrix} \mathbb{E}_{F^*}[Z_{1,i}U_i] \\ \widetilde{\delta} \end{pmatrix} \text{ and } \mathbb{E}_F[Z_{2,i}(Y_i - X_i'\theta_F)] = \begin{pmatrix} \mathbb{E}_{F^*}[Z_{1,i}U_i] \\ 0_{r^*\times 1} \end{pmatrix} \tag{3.A.16}$$

which implies that

$$\begin{aligned} &\left\| \begin{array}{c} \mathbb{E}_{\widetilde{F}}[Z_{2,i}(Y_i - X_i'\theta_{\widetilde{F}})]\mathbb{E}_{\widetilde{F}}[Z_{2,i}'(Y_i - X_i'\theta_{\widetilde{F}})] \\ -\mathbb{E}_F[Z_{2,i}(Y_i - X_i'\theta_F)]\mathbb{E}_F[Z_{2,i}'(Y_i - X_i'\theta_F)] \end{array} \right\| \\ &= \left\| \begin{pmatrix} 0_{r_1\times r_1} & \mathbb{E}_{F^*}[Z_{1,i}U_i]\widetilde{\delta}' \\ \widetilde{\delta}\mathbb{E}_{F^*}[Z_{1,i}'U_i] & \widetilde{\delta}\widetilde{\delta}' \end{pmatrix} \right\| \\ &\leq \left\|\mathbb{E}_{F^*}[Z_{1,i}U_i]\widetilde{\delta}'\right\| + \left\|\widetilde{\delta}\mathbb{E}_{F^*}[Z_{1,i}'U_i]\right\| + ||\widetilde{\delta}||^2 \\ &\leq C(||\widetilde{\delta}||^2 + ||\widetilde{\delta}||) \leq C||\widetilde{\delta}||, \end{aligned} \tag{3.A.17}$$

where the first inequality is by the triangle inequality, the second inequality is by the Cauchy-Schwarz inequality, the Hölder inequality and Assumption 3.3.4.(iii), and the last inequality is by $||\widetilde{\delta}|| \leq C$. By the definition of $\Omega_{2,F}$ in (3.3.6), we can use the triangle inequality and the results in (3.A.15) and (3.A.17) to deduce that

$$\left\|\Omega_{2,\widetilde{F}} - \Omega_{2,F}\right\| \leq C||\widetilde{\delta}||, \tag{3.A.18}$$

which combined with (3.A.10) implies that $||\overline{v}_F - \overline{v}_{\widetilde{F}}|| \leq C||\widetilde{\delta}||$. This verifies Assumption

3.3.3.(ii) with $\kappa = 1$. □

3.B Proof of the Results in Section 3.4

Let $\mu_n(g_2(W,\theta)) = n^{-1/2}\sum_{i=1}^{n}(g_2(W_i,\theta) - \mathbb{E}_{F_n}[g_2(W_i,\theta)])$.

Lemma 3.B.1. *Suppose that Assumption 3.3.2.(i) holds and Θ is compact. Then we have*

(i) $\sup_{\theta\in\Theta}\|\overline{g}_2(\theta) - \mathbb{E}_{F_n}[g_2(W_i,\theta)]\| = o_p(1)$;

(ii) $\sup_{\theta\in\Theta}\|n^{-1}\sum_{i=1}^{n} g_2(W_i,\theta)g_2(W_i,\theta)' - \mathbb{E}_{F_n}[g_2(W_i,\theta)g_2(W_i,\theta)']\| = o_p(1)$;

(iii) $\sup_{\theta\in\Theta}\|n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\theta) - \mathbb{E}_{F_n}[g_{2,\theta}(W_i,\theta)]\| = o_p(1)$;

(iv) $\mu_n(g_2(W,\theta))$ *is stochastic equicontinuous over $\theta \in \Theta$;*

(v) $\Omega_{2,F_n}^{-1/2}\mu_n(g_2(W,\theta_{F_n})) \to_d N(0_{r_2\times 1}, I_{r_2})$.

Proof of Lemma 3.B.1. See Lemma 11.3-11.5 of Andrews and Cheng (2013). □

Define $M_{2,F}(\theta) = \mathbb{E}_F\left[g_2(W,\theta)\right]$, $G_{2,F}(\theta) = \mathbb{E}_F\left[g_{2,\theta}(W,\theta)\right]$ and $\Omega_{2,F}(\theta) = \mathrm{Var}_F\left[g_2(W,\theta)\right]$ for any $\theta \in \Theta$. The next lemma shows that $M_{2,F}(\cdot)$, $G_{2,F}(\cdot)$ and $\Omega_{2,F}(\cdot)$ are Lipschitz continuous uniformly over $F \in \mathcal{F}$.

Lemma 3.B.2. *Under Assumption 3.3.2.(i), for any $F \in \mathcal{F}$ and any $\theta_1, \theta_2 \in \Theta$, we have:*

(i) $\|M_{2,F}(\theta_1) - M_{2,F}(\theta_2)\| \leq C\|\theta_1 - \theta_2\|$;

(ii) $\|G_{2,F}(\theta_1) - G_{2,F}(\theta_2)\| \leq C\|\theta_1 - \theta_2\|$;

(iii) $\|\Omega_{2,F}(\theta_1) - \Omega_{2,F}(\theta_2)\| \leq C\|\theta_1 - \theta_2\|$.

Proof of Lemma 3.B.2. (i) Let $g_{2,j}(w,\theta)$ denote the j-th ($j = 1,\ldots,r_2$) component of $g_2(w,\theta)$. By the mean value expansion,

$$g_{2,j}(w,\theta_1) - g_{2,j}(w,\theta_2) = g_{2,j,\theta}(w,\widetilde{\theta}_{1,2})(\theta_1 - \theta_2) \tag{3.B.1}$$

for any $j = 1,\ldots,r_2$, where $\widetilde{\theta}_{1,2}$ is some vector between θ_1 and θ_2. By (3.B.1) and the Cauchy-Schwarz inequality

$$\left|\mathbb{E}_F\left[g_{2,j}(w,\theta_1) - g_{2,j}(w,\theta_2)\right]\right| \leq \mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1 - \theta_2\|, \tag{3.B.2}$$

for any $j = 1,\ldots,r_2$. By Assumption 3.3.2.(i), (3.B.2) and the triangle inequality, we deduce that

$$\|M_{2,F}(\theta_1) - M_{2,F}(\theta_2)\| \leq \sqrt{r_2}\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1 - \theta_2\| \leq C\sqrt{r_2}\|\theta_1 - \theta_2\| \tag{3.B.3}$$

which immediately proves the claim in (i). The claim in (ii) follows by similar argument and its proof is omitted.

(iii) By the mean value expansion,

$$\begin{aligned} & g_{2,j_1}(w,\theta_1)g_{2,j_2}(w,\theta_1) - g_{2,j_1}(w,\theta_2)g_{2,j_2}(w,\theta_2) \\ &= \left[g_{2,j_1,\theta}(w,\widetilde{\theta}_{1,2})g_{2,j_2}(w,\widetilde{\theta}_{1,2}) + g_{2,j_1}(w,\widetilde{\theta}_{1,2})g_{2,j_2,\theta}(w,\widetilde{\theta}_{1,2})\right](\theta_1 - \theta_2) \end{aligned} \tag{3.B.4}$$

for any $j_1, j_2 = 1,\ldots,r_2$, where $\widetilde{\theta}_{1,2}$ is some vector between θ_1 and θ_2 and may take different values from the $\widetilde{\theta}_{1,2}$ in (3.B.1). By (3.B.4), the triangle inequality and the Cauchy-Schwarz inequality

$$\begin{aligned}
&|\mathbb{E}_F\left[g_{2,j_1}(w,\theta_1)g_{2,j_2}(w,\theta_1) - g_{2,j_1}(w,\theta_2)g_{2,j_2}(w,\theta_2)\right]| \\
&\leq 2\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_2(W,\theta)\|\,\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1-\theta_2\| \\
&\leq \mathbb{E}_F\left[\sup_{\theta\in\Theta}(\|g_2(W,\theta)\|^2 + \|g_{2,\theta}(W,\theta)\|^2)\right]\|\theta_1-\theta_2\| \qquad (3.B.5)
\end{aligned}$$

for any $j_1, j_2 = 1, \ldots, r_2$, where the second inequality is by the simple inequality that $|ab| \leq (a^2+b^2)/2$. By (3.B.5) and Assumption 3.3.2.(i),

$$\begin{aligned}
&\|\mathbb{E}_F\left[g_2(W,\theta_1)g_2(W,\theta_1)' - g_2(W,\theta_2)g_2(W,\theta_2)'\right]\| \\
&\leq r_2\mathbb{E}_F\left[\sup_{\theta\in\Theta}(\|g_2(W,\theta)\|^2 + \|g_{2,\theta}(W,\theta)\|^2)\right]\|\theta_1-\theta_2\| \\
&\leq r_2 C\|\theta_1-\theta_2\|. \qquad (3.B.6)
\end{aligned}$$

Using the triangle inequality, and the inequality in (3.B.2), we deduce that

$$\begin{aligned}
&|\mathbb{E}_F[g_{2,j_1}(w,\theta_1)]\mathbb{E}_F[g_{2,j_2}(w,\theta_1)] - \mathbb{E}_F[g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_2)]| \\
&\leq |\mathbb{E}_F[g_{2,j_1}(w,\theta_1) - g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_1)]| \\
&+ |\mathbb{E}_F[g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_2) - g_{2,j_2}(w,\theta_1)]| \\
&\leq 2\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_2(W,\theta)\|\right]\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1-\theta_2\| \qquad (3.B.7)
\end{aligned}$$

for any $j_1, j_2 = 1, \ldots, r_2$. By (3.B.7) and Assumption 3.3.2.(i),

$$\|\mathbb{E}_F[g_2(w,\theta_1)]\mathbb{E}_F[g_2(w,\theta_1)'] - \mathbb{E}_F[g_2(w,\theta_2)]\mathbb{E}_F[g_2(w,\theta_2)']\| \leq 2r_2C^2 \|\theta_1 - \theta_2\|. \quad (3.B.8)$$

By the definition of $\Omega_{2,F}(\theta)$, the triangle inequality and the results in (3.B.6) and (3.B.8)

$$\|\Omega_{2,F}(\theta_1) - \Omega_{2,F}(\theta_2)\| \leq r_2(C + 2C^2) \|\theta_1 - \theta_2\|, \quad (3.B.9)$$

which immediately proves the claim in (iii). □

Lemma 3.B.3. *Suppose that Assumptions 3.3.1.(i), 3.3.1.(ii) and 3.3.2.(i) hold. Then for any sequence of DGPs $\{F_n\}$, we have*

$$\widetilde{\theta}_1 - \theta_{F_n} = o(1) \text{ and } \overline{\Omega}_2 = \Omega_{2,F_n} + o_p(1), \quad (3.B.10)$$

where $\widetilde{\theta}_1$ is a preliminary estimator defined as

$$\widetilde{\theta}_1 = \arg\min_{\theta \in \Theta} \overline{g}_1(\theta)'\overline{g}_1(\theta). \quad (3.B.11)$$

Proof of Lemma 3.B.3. By Lemma 3.B.1.(i),

$$\overline{g}_2(\theta) = M_{2,F_n}(\theta) + \left[n^{-1}\sum_{i=1}^{n} g_2(W_i,\theta) - M_{2,F_n}(\theta)\right] = M_{2,F_n}(\theta) + o_p(1), \quad (3.B.12)$$

uniformly over $\theta \in \Theta$. As $g_1(W,\theta)$ is a subvector of $g_2(W,\theta)$, by (3.B.12) and Assumption

3.3.2.(i),

$$\overline{g}_1(\theta)'\overline{g}_1(\theta) = M_{1,F_n}(\theta)'M_{1,F_n}(\theta) + o_p(1) \tag{3.B.13}$$

uniformly over $\theta \in \Theta$. By Assumptions 3.3.1.(i)-(ii) and $F_n \in \mathcal{F}$, $M_{1,F_n}(\theta)'M_{1,F_n}(\theta)$ is uniquely minimized at θ_{F_n}, which together with the uniform convergence in (3.B.13) implies that

$$\widetilde{\theta}_1 - \theta_{F_n} \to_p 0. \tag{3.B.14}$$

To show the consistency of $\overline{\Omega}_2$, note that

$$\begin{aligned}
\overline{\Omega}_2 &= n^{-1}\sum_{i=1}^{n} g_2(W_i,\widetilde{\theta}_1)g_2(W_i,\widetilde{\theta}_1)' - \overline{g}_2(\widetilde{\theta}_1)\overline{g}_2(\widetilde{\theta}_1)' \\
&= \mathbb{E}_{F_n}[g_2(W,\widetilde{\theta}_1)g_2(W,\widetilde{\theta}_1)'] - M_{2,F_n}(\widetilde{\theta}_1)'M_{2,F_n}(\widetilde{\theta}_1) + o_p(1) \\
&= \Omega_{2,F_n}(\widetilde{\theta}_1) + o_p(1) = \Omega_{2,F_n} + o_p(1),
\end{aligned} \tag{3.B.15}$$

where the first equality is by the definition of $\overline{\Omega}_2$, the second equality holds by (3.B.12), Lemma 3.B.1.(ii) and Assumption 3.3.2.(i), the third equality follows from the definition of $\Omega_{2,F_n}(\theta)$, and the last equality holds by Lemma 3.B.2.(iii) and (3.B.14). This shows the consistency of $\overline{\Omega}_2$. □

Lemma 3.B.4. *Suppose that Assumptions 3.3.1.(i)-(ii) and 3.3.2 hold. Then for any sequence of DGPs $\{F_n\}$, we have*

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = \Gamma_{1,F_n}\mu_n(g_1(W,\theta_{F_n})) + o_p(1), \tag{3.B.16}$$

where $\Gamma_{1,F_n}\mu_n(g_1(W,\theta_{F_n})) = O_p(1)$.

Proof of Lemma 3.B.4. As $\overline{g}_1(\theta)$ is a subvector of $\overline{g}_2(\theta)$, and $\overline{\Omega}_{1,n}$ is a submatrix of $\overline{\Omega}_{2,n}$, using (3.B.12), (3.B.15) and Assumptions 3.3.2.(i)-(ii), we have

$$\overline{g}_1(\theta)'(\overline{\Omega}_1)^{-1}\overline{g}_1(\theta) = M_{1,F_n}(\theta)'\Omega_{1,F_n}^{-1}M_{1,F_n}(\theta) + o_p(1), \tag{3.B.17}$$

uniformly over Θ. By Assumptions 3.3.2.(i)-(ii),

$$C^{-1} \leq \rho_{\min}(\Omega_{1,F_n}^{-1}) \leq \rho_{\max}(\Omega_{1,F_n}^{-1}) \leq C \tag{3.B.18}$$

which together with Assumptions 3.3.1.(i)-(ii) implies that $M_{1,F_n}(\theta)'\Omega_{1,F_n}^{-1}M_{1,F_n}(\theta)$ is uniquely minimized at θ_{F_n}. By the standard arguments for the consistency of an extremum estimator, we have

$$\widehat{\theta}_1 - \theta_{F_n} = o_p(1). \tag{3.B.19}$$

Using (3.B.19), Lemma 3.B.1.(iv) and Assumption 3.3.2.(i), we have

$$\begin{aligned}\overline{g}_1(\widehat{\theta}_1) &= \overline{g}_1(\theta_{F_n}) + \left[M_{1,F_n}(\widehat{\theta}_1) - M_{1,F_n}(\theta_{F_n})\right] + o_p(n^{-1/2}) \\ &= \overline{g}_1(\theta_{F_n}) + [G_{1,F_n}(\theta_{F_n}) + o_p(1)]\,(\widehat{\theta}_1 - \theta_{F_n}) + o_p(n^{-1/2}).\end{aligned} \tag{3.B.20}$$

Similarly,

$$n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i,\widehat{\theta}_1) = G_{1,F_n}(\widehat{\theta}_1) + o_p(1) = G_{1,F_n} + o_p(1), \tag{3.B.21}$$

where the first equality follows from Lemma 3.B.1.(iii) and the second equality follows by

(3.B.19) and Lemma 3.B.2.(ii). From the first order condition for the GMM estimator $\widehat{\theta}_1$, we deduce that

$$0 = \left[n^{-1} \sum_{i=1}^{n} g_{1,\theta}(W_i, \widehat{\theta}_1) \right]' (\overline{\Omega}_1)^{-1} \overline{g}_1(\widehat{\theta}_1)$$
$$= (G'_{1,F_n} \Omega^{-1}_{1,F_n} + o_p(1)) \left[\overline{g}_1(\theta_{F_n}) + (G_{1,F_n} + o_p(1))(\widehat{\theta}_1 - \theta_{F_n}) + o_p(n^{-1/2}) \right] \tag{3.B.22}$$

where the second equality follows from Assumptions 3.3.2.(i)-(ii), (3.B.15), (3.B.20) and (3.B.21). By (3.B.22), $\mathbb{E}_{F_n}[g_1(W, \theta_{F_n})] = 0$ and Assumption 3.3.2,

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = (\Gamma_{1,F_n} + o_p(1))\mu_n(g_1(W, \theta_{F_n})) + o_p(1). \tag{3.B.23}$$

By Assumptions 3.3.2 and Lemma 3.B.1.(v), $\Gamma_{1,F_n} = O(1)$ and $\mu_n(g_1(W, \theta_{F_n})) = o_p(1)$, which together with (3.B.23) implies that

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = \Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) + o_p(1),$$

where $\Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) = O_p(1)$. This finishes the proof. □

Lemma 3.B.5. *Suppose that Assumptions 3.3.1.(iii) and 3.3.2.(i)-(ii) hold. Then for any sequence of DGPs $\{F_n\}$, we have*

$$\widehat{\theta}_2 - \theta^*_{F_n} = o_p(1), \tag{3.B.24}$$

where $\theta^*_{F_n}$ denotes the minimizer of $Q_{F_n}(\theta) \equiv M_{2,F_n}(\theta)'\Omega^{-1}_{2,F_n} M_{2,F_n}(\theta)$.

Proof of Lemma 3.B.5. By (3.B.12), (3.B.15) and Assumptions 3.3.2.(i)-(ii), we have

$$\overline{g}_2(\theta)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta) = M_{2,F_n}(\theta)'\Omega^{-1}_{2,F_n} M_{2,F_n}(\theta) + o_p(1) = Q_{F_n}(\theta) + o_p(1) \tag{3.B.25}$$

uniformly over Θ. By Assumption 3.3.1.(iii), $Q_{F_n}(\theta)$ is uniquely minimized at $\theta^*_{F_n}$. The consistency result $\widehat{\theta}_2 - \theta^*_{F_n} \rightarrow_p 0$ follows from standard arguments for the consistency of an extremum estimator. □

Lemma 3.B.6. *Suppose that Assumptions 3.3.1.(i)-(ii) and 3.3.2.(i)-(ii) hold. Consider any sequence of DGPs $\{F_n\}$ such that $\delta_{F_n} = o(1)$. Then we have*

$$\widehat{\theta}_2 - \theta_{F_n} = o_p(1). \tag{3.B.26}$$

If we further have Assumption 3.3.2.(iii), then

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1))\left\{\mu_n(g_2(W, \theta_{F_n})) + n^{1/2}\delta_{0,F_n}\right\} + o_p(1), \tag{3.B.27}$$

where $\Gamma_{2,F_n} = -\left(G'_{2,F_n}\Omega^{-1}_{2,F_n} G_{2,F_n}\right)^{-1} G'_{2,F_n}\Omega^{-1}_{2,F_n}$ and $\delta_{0,F_n} = (\mathbf{0}_{r_1\times 1}, \delta'_{F_n})'$.

Proof of Lemma 3.B.6. By the definition of $\widehat{\theta}_2$,

$$\overline{g}_2(\widehat{\theta}_2)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \leq \overline{g}_2(\theta_{F_n})'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta_{F_n}), \tag{3.B.28}$$

which implies that

$$||\overline{g}_2(\widehat{\theta}_2)||^2 \leq \rho_{\max}(\overline{\Omega}_2)\rho_{\min}^{-1}(\overline{\Omega}_2) \left\|\overline{g}_2(\theta_{F_n})\right\|^2. \tag{3.B.29}$$

By (3.B.15) and Assumptions 3.3.2.(i)-(ii),

$$C^{-1} \leq \rho_{\min}(\overline{\Omega}_2) \leq \rho_{\max}(\overline{\Omega}_2) \leq C \tag{3.B.30}$$

with probability approaching 1. By Lemma 3.B.1.(i), $M_{1,F_n}(\theta_{F_n}) = 0_{r_1\times 1}$ and $\delta_{F_n} = o(1)$,

$$\left\|\overline{g}_2(\theta_{F_n})\right\|^2 = o_p(1) \tag{3.B.31}$$

which combined with (3.B.29) and (3.B.30) implies that

$$||\overline{g}_2(\widehat{\theta}_2)|| = o_p(1). \tag{3.B.32}$$

Moreover, by (3.B.32), Lemma 3.B.1.(i) and the triangle inequality,

$$||M_{2,F_n}(\widehat{\theta}_2)|| \leq ||\overline{g}_2(\widehat{\theta}_2) - M_{2,F_n}(\widehat{\theta}_2)|| + ||\overline{g}_2(\widehat{\theta}_2)|| = o_p(1) \tag{3.B.33}$$

which immediately implies that

$$||M_{1,F_n}(\widehat{\theta}_2)|| = o_p(1). \tag{3.B.34}$$

The claimed result follows by (3.B.34) and the unique identification of θ_{F_n} maintained by

Assumptions 3.3.1.(i)-(ii).

Using $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$, Lemma 3.B.1.(iv) and Assumption 3.3.2.(i), we have

$$\begin{aligned}\overline{g}_2(\widehat{\theta}_2) &= \overline{g}_2(\theta_{F_n}) + \left[M_{2,F_n}(\widehat{\theta}_2) - M_{2,F_n}(\theta_{F_n})\right] + o_p(n^{-1/2}) \\ &= \overline{g}_2(\theta_{F_n}) + \left[G_{2,F_n}(\theta_{F_n}) + o_p(1)\right](\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2}). \end{aligned} \tag{3.B.35}$$

Similarly,

$$n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i, \widehat{\theta}_2) = G_{2,F_n}(\widehat{\theta}_2) + o_p(1) = G_{2,F_n}(\theta_{F_n}) + o_p(1), \tag{3.B.36}$$

where the first equality follows from Lemma 3.B.1.(iii) and the second equality follows by $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$ and Lemma 3.B.2.(ii). From the first order condition for the GMM estimator $\widehat{\theta}_2$, we deduce that

$$\begin{aligned} 0 &= \left[n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i, \widehat{\theta}_2)\right]' (\overline{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \\ &= (G'_{2,F_n}\Omega^{-1}_{2,F_n} + o_p(1))\left[\overline{g}_2(\theta_{F_n}) + (G_{2,F_n} + o_p(1))(\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2})\right] \end{aligned} \tag{3.B.37}$$

where the second equality follows from Assumptions 3.3.2.(i)-(ii), (3.B.15), (3.B.35) and (3.B.36). By (3.B.37) and Assumption 3.3.2,

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1))\left\{\mu_n(g_2(W, \theta_{F_n})) + n^{1/2}\mathbb{E}_{F_n}\left[g_2(W, \theta_{F_n})\right]\right\} + o_p(1), \tag{3.B.38}$$

where $\Gamma_{2,F_n} = -\left(G'_{2,F_n}\Omega^{-1}_{2,F_n}G_{2,F_n}\right)^{-1} G'_{2,F_n}\Omega^{-1}_{2,F_n}$. □

Lemma 3.B.7. *Under Assumptions 3.3.2.(i) and 3.3.3.(i), for any sequence of DGPs* $\{F_{p_n}\}$ *with* $F_{p_n} \in \mathcal{F}$ *where* $\{p_n\}$ *is a subsequence of* $\{n\}$*, there is a subsequence* $\{p_n^*\}$ *of* $\{p_n\}$ *such that* $v_{F_{p_n^*}}(\theta_{F_{p_n^*}}) \to v_F(\theta_F)$ *as* $p_n^* \to \infty$*, where* $F \in \mathcal{F}$.

Proof of Lemma 3.B.7. Recall that $\Lambda = \{v_F : \ F \in \mathcal{F}\}$. By Assumptions 3.3.2.(i) and 3.3.3.(i), Λ is compact. Hence for any sequence $\left\{v_{F_{p_n}}(\theta_{F_{p_n}})\right\}$ in Λ, it has a convergent subsequence $\{v_{F_{p_n^*}}(\theta_{F_{p_n^*}})\}$ such that $v_{F_{p_n^*}}(\theta_{F_{p_n^*}}) \to v_F(\theta_F)$ as $p_n^* \to \infty$, where $F \in \mathcal{F}$. $\square$

Lemma 3.B.8. *Suppose that Assumptions 3.3.1.(i)-(ii) and 3.3.2 hold. Consider any sequence of DGPs* $\{F_n\}$ *such that* $\overline{v}_{F_n} \to \overline{v}_F$ *for some* $F \in \mathcal{F}$*, and* $n^{1/2}\delta_{F_n} \to d$ *for* $d \in \mathbb{R}^{r^*}$. *Then*

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_d \begin{pmatrix} \xi_{1,F} \\ \xi_{2,F} \end{pmatrix} \equiv \begin{pmatrix} \Gamma_{1,F}\mathcal{Z}_{1,F} \\ \Gamma_{2,F}\left(\mathcal{Z}_{2,F} + d_0\right) \end{pmatrix},$$

where $d_0 = (\mathbf{0}_{1\times r_1}, d')'$.

Proof of Lemma 3.B.8. In the proof, we use

$$G_{2,F_n} \to G_{2,F} \text{ and } \Omega_{2,F_n} \to \Omega_{2,F} \tag{3.B.39}$$

for some $F \in \mathcal{F}$, which is assumed in the lemma. Under Assumptions 3.3.1.(i)-(ii) and 3.3.2, for the sequence of DGPs $\{F_n\}$ considered in the lemma, we can apply Lemma 3.B.4 and Lemma 3.B.6 to deduce that

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} = \begin{pmatrix} \Gamma_{1,F_n}\mu_n(g_1(W,\theta_{F_n})) \\ (\Gamma_{2,F_n} + o_p(1))\left\{\mu_n(g_2(W,\theta_{F_n})) + n^{1/2}\delta_{0,F_n}\right\} \end{pmatrix} + o_p(1), \tag{3.B.40}$$

where $\delta_{0,F_n} = (\mathbf{0}_{r_1\times 1}, \delta'_{F_n})'$. By (3.B.39) and Assumption 3.3.2, we have

$$\Gamma_{1,F_n} = \Gamma_{1,F} + o(1) \text{ and } \Gamma_{2,F_n} = \Gamma_{2,F} + o(1) \tag{3.B.41}$$

where $\Gamma_{k,F} = -\left(G'_{k,F}\Omega^{-1}_{k,F}G_{k,F}\right)^{-1} G'_{k,F}\Omega^{-1}_{k,F}$ for $k = 1, 2$. Collecting the results in Lemma 3.B.1.(v), (3.B.40) and (3.B.41), and then applying the continuous mapping theorem (CMT), we have

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_d \begin{pmatrix} \Gamma^*_{1,F} \\ \Gamma_{2,F} \end{pmatrix} (\mathcal{Z}_{2,F} + d_0)\,, \tag{3.B.42}$$

where $\mathcal{Z}_{2,F} \sim N(\mathbf{0}_{r_2\times 1}, \Omega_{2,F})$, $\Gamma^*_{1,F} = (\Gamma_{1,F}, \mathbf{0}_{d_\theta\times r^*})$ and $d_0 = (\mathbf{0}_{1\times r_1}, d')'$. The claimed result follows from (3.B.42) and the definitions of $\Gamma^*_{1,F}$ and $\mathcal{Z}_{2,F}$. □

Proof of Lemma 3.4.1. The claimed result in Part (a) has been proved in Lemma 3.B.8.

We next consider the case that $n^{1/2}\delta_{F_n} \to d$ with $||d|| = \infty$. Note that the results in (3.B.27) and (3.B.41) do not depend on $||d|| < \infty$ or $||d|| = \infty$. Using (3.B.27), (3.B.41), Lemma 3.B.1.(v) and the CMT, we have

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_d \Gamma_{1,F}\mathcal{Z}_{1,F}. \tag{3.B.43}$$

To study the properties of $\widehat{\theta}_2$, we have to consider two separate scenarios: (1) $\delta_{F_n} = o(1)$; and (2) $\|\delta_{F_n}\| > C^{-1}$. In scenario (1), Assumption 3.3.2, Lemma 3.B.1.(v) and Lemma 3.B.6 imply that

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1))n^{1/2}\delta_{F_n} + O_p(1). \tag{3.B.44}$$

By Assumption 3.3.1.(iv) and $||n^{1/2}\delta_{F_n}|| \to \infty$,

$$n\delta'_{F_n}\Gamma'_{2,F_n}\Gamma_{2,F_n}\delta_{F_n} \geq C^{-2}n\delta'_{F_n}\delta_{F_n} \to \infty \tag{3.B.45}$$

which together with (3.B.44) implies that $||n^{1/2}(\widehat{\theta}_2 - \theta_{F_n})|| \to_p \infty$.

Finally, we consider the scenario (2) where $\|\delta_{F_n}\| > C^{-1}$. By Assumption 3.3.1.(iv),

$$||G'_{2,F_n}\Omega^{-1}_{2,F_n}\delta_{F_n}|| > C^{-1}\|\delta_{F_n}\| > C^{-2}. \tag{3.B.46}$$

As $\theta^*_{F_n}$ is the minimizer of $Q_{F_n}(\theta)$, it has the following first order condition

$$0_{d_\theta \times 1} = G_{2,F_n}(\theta^*_{F_n})'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta^*_{F_n}), \tag{3.B.47}$$

which implies that

$$\begin{aligned} G'_{2,F_n}\Omega^{-1}_{2,F_n}\delta_{F_n} &= G_{2,F_n}(\theta_{F_n})'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta^*_{F_n}) \\ &= \left[G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right]'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta_{F_n}) \\ &+ G_{2,F_n}(\theta_{F_n})'\Omega^{-1}_{2,F_n}\left[M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n})\right]. \end{aligned} \tag{3.B.48}$$

By Lemma 3.B.2, the Cauchy-Schwarz inequality and Assumption 3.3.2.(i)-(ii), we have

$$\begin{aligned} &\left\|\left[G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right]'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta_{F_n})\right\| \\ &\leq \left\|G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right\|\left\|\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta_{F_n})\right\| \end{aligned}$$

$$\leq C \left\| \theta_{F_n} - \theta^*_{F_n} \right\|, \tag{3.B.49}$$

where C is a fixed constant. Similarly, we have

$$\begin{aligned}
&\left\| G_{2,F_n}(\theta_{F_n})' \Omega^{-1}_{2,F_n} \left[M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n}) \right] \right\| \\
&\leq \left\| M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n}) \right\| \left\| \Omega^{-1}_{2,F_n} G_{2,F_n}(\theta_{F_n}) \right\| \\
&\leq C \left\| \theta_{F_n} - \theta^*_{F_n} \right\|. \qquad (3.B.50)
\end{aligned}$$

Combining the results in (3.B.48), (3.B.49) and (3.B.50), and using the triangle inequality, we have

$$\left\| \theta_{F_n} - \theta^*_{F_n} \right\| \geq C \tag{3.B.51}$$

for some fixed constant C. Using $\widehat{\theta}_2 = \theta^*_{F_n} + o_p(1)$ (which is proved in Lemma 3.B.5) and the triangle inequality, we obtain

$$\left\| \widehat{\theta}_2 - \theta_{F_n} \right\| \geq \left| ||\widehat{\theta}_2 - \theta^*_{F_n}|| - \left\| \theta^*_{F_n} - \theta_{F_n} \right\| \right| = \left\| \theta^*_{F_n} - \theta_{F_n} \right\| (1 + o_p(1)), \tag{3.B.52}$$

which together with (3.B.51) implies that $n^{1/2}||\widehat{\theta}_2 - \theta_{F_n}|| \to_p \infty$. This finishes the proof. □

Proof for the claim in equation (3.4.3). Consider the case $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$. By Lemma 3.4.1,

$$n^{1/2} \left[\widehat{\theta}(\omega) - \theta_{F_n} \right] = n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \omega \left[n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) - n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \right]$$

$$\to_d \Gamma^*_{1,F}\mathcal{Z}_{d,2,F} + \omega(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}. \tag{3.B.53}$$

This implies that

$$\ell(\widehat{\theta}(\omega)) = n\left[\widehat{\theta}_n(\omega) - \theta_{F_n}\right]' \Upsilon \left[\widehat{\theta}_n(\omega) - \theta_{F_n}\right] \to_d \lambda_F(\omega) \tag{3.B.54}$$

where

$$\begin{aligned}\lambda_F(\omega) = {} & \mathcal{Z}'_{d,2,F}\Gamma^{*\prime}_{1,F}\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F} + 2\omega\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F} \\ & + \omega^2\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}.\end{aligned}$$

Now we consider $\mathbb{E}[\lambda_F(\omega)]$ using the equalities in Lemma 3.B.9 below. First,

$$\mathbb{E}[\mathcal{Z}'_{d,2,F}\Gamma^{*\prime}_{1,F}\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F}] = \mathrm{tr}(\Upsilon\Sigma_{1,F}) \tag{3.B.55}$$

because $\Gamma^*_1\mathcal{Z}_{d,2,F} = \Gamma_{1,F}\mathcal{Z}_{1,F}$ and $\Gamma_{1,F}\mathbb{E}[\mathcal{Z}_{1,F}\mathcal{Z}'_{1,F}]\Gamma'_{1,F} = \Sigma_{1,F}$ by definition. Second,

$$\begin{aligned}& \mathbb{E}\left[\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F}\right] \\ & = \mathrm{tr}(\Upsilon\Gamma^*_{1,F}\mathbb{E}\left[\mathcal{Z}_{d,2,F}\mathcal{Z}'_{d,2,F}\right](\Gamma_{2,F} - \Gamma^*_{1,F})') \\ & = \mathrm{tr}(\Upsilon\Gamma^*_{1,F}\left[d_0 d_0' + \Omega_{2,F}\right](\Gamma_{2,F} - \Gamma^*_{1,F})') \\ & = \mathrm{tr}(\Upsilon(\Sigma_{2,F} - \Sigma_{1,F})),\end{aligned} \tag{3.B.56}$$

where the last equality holds by Lemma 3.B.9. Third,

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}'_{d,2,F}(\Gamma_{2,F}-\Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F}-\Gamma^*_{1,F})\mathcal{Z}_{d,2,F}\right] \\
&= \mathrm{tr}(\Upsilon(\Gamma_{2,F}-\Gamma^*_{1,F})\left[d_0d_0'+\Omega_{2,F}\right](\Gamma_{2,F}-\Gamma^*_{1,F})') \\
&= d_0'\Gamma'_{2,F}\Upsilon\Gamma_{2,F}d_0+\mathrm{tr}(\Upsilon(\Sigma_{2,F}-\Sigma_{1,F})) \qquad (3.B.57)
\end{aligned}
$$

by Lemma 3.B.9. Combining the results in (3.B.55)-(3.B.57), we obtain

$$
\begin{aligned}
\mathbb{E}[\lambda_F(\omega)] = {}& \mathrm{tr}(\Upsilon\Sigma_{1,F}) - 2\omega\mathrm{tr}\left(\Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)\right) \\
&+ \omega^2\left[d_0'\Gamma'_{2,F}\Upsilon\Gamma_{2,F}d_0 + tr\left(\Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)\right)\right]. \qquad (3.B.58)
\end{aligned}
$$

Note that $d_0'\Gamma'_{2,F}\Upsilon\Gamma_{2,F}d_0 = d_0'(\Gamma_{2,F}-\Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F}-\Gamma^*_{1,F})d_0$ because $\Gamma^*_{1,F}d_0=\mathbf{0}_{d_\theta}$. It is clear that the optimal weight ω^*_F in (3.4.3) minimizes the quadratic function of ω in (3.B.58). □

Lemma 3.B.9. (a) $\Gamma^*_{1,F}d_0=\mathbf{0}_{d_\theta\times 1}$; (b) $\Gamma^*_{1,F}\Omega_{2,F}\Gamma^*_{1,F}=\Sigma_{1,F}$; (c) $\Gamma^*_{1,F}\Omega_{2,F}\Gamma'_{2,F}=\Sigma_{2,F}$; (d) $\Gamma_{2,F}\Omega_{2,F}\Gamma'_{2,F}=\Sigma_{2,F}$.

Proof of Lemma 3.B.9. By construction, $\Gamma^*_{1,F}d_0=\mathbf{0}_{d_\theta\times 1}$. For ease of notation, we write $\Omega_{2,F}$ and $G_{2,F}$ as

$$
\Omega_{2,F} = \begin{pmatrix} \Omega_{1,F} & \Omega_{1r^*} \\ \Omega_{r^*1,F} & \Omega_{r^*,F} \end{pmatrix} \text{ and } G_{2,F} = \begin{pmatrix} G_{1,F} \\ G_{r^*,F} \end{pmatrix}. \qquad (3.B.59)
$$

To prove part (b), we have

$$\Gamma^*_{1,F}\Omega_{2,F}\Gamma^*_{1,F} = [\Gamma_{1,F}, \mathbf{0}_{d_\theta \times r^*}] \begin{pmatrix} \Omega_{1,F} & \Omega_{1r^*} \\ \\ \Omega_{r^*1,F} & \Omega_{r^*,F} \end{pmatrix} [\Gamma_{1,F}, \mathbf{0}_{d_\theta \times r^*}]$$

$$= \Gamma_{1,F}\Omega_{1,F}\Gamma'_{1,F} = \left(G'_{1,F}\Omega^{-1}_{1,F}G_{1,F}\right)^{-1} = \Sigma_{1,F}. \tag{3.B.60}$$

To show part (c), note that

$$\Gamma^*_{1,F}\Omega_{2,F}\Gamma'_{2,F} = -\left[\Gamma_{1,F}, \mathbf{0}_{d_\theta \times r^*}\right]\Omega_{2,F}\Omega^{-1}_{2,F}G_{2,F}\left(G'_{2,F}\Omega^{-1}_{2,F}G_{2,F}\right)^{-1}$$

$$= -\Gamma_{1,F}G_{1,F}\left(G'_{2,F}\Omega^{-1}_{2,F}G_{2,F}\right)^{-1} = \left(G'_{2,F}\Omega^{-1}_{2,F}G_{2,F}\right)^{-1} = \Sigma_{2,F} \tag{3.B.61}$$

because $-\Gamma_{1,F}G_{1,F} = I_{d_\theta \times d_\theta}$. Part (d) follows from the definition of $\Gamma_{2,F}$. □

Proof of Lemma 3.4.2. We first prove the consistency of $\widehat{\Omega}_k$, $\widehat{G}_k$ and $\widehat{\Sigma}_k$ for $k = 1, 2$. By Lemma 3.4.1, we have $\widehat{\theta}_1 = \theta_{F_n} + o_p(1)$. Using the same arguments in showing (3.B.15), we can show that

$$\widehat{\Omega}_2 = \Omega_{2,F_n} + o_p(1) = \Omega_{2,F} + o_p(1), \tag{3.B.62}$$

where the second equality is by (3.B.39) which is assumed in the lemma. As $\widehat{\Omega}_1$ is a submatrix of $\widehat{\Omega}_2$, by (3.B.62) we have

$$\widehat{\Omega}_1 = \Omega_{1,F_n} + o_p(1) = \Omega_{1,F} + o_p(1). \tag{3.B.63}$$

By the consistency of $\widehat{\theta}_1$ and the same arguments used to show (3.B.36), we have

$$n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\widehat{\theta}_1) = G_{2,F_n}(\theta_{F_n}) + o_p(1) = G_{2,F} + o_p(1), \tag{3.B.64}$$

where the second equality is by (3.B.39) which is assumed in the lemma. As $n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i,\widehat{\theta}_1)$ is a submatrix of $n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\widehat{\theta}_1)$, by (3.B.64) we have

$$n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i,\widehat{\theta}_1) = G_{1,F_n}(\theta_{F_n}) + o_p(1) = G_{1,F} + o_p(1). \tag{3.B.65}$$

From Assumption 3.3.2, (3.B.62), (3.B.63), (3.B.64) and (3.B.65), we see that $\widehat{\Omega}_k$ and $\widehat{G}_k$ are consistent estimators of $\Omega_{k,F}$ and $G_{k,F}$ respectively for $k = 1, 2$. By the Slutsky theorem, we know that $\widehat{\Sigma}_k$ is a consistent estimator of $\Sigma_{k,F}$ for $k = 1, 2$.

In the case $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$, the desired result follows from Lemma 3.4.1, the consistency of $\widehat{\Sigma}_{1,F}$ and $\widehat{\Sigma}_{2,F}$, and the CMT. In the case $||n^{1/2}\delta_{F_n}|| \to \infty$, $\widetilde{\omega}_{eo} \to_p 0$ because $n^{1/2}||\widehat{\theta}_2 - \widehat{\theta}_1|| \to_p \infty$ and

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) &= n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \widetilde{\omega}_{eo} n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) \\ &= n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \frac{n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) tr\left[\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]}{n(\widehat{\theta}_2 - \widehat{\theta}_1)'\Upsilon(\widehat{\theta}_2 - \widehat{\theta}_1) + tr\left[\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]} \to_d \xi_{1,F} \end{aligned} \tag{3.B.66}$$

by Lemma 3.4.1. □

3.C Proof of the Results in Section 3.5

We first present some generic results on the bounds of asymptotic risk difference between two estimators under some high-level conditions. Then we apply these generic results to the two specific estimators we consider in this paper: $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$. The proof uses the subsequence techniques used to show the asymptotic size of a test in Andrews, Cheng, and Guggenberger (2011) but we adapt the proof and notations to the current setup and extend results from test to estimators.

Recall that $h_{F,d} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})')$ and $\overline{v}_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})')$ for any $F \in \mathcal{F}$ and any $d \in \mathbb{R}_\infty^{r^*}$. We have defined

$$H = \{h_{F,d} : d \in \mathbb{R}^{r^*} \text{ and } F \in \mathcal{F} \text{ with } \delta_F = 0_{r^* \times 1}\} \tag{3.C.1}$$

where δ_F is defined by (3.1.4) for a given F. Define

$$H_\infty^* = \{h_{F,d} : d \in \mathbb{R}_\infty^{r^*} \text{ with } ||d|| = \infty \text{ and } F \in \mathcal{F}\}. \tag{3.C.2}$$

Let $d_h = r^* + d_\theta r_2 + (r_2+1)r_2/2$. It is clear that $h_{F,d}$ is a d_h-dimensional vector.

Condition 3.C.1. (i) *For any sequence of DGPs $\{F_{p_n}\}$ with $F_{p_n} \in \mathcal{F}$ where $\{p_n\}$ is a subsequence of $\{n\}$, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ and some $F \in \mathcal{F}$ such that $v_{F_{p_n^*}} \to v_F$ as $p_n^* \to \infty$;* (ii) *$M_{1,F}(\theta) = 0_{r_1 \times 1}$ has a unique solution at $\theta_F \in \Theta$ for any $F \in \mathcal{F}$;* (iii) *$M_{2,F}(\cdot)$ is uniform equicontinuous over $F \in \mathcal{F}$;* (iv) *for any subsequence $\{p_n\}$ of $\{n\}$, if*

$(p_n)^{1/2}\delta_{F_{p_n}} \to d$ *for* $d \in \mathbb{R}^{r^*}_\infty$ *and* $v_{F_{p_n}} \to v_F$*, then*

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})] = R_\zeta(h_{F,d}) \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widetilde{\theta})] = \widetilde{R}_\zeta(h_{F,d})$$

where $R_\zeta(h_{F,d})$ *and* $\widetilde{R}_\zeta(h_{F,d})$ *are non-negative and bounded from above by* ζ *for any* $F \in \mathcal{F}$ *and any* $d \in \mathbb{R}^{r^*}_\infty$*;* (v) *for any* $F \in \mathcal{F}$ *with* $\delta_F = 0_{r^*\times 1}$*, there exists a constant* $\varepsilon_F > 0$ *such that for any* $\widetilde{\delta} \in \mathbb{R}^{r^*}$ *with* $0 \leq ||\widetilde{\delta}|| < \varepsilon_F$*, there is* $\widetilde{F} \in \mathcal{F}$ *with* $\delta_{\widetilde{F}} = \widetilde{\delta}$ *and* $||\overline{v}_F - \overline{v}_{\widetilde{F}}|| \leq C||\widetilde{\delta}||^\kappa$ *for some* $\kappa > 0$*;* (vi) *for any* $h_{F,d} \in H^*_\infty$ *and* $h_{F,\widetilde{d}} \in H^*_\infty$*, we have*

$$R_\zeta(h_{F,d}) = R_\zeta(h_{F,\widetilde{d}}) \text{ and } \widetilde{R}_\zeta(h_{F,d}) = \widetilde{R}_\zeta(h_{F,\widetilde{d}})$$

for any $\zeta > 0$*.*

Condition 3.C.1.(i) requires that for any sequence of $\{v_{F_{p_n}}\}$, it has a convergent subsequence $\{v_{F_{p^*_n}}\}$ with limit being v_F for some $F \in \mathcal{F}$. This condition is verified under Assumptions 3.3.2.(i) and 3.3.3.(i) in Lemma 3.B.7. Condition 3.C.1.(ii) is the unique identification condition of θ_F which holds under Assumptions 3.3.1.(i)-(ii). Condition 3.C.1.(iii) holds under Assumption 3.3.2.(i) by Lemma 3.B.2. Condition 3.C.1.(iv) is a key assumption to derive an explicit upper bound of asymptotic risk. This condition can be verified by using Lemma 3.4.1 as we shall show in the proof of Theorem 3.5.1. Condition 3.C.1.(v) enables us to show that the upper bound we derived for the asymptotic risk is also a lower bound. This condition is assumed in Assumption 3.3.3.(ii). Condition 3.C.1.(vi) requires that the asymptotic (truncated) risk of $\widehat{\theta}$ (or $\widetilde{\theta}$) under the subsequences of DGPs $\{F_{p_n}\}$ satisfying the restrictions in Condition 3.C.1.(iv) are identical whenever $(p_n)^{1/2}\delta_{F_{p_n}} \to d$ with $||d|| = \infty$. This condition holds in the GMM framework when severely misspecified moment condition(s) in $\mathbb{E}_F[g^*(W_i, \theta_F)] = 0_{r^* \times 1}$ leads to slower than root-n or inconsistent estimation of θ_F, which is ensured by Assumption 3.3.1.(iv).

Lemma 3.C.2. *Under Conditions 3.C.1.(i) - 3.C.1.(iv), we have*

$$AsyR_\zeta(\widehat{\theta}) \leq \max\left\{\sup_{h \in H} R_\zeta(h), \sup_{h \in H^*_\infty} R_\zeta(h)\right\}. \tag{3.C.3}$$

Proof of Lemma 3.C.2. Let $\{F_n\}$ be a sequence such that

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta})] = \limsup_{n\to\infty}\left(\sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta})]\right). \tag{3.C.4}$$

Such a sequence always exists by the definition of supremum. The sequence $\{\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta})]\colon n \geq 1\}$ may not converge. However, by the definition of limsup, there exists a subsequence of $\{n\}$, say $\{p_n\}$, such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})]\colon n \geq 1\}$ converges and

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})] = AsyR_\zeta(\widehat{\theta}). \tag{3.C.5}$$

Below we show that for any subsequence $\{p_n\}$ of $\{n\}$ such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})]\colon n \geq 1\}$ is convergent, there exists a subsequence $\{p^*_n\}$ of $\{p_n\}$ such that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p^*_n}}[\ell_\zeta(\widehat{\theta})] = R_\zeta(h) \text{ for some } h \in H \text{ or } H^*_\infty \tag{3.C.6}$$

Because $\lim_n \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta})] = \lim_n \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})]$, which combined with (3.C.5) and (3.C.6) implies that

$$AsyR_\zeta(\widehat{\theta}) = R_\zeta(h) \text{ for some } h \in H \text{ or } H_\infty^*. \tag{3.C.7}$$

The desired result in (3.C.3) follows immediately by (3.C.7).

To show that there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ such that (3.C.6) holds, it suffices to show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}$, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ for which we have

$$(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to d \text{ for } d \in \mathbb{R}_\infty^{r^*} \text{ and } \upsilon_{F_{p_n^*}} \to \upsilon_F \tag{3.C.8}$$

for some $F \in \mathcal{F}$. By (3.C.8), we can use Condition 3.C.1.(iv) to deduce that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta})] = R_\zeta(h_{F,d}) \tag{3.C.9}$$

for the sequence of DGPs $\{F_{p_n^*}\}$ satisfies (3.C.8). As $d \in \mathbb{R}_\infty^{r^*}$, we have either $\|d\| < \infty$ or $\|d\| = \infty$. In the first case, $\|d\| < \infty$ together with $(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to d$ and $\delta_{F_{p_n^*}} \to \delta_F$ (which is implied by $\upsilon_{F_{p_n^*}} \to \upsilon_F$) implies that $\delta_F = 0_{r^*\times 1}$, which implies that $h_{F,d} \in H$ by the definition of H. In the second case, $h_{F,d} \in H_\infty^*$ by the definition of H_∞^*. We have proved that $h_{F,d}$ in (3.C.9) belongs either to H or H_∞^* which together with (3.C.9) proves (3.C.6).

Finally, we show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}$, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ for which (3.C.8) holds. Let $\delta_{p_n,j}$ denote the j-th component of δ_{p_n} and $p_{1,n} = p_n$ for any $n \geq 1$. For $j = 1$, either (a) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| < \infty$; or

(b) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| = \infty$. If (a) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\delta_{p_{j+1,n},j} \to d_j$ for some $d_j \in \mathbb{R}$. If (b) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\delta_{p_{j+1,n},j} \to \infty$ or $-\infty$. As r^* is a fixed positive integer, we can apply the same arguments successively for $j = 1, ..., r^*$ to obtain a subsequence $\{p_{r^*,n}\}$ of $\{p_n\}$ such that $(p_{r^*,n})^{1/2}\delta_{p_{r^*,n}} \to d \in \mathbb{R}_{\infty}^{r^*}$. By Condition 3.C.1.(i), we know that there exists a subsequence $\{p_n^*\}$ of $\{p_{r^*,n}\}$ such that $v_{p_n^*} \to v_F$ for some $F \in \mathcal{F}$, which finishes the proof of (3.C.8). $\square$

Lemma 3.C.3. *Suppose that Condition 3.C.1.(v) holds. Then (i) for any $h_{F,d} \in H$, there exists a sequence of DGPs $\{F_n\}$ with $F_n \in \mathcal{F}$ such that*

$$n^{1/2}\delta_{F_n} \to d,\ G_{2,F_n} \to G_{2,F} \text{ and } \Omega_{2,F_n} \to \Omega_{2,F}; \tag{3.C.10}$$

(ii) for any $h_{F,d} \in H_{\infty}^*$, there exists a sequence of DGPs $\{F_n\}$ with $F_n \in \mathcal{F}$ such that

$$n^{1/2}\delta_{F_{p_n}} \to \widetilde{d} \text{ with } ||\widetilde{d}|| = \infty,\ G_{2,F_{p_n}} \to G_{2,F} \text{ and } \Omega_{2,F_{p_n}} \to \Omega_{2,F} \tag{3.C.11}$$

for any subsequence $\{p_n\}$ of $\{n\}$, where $\widetilde{d}$ may depend on $\{p_n\}$.

Proof of Lemma 3.C.3. (i) By the definition of H, we have $\delta_F = 0_{r^*\times 1}$ for the F generating $h_{F,d}$. Let N_{ε_F} be the smallest n such that $\|d\|\, n^{-1/2} < \varepsilon_F$. By Condition 3.C.1.(v), for any $n > N_{\varepsilon_F}$ we can find a DGP F_n such that

$$\delta_{F_n} = n^{-1/2}d \text{ and } \|\overline{v}_{F_n} - \overline{v}_F\| \leq n^{-\kappa/2}C||d||^{\kappa}. \tag{3.C.12}$$

For any n such that $\|d\| n^{-1/2} \geq \varepsilon_F$, we let $F_n = F$. The desired properties in (3.C.10) holds under the constructed sequence of DGPs $\{F_n\}$ by (3.C.12), because C is a fixed constant and $\kappa > 0$.

(ii) For any $h_{F,d} \in H_\infty^*$, we have either $\delta_F = 0_{r^* \times 1}$ or $||\delta_F|| > 0$. We first consider the case that $\delta_F = 0_{r^* \times 1}$. Let $1_{r^* \times 1}$ denote the $r^* \times 1$ vector of ones. Let N_{ε_F} be the smallest n such that $(p_n)^{-1/4} r^* < \varepsilon_F$. By Condition 3.C.1.(v), for any $n > N_{\varepsilon_F}$ we can find a DGP F_n such that

$$\delta_{F_{p_n}} = (p_n)^{-1/4} 1_{r^* \times 1} \text{ and } \left\| \overline{v}_{F_{p_n}} - \overline{v}_F \right\| \leq (p_n)^{-\kappa/4} C r^*. \tag{3.C.13}$$

For any n such that $(p_n)^{-1/4} r^* \geq \varepsilon_F$, we let $F_n = F$. The desired properties in (3.C.11) holds under the constructed sequence of DGPs $\{F_n\}$ by (3.C.13), because C is a fixed constant and $\kappa > 0$. When $||\delta_F|| > 0$, we define a trivial sequence of DGPs $\{F_n\}$ as $F_{p_n} = F$ for any n. It is clear that (3.C.11) holds trivially in this case. □

Lemma 3.C.4. *Under Condition 3.C.1, we have*

$$AsyR_\zeta(\widehat{\theta}) = \max \left\{ \sup_{h \in H} R_\zeta(h), \sup_{h \in H_\infty^*} R_\zeta(h) \right\}. \tag{3.C.14}$$

Proof of Lemma 3.C.4. In view of the upper bound in (3.C.3), it is sufficient to show that

$$AsyR_\zeta(\widehat{\theta}) \geq \max \left\{ \sup_{h \in H} R_\zeta(h), \sup_{h \in H_\infty^*} R_\zeta(h) \right\}. \tag{3.C.15}$$

First, we note that for any $h_{d,F} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \in H$, there exists a sequence

$\{F_n \in \mathcal{F} : n \geq 1\}$ such that

$$n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*},\ \delta_{F_n} \to 0_{r^* \times 1} \text{ and } \overline{v}_{F_n} \to (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \tag{3.C.16}$$

by Lemma 3.C.3.(i). The sequence $\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta})]$ may not be convergent, but there exists a subsequence $\{p_n\}$ of n such that $\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta})]$ is convergent and

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta})] = \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})]. \tag{3.C.17}$$

As $\{p_n\}$ is a subsequence of $\{n\}$, by (3.C.16)

$$n^{1/2}\delta_{F_{p_n}} \to d \in \mathbb{R}^{r^*} \text{ and } \overline{v}_{F_{p_n}} \to (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \tag{3.C.18}$$

for some $F \in \mathcal{F}$. Moreover, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ such that

$$n^{1/2}\delta_{F_{p_n^*}} \to d \in \mathbb{R}^{r^*} \text{ and } v_{F_{p_n^*}} \to v_F \tag{3.C.19}$$

for the F in (3.C.18). Hence by (3.C.19) and Condition 3.C.1.(iv), we deduce that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell(\widehat{\theta})] = R_\zeta(h_{F,d}). \tag{3.C.20}$$

As $\{p_n^*\}$ is a subsequence of $\{p_n\}$, $\lim_n \mathbb{E}_{F_{p_n^*}}[\ell(\widehat{\theta})] = \lim_n \mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta})]$ which combined with

(3.C.17), (3.C.20) and the definition of $AsyR_{\zeta}(\widehat{\theta})$ implies that

$$AsyR_{\zeta}(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_{\zeta}(\widehat{\theta})] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R_{\zeta}(h_{F,d}) \tag{3.C.21}$$

for any $h_{F,d} \in H$.

Second, consider any $h_{d,F} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \in H_{\infty}^{*}$. By Lemma 3.C.4.(ii), there exists a sequence of DGPs $\{F_n\}$ such that

$$n^{1/2}\delta_{F_n} \to \widetilde{d} \text{ with } ||\widetilde{d}|| = \infty,\ G_{2,F_n} \to G_{2,F} \text{ and } \Omega_{2,F_n} \to \Omega_{2,F}. \tag{3.C.22}$$

Using the same arguments in proving (3.C.17) and (3.C.20), we can show that

$$v_{p_n^*} \to v_F \text{ for some } F \in \mathcal{F} \tag{3.C.23}$$

and

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = \lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell(\widehat{\theta})] \tag{3.C.24}$$

where $\{p_n^*\}$ is a subsequence of $\{n\}$. By Lemma 3.C.4.(ii),

$$(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to \widetilde{d}^* \text{ with } ||\widetilde{d}^*|| = \infty, \tag{3.C.25}$$

under the sequence of DGPs $\left\{F_{p_n^*}\right\}$. Using (3.C.23), (3.C.25) and Condition 3.C.1.(iv), we deduce that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell(\widehat{\theta})] = R_{\zeta}(h_{F,\widetilde{d}^*}). \tag{3.C.26}$$

As $||\widetilde{d}^*|| = \infty$ and $||d|| = \infty$, by Conditions 3.C.1.(vi), we have $R_\zeta(h_{F,\widetilde{d}^*}) = R_\zeta(h_{F,d})$ which together with (3.C.24) and (3.C.26) implies that

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R_\zeta(h_{F,d}). \tag{3.C.27}$$

By the definition of $AsyR_\zeta(\widehat{\theta})$ and (3.C.27),

$$AsyR_\zeta(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta})] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R_\zeta(h_{F,d}) \tag{3.C.28}$$

for any $h_{F,d} \in H^*_\infty$. Combining the results in (3.C.21) and (3.C.28), we immediately get (3.C.14). □

Lemma 3.C.5. *Under Conditions 3.C.1.(i) - 3.C.1.(iv), the upper and lower bounds of the asymptotic risk difference between* $\widehat{\theta}$ *and* $\widetilde{\theta}$ *satisfy*

$$Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta}) \leq \lim_{\zeta\to\infty} \left(\max \left\{ \sup_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \sup_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\} \right), \tag{3.C.29}$$

$$Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta}) \geq \lim_{\zeta\to\infty} \left(\min \left\{ \inf_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \inf_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\} \right). \tag{3.C.30}$$

Proof of Lemma 3.C.5. Define

$$\overline{R}_\zeta(H, H^*_\infty) = \max \left\{ \sup_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \sup_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\}, \tag{3.C.31}$$

$$\underline{R}_\zeta(H, H^*_\infty) = \min \left\{ \inf_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \inf_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\}. \tag{3.C.32}$$

By the definition of $Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta})$, for (3.C.29) it is sufficient to show that for any $\zeta > 0$

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})] \le \overline{R}_\zeta(H, H_\infty^*), \tag{3.C.33}$$

which can be proved using the same arguments in the proof of Lemma 3.C.2 (but replacing $\ell_\zeta(\widehat{\theta})$ and $R_\zeta(h)$ by $\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})$ and $R_\zeta(h) - \widetilde{R}_\zeta(h)$ respectively). Similarly by the definition of $Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta})$, for (3.C.30) it is sufficient to show that for any $\zeta > 0$

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})] \ge \underline{R}_\zeta(H, H_\infty^*), \tag{3.C.34}$$

which can be proved using the same arguments in the proof of Lemma 3.C.2 (but replacing $\limsup_n$, $\sup_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta})$ and $R_\zeta(h)$ by $\liminf_n$, $\inf_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})$ and $R_\zeta(h) - \widetilde{R}_\zeta(h)$ respectively). □

Lemma 3.C.6. *Under Condition 3.C.1, the upper and lower bounds of the asymptotic risk difference between $\widehat{\theta}$ and $\widetilde{\theta}$ have the following representations:*

$$Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta}) = \lim_{\zeta\to\infty} \left(\max\left\{ \sup_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \sup_{h\in H_\infty^*} \left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right] \right\} \right), \tag{3.C.35}$$

$$Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta}) = \lim_{\zeta\to\infty} \left(\min\left\{ \inf_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \inf_{h\in H_\infty^*} \left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right] \right\} \right). \tag{3.C.36}$$

Proof of Lemma 3.C.6. By Lemma 3.C.5, it is sufficient to show that

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})] \ge \overline{R}_\zeta(H, H_\infty^*), \tag{3.C.37}$$

$$\liminf_{n\to\infty}\inf_{F\in\mathcal{F}}\mathbb{E}_F[\ell_\zeta(\widehat{\theta})-\ell_\zeta(\widetilde{\theta})]\leq\underline{R}_\zeta(H,H_\infty^*),\tag{3.C.38}$$

for any $\zeta>0$. (3.C.37) can be proved using the same arguments in the proof of Lemma 3.C.4 by replacing $\ell_\zeta(\widehat{\theta})$ and $R_\zeta(h)$ by $\ell_\zeta(\widehat{\theta})-\ell_\zeta(\widetilde{\theta})$ and $R_\zeta(h)-\widetilde{R}_\zeta(h)$ respectively. Similarly, (3.C.38) can be proved using the same arguments in the proof of Lemma 3.C.4 by replacing $\limsup_n$, $\sup_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta})$ and $R_\zeta(h)$ by $\liminf_n$, $\inf_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta})-\ell_\zeta(\widetilde{\theta})$ and $R_\zeta(h)-\widetilde{R}_\zeta(h)$ respectively. □

Lemma 3.C.7. *Under Assumptions 3.3.2.(i) and 3.3.2.(iii), we have*

$$\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq C \text{ and } \sup_{h\in H}\mathbb{E}[(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^2]\leq C.\tag{3.C.39}$$

Proof of Lemma 3.C.7. By definition,

$$\xi_{1,F}'\Upsilon\xi_{1,F}=\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F}=\mathcal{Z}_1'\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2}\mathcal{Z}_1\tag{3.C.40}$$

where $\mathcal{Z}_1\sim N(\mathbf{0}_{r_1},I_{r_1\times r_1})$. By Assumptions 3.3.2.(i) and 3.3.2.(iii), and the fact that Υ is a fixed matrix,

$$\sup_{F\in\mathcal{F}}\rho_{\max}(\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2})\leq C.\tag{3.C.41}$$

By (3.C.41),

$$\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq\sup_{h\in H}\rho_{\max}^2(\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2})\mathbb{E}[(\mathcal{Z}_1'\mathcal{Z}_1)^2]\leq 3r_1C\tag{3.C.42}$$

where the second inequality is by $\mathbb{E}[(\mathcal{Z}_1'\mathcal{Z}_1)^2] \leq 3r_1$ which is implied by the assumption that $\mathcal{Z}_1$ is a r_1-dimensional standard normal random vector. The first inequality in (3.C.39) follows as the upper bound in (3.C.42) does not depend on F.

By the Cauchy-Schwarz inequality and the simple inequality $|ab| \leq (a^2 + b^2)/2$ (for any real numbers a and b),

$$\begin{aligned}\overline{\xi}_F'\Upsilon\overline{\xi}_F &\leq 2\left(\mathcal{Z}_{d,2,F}'\Gamma_{1,F}^{*\prime}\Upsilon\Gamma_{1,F}^{*}\mathcal{Z}_{d,2,F} + \overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}\right) \\ &= 2\left(\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F} + \overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}\right) \qquad (3.C.43)\end{aligned}$$

where the equality is by $\Gamma_{1,F}^{*}d_0 = \mathbf{0}_{d_\theta\times 1}$ (which is proved in Lemma 3.B.9). By (3.C.43) and the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$ (for any real numbers a and b),

$$(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^2 \leq 8(\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F})^2 + 8(\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2. \qquad (3.C.44)$$

By the first inequality in (3.C.39), we have $\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2] \leq C$. Hence by (3.C.44), to show the second inequality in (3.C.39), it is sufficient to prove that

$$\sup_{h\in H}\mathbb{E}[(\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2] \leq C. \qquad (3.C.45)$$

For any $F \in \mathcal{F}$, define

$$B_F = (\Gamma_{2,F} - \Gamma_{1,F}^{*})'\Upsilon(\Gamma_{2,F} - \Gamma_{1,F}^{*}) \qquad (3.C.46)$$

Recall that we have defined $A_F = \Upsilon\left(\Sigma_{1,F} - \Sigma_{2,F}\right)$ in Theorem 3.5.2. By the definition,

$$\begin{aligned}\overline{\omega}_F^2 \mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} &= \frac{(\mathrm{tr}(A_F))^2 \mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F}}{(\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F))^2} \\ &= \mathrm{tr}(A_F) \frac{\mathrm{tr}(A_F)}{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)} \frac{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F}}{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)}. \end{aligned} \quad (3.C.47)$$

By Lemma 2.1 in Cheng and Liao (2015), $\mathrm{tr}(A_F) \geq 0$ for any $F \in \mathcal{F}$. This together with $\mathcal{Z}_{d,2,F}' D_F \mathcal{Z}_{d,2,F} \geq 0$ implies that

$$\frac{\mathrm{tr}(A_F)}{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)} \leq 1 \text{ and } \frac{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F}}{\mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)} \leq 1. \quad (3.C.48)$$

By (3.C.48) and $\mathrm{tr}(A_F) \geq 0$,

$$\overline{\omega}_F^2 \mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F} \leq \mathrm{tr}(A_F) = \mathrm{tr}(\Upsilon \Sigma_{1,F}) - \mathrm{tr}(\Upsilon \Sigma_{2,F}), \quad (3.C.49)$$

where the equality is by $A_F = \Upsilon(\Sigma_{1,F} - \Sigma_{2,F})$. By (3.C.49) and the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$,

$$\mathbb{E}[(\overline{\omega}_F^2 \mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F})^2] \leq 2(\mathrm{tr}(\Upsilon \Sigma_{1,F}))^2 + 2(\mathrm{tr}(\Upsilon \Sigma_{2,F}))^2. \quad (3.C.50)$$

By Assumptions 3.3.2.(i) and 3.3.2.(iii),

$$\rho_{\min}(G_{k,F}' \Omega_{k,F}^{-1} G_{k,F}) \geq \rho_{\min}(\Omega_{k,F}^{-1}) \rho_{\min}(G_{k,F}' G_{k,F}) = \rho_{\min}(G_{k,F}' G_{k,F}) / \rho_{\max}(\Omega_{k,F}) \geq C^{-1} \quad (3.C.51)$$

for $k = 1, 2$. By (3.C.51) and the definition of $\Sigma_{k,F}$ $(k = 1, 2)$,

$$\rho_{\max}(\Sigma_{k,F}) = \rho_{\min}^{-1}(G'_{k,F}\Omega_{k,F}^{-1}G_{k,F}) \leq C \tag{3.C.52}$$

for any $F \in \mathcal{F}$. As Υ and $\Sigma_{k,F}$ are positive definite symmetric matrix, by the standard trace inequality ($tr(AB) \leq tr(A)\rho_{\max}(B)$ for Hermitian matrices $A \geq 0$ and $B \geq 0$),

$$\mathrm{tr}(\Upsilon\Sigma_{k,F}) \leq \mathrm{tr}(\Upsilon)\rho_{\max}(\Sigma_{k,F}) \leq C \text{ for } k = 1, 2, \tag{3.C.53}$$

for any $F \in \mathcal{F}$. Collecting the results in (3.C.50) and (3.C.53), we immediately get (3.C.45). This finishes the proof. □

Lemma 3.C.8. *Under Assumptions 3.3.2.(i) and 3.3.2.(iii), we have*

$$\lim_{\zeta \to \infty} \sup_{h \in H} [|g_\zeta(h) - g(h)|] = 0 \tag{3.C.54}$$

where $\sup_{h \in H} [|g(h)|] \leq C$.

Proof of Lemma 3.C.8. First note that

$$\min\{x, \zeta\} - x = (\zeta - x)I\{x > \zeta\}. \tag{3.C.55}$$

Hence we have

$$
\begin{aligned}
&\sup_{h_{d,F}\in H}\left|\mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\}-\overline{\xi}_F'\Upsilon\overline{\xi}_F\right]\right| \\
&\le \sup_{h_{d,F}\in H}\mathbb{E}\left[\left|\zeta-\overline{\xi}_F'\Upsilon\overline{\xi}_F\right| I\{\overline{\xi}_F'\Upsilon\overline{\xi}_F>\zeta\}\right] \\
&\le \zeta\sup_{h_{d,F}\in H}\mathbb{E}\left[I\{\overline{\xi}_F'\Upsilon\overline{\xi}_F>\zeta\}\right]+\sup_{h_{d,F}\in H}\mathbb{E}\left[\overline{\xi}_F'\Upsilon\overline{\xi}_F I\{\zeta^{-1}>(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^{-1}\}\right] \\
&\le 2\zeta^{-1}\sup_{h_{d,F}\in H}\mathbb{E}\left[(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^2\right]\le 2C\zeta^{-1} \qquad (3.C.56)
\end{aligned}
$$

where the first inequality is by the Jensen's inequality, the second inequality is by the Markov inequality, the third inequality is by the monotonicity of expectation and the last inequality is by Lemma 3.C.7. Using the same arguments, we can show that

$$
\sup_{h_{d,F}\in H}\left|\mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}-\xi_{1,F}'\Upsilon\xi_{1,F}\right]\right|\le 2C\zeta^{-1}. \qquad (3.C.57)
$$

Collecting the results in (3.C.56) and (3.C.57), and applying the triangle inequality, we deduce that

$$
\sup_{h\in H}\left[|g_\zeta(h)-g(h)|\right]\le 4C\zeta^{-1}. \qquad (3.C.58)
$$

The claimed result in (3.C.54) follows by (3.C.58) as C is a fixed constant.

By the triangle inequality, the Jensen's inequality and Lemma 3.C.7,

$$
\begin{aligned}
\sup_{h\in H}|g(h)| &\le \sup_{h\in H}\left|\mathbb{E}[\overline{\xi}_F'\Upsilon\overline{\xi}_F-\xi_{1,F}'\Upsilon\xi_{1,F}]\right| \\
&\le \sup_{h\in H}\mathbb{E}[\overline{\xi}_F'\Upsilon\overline{\xi}_F]+\sup_{h\in H}\mathbb{E}[\xi_{1,F}'\Upsilon\xi_{1,F}]\le C
\end{aligned}
$$

which finishes the proof of the lemma. □

Proof of Theorem 3.5.1. We first apply Lemma 3.C.6 with $\widehat{\theta} = \theta_{eo}$ and $\widetilde{\theta} = \widehat{\theta}_1$ to show that

$$AsyRD(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\left[g_\zeta(h)\right], 0\right\}, \text{ and} \tag{3.C.59}$$

$$Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) = \lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\left[g_\zeta(h)\right], 0\right\}. \tag{3.C.60}$$

Now we verify Condition 3.C.1 under Assumptions 3.3.1-3.3.3. Condition 3.C.1.(i) is verified by Lemma 3.B.7 under Assumptions 3.3.2.(i) and 3.3.3.(i). Condition 3.C.1.(ii) is implied by Assumptions 3.3.1.(i) and 3.3.1.(ii). Condition 3.C.1.(iii) is implied by the continuous differentiability of $g_2(w, \theta)$ with respect to θ for any $w \in \mathcal{W}$, Assumptions 3.3.2.(i) and the dominated convergence theorem (DCT). Condition 3.C.1.(v) is assumed in Assumption 3.3.1.(ii). We next verify Conditions 3.C.1.(iv) and 3.C.1.(vi).

Consider any sequence of DGPs $\{F_{p_n}\}$ with

$$(p_n)^{1/2}\delta_{F_{p_n}} \to d \text{ for } d \in \mathbb{R}_\infty^{r^*} \text{ and } v_{F_{p_n}} \to v_F \tag{3.C.61}$$

for some $F \in \mathcal{F}$, where $\{p_n\}$ is a subsequence of $\{n\}$. First, we consider the case that $d \in \mathbb{R}^{r^*}$. By Lemma 3.4.1.(a) and 3.4.2.(a),

$$(p_n)^{1/2}(\widehat{\theta}_1 - \theta_{F_{p_n}}) \to_d \xi_{1,F} \text{ and } (p_n)^{1/2}(\widehat{\theta}_{eo} - \theta_{F_{p_n}}) \to_d \overline{\xi}_F \tag{3.C.62}$$

which combined with the continuous mapping theorem implies that

$$\ell(\widehat{\theta}_1, \theta_{F_{p_n}}) \to_d \xi_{1,F}' \Upsilon \xi_{1,F} \text{ and } \ell(\widehat{\theta}_{eo}, \theta_{F_{p_n}}) \to_d \overline{\xi}_F' \Upsilon \overline{\xi}_F. \tag{3.C.63}$$

The function $f_\zeta(x) = \min\{x, \zeta\}$ is a bounded continuous function. By (3.C.63) and the Portmanteau Lemma,

$$\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo})] \to \mathbb{E}\left[\min\{\overline{\xi}_F' \Upsilon \overline{\xi}_F, \zeta\}\right] \text{ and } \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_1)] \to \mathbb{E}\left[\min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right]. \tag{3.C.64}$$

Second, we consider the case that $\|d\| = \infty$. Then under Lemma 3.4.1.(b) and 3.4.2.(b),

$$(p_n)^{1/2}(\widehat{\theta}_1 - \theta_{F_{p_n}}) \to_d \xi_{1,F} \text{ and } (p_n)^{1/2}(\widehat{\theta}_{eo} - \theta_{F_{p_n}}) \to_d \xi_{1,F}. \tag{3.C.65}$$

Using the same arguments in showing (3.C.64), we get

$$\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo})] \to \mathbb{E}\left[\min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right] \text{ and } \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_1)] \to \mathbb{E}\left[\min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right]. \tag{3.C.66}$$

Define

$$\widetilde{R}_\zeta(h_{F,d}) = \mathbb{E}\left[\min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right] \text{ and } R_\zeta(h_{F,d}) = \begin{cases} \mathbb{E}[\min\{\overline{\xi}_F' \Upsilon \overline{\xi}_F, \zeta\}], & \|d\| < \infty \\ \mathbb{E}\left[\min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right], & \|d\| = \infty \end{cases}. \tag{3.C.67}$$

Collecting the results in (3.C.64) and (3.C.66), we deduce that under the sequence of DGPs

$\{F_{p_n}\}$ satisfying (3.C.61),

$$\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo})] \to R_\zeta(h_{F,d}) \text{ and } \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_1)] \to \widetilde{R}_\zeta(h_{F,d}), \tag{3.C.68}$$

where $R_\zeta(h_{F,d})$ and $\widetilde{R}_\zeta(h_{F,d})$ are non-negative and bounded from above by ζ for any $d \in \mathbb{R}^{r^*}_\infty$ and any $F \in \mathcal{F}$. This verifies Condition 3.C.1.(iv).

By definition, $\widetilde{R}_\zeta(h_{F,d})$ in (3.C.67) does not depend on d for any F. Moreover, for any d and $\widetilde{d}$ with $||d|| = \infty$ and $||\widetilde{d}|| = \infty$, by the definition of $R_\zeta(h_{F,d})$ in (3.C.68),

$$R_\zeta(h_{F,d}) = \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right] = R_\zeta(h_{F,\widetilde{d}}). \tag{3.C.69}$$

Hence, Condition 3.C.1.(vi) is also verified.

We next apply Lemma 3.C.6 to get (3.C.59) and (3.C.60) above. By (3.C.67),

$$R_\zeta(h) - \widetilde{R}_\zeta(h) = \mathbb{E}[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F, \zeta\}] - \mathbb{E}[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}] \text{ for any } h \in H \tag{3.C.70}$$

and

$$R_\zeta(h) - \widetilde{R}_\zeta(h) = \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right] - \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right] = 0 \text{ for any } h \in H^*_\infty. \tag{3.C.71}$$

By Lemma 3.C.6, (3.C.70) and (3.C.71), we have

$$Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta}) = \lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \sup_{h\in H^*_\infty}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right]\right\}$$

$$= \lim_{\zeta\to\infty} \max\left\{\sup_{h\in H} \mathbb{E}\left[\min\{\overline{\xi}_F' \Upsilon \overline{\xi}_F, \zeta\} - \min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right], 0\right\} \tag{3.C.72}$$

and

$$AsyRD\underline{RD}(\widehat{\theta}, \widetilde{\theta}) = \lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \inf_{h\in H_\infty^*}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right]\right\}$$

$$= \lim_{\zeta\to\infty} \min\left\{\inf_{h\in H} \mathbb{E}\left[\min\{\overline{\xi}_F' \Upsilon \overline{\xi}_F, \zeta\} - \min\{\xi_{1,F}' \Upsilon \xi_{1,F}, \zeta\}\right], 0\right\}, \tag{3.C.73}$$

which proves (3.C.59) and (3.C.60).

We next show that

$$\lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\left[g_\zeta(h)\right], 0\right\} = \max\left\{\sup_{h\in H}\left[g(h)\right], 0\right\}, \text{ and} \tag{3.C.74}$$

$$\lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\left[g_\zeta(h)\right], 0\right\} = \min\left\{\inf_{h\in H}\left[g(h)\right], 0\right\}. \tag{3.C.75}$$

By Lemma 3.C.8,

$$\lim_{\zeta\to\infty} \sup_{h\in H}\left[g_\zeta(h)\right] = \sup_{h\in H}\left[g(h)\right] \text{ and } \lim_{\zeta\to\infty} \inf_{h\in H}\left[g_\zeta(h)\right] = \inf_{h\in H}\left[g(h)\right], \tag{3.C.76}$$

where $\sup_{h\in H}[g(h)]$ and $\inf_{h\in H}[g(h)]$ are finite real numbers. Let $\overline{f}(x) = \max(x, 0)$ and $\underline{f}(x) = \min(x, 0)$. It is clear that $\overline{f}(x)$ and $\underline{f}(x)$ are continuos function on $\mathbb{R}$. The asserted results in (3.C.74) and (3.C.75) follow by (3.C.76), and the continuity of $\overline{f}(x)$ and $\underline{f}(x)$. □

Proof of Theorem 3.5.2. For any $F \in \mathcal{F}$, define

$$D_F = (\Gamma_{2,F} - \Gamma^*_{1,F})' \Upsilon \Gamma^*_{1,F}. \tag{3.C.77}$$

Recall that we have defined

$$A_F = \Upsilon \left(\Sigma_{1,F} - \Sigma_{2,F}\right) \text{ and } B_F = (\Gamma_{2,F} - \Gamma^*_{1,F})' \Upsilon (\Gamma_{2,F} - \Gamma^*_{1,F})$$

in Theorem 3.5.2 and (3.C.46) respectively. By the definition of $\overline{\xi}_F$,

$$\mathbb{E}[\overline{\xi}'_F \Upsilon \overline{\xi}_F] = \text{tr}(\Upsilon \Sigma_{1,F}) + 2\text{tr}(A_F) J_{1,F} + \text{tr}(A_F)^2 J_{2,F} \tag{3.C.78}$$

where

$$J_{1,F} = \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F} D_F \mathcal{Z}_{d,2,F}}{\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} + \text{tr}(A_F)}\right] \text{ and } J_{2,F} = \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F}}{(\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} + \text{tr}(A_F))^2}\right]. \tag{3.C.79}$$

We provide a upper bound for $J_{1,F}$ defined in (3.C.79). Let

$$\eta(x) = \frac{x}{x' B_F x + \text{tr}(A_F)}, \text{ where } x = \mathcal{Z}_{d,2}. \tag{3.C.80}$$

Its derivative is

$$\frac{\partial \eta(x)'}{\partial x} = \frac{1}{x' B_F x + \text{tr}(A_F)} I_{r_2} - \frac{2 B_F}{\left(x' B_F x + \text{tr}(A_F)\right)^2} x x'. \tag{3.C.81}$$

Define

$$D_F = (\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}, \tag{3.C.82}$$

which satisfies $D_F\mathcal{Z}_{d,2,F} = D_F\mathcal{Z}_{2,F}$ by construction because the last r^* rows of $\Gamma^*_{1,F}$ are zeros. Applying Lemma 3.B.9 yields

$$\begin{aligned}
\operatorname{tr}(D_F\Omega_{2,F}) &= \operatorname{tr}\left((\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\Omega_{2,F}\right) \\
&= \operatorname{tr}(\Upsilon\left(\Gamma^*_{1,F}\Omega_{2,F}\Gamma'_{2,F} - \Gamma^*_{1,F}\Omega_{2,F}\Gamma^*_{1,F}\right)) \\
&= \operatorname{tr}(\Upsilon(\Sigma_{2,F} - \Sigma_{1,F})) = -\operatorname{tr}(A_F).
\end{aligned} \tag{3.C.83}$$

By Lemma 1 of Hansen (2016), which is a matrix version of the Stein's Lemma (Stein, 1956),

$$J_1 = \mathbb{E}\left(\eta(\mathcal{Z}_{d,2,F})'D_F\mathcal{Z}_{d,2,F}\right) = \mathbb{E}\left(\eta(\mathcal{Z}_{d,2,F})'D_F\mathcal{Z}_{d,2,F}\right) = \mathbb{E}\left[\operatorname{tr}\left(\frac{\partial\eta(\mathcal{Z}_{d,2,F})'}{\partial x}D_F\Omega_{2,F}\right)\right]. \tag{3.C.84}$$

Plugging (3.C.80)-(3.C.82) into (3.C.84), we have

$$\begin{aligned}
J_1 &= \mathbb{E}\left[\frac{\operatorname{tr}(D_F\Omega_{2,F})}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \operatorname{tr}(A_F)}\right] - 2\mathbb{E}\left[\frac{\operatorname{tr}\left(B_F\mathcal{Z}_{d,2,F}\mathcal{Z}'_{d,2,F}D_F\Omega_{2,F}\right)}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \operatorname{tr}(A_F)\right)^2}\right] \\
&= \mathbb{E}\left[\frac{-\operatorname{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \operatorname{tr}(A_F)}\right] + 2\mathbb{E}\left[\frac{-\mathcal{Z}'_{d,2,F}D_F\Omega_{2,F}B_F\mathcal{Z}_{d,2,F}}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \operatorname{tr}(A_F)\right)^2}\right]
\end{aligned} \tag{3.C.85}$$

where the second equality is by (3.C.83). By definition,

$$\begin{aligned}
&-\mathcal{Z}'_{d,2,F}D_F\Omega_{2,F}B_F\mathcal{Z}_{d,2,F} \\
&= -\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\Omega_{2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}
\end{aligned}$$

$$= \mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Sigma_{1,F} - \Sigma_{2,F})\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}$$

$$\leq \rho_{\max}(\Upsilon^{1/2}(\Sigma_{1,F} - \Sigma_{2,F})\Upsilon^{1/2})(\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F})$$

$$= \rho_{\max}(A_F)\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}, \tag{3.C.86}$$

where the last equality is by $\rho_{\max}(\Upsilon^{1/2}(\Sigma_{1,F}-\Sigma_{2,F})\Upsilon^{1/2}) = \rho_{\max}(\Upsilon(\Sigma_{1,F}-\Sigma_{2,F}))$. Combining the results in (3.C.85) and (3.C.86), we get

$$\begin{aligned}
J_{1,F} \leq & \mathbb{E}\left[\frac{-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)}\right] + 2\mathbb{E}\left[\frac{\rho_{\max}(A_F)\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right)^2}\right] \\
= & \mathbb{E}\left[\frac{-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)}\right] \\
& + 2\mathbb{E}\left[\frac{\left[\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A)\right]\rho_{\max}(A_F) - \mathrm{tr}(A_F)\rho_{\max}(A_F)}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right)^2}\right] \\
= & \mathbb{E}\left[\frac{2\rho_{\max}(A_F) - \mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{2\rho_{\max}(A_F)\mathrm{tr}(A_F)}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right)^2}\right].
\end{aligned} \tag{3.C.87}$$

Next, note that

$$\begin{aligned}
J_{2,F} = & \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right|^2}\right] \\
= & \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F) - \mathrm{tr}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right|^2}\right] \\
= & \mathbb{E}\left[\frac{1}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \mathrm{tr}(A_F)\right|^2}\right].
\end{aligned} \tag{3.C.88}$$

Combining (3.C.87) and (3.C.88), we obtain that

$$
\begin{aligned}
g(h_{d,F}) =& 2\mathrm{tr}(A_F)J_{1,F} + \mathrm{tr}(A_F)^2 J_{2,F} \\
\leq& 2\mathrm{tr}(A_F)\left(\mathbb{E}\left[\frac{2\rho_{\max}(A_F)-\mathrm{tr}(A_F)}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right]-\mathbb{E}\left[\frac{2\mathrm{tr}(A_F)\rho_{\max}(A_F)}{\left|\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]\right) \\
&+\mathrm{tr}(A)^2\left(\mathbb{E}\left[\frac{1}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right]-\mathbb{E}\left[\frac{\mathrm{tr}(A_F)}{\left|\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]\right) \\
=& \mathbb{E}\left[\frac{\mathrm{tr}(A_F)\left(4\rho_{\max}(A_F)-\mathrm{tr}(A_F)\right)}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right]-\mathbb{E}\left[\frac{\mathrm{tr}(A_F)^2\left(4\rho_{\max}(A_F)+\mathrm{tr}(A_F)\right)}{\left|\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right].
\end{aligned}
\tag{3.C.89}
$$

For all G_2 and Ω_2 such that $h = (d, \mathrm{vec}(G_2)', \mathrm{vech}(\Omega_2)') \in H$, we have $G_2 = G_{2,F}$ and $\Omega_2 = \Omega_{2,F}$ for some $F \in \mathcal{F}$ by the definition of H. Hence, we have desired result if $\mathrm{tr}(A_F) > 0$ and $4\rho_{\max}(A_F) - \mathrm{tr}(A_F) \leq 0$ for $\forall F \in \mathcal{F}$. □

3.D Proof of the Results in Section 3.6

Lemma 3.D.1. *Suppose that Assumption 3.6.1 holds. Consider $\{F_n\}$ such that $\overline{v}_{F_n} \to \overline{v}_F$ for some $\overline{v}_F \in \overline{\Lambda}$ and $n^{1/2}\delta_{F_n} \to d$ with $\|d\| \leq r_2^{1/2}D$. We have*

$$\lim_{\zeta\to\infty}\lim_{n\to\infty}\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] = g(h)$$

where $g(h) = \mathbb{E}\left[\overline{\xi}_F'\Upsilon\overline{\xi}_F - \xi_{1,F}'\Upsilon\xi_{1,F}\right]$.

Proof of Lemma 3.D.1. For the sequence of DGPs $\{F_n\}$ considered in the lemma, by Assumptions 3.6.1.(i), 3.6.1.(ii) and 3.6.1.(iv), we can use Lemma 3.B.8 to deduce that

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_d \begin{pmatrix} \xi_{1,F} \\ \xi_{2,F} \end{pmatrix} \tag{3.D.1}$$

where $d_0 = (\mathbf{0}_{1\times r_1}, d')'$. In the proof of Lemma 3.4.2, we have show that

$$\widehat{\Omega}_k = \Omega_{2,F} + o_p(1) \text{ and } \widehat{G}_k = G_{k,F} + o_p(1) \tag{3.D.2}$$

under $\overline{v}_{F_n} \to \overline{v}_F$, Assumptions 3.6.1.(i), 3.6.1.(ii) and 3.6.1.(iv). By (3.D.2), Assumption 3.6.1.(iv) and the Slutsky Theorem, $\widehat{\Sigma}_{1,F}$ and $\widehat{\Sigma}_{2,F}$ are consistent estimators of $\Sigma_{1,F}$ and $\Sigma_{2,F}$ respectively. By the consistency of $\widehat{\Sigma}_{1,F}$ and $\widehat{\Sigma}_{2,F}$, the weak convergence in (3.D.1) and the CMT, we deduce that

$$n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) \to_d \overline{\xi}_F. \tag{3.D.3}$$

Collecting the results in (3.D.1) and (3.D.3), and then applying the CMT and the Portmanteau Lemma, we get

$$\lim_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] \to g_\zeta(h) \tag{3.D.4}$$

where $g_\zeta(h) = \mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F, \zeta\} - \min\left\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\right\}\right]$. The asserted result follows by Lemma 3.C.8 and (3.D.4). □

Proof of Lemma 3.6.1. Let $\{F_n\}$ be a sequence such that

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] = \limsup_{n\to\infty}\left(\sup_{F\in\mathcal{F}_n} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]\right). \tag{3.D.5}$$

Such a sequence always exists by the definition of supremum. The sequence $\{\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]\colon n \geq 1\}$ may not converge. However, by the definition of limsup, there exists a subsequence of $\{n\}$, say $\{p_n\}$, such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]\colon n \geq 1\}$ converges and

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] = \limsup_{n\to\infty}\left(\sup_{F\in\mathcal{F}_n} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]\right). \tag{3.D.6}$$

Below we show that for any subsequence $\{p_n\}$ of $\{n\}$ such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo})-\ell_\zeta(\widehat{\theta}_1)]\colon n \geq 1\}$ is convergent, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ such that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] = R_\zeta(h) \text{ for some } h \in H_D \tag{3.D.7}$$

Because $\lim_n \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta}_{eo})-\ell_\zeta(\widehat{\theta}_1)] = \lim_n \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo})-\ell_\zeta(\widehat{\theta}_1)]$, which combined with (3.D.6) and (3.D.7) implies that

$$\limsup_{n\to\infty}\left(\sup_{F\in\mathcal{F}_n} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)]\right) = R_\zeta(h) \text{ for some } h \in H_D. \tag{3.D.8}$$

The desired result in (3.6.5) follows immediately by (3.D.8).

To show that there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ such that (3.D.7) holds, it suffices to show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}$, there exists a subsequence

$\{p_n^*\}$ of $\{p_n\}$ for which we have

$$(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to d \text{ for } \|d\| \leq D \text{ and } \overline{v}_{F_{p_n^*}} \to \overline{v}_F \tag{3.D.9}$$

for some $F \in \mathcal{F}$. By (3.D.9), we can use Lemma 3.D.1 to deduce that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta}_{eo}) - \ell_\zeta(\widehat{\theta}_1)] = R_\zeta(h_{F,d}) \tag{3.D.10}$$

for the sequence of DGPs $\{F_{p_n^*}\}$ satisfies (3.D.9). Moreover, we have $h_{F,d} \in H_D$ by the definition of H_D, which together with (3.D.10) proves (3.D.7).

Finally, we show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}$, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ for which (3.D.9) holds. Let $\delta_{p_n,j}$ denote the j-th component of δ_{p_n} and $p_{1,n} = p_n$ for any $n \geq 1$. For $j = 1$, we have $|p_{j,n}^{1/2}\delta_{p_{j,n},j}| \leq D$ for any n by Assumption 3.6.1.(vi). Hence there is some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\delta_{p_{j+1,n},j} \to d_j$ for some $|d_j| \leq D$. As r^* is a fixed positive integer, we can apply the same arguments successively for $j = 1, ..., r^*$ to obtain a subsequence $\{p_{r^*,n}\}$ of $\{p_n\}$ such that $(p_{r^*,n})^{1/2}\delta_{p_{r^*,n}} \to d$ with $\|d\| \leq D$. By Assumptions 3.3.2.(i) and 3.6.1.(v), $\overline{\Lambda}$ is a compact set. Hence, there is a subsequence $\{p_n^*\}$ of $\{p_{r^*,n}\}$ such that $\overline{v}_{F_{p_n^*}} \to \overline{v}_F$, which finishes the proof of (3.D.9). $\square$

Proof of Theorem 3.6.2. Let $\tau_D = D$. By (3.C.89) in the proof of Theorem 3.5.2

$$g(h_{d,F}) \leq \mathbb{E}\left[\frac{\text{tr}(A_F)\,(4\rho_{\max}(A_F) - \text{tr}(A_F))}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \text{tr}(A_F)}\right] - \mathbb{E}\left[\frac{\text{tr}(A_F)^2\,(4\rho_{\max}(A_F) + \text{tr}(A_F))}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F} + \text{tr}(A_F)\right|^2}\right]. \tag{3.D.11}$$

By Jensen's inequality,

$$\mathbb{E}\left[\frac{1}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right]\geq\frac{1}{\mathrm{tr}(d'B_Fd)+2\mathrm{tr}(A_F)},\tag{3.D.12}$$

and similarly

$$\mathbb{E}\left[\frac{1}{\left|\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]\geq\frac{1}{\left|\mathrm{tr}(d'B_Fd)+2\mathrm{tr}(A_F)\right|^2}.\tag{3.D.13}$$

Hence

$$\begin{aligned}g(h)\leq&\frac{\mathrm{tr}(A_F)\left(4\rho_{\max}(A_F)-\mathrm{tr}(A_F)\right)}{\mathrm{tr}(d'B_Fd)+2\mathrm{tr}(A_F)}-\frac{\mathrm{tr}(A_F)^2\left(4\rho_{\max}(A_F)+\mathrm{tr}(A_F)\right)}{\left|\mathrm{tr}(d'B_Fd)+2\mathrm{tr}(A_F)\right|^2}\\\leq&\frac{\mathrm{tr}(A_F)\left(4\rho_{\max}(A_F)-\mathrm{tr}(A_F)\right)}{\tau_D^2\rho_{\max}(B_F)+2\mathrm{tr}(A_F)}-\frac{\mathrm{tr}(A_F)^2\left(4\rho_{\max}(A_F)+\mathrm{tr}(A_F)\right)}{\left|\tau_D^2\rho_{\max}(B_F)+2\mathrm{tr}(A_F)\right|^2}.\end{aligned}\tag{3.D.14}$$

As $\mathrm{tr}(A_F)>4\rho_{\max}(A_F)$, $\rho_{\max}(A_F)>0$, $\rho_{\max}(B_F)<\infty$ and $\tau_D^2<\infty$, (3.D.14) immediately implies that $g(h)<0$ for any $h\in H_D$. □

3.E Asymptotic Risk of the Pre-test GMM Estimator

In this section, we establish similar results in Theorem 3.5.1 for the pre-test GMM estimator based on the J-test statistic. The pre-test estimator is defined as

$$\widehat{\theta}_{pre}=1\{J_n>c_\alpha\}\widehat{\theta}_1+1\{J_n\leq c_\alpha\}\widehat{\theta}_2,\tag{3.E.1}$$

where $J_n = n\overline{g}_2(\widehat{\theta}_2)'(\widehat{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2)$ and c_α is the $(1-\alpha)$th quantile of the chi-squared distribution with degree of freedom $r_2 - d_\theta$.

Theorem 3.E.1. *Suppose that Assumptions 3.3.1-3.3.3 hold. The bounds of the asymptotic risk difference satisfy*

$$
\begin{aligned}
Asy\underline{RD}(\widehat{\theta}_{pre}, \widehat{\theta}_1) &= \min\left\{\inf_{h\in H}\left[g_p(h)\right], 0\right\}, \\
Asy\overline{RD}(\widehat{\theta}_{pre}, \widehat{\theta}_1) &= \max\left\{\sup_{h\in H}\left[g_p(h)\right], 0\right\},
\end{aligned}
$$

where $g_{p,\zeta}(h) \equiv \mathbb{E}\left[\min\{\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F}, \zeta\} - \min\left\{\xi_{1,F}\Upsilon\xi'_{1,F}, \zeta\right\}\right]$ *and* $g_p(h) \equiv \mathbb{E}[\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F} - \xi_{1,F}\Upsilon\xi'_{1,F}]$.

Proof of Theorem 3.E.1. The two equalities and inequalities in the theorem follow by the same arguments in the proof of Theorem 3.5.1 with Lemma 3.4.2 for $\widehat{\theta}_{eo}$ replaced by Lemma 3.E.2 for $\widehat{\theta}_{pre}$, and Lemma 3.C.8 replaced by Lemma 3.E.4. Its proof is hence omitted. □

By Definition,

$$
\begin{aligned}
\mathbb{E}[\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F}] =& \mathbb{E}[\mathcal{Z}'_{d,2,F}\Gamma^{*\prime}_{1,F}\Upsilon\Gamma^{*}_{1,F}\mathcal{Z}_{d,2,F}] \\
&+ \mathbb{E}[\overline{\omega}^2_{p,F}\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}] \\
&+ 2\mathbb{E}[\overline{\omega}_{p,F}\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F}] \\
=& tr(\Upsilon\Sigma_{1,F}) + 2\mathbb{E}[\overline{\omega}_{p,F}\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F}] \\
&+ \mathbb{E}[\overline{\omega}^2_{p,F}\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}] \qquad (3.E.2)
\end{aligned}
$$

The asymptotic risk of the pre-test estimator $\widehat{\theta}_p$ in Figure 2 is simulated based on the formula in (3.E.2).

The following lemma provides the asymptotic distribution of the pre-test GMM estimator under various sequence of DGPs, which is used to show Theorem 3.E.1.

Lemma 3.E.2. *Suppose that Assumptions 3.3.1-3.3.3 hold. Consider* $\{F_n\}$ *such that* $G_{2,F_n} \to G_{2,F}$*,* $\Omega_{2,F_n} \to \Omega_{2,F}$ *and* $M_{2,F_n} \to M_{2,F}$ *for some* $F \in \mathcal{F}$*.*

(a) *If* $n^{1/2}\delta_{F_n} \to d$ *for some* $d \in \mathbb{R}^{r^*}$*, then*

$$J_n \to_d J_\infty(h_{d,F}) \equiv (\mathcal{Z}_{2,F} + d_0)' L_F (\mathcal{Z}_{2,F} + d_0),$$

where $L_F \equiv \Omega_{2,F}^{-1} - \Omega_{2,F}^{-1} G_{2,F} \left(G_{2,F}' \Omega_{2,F}^{-1} G_{2,F} \right)^{-1} G_{2,F}' \Omega_{2,F}^{-1}$ *and* $d_0 = (0_{r_1 \times 1}, d')'$*, and*

$$n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_d \overline{\xi}_{p,F} \equiv (1 - \overline{\omega}_{p,F})\xi_{1,F} + \overline{\omega}_{p,F}\xi_{2,F}$$

where $\overline{\omega}_{p,F} = 1\{J_\infty(h_{d,F}) \le c_\alpha\}$*.*

(b) *If* $||n^{1/2}\delta_{F_n}|| \to \infty$*, then* $\overline{\omega}_{p,F} \to_p 0$ *and* $n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_d \xi_{1,F}$*.*

Proof of Lemma 3.E.2. (a) By Assumption 3.3.2.(i), (3.B.35) and (3.B.38),

$$\begin{aligned}
\overline{g}_2(\widehat{\theta}_2) &= \overline{g}_2(\theta_{F_n}) + [G_{2,F_n}(\theta_{F_n}) + o_p(1)]\,(\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2}) \\
&= \overline{g}_2(\theta_{F_n}) + G_{2,F_n}\Gamma_{2,F_n}\overline{g}_2(\theta_{F_n}) + o_p(n^{-1/2}) \\
&= (I_{r_2} + G_{2,F_n}\Gamma_{2,F_n})\overline{g}_2(\theta_{F_n}) + o_p(n^{-1/2}),
\end{aligned} \tag{3.E.3}$$

which implies that

$$J_n = n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) + o_p(1) \tag{3.E.4}$$

By $n^{1/2}\delta_{F_n} \to d$ and Lemma 3.B.1.(v),

$$n^{-1/2}\Omega_{2,F_n}^{-1/2}\overline{g}_2(\theta_{F_n}) = \Omega_{2,F_n}^{-1/2}\mu_n(g_2(W,\theta_{F_n})) + \Omega_{2,F_n}^{-1/2} n^{1/2}\delta_{F_n} \to_d \mathcal{Z} + \Omega_{2,F}^{-1/2} d_0 \tag{3.E.5}$$

where $d_0' = (0_{1\times r_1}, d')$ and $\mathcal{Z}$ is a $r_2 \times 1$ standard normal random vector. By $v_{F_n} \to v_F$, (3.E.4), (3.E.5) and the CMT,

$$J_n \to_d (\mathcal{Z}_{2,F} + d_0)' L_F (\mathcal{Z}_{2,F} + d_0). \tag{3.E.6}$$

By Lemma 3.4.1.(a) implies that

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_d \xi_{1,F} \text{ and } n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \to_d \xi_{2,F}, \tag{3.E.7}$$

which together with (3.E.6) and the CMT implies that

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) &= 1\{J_n > c_\alpha\} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + 1\{J_n \leq c_\alpha\} n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \\ &\to_d (1 - \overline{\omega}_{p,F})\xi_{1,F} + \overline{\omega}_{p,F}\xi_{2,F}, \end{aligned} \tag{3.E.8}$$

which finishes the proof of the claim in (a).

(b) There are two cases to consider: (i) $||\delta_{F_n}|| > C^{-1}$; and (ii) $||\delta_{F_n}|| \to 0$. We first consider case (i). As $\overline{g}_1(\widehat{\theta}_2)$ is a subvector of $\overline{g}_2(\widehat{\theta}_2)$,

$$\begin{aligned} J_n &= n\overline{g}_2(\widehat{\theta}_2)'(\widehat{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \\ &\geq n\rho_{\max}^{-1}(\widehat{\Omega}_2)\overline{g}_2(\widehat{\theta}_2)'\overline{g}_2(\widehat{\theta}_2) \\ &\geq n\rho_{\max}^{-1}(\widehat{\Omega}_2)\overline{g}_1(\widehat{\theta}_2)'\overline{g}_1(\widehat{\theta}_2). \end{aligned} \tag{3.E.9}$$

By (3.B.51) and (3.B.52),

$$\left\|\widehat{\theta}_2 - \theta_{F_n}\right\| \geq C^{-1} \text{ with probability approaching 1,} \tag{3.E.10}$$

which together with Assumption 3.3.1.(ii) and Lemma 3.B.1.(i) implies that

$$\overline{g}_1(\widehat{\theta}_2) = M_{1,F}(\widehat{\theta}_2) + o_p(1) \geq C \tag{3.E.11}$$

with probability approaching 1. By (3.B.62) and Assumption 3.3.2.(i), we have

$$\rho_{\max}(\widehat{\Omega}_2) \leq C \text{ with probability approaching 1.} \tag{3.E.12}$$

Combining the results in (3.E.9), (3.E.11) and (3.E.12), we deduce that

$$J_n \geq nC^{-1} \text{ with probability approaching 1,} \tag{3.E.13}$$

which immediately implies that

$$\overline{\omega}_{p,F} = 1\{J_n \leq c_\alpha\} = 0 \tag{3.E.14}$$

with probability approaching 1, as c_α is a fixed constant. By Lemma 3.4.1.(b), (3.E.14) and the assumption that Θ is bounded, we have

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) &= 1\{J_n > c_\alpha\} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + 1\{J_n \leq c_\alpha\} n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \\ &= 1\{J_n > c_\alpha\} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + o_p(1) \rightarrow_d \xi_{1,F} \end{aligned} \tag{3.E.15}$$

where the convergence in distribution is by the CMT.

We next consider the case that $||\delta_{F_n}|| \rightarrow 0$ and $||n^{1/2}\delta_{F_n}|| \rightarrow \infty$. In the proof of Lemma 3.4.1, we have shown that $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$ and (3.E.3) hold in this case. It is clear that

$$n^{1/2}\overline{g}_2(\theta_{F_n}) = \mu_n(g_2(W, \theta_{F_n}) + \begin{pmatrix} 0_{r_1 \times 1} \\ n^{1/2}\delta_{F_n} \end{pmatrix} \tag{3.E.16}$$

which implies that

$$\begin{aligned} n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) &= [\mu_n(g_2(W, \theta_{F_n})]' L_{F_n} [\mu_n(g_2(W, \theta_{F_n})] \\ &+ 2 \begin{pmatrix} 0_{1 \times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n} [\mu_n(g_2(W, \theta_{F_n})] \\ &+ \begin{pmatrix} 0_{1 \times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n} \begin{pmatrix} 0_{1 \times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix}'. \end{aligned} \tag{3.E.17}$$

By Lemma 3.B.1.(v) and Assumptions 3.3.2.(i)-(ii),

$$[\mu_n(g_2(W,\theta_{F_n})]' H_{F_n}[\mu_n(g_2(W,\theta_{F_n})] = O_p(1). \tag{3.E.18}$$

We shall show that for any $d_0 = (0_{1\times r_1}, d')'$ for $d \in \mathbb{R}^{r_*}$ with $\|d\| = 1$,

$$d_0' L_{F_n} d_0 \geq C^{-1} \tag{3.E.19}$$

By definition, L_{F_n} has d_θ many zero eigenvalues and $r_2 - d_\theta$ many of eigenvalues of ones. The matrix G_{2,F_n} contains the d_θ many eigenvectors of the zero eigenvalues of L_{F_n}, because

$$L_{F_n} G_{2,F_n} = 0_{r_2\times 1} \text{ and } \rho_{\min}(G_{2,F_n}' G_{2,F_n}) \geq C^{-1}. \tag{3.E.20}$$

Let $G_{\perp,F_n}$ denote the orthogonal complement of G_{2,F_n} with $G_{\perp,F_n}' G_{\perp,F_n} = I_{r_2-d_\theta}$. Then we have

$$\begin{pmatrix} G_{1,F_n} \\ G_{r^*,F_n} \end{pmatrix} a_1 + \begin{pmatrix} G_{1,\perp,F_n} \\ G_{r^*,\perp,F_n} \end{pmatrix} a_2 = \begin{pmatrix} 0_{r_1\times 1} \\ d \end{pmatrix}. \tag{3.E.21}$$

As $\rho_{\min}(G_{1,F_n}' G_{1,F_n}) \geq C^{-1}$ by Assumptions 3.3.2.(ii), we have

$$a_1 = -(G_{1,F_n}' G_{1,F_n})^{-1} G_{1,F_n}' G_{1,\perp,F_n} a_2 \tag{3.E.22}$$

and

$$(G_{r^*,\perp,F_n} - G_{r^*,F_n}(G_{1,F_n}' G_{1,F_n})^{-1} G_{1,F_n}' G_{1,\perp,F_n}) a_2 = d. \tag{3.E.23}$$

Let $L_{F_n} = G_{r^*,\perp,F_n} - G_{r^*,F_n}(G'_{1,F_n}G_{1,F_n})^{-1}G'_{1,F_n}G_{1,\perp,F_n}$. By $\rho_{\min}(G'_{1,F_n}G_{1,F_n}) \geq C^{-1}$, Assumptions 3.3.2.(i), (3.E.23) and the Cauchy-Schwarz inequality,

$$\|d\|^2 = a_2' L_{F_n} L'_{F_n} a_2 \leq C \|a_2\|^2 \tag{3.E.24}$$

which together with $\|d\| = 1$ implies that

$$\|a_2\|^2 \geq C^{-1}. \tag{3.E.25}$$

Using (3.E.20), (3.E.21) and (3.E.25), we deduce that

$$\begin{aligned} d_0' L_{F_n} d_0 &= (G_{2,F_n}a_1 + G_{\perp,F_n}a_2)' L_{F_n} (G_{2,F_n}a_1 + G_{\perp,F_n}a_2) \\ &= a_2' G'_{\perp,F_n} L_{F_n} G_{\perp,F_n} a_2 = \|a_2\|^2 \geq C^{-1} \end{aligned} \tag{3.E.26}$$

which proves (3.E.19). By (3.E.19),

$$\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n} \begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix}' \geq C^{-1} n ||\delta_{F_n}||^2 \tag{3.E.27}$$

which together with $n||\delta_{F_n}||^2 \to \infty$ implies that

$$\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n} \begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix}' \to \infty. \tag{3.E.28}$$

Collecting the results in (3.E.17), (3.E.18) and (3.E.18), we deduce that $n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) \to$

∞, which together with (3.E.4) implies that

$$n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) \to \infty. \tag{3.E.29}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} n\overline{g}_2(\widehat{\theta}_2)' \Omega_{2,F_n}^{-1} \overline{g}_2(\widehat{\theta}_2) &\geq n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) \\ &\quad - (n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}))^{1/2} o_p(n^{-1/2}) + o_p(n^{-1}) \end{aligned}$$

which together with (3.E.29) implies that

$$n\overline{g}_2(\widehat{\theta}_2)' \Omega_{2,F_n}^{-1} \overline{g}_2(\widehat{\theta}_2) \to_p \infty. \tag{3.E.30}$$

It is clear that

$$J_n = n\overline{g}_2(\widehat{\theta}_2)'(\widehat{\Omega}_2)^{-1} \overline{g}_2(\widehat{\theta}_2) \geq \rho_{\max}^{-1}(\widehat{\Omega}_2) \rho_{\min}(\Omega_{2,F_n}) n\overline{g}_2(\widehat{\theta}_2)' \Omega_{2,F_n}^{-1} \overline{g}_2(\widehat{\theta}_2) \tag{3.E.31}$$

which together with Assumptions 3.3.2.(ii), (3.E.12) and (3.E.30) implies that $J_n \geq nC^{-1}$ with probability approaching 1. Using the same arguments in showing (3.E.15), we deduce that

$$n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_d \xi_{1,F}. \tag{3.E.32}$$

This finishes the proof. □

Lemma 3.E.3. *Under Assumptions 3.3.2, we have*

$$\sup_{h\in H}\mathbb{E}[(\overline{\xi}_{p,F}'\Upsilon\overline{\xi}_{p,F})^2]\leq C. \tag{3.E.33}$$

Proof of Lemma 3.E.3. By the same arguments in showing (3.C.44), we have

$$(\overline{\xi}_{p,F}'\Upsilon\overline{\xi}_{p,F})^2\leq 8(\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F})^2+8(\overline{\omega}_{p,F}^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2. \tag{3.E.34}$$

By the first inequality in (3.C.39), we have $\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq C$. Hence by (3.E.34), to show the inequality in (3.E.33), it is sufficient to prove that

$$\sup_{h\in H}\mathbb{E}[(\overline{\omega}_{p,F}^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2]\leq C. \tag{3.E.35}$$

By definition,

$$\overline{\omega}_{p,F}=I\{J_\infty(h_{d,F})\leq c_\alpha\}=I\{\mathcal{Z}_{d,2,F}'L_F\mathcal{Z}_{d,2,F}\leq c_\alpha\}. \tag{3.E.36}$$

By the simple inequality $(a+b)^2\geq a^2/2-2b^2$,

$$(z+d_0)'L_F(z+d_0)\geq d_0'L_Fd_0/2-2z'L_Fz \tag{3.E.37}$$

for any $z\in\mathbb{R}$, which together with Assumptions 3.3.2 and (3.E.19) implies that

$$(z+d_0)'L_F(z+d_0)\geq\|d\|^2/C-2z'L_Fz\geq\|d\|^2/C-C\|z\|^2. \tag{3.E.38}$$

Under Assumptions 3.3.2, $\|B_F\| \leq C$ for any $F \in \mathcal{F}$ which together with the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$ implies that

$$(z+d_0)'B_F(z+d_0) \leq 2C(\|d\|^2 + \|z\|^2) \tag{3.E.39}$$

for any $z \in \mathbb{R}$. Collecting the results in (3.E.37) and (3.E.39), we get

$$\begin{aligned}
&I\{(z+d_0)'L_F(z+d_0) \leq c_\alpha\}z'B_F z \\
&\leq 2I\{\|d\|^2 \leq c_\alpha C + C^2\|z\|^2\}(\|d\|^2 + \|z\|^2) \\
&\leq 2(c_\alpha C + (C^2+1)\|z\|^2)
\end{aligned} \tag{3.E.40}$$

which implies that

$$\begin{aligned}
\sup_{h\in H} \mathbb{E}[(\overline{\omega}_{p,F}^2 \mathcal{Z}_{d,2,F}' B_F \mathcal{Z}_{d,2,F})^2] &\leq 4\mathbb{E}[(c_\alpha C + (C^2+1)\mathcal{Z}'\mathcal{Z})^2] \\
&\leq C(c_\alpha + \mathbb{E}[(\mathcal{Z}'\mathcal{Z})^2]) = C(c_\alpha + 3r_2).
\end{aligned} \tag{3.E.41}$$

This finishes the proof. □

Lemma 3.E.4. *Under Assumptions 3.3.2, we have*

$$\lim_{\zeta\to\infty} \sup_{h\in H} [|g_{p,\zeta}(h) - g_p(h)|] = 0 \tag{3.E.42}$$

where $\sup_{h\in H}[|g_p(h)|] \leq C$.

Proof of Lemma 3.E.4. The proof follows the same arguments of the proof of 3.C.8 with the second inequality in (3.C.39) replaced by (3.E.33). □

3.F Illustration in Gaussian Location Model

Suppose that we have one observation $(X_1', X_2')'$ from the normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta + d \end{pmatrix}, I_{2k}\right) \tag{3.F.1}$$

where θ and d are $k \times 1$ vectors and I_{2k} is a $2k \times 2k$ identity matrix. We are interested in estimating θ.

Let Υ be the $k \times k$ identity matrix. The conservative GMM estimator $\widehat{\theta}_1 = X_1$ has risk $\mathrm{tr}(\Upsilon I_k) = k$. On the other hand, the aggressive GMM estimator is $\widehat{\theta}_2 = (X + Y)/2$, which has risk $k/2 + \|d\|^2/4$. The empirical optimal weight defined in (3.4.7) becomes

$$\widetilde{\omega}_{eo} = \frac{2k}{2k + (Y - X)'(Y - X)}, \tag{3.F.2}$$

which together with the conservative and aggressive GMM estimators leads to the averaging estimator

$$\widehat{\theta}_{eo} = X + \frac{k}{2k + (Y - X)'(Y - X)}(Y - X). \tag{3.F.3}$$

Lemma 3.F.1. *The averaging estimator* $\widehat{\theta}_o$ *defined in (3.F.3) satisfies*

$$\mathbb{E}\left[||\widehat{\theta}_{eo}-\theta||^2-||\widehat{\theta}_1-\theta||^2\right] \leq \frac{-2k^2(k+4)}{(4k+\|d\|^2)^2}+\mathbb{E}\left[\frac{k(4-k)}{2k+(Y-X)'(Y-X)}\right]. \tag{3.F.4}$$

The inequality (3.F.4) shows that the risk of the averaging estimator is strictly smaller than the conservative estimator if $k \geq 4$, for any θ and any d.

Proof of Lemma 3.F.1. By definition

$$\begin{aligned}&\mathbb{E}\left[||\widehat{\theta}_{eo}-\theta||^2\right]-\mathbb{E}\left[||\widehat{\theta}_1-\theta||^2\right]\\ &=\mathbb{E}\left[\frac{k^2(Y-X)'(Y-X)}{(2k+(Y-X)'(Y-X))^2}\right]\\ &+2\mathbb{E}\left[\frac{k(X-\theta)'(Y-X)}{2k+(Y-X)'(Y-X)}\right].\end{aligned} \tag{3.F.5}$$

Let $X^*=X-\theta$, $Y=Y-\theta$ and $Z^*=(X^{*\prime},Y^{*\prime})'$. Then we can write

$$\mathbb{E}\left[\frac{(X-\theta)'(Y-X)}{2k+(Y-X)'(Y-X)}\right]=\mathbb{E}\left[\frac{X^{*\prime}(Y^*-X^*)}{2k+(Y^*-X^*)'(Y^*-X^*)}\right]=\mathbb{E}\left[\frac{Z^{*\prime}D_1Z^*}{2k+Z^{*\prime}D_2Z^*}\right] \tag{3.F.6}$$

where

$$D_1=\begin{pmatrix}-I_k & 0_k\\ I_k & 0_k\end{pmatrix} \text{ and } D_2=\begin{pmatrix}I_k & -I_k\\ -I_k & I_k\end{pmatrix}. \tag{3.F.7}$$

Note that

$$\mathbb{E}\left[D_1Z^*Z^{*\prime}D_1'\right]=D_2 \tag{3.F.8}$$

by definition and the Gaussian assumption. Let $\eta(x) = x/(x'D_2x + 2k)$. Then its derivative is

$$\frac{\partial \eta(x)'}{\partial x} = \frac{1}{x'D_2x + 2k} I_k - \frac{2}{(x'D_2x + 2k)^2} D_2xx'. \tag{3.F.9}$$

By Lemma 1 of Hansen (2016), which is a matrix version of the Stein's Lemma (Stein, 1956),

$$\begin{aligned}
J_1 = \mathbb{E}\left(\eta(Z^*)'D_1Z^*\right) &= \mathbb{E}\left[\operatorname{tr}\left(\frac{\partial \eta(Z^*)'}{\partial x} D_1\right)\right] \\
&= \mathbb{E}\left[\frac{\operatorname{tr}(D_1)}{2k + Z^{*\prime}D_2Z^*}\right] - 2\mathbb{E}\left[\frac{\operatorname{tr}(D_2Z^*Z^{*\prime}D_1)}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{-k}{2k + Z^{*\prime}D_2Z^*}\right] - 2\mathbb{E}\left[\frac{Z^{*\prime}D_1D_2Z^*}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{-k}{2k + Z^{*\prime}D_2Z^*}\right] + 2\mathbb{E}\left[\frac{Z^{*\prime}D_2Z^*}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{2-k}{2k + Z^{*\prime}D_2Z^*}\right] + \mathbb{E}\left[\frac{-4k}{(2k + Z^{*\prime}D_2Z^*)^2}\right]
\end{aligned}$$

where fourth equality is by the following result

$$D_1D_2 = \begin{pmatrix} -I_k & I_k \\ I_k & -I_k \end{pmatrix} = -D_2. \tag{3.F.10}$$

Moreover,

$$\mathbb{E}\left[\frac{k^2(Y-X)'(Y-X)}{(2k + (Y-X)'(Y-X))^2}\right] = \mathbb{E}\left[\frac{k^2}{2k + Z^{*\prime}D_2Z^*}\right] - \mathbb{E}\left[\frac{2k^3}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \tag{3.F.11}$$

which together with (3.F.10) implies that

$$\begin{aligned}
&\mathbb{E}\left[||\widehat{\theta}_{eo}-\theta||^2\right]-\mathbb{E}\left[||\widehat{\theta}_1-\theta||^2\right] \\
&=\mathbb{E}\left[\frac{2k(2-k)+k^2}{2k+Z^{*\prime}D_2Z^*}\right]-\mathbb{E}\left[\frac{2k^3+8k^2}{(2k+Z^{*\prime}D_2Z^*)^2}\right] \\
&=\mathbb{E}\left[\frac{k(4-k)}{2k+Z^{*\prime}D_2Z^*}\right]-\mathbb{E}\left[\frac{2k^2(k+4)}{(2k+Z^{*\prime}D_2Z^*)^2}\right].
\end{aligned} \tag{3.F.12}$$

By Jensen's inequality,

$$\mathbb{E}\left[\frac{2k^2(k+4)}{(2k+Z^{*\prime}D_2Z^*)^2}\right]\geq\frac{2k^2(k+4)}{(2k+\mathbb{E}\left[Z^{*\prime}D_2Z^*\right])^2}=\frac{2k^2(k+4)}{(4k+\|d\|^2)^2}. \tag{3.F.13}$$

The asserted result follows by combining the results in (3.F.12) and (3.F.13). □

Chapter 4

Synthetic Control and Inference

4.1 Introduction

Synthetic control method, proposed and discussed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010), is a very useful way of conducting comparison studies when exact matches are unavailable. Estimation of treatment effects usually takes the form of comparing outcomes between the treated unit and the control unit. Common sense suggests that, for the comparison to be meaningful, the control unit needs to be similar to the treatment unit in various dimensions. Such a requirement may not be satisfied in many observational studies. In some cases, availability of panel data makes such comparisons reasonable, the difference-in-difference method being a very well-known example. The difference-in-difference method requires a very specific set of assumptions, i.e., the common trend assumption, which may not be plausible for many applications. The synthetic control method offers a sensible generalization of the difference-in-difference. The synthetic control is a combination of control outcomes, where the combination is manufactured by analyzing the pre-intervention outcomes.

For the purpose of statistical inference with synthetic control, i.e., confidence interval and hypothesis testing, various versions of placebo test are often adopted. The idea underlying the placebo tests is similar to the usual permutation tests, where the critical value of a test statistic is computed under all possible permutations of the "treatment" assignments in the control unit. The idea of permutation test is very intuitive and attractive. Applying the synthetic control method to every potential control unit presumably allows researchers to assess the distribution of a test statistic under the null hypothesis of no treatment effects, and the inference is seemingly exact in the sense that the burden of asymptotic approximation can be obviated.

The purpose of this paper is very specific. We ask whether the permutation test is a reasonable idea in the context of the synthetic control method, and argue that the intuitive appeal of the permutation test is misplaced. The validity of permutation tests usually requires a certain symmetry assumption, which is often violated in the context of synthetic control studies. Using Monte Carlo simulations, we document the size distortion of the permutation tests. We go on to discuss a few alternative methods of inference.

Alberto Abadie kindly pointed out that the placebo test in synthetic control is often based on randomization inference idea, under which the symmetry restriction is built-in, while our analysis is predicated on the usual random sampling perspective, which leads to the violation of symmetry. It would be useful to understand the exact mechanism through which such difference manifest itself.

4.2 Synthetic Control: An Overview

In this section, we provide an overview of the synthetic control method. We heavily borrow discussions in Abadie, Diamond, and Hainmueller (2010) and Doudchenko and Imbens (2016). Our discussion does address the source of identification, which seems absent in the current literature, with the exception of Ferman and Pinto (2016), who showed that the estimated synthetic control weights will generally not converge to the weights that reconstruct the factor loadings of the treated unit, even when the number of pre-intervention periods goes to infinity.

Consider a panel data with $J+1$ cross sectional units observed over the time periods $t=1,\ldots,T$. Units $j=1,\ldots,J$ are the control units which do not receive the treatment in either of the time periods. The unit $j=0$ receives no treatment in periods $1,\ldots,T_0$ and the active treatment in time periods $t=T_0+1,\ldots,T$. For simplicity, we will often assume that $T=T_0+1$. The outcome variable $Y_{j,t}$ is such that $Y_{j,t}=Y_{j,t}(1)$ if the jth unit receives treatment in time t, and $Y_{j,t}=Y_{j,t}(0)$ otherwise. Obviously

$$\begin{aligned} Y_{j,t} &= Y_{j,t}(0), \quad j=1,\ldots,J; t=1,\ldots,T, \\ Y_{0,t} &= Y_{0,t}(0), \quad t=1,\ldots,T_0, \\ Y_{0,t} &= Y_{0,t}(1), \quad t=T_0+1,\ldots,T. \end{aligned}$$

The idea underlying the synthetic control is that if there were some weights[1] $\hat{\omega}_1,\ldots,\hat{\omega}_J$ such that

$$Y_{0,t} \approx \sum_{j=1}^{J} \hat{\omega}_j Y_{j,t} \tag{4.2.1}$$

during the pre-intervention periods ($t=1,\ldots,T_0$), then $\sum_{j=1}^{J} \hat{\omega}_j Y_{j,t}$ can be used as a (synthetic) control for $Y_{0,t}$ during the post-intervention period ($t=T_0+1,\ldots,T$). Abadie, Diamond, and Hainmueller (2010) and Doudchenko and Imbens (2016) discuss various methods of finding the $\hat{\omega}$'s so that the requirement (4.2.1) is satisfied.

[1] Doudchenko and Imbens (2016) also consider a slightly general requirement $Y_{0,t} \approx \alpha + \sum_{j=1}^{J} w_j Y_{j,t}$. This is a sensible way to enhance accuracy of synthetic control viewed as a point estimator. It also provides a link to the difference-in-difference estimator. Because our focus is on inferential aspects of the problem, we simplify notation and analysis by abstracting away from the intercept term.

We now present our own understanding/interpretation of the synthetic control method. Our interpretation is not the only possible one. Doudchenko and Imbens (2016) provide an in-depth analysis of many possible methods. On the other hand, our interpretation is helpful for two reasons. First, it makes a concrete interpretation of $\hat{\omega}$'s as estimates of some pseudo-parameter, say ω_*'s, along with analytic expressions of the ω_*'s, which makes it easy to understand the potential pitfalls of permutation methods afterwards. Second, it helps us to motivate a completely different method of inference exploiting a time series variation.

The nature of approximation in (4.2.1) seems to be implicitly related to the asymptotics where $T_0 \to \infty$. Abadie, Diamond, and Hainmueller (2010) (p. 504) discusses the theoretical motivation in the appendix of their paper, and shows that some term "goes to zero as the number of pretreatment periods increases."[2] The time-series type motivation, which is only implicit in their discussion, is a little more explicit in the form of the estimator. For example, a special case of the estimator discussed by Abadie, Diamond, and Hainmueller (2010) (p. 496) solves

$$\min_{\omega_1,\ldots,\omega_J} \left(\bar{Y}_0 - \sum_{j=1}^{J} \omega_j \bar{Y}_j \right)^2 \tag{4.2.2}$$

where $\bar{Y}_j = T_0^{-1} \sum_{t=1}^{T_0} Y_{j,t}$. Obviously there are multiple solutions, because any ω's with $\bar{Y}_0 = \sum_{j=1}^{J} \omega_j \bar{Y}_j$ will minimize the objective function above. Although they do not make it explicit, Abadie, Diamond, and Hainmueller (2010) avoided the problem by insisting that $\hat{\omega}$'s be chosen such that some of the other observed characteristic, say $Z_{j,t}$, also satisfy a

[2] On the same page, they also assume that their equation (2) is satisfied, i.e., $Y_{0,t} = \sum_{j=1}^{J} \hat{w}_j Y_{j,t}$ for all $t = 1, \ldots, T_0$ in our notation above. Because there are J elements of $\hat{w}$'s which is mathematically impossible if $T_0 > J$. This is a minor problem because the problem disappears if the equality is understood to hold approximately (in some informal sense).

requirement similar to (4.2.1).[3]

In order to understand the role of taking the sample average, it makes sense to consider the linear factor structure[4] as in Abadie, Diamond, and Hainmueller (2010):

$$Y_{j,t} = Y_{j,t}(0) = \alpha_j + \theta_t + \gamma_j'\delta_t + \epsilon_{j,t}, \quad t = 1, \ldots, T_0 \tag{4.2.3}$$

The factor structure is interesting because it invalidates the usual difference-in-difference methods. Suppose that θ_t, δ_t, and $\epsilon_{j,t}$ satisfy strict stationarity. Without loss of generality, we also assume that $E[\delta_t] = 0$ and $E[\epsilon_{j,t}] = 0$. We would have $\bar{Y}_j \to \alpha_j + E[\theta_t]$ in probability as $T_0 \to \infty$. Assuming that ω_{*j} satisfy

$$\alpha_0 + E[\theta_t] = \sum_{j=1}^{J} \omega_{*j}\left(\alpha_j + E[\theta_t]\right), \tag{4.2.4}$$

we can understand that the population version of the synthetic control $\sum_{j=1}^{J} \omega_{*j} Y_{j,t}$ is such that the difference $Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_{*j} Y_{j,T_0+1}(0)$ is designed to have a mean zero.

Obviously there are multiple possible values of ω_* that guarantees the zero mean property. See Doudchenko and Imbens (2016) for related discussion. In Abadie, Diamond, and Hainmueller (2010), the non-uniqueness problem seems to be solved by using a side con-

[3] Abadie, Diamond, and Hainmueller (2010) (Section 2.3) in fact consider the weighted averages $\bar{Y}_j^{(m)} \equiv \sum_t k_t^{(m)} Y_{j,t}$ $m = 1, \ldots, M$ instead of $\bar{Y}_j = T_0^{-1} \sum_{t=1}^{T_0} Y_{j,t}$ with the understanding that $\sum_t k_t^{(m)} = 1$. Because $\bar{Y}_j^{(m)}$ have the same expectation as $\bar{Y}_j$, they would have the same probability limit as long as $\bar{Y}_j^{(m)}$ converges in probability to a nonrandom constant. Therefore, their use of multiple weighted averages as discussed in Abadie, Diamond, and Hainmueller (2010) (Section 2.3) should not be viewed as a way of overcoming the non-uniqueness problem in the population. Rather, it should be viewed as a device that may stabilize the finite sample performance.

[4] Their equation (1) takes the form $Y_{j,t}(0) = \alpha_j + \theta_t' Z_j + \gamma_j'\delta_t + \epsilon_{j,t}$, so the factor structure in (4.2.3) is a special case of their equation (1) where $Z_i = 1$, i.e., it is a special case where the Z_i does not exist.

dition using some Z_j; their equation (1) takes the form $Y_{j,t}(0) = \alpha_j + Z_j'\theta_t + \gamma_j'\delta_t + \epsilon_{j,t}$, and they proposed minimizing $(X_0 - X\omega)' V (X_0 - X\omega)$ for some positive definite V, where $X_0 = \left(Z_0', \bar{Y}_0\right)'$ and X is the collection of corresponding objects for the controls.[5] The implementation makes sense as a way of requiring that $Z_0 - \sum_{j=1}^{J} \omega_{*j} Z_j$ be as close to zero as possible, i.e., it can be understood to be a reflection of the constraint that $Z_0 - \sum_{j=1}^{J} \omega_{*j} Z_j = 0$ as well as $E[Y_{0,T_0+1}(0)] - \sum_{j=1}^{J} \omega_{*j} E[Y_{j,T_0+1}(0)] = 0$. The requirement helps us avoid the non-uniqueness problem if the dimension of Z is sufficiently large relative to J. If the Z consists of only a small number of components, the non-uniqueness problem persists.

In order to achieve uniqueness, we may adopt various methods. See Doudchenko and Imbens (2016). We will consider the method of exact balancing, and minimize $\sum_{j=1}^{J} \omega_j^2$ subject to $E[Y_{0,t}(0)] = \sum_{j=1}^{J} \omega_j E[Y_{j,t}(0)]$ and $1 = \sum_{j=1}^{J} \omega_j$.[6] We make three observations on the method of exact balancing. First, the method is probably best motivated when $\delta_t = 0$. We can see that

$$
\begin{aligned}
& Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_j Y_{j,T_0+1}(0) \\
&= \left(\alpha_0 - \sum_{j=1}^{J} \omega_j \alpha_j\right) + \left(1 - \sum_{j=1}^{J} \omega_j\right) \theta_{T_0+1} + \left(\epsilon_{0,T_0+1} - \sum_{j=1}^{J} \omega_j \epsilon_{j,T_0+1}\right) \qquad (4.2.5)
\end{aligned}
$$

It is straightforward to see that the constraint $E[Y_{0,t}(0)] = \sum_{j=1}^{J} \omega_j E[Y_{j,t}(0)]$ is equivalent to $\alpha_0 = \sum_{j=1}^{J} \omega_j \alpha_j$. Therefore, the variance of $Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_j Y_{j,T_0+1}(0)$ is equal to

[5]We write X_0 and X, instead of X_1 and X_0 as in Abadie, Diamond, and Hainmueller (2010) (Section 2.3). They also propose using some other linear combinations of $Y_{j,t}, \ldots, Y_{j,T_0}$, in addition to $\bar{Y}_j$, but it does not affect anything in $T_0 \to \infty$ asymptotics.

[6]Abadie, Diamond, and Hainmueller (2010) also impose the positivity restriction, i.e., $\omega_j \geq 0$ for all J.

$\left(1+\sum_{j=1}^{J}\omega_j^2\right)\sigma_\epsilon^2$ under the constraints $E\left[Y_{0,t}\left(0\right)\right]=\sum_{j=1}^{J}\omega_j E\left[Y_{j,t}\left(0\right)\right]$ and $1=\sum_{j=1}^{J}\omega_j$ if $\epsilon_{j,t}$ were i.i.d. over j.[7] Here, the σ_ϵ^2 denotes the common variance of $\epsilon_{j,t}$. Therefore, the method of exact balancing may be understood to be a way of minimizing variance of $Y_{0,T_0+1}\left(0\right)-\sum_{j=1}^{J}\omega_j Y_{j,T_0+1}\left(0\right)$. Second, the motivation is fragile, yet the method is useful regardless. The simple motivation does not carry over to the case when δ_t is not fixed at zero, because we have

$$\begin{aligned}
&Y_{0,T_0+1}\left(0\right)-\sum_{j=1}^{J}\omega_j Y_{j,T_0+1}\left(0\right)\\
&=\left(\alpha_0-\sum_{j=1}^{J}\omega_j\alpha_j\right)+\left(1-\sum_{j=1}^{J}\omega_j\right)\theta_{T_0+1}+\left(\gamma_0-\sum_{j=1}^{J}\omega_j\gamma_j\right)'\delta_{T_0+1}+\left(\epsilon_{0,T_0+1}-\sum_{j=1}^{J}\omega_j\epsilon_{j,T_0+1}\right)
\end{aligned} \tag{4.2.6}$$

and the minimization requires knowledge of the variance covariance matrix of δ_t as well as the variance of $\epsilon_{j,t}$. Regardless, the method of exact balancing does guarantee uniqueness, and it serves a useful purpose in this regard. Third, a feasible version of the method of exact balancing is a solution to

$$\min_{\omega_1,\ldots,\omega_J}\sum_{j=1}^{J}\omega_j^2 \tag{4.2.7}$$

[7] Note that with Abadie, Diamond, and Hainmueller (2010) equation (1), we get a slightly different version of (4.2.6):

$$\begin{aligned}
&Y_{0,T_0+1}\left(0\right)-\sum_{j=1}^{J}\omega_j Y_{j,T_0+1}\left(0\right)\\
&=\left(\alpha_0-\sum_{j=1}^{J}\omega_j\alpha_j\right)+\left(Z_0-\sum_{j=1}^{J}\omega_j Z_j\right)'\theta_{T_0+1}+\left(\gamma_0-\sum_{j=1}^{J}\omega_j\gamma_j\right)'\delta_{T_0+1}+\left(\epsilon_{0,T_0+1}-\sum_{j=1}^{J}\omega_j\epsilon_{j,T_0+1}\right).
\end{aligned}$$

Therefore, the side-condition in their implementation can be understood to be a way of insisting $Z_0-\sum_{j=1}^{J}\omega_{*j}Z_j=0$, which can be understood to be a counterpart of the "adding-up" condition.

subject to

$$\bar{Y}_0 = \sum_{j=1}^{J} \omega_j \bar{Y}_j$$

and the adding-up condition $1 = \sum_{j=1}^{J} \omega_j$.

To summarize, the synthetic control can be understood to be a way of imposing the restriction $E\left[Y_{0,t}\left(0\right)\right] = \sum_{j=1}^{J} \omega_j E\left[Y_{j,t}\left(0\right)\right]$ without explicitly specifying the factor structure such as the dimension of δ_t. In fact, it can accommodate even more general nonlinear time series structure. Other than the crucial requirement that the time series is strictly stationary, there seems to be no other restriction, which makes it a convenient and attractive method of estimation.

4.3 Placebo Test and Synthetic Control

We now discuss how placebo tests can be used in the context of synthetic control. For this purpose, we first present a summary of the placebo test/permutation tests. The tests are motivated to deal with the case where the number of the treated is small and the number of controls is relatively large. In order to focus on the salient feature of the tests, we will consider an extreme case and assume that there is only one treated unit.

The basic intuition underlying the general placebo test can be gleaned by examining a standard textbook case of randomized treatments. Suppose that there is a cross sectional data with $J+1$ units, where the units $j = 1, \ldots, J$ are the control units and the unit $j = 0$ receives the active treatment. A reasonable estimator of the treatment effect is the difference $Y_0 - \bar{Y}$, where Y_0 is the outcome of the unit $j = 0$, and $\bar{Y} = J^{-1} \sum_{j=1}^{J}$ denotes the average of the outcomes of the controls. Suppose that we are interested in testing whether the treatment had impact, i.e., suppose that we are interested in such a hypothesis testing. Given that there is only one treated unit, the standard t-test comparing the difference of the mean outcomes is not applicable. On the other hand, common sense suggests that we may implement such a test by "assigning" each control units to fictitious treatments. More precisely, one can estimate the empirical distribution of $Y_k - (J-1)^{-1} \sum_{j \neq k} Y_j$ for $k = 1, \ldots, J$, and use it as if it were the distribution of the treatment effects under the null hypothesis.[8]

Implementation of the placebo test with synthetic control requires a bit more notation. First let $\hat{\omega} = (\hat{\omega}_1, \ldots, \hat{\omega}_J)$ denote the estimator of ω_*. Although we will later use the estimator in (4.2.7) in our Monte Carlo simulations, we do not need to restrict ourselves to this particular estimator. For now, we can view $\hat{\omega}$ as an output from a blackbox and let ω_* denote its probability limit as $T_0 \to \infty$. Second, let $\hat{\omega}^{(-k)}$ denote the outcome of the same blackbox except that we use the kth unit as the outcome of the treated unit, and $Y_{j,t}$ with $j \neq k$ as our control units. The placebo test then uses the empirical distribution of $Y_{k,T_0+1} - \sum_{j \neq k} \hat{\omega}_j^{(-k)} Y_{j,T_0+1}$ for $k = 1, \ldots, J$ as if it were the distribution of the treatment effects under the null hypothesis of no treatment effect. If the estimated effect $Y_{0,T_0+1} - \sum_{j=1}^{J} \hat{\omega}_j Y_{j,T_0+1}$ belongs to the extreme tails of the empirical distribution, it is understood to be the evidence that the null hypothesis is incorrect.

This summary immediately raises two questions about the validity of the placebo test in the context of synthetic control. First, given that ω's are estimated, would it affect the finite

[8] Conley and Taber (2011), who proposed a similar test, cite Bertrand, Duflo, and Mullainathan (2004) when they discuss placebo tests. Abadie, Diamond, and Hainmueller (2010) reference many other papers that precede Bertrand, Duflo, and Mullainathan (2004).

sample properties of the test? Second, does the permutation test have the correct size even if the true ω's were used?

The second concern may sound somewhat strange because the placebo test here seems so intuitive. In order to understand why it is not so straightforward, it helps to recall that the placebo test is a version of the permutation test, which requires for its validity what may be called the symmetry assumption. For review of this property, we will borrow the short discussion in Canay, Romano, and Shaikh (2014). Suppose that a researcher observes a vector of observations X, whose joint distribution is P. The objective is to test whether $P \in \mathbf{P}_0$, where $\mathbf{P}_0$ is a collection of probability distributions such that the distribution of X is equal to that of gX for every g in $\mathbf{G}$, where $\mathbf{G}$ is a finite collection of transformations. The permutation test has the exact size if for the test statistic $T(X)$, the critical value is taken from the multinomial distribution of $T(gX)$ for every g in $\mathbf{G}$. In the context of the placebo test above, one can understand X to be the vector $(Y_1, \ldots, Y_J)$, and gX to be the permutation of the Ys.

We note that the symmetry is not mathematically obvious in the context of synthetic control. In order for the permutation test to be valid, we need the distribution of $Y_{0,T_0+1} - \sum_{j=1}^{J} \omega_{*j} Y_{j,T_0+1}$ and those of $Y_{k,T_0+1} - \sum_{j \neq k} \omega_{*,j}^{(-k)} Y_{j,T_0+1}$ for $k = 1, \ldots, J$ to be identical. Even for the relatively simple model (4.2.3), the nature of the synthetic control (4.2.6) is such that the symmetry does not naturally follow. Using the restriction (4.2.4), we may write

$$Y_{0,T} - \sum_{j=1}^{J} \omega_{*j} Y_{j,T} = \theta_T \left(1 - \sum_{j=1}^{J} \omega_{*j}\right) + \left(\gamma_0 - \sum_{j=1}^{J} \omega_{*j} \gamma_j\right)' \delta_T + \left(\epsilon_{0,T} - \sum_{j=1}^{J} \omega_{*j} \epsilon_{j,T}\right)$$

Even if the first two terms were identically equal to zero over the permutations, we believe that the third term above is not likely to satisfy the symmetry property. It is because we believe that under the further restriction that the ϵ's have a finite variance, the term can be symmetric only when they are normally distributed. We in fact argue that normality is necessary if the distribution of $\epsilon_{0,T} - \omega' \epsilon_T$ is to be symmetric up to normalization, where $\epsilon_T = (\epsilon_{1,T}, \ldots, \epsilon_{J,T})'$ and $\omega = (\omega_{*1}, \ldots, \omega_{*T})'$. This is for the following reason. Suppose that $\epsilon_{0,T}, \ldots, \epsilon_{J,T}$ are i.i.d., and their common distribution is such that the variance is finite and the characteristic function does not disappear. Also suppose that there exists a random variable $\mathcal{Y}$ such that the distribution of $\epsilon_{0,T} - \omega' \epsilon_T$ is the same as that of $c\mathcal{Y}$ for some scalar c. Because the standard deviation of $\epsilon_{0,T} - \omega' \epsilon_T$ is proportional to $\sqrt{1 + \omega' \omega}$, we may without loss of generality take $c = \sqrt{1 + \omega' \omega}$. This implies that the distribution of $\omega' \epsilon_T$ only depends on $\omega' \omega$. In particular, we can take ω such that all the components are zero except for the first one, and conclude that the distribution of $\omega' \epsilon_T$ is identical to that of $\sqrt{\omega' \omega} \epsilon_{1,T}$. This implies that $\epsilon_{j,T}$ should have a stable distribution.[9] Because the only stable distribution with a finite variance is the normal distribution, the symmetry (up to normalization) can be achieved only with normality. Note that the third term above arises in an ideal situation where the weights ω do not need to be estimated and the first two terms completely disappear. Our analysis suggests that even if we normalize the third term by its standard deviation, the symmetry can be achieved only with normal distribution. Obviously with normality, we can simply use the standard normal distribution table, and avoid permutation tests which sometimes requires a random tie-breaking for size control.

[9]See Nolan (2003), or https://en.wikipedia.org/wiki/Stable_distribution

4.4 Monte Carlo

In the previous section, we raised two concerns about the placebo test for use with synthetic control, i.e., (i) the $\hat{\omega}$ are estimated; and (ii) the symmetry assumption is violated. The concern about the $\hat{\omega}$ being estimated can be quantitatively assessed by Monte Carlo. Likewise, Monte Carlo can answer the quantitative importance of the violation of the symmetry assumption, which is a sufficient condition but not necessary condition for validity of the permutation test.

For our Monte Carlo analysis, we adopted a simplified version of the factor model (4.2.3)

$$Y_{j,t} = \alpha_j + \theta_t + \gamma_j' \delta_t + \epsilon_{j,t}, \quad t = 1, \ldots, T_0$$

such that (i) δ_t is a scalar; (ii) $\theta_t \sim \mathcal{N}(0, 1)$; (iii) $\delta_t \sim \mathcal{N}(0, \sigma_\delta^2)$; (iv) $\epsilon_{j,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is i.i.d. over j and t. Our synthetic control did not impose the adding-up condition for simplicity, i.e., it solved (4.2.7). In matrix notations, our estimator solves

$$\min \omega' \omega \quad \text{s.t.} \quad \bar{Y}' \omega = \bar{Y}_0$$

It is straightforward to show that the analytic form of the estimator is given by

$$\hat{\omega} = \frac{\bar{Y}_0}{\bar{Y}' \bar{Y}} \cdot \bar{Y}, \tag{4.4.1}$$

where $\bar{Y} = \left(\bar{Y}_1, \ldots, \bar{Y}_J \right)'$. Because $E[Y_{j,t}] = \alpha_j$, we can see that the population counterpart is given by

$$\omega_* = \frac{\alpha_0}{\alpha'\alpha} \cdot \alpha, \tag{4.4.2}$$

where $\alpha = (\alpha_1, \ldots, \alpha_J)'$.

We now write

$$Y_{0,T} - \sum_{j=1}^{J} \hat{\omega}_j Y_{j,T} = I + II + III + IV,$$

where

$$\begin{aligned}
I &= \sum_{j=1}^{J} \left(\frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} - \hat{\omega}_j \right) \alpha_j \\
II &= \underbrace{\theta_T \left(1 - \sum_{j=1}^{J} \frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} \right)}_{II(i)} + \underbrace{\theta_T \sum_{j=1}^{J} \left(\frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} - \hat{\omega}_j \right)}_{II(ii)} \\
III &= \underbrace{\left(\gamma_0 - \sum_{j=1}^{J} \frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} \gamma_j \right)' \delta_T}_{III(i)} + \underbrace{\left(\sum_{j=1}^{J} \left(\frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} - \hat{\omega}_j \right) \gamma_j \right)' \delta_T}_{III(ii)} \\
IV &= \underbrace{\left(\epsilon_{0,T} - \sum_{j=1}^{J} \frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} \epsilon_{j,T} \right)}_{IV(i)} + \underbrace{\left(\sum_{j=1}^{J} \left(\frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} - \hat{\omega}_j \right) \epsilon_{j,T} \right)}_{IV(ii)}.
\end{aligned}$$

The sum of the components of ω_* is equal to $\left(\alpha_0 \sum_{k=1}^{J} \alpha_k \right) \Big/ \left(\sum_{k=1}^{J} \alpha_k^2 \right)$, which may or may not be 1. In our Monte Carlo, we only consider the DGP where the adding-up condition is satisfied in the population, so the term II(i) is equal to 0.[10] We speculate that the placebo test is used in the hope that (i) $Y_{0,T} - \sum_{j=1}^{J} \hat{\omega}_j Y_{j,T}$ is is dominated by the term IV(i) above; (ii) the four terms I, II(ii), III(ii) and IV(ii) above, which reflect the noise of estimating ω_*

[10]See Appendix 4.A for discussion of the form of (4.4.1) as well as the adding-up condition.

Table 4.1: Different DGP's with Deterministic α's

DGP No.	α's	γ's	Variations
1	$\alpha_0 = (2J+1)/3, \alpha_1 = 1, \cdots, \alpha_J = J$	$\gamma_0 = \gamma_1 = \cdots = \gamma_J = 0$	1(a), 2(a)
2	$\alpha_0 = (2J+1)/3, \alpha_1 = 1, \cdots, \alpha_J = J$	$\gamma_0 = 2, \gamma_1 = \cdots = \gamma_J = 1$	1(a), 2(b)
3	$\alpha_0 = 5/3, \alpha_1 = 1, \alpha_2 = 2, \alpha_3 = \cdots = \alpha_J = 0$	$\gamma_0 = \gamma_1 = \cdots = \gamma_J = 0$	1(b), 2(a)
4	$\alpha_0 = 5/3, \alpha_1 = 1, \alpha_2 = 2, \alpha_3 = \cdots = \alpha_J = 0$	$\gamma_0 = 2, \gamma_1 = \cdots = \gamma_J = 1$	1(b), 2(b)

by $\hat{\omega}$, are ignorable; and (iii) the two terms III(i) and IV(i) more or less satisfy the symmetry property.[11] We argued in the previous section that the term IV(i) is likely to violate the symmetry property.

In order to assess the impacts of other terms, we consider the following variations in DGP's:

1. Vary the values of α's such that (a) none of the components of ω_* dominates; (b) only two of the elements are non-zero.

2. Vary the values of γ's such that the unbalanced unobservable factors III(i) (a) disappear; and (b) present.

3. Vary T_0 such that the estimation errors in the weights are (a) prominent; and (b) negligible.

Combinations of the first two variations give us four different DGP's, as shown in Table 4.1.

[11] Given the i.i.d. assumption on ϵ, a sufficient condition for approximate symmetry in IV(i) is that $\sum_{j=1}^{J} \frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} \epsilon_{j,T}$ is negligible, which is implied if $J \to \infty$ and $\sum_{j=1}^{J} \left(\frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} \right)^2 \longrightarrow 0$. It is probably difficult to claim that this requirement is satisfied if the estimated weights are zero for most control units.

We considered four tests: The first one is what might be called a feasible asymmetric test. Formally, for $j = 1, \ldots, J$, let y_j be a $T \times 1$ vector of outcomes for the jth control unit, and let Y_{-j} be a $T \times (J-1)$ matrix that deletes the jth column from Y. Similar to (4.4.1), define the leave-one-out synthetic control weights $\hat{w}_{-j}$ for the jth control unit as:

$$\hat{\omega}_{-j} \equiv \frac{\bar{Y}_j}{\bar{Y}'_{-j\bar{Y}-j}} \cdot \bar{Y}_{-j} \tag{4.4.3}$$

where $\bar{Y}_{-j}$ is to delete the jth element from $\bar{Y}$. For $j = 1, \ldots, J$ and $k \neq j$, let $\hat{\omega}_{-j,k}$ be the element in $\hat{\omega}_{-j}$ that corresponds to the kth control unit. Also define $\hat{\omega}_{-j,j} \equiv 0$ for $j = 1, \ldots, J$. Then for $j = 1, \ldots, J$, we can compute

$$S(Y_j, Y_{-j}) \equiv Y_{j,T} - \sum_{k=1}^{J} \hat{\omega}_{-j,k} Y_{k,T} \tag{4.4.4}$$

Let $S^{(1)}, \ldots, S^{(J)}$ be the order statistics of $S(Y_j, Y_{-j})$'s. We reject H_0 if $S(Y_0, Y) > S^{(J(1-\alpha))}$ or $S(Y_0, Y) < S^{(J\alpha)}$.

The second test is identical to the first test, except that we use the true value of ω_*, i.e., we use

$$S^{true}(Y_0, Y) \equiv Y_{0,T} - \sum_{j=1}^{J} \frac{\alpha_0 \alpha_j}{\sum_{k=1}^{J} \alpha_k^2} Y_{j,T}$$
$$S^{true}(Y_j, Y_{-j}) \equiv Y_{j,T} - \sum_{k=1} \frac{\alpha_j \alpha_{-j,k}}{\sum_{l=1}^{J} \alpha_{-j,l}^2} Y_{k,T} = Y_{j,T} - \sum_{k \neq j}^{J} \frac{\alpha_j \alpha_k}{\sum_{l \neq j}^{J} \alpha_l^2} Y_{k,T}$$

and we reject H_0 if $S^{true}(Y_0, Y) > S^{true,(J(1-\alpha))}$ or $S^{true}(Y_0, Y) < S^{true,(J\alpha)}$.

The third test is identical to the first test, except that we compare the absolute value of $S(Y_0, Y)$ with the order statistic of the absolute values of $S(Y_j, Y_{-j})$'s. The fourth test is identical to the second test, except that we compare the absolute value of $S^{true}(Y_0, Y)$ with the order statistic of the absolute values of $S^{true}(Y_j, Y_{-j})$.

For each DGP, we try $T_0 \in \{40, 80, 400, 800\}$ and $\sigma_\epsilon^2 \in \{.1, 1, 10\}$. For all designs, we set $\alpha = 10\%$.

The results are summarized in Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7.[12] Although the actual size of the tests is seemingly close to the nominal size, we do see discrepancies in Tables 4.2 and 4.3. See DGP #2 and #4 there. The size distortion there cannot be attributed to the noise of estimating ω. First, the problem persists even as T_0 approaches unrealistically large values. Second, the size distortion is similar over the feasible and infeasible versions of the test. We suspect that the problem is a fundamental problem that may have something to do with the violation of symmetry.

Our analysis in the previous section indicates that the placebo test does have the size distortion problem. The results in Tables 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7 suggest that the size problem is potentially bigger in DGP #2 and #4. DGP #2 and #4 differ from #1 and #3 in that the γ's are nonzero and the aggregate shock δ_t plays a role as a consequence. Therefore, it is of interest to investigate this further source of asymmetry. For this purpose, we revisit the decomposition (4.2.6) of $Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_{*j} Y_{j,T_0+1}(0)$, assuming that the

[12] We set $\theta_t = 0$ in Tables 4.3, 4.5, 4.7, in order to reflect some of the concerns raised in Appendix 4.A.

first and second terms are not present[13]:

$$Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_{*j} Y_{j,T_0+1}(0) = \left(\gamma_0 - \sum_{j=1}^{J} \omega_{*j}\gamma_j\right)' \delta_{T_0+1} + \left(\epsilon_{0,T_0+1} - \sum_{j=1}^{J} \omega_{*j}\epsilon_{j,T_0+1}\right).$$

This implies that the variance of $Y_{0,T_0+1}(0) - \sum_{j=1}^{J} \omega_{*j} Y_{j,T_0+1}(0)$ can be written as

$$\left(\gamma_0 - \sum_{j=1}^{J} \frac{\alpha_0\alpha_j}{\sum_{k=1}^{J}\alpha_k^2}\gamma_j\right)' \Sigma(\delta_t) \left(\gamma_0 - \sum_{j=1}^{J} \frac{\alpha_0\alpha_j}{\sum_{k=1}^{J}\alpha_k^2}\gamma_j\right) + \mathrm{var}(\epsilon_{0,t}) + \sum_{j=1}^{J}\left(\frac{\alpha_0\alpha_j}{\sum_{k=1}^{J}\alpha_k^2}\right)^2 \mathrm{var}(\epsilon_{j,t})$$
$$= \left(\gamma_0 - \sum_{j=1}^{J} \frac{\alpha_0\alpha_j}{\sum_{k=1}^{J}\alpha_k^2}\gamma_j\right)' \Sigma(\delta_t) \left(\gamma_0 - \sum_{j=1}^{J} \frac{\alpha_0\alpha_j}{\sum_{k=1}^{J}\alpha_k^2}\gamma_j\right) + \mathrm{var}(\epsilon_{j,t})\left(1 + \frac{\alpha_0^2}{\sum_{k=1}^{J}\alpha_k^2}\right)$$

under the assumptions of the DGP, where $\Sigma(\delta_t)$ is the covariance matrix of the vector δ_t. Likewise, the variances of the permutation statistics

$$\left(\gamma_j - \sum_{k\neq j}^{J} \frac{\alpha_j\alpha_k}{\sum_{k\neq j}^{J}\alpha_k^2}\gamma_k\right)' \Sigma(\delta_t) \left(\gamma_j - \sum_{k\neq j}^{J} \frac{\alpha_j\alpha_k}{\sum_{k\neq j}^{J}\alpha_k^2}\gamma_k\right) + \mathrm{var}(\epsilon_{j,t})\left(1 + \frac{\alpha_j^2}{\sum_{k\neq j}^{J}\alpha_k^2}\right)$$

This implies that the symmetry assumption is not a very natural assumption. Depending on the relative magnitudes of γs, we can easily come up with examples that violate the symmetry. In the examples below, we assume that δ_t is a scalar.

First, we considered the case where $\alpha_0 = J$, $\alpha_1 = \cdots = \alpha_J = 1$, $\gamma_0 = 2$ and $\gamma_1 = \cdots = \gamma_J = 1$. We then have

$$\mathrm{var}\left(S(Y_0, Y)\right) = (J-2)^2\mathrm{var}(\delta_t) + (J+1)\mathrm{var}(\epsilon_{j,t}),$$

[13]This can be done by assuming that $\alpha_0 = \sum_{j=1}^{J} \omega_j\alpha_j$ and $\theta_t = 0$.

$$\text{var}\left(S(Y_j, Y_{-j})\right) = \frac{J}{J-1}\text{var}(\epsilon_{j,t}),$$

which implies that the severe size distortion can be easily found depending on the magnitudes of $\text{var}(\delta_t)$ and $\text{var}(\epsilon_{j,t})$. See Tables 4.8 to 4.13. We note that this is a case where the adding-up condition is not satisfied; from (4.4.2), we can see that $\omega_{*j} = 1$, so they do not add up to 1. In Tables 4.9, 4.11, 4.13, we set $\theta_t = 0$, so the second term in (4.2.6) is still zero. Tables 4.8, 4.10, and 4.12 are for the case where $\theta_t \sim \mathcal{N}(0, 1)$, so it is technically outside the scope of our earlier analysis.

Second, we considered the case where $\alpha_0 = \alpha_1 = \cdots = \alpha_J = 1$, $\gamma_0 = 2$, and $\gamma_1 = \cdots = \gamma_J = 1$. Then $\text{var}(S(Y_0, Y)) = \sigma_\delta^2 + \frac{21}{20}\sigma_\epsilon^2$ and $\text{var}(S(Y_j, Y_{-j})) = \frac{20}{19}\sigma_\epsilon^2$, which implies that the severe size distortion can be easily found depending on the magnitudes of $\text{var}(\delta_t)$ and $\text{var}(\epsilon_{j,t})$. See Table 4.14 to 4.19. We note that this is a case where the adding-up condition is satisfied; from (4.4.2), we can see that $\omega_{*j} = \frac{1}{J}$, so they do add up to 1. Therefore, the second term in (4.2.6) is zero whether $\theta_t = 0$ or not. We did consider the case with $\theta_t = 0$ in Tables 4.15, 4.17, and 4.19, for the purpose of providing a similar set of analyses as in Tables 4.8 - 4.13.

Third, we considered the case where $\alpha_0 = \alpha_1 = \cdots = \alpha_J = 1$, $\gamma_j = J - j$ for $j = 0, 1, \ldots, 20$. Again, we can find cases of severe size distortion depending on the magnitudes of $\text{var}(\delta_t)$ and $\text{var}(\epsilon_{j,t})$. See Table 4.20 to 4.25. We note that this is a case where the adding-up condition is satisfied; from (4.4.2), we can see that $\omega_{*j} = \frac{1}{J}$, so they do add up to 1. Therefore, the second term in (4.2.6) is zero whether $\theta_t = 0$ or not. We did consider the case with $\theta_t = 0$ in Tables 4.21, 4.23, and 4.25, for the purpose of providing a similar set of analyses as in Tables 4.8 - 4.13.

4.5 Possible Alternatives to Placebo Tests

If we take the time series asymptotics ($T_0 \to \infty$) seriously, the problem can be avoided by using the same idea as in Andrews (2003). The hypothesis of no treatment effects can be understood to be a hypothesis of stationarity of the time series $W_t \equiv Y_{0,t} - \sum_{j=1}^{J} \omega_{*j} Y_{j,t}$. In particular, the researcher is interested whether the distribution of $W_{T_0+1}, \ldots, W_T$ is the same as that of $W_1, \ldots, W_{T_0}$, for which the Andrews (2003)'s test is well-suited. In the simple case that we consider where $T = T_0 + 1$, one rejects the null if W_{T_0+1} belongs to the extreme tails of the empirical distribution of $W_1, \ldots, W_{T_0}$. We conducted Monte Carlo simulations for all the DGP's considered in the previous section, and verified that Andrews' test suffered no size distortion.[14] Andrews (2003)'s test is geared for application in time series, and as such, robust to certain heteroscedasticity. If the variance of $\epsilon_{j,t}$ in (4.2.3) were different across j's, most of the available methods exploiting cross sectional variation may need to be used with caution, as noted by Ferman and Pinto (2016). The end-of-sample instability test being a test of stationarity of $Y_{0,t} - \sum_{j=1}^{J} \omega_{*j} Y_{j,t}$, its validity does not depend on whether the $\epsilon_{j,t}$'s have identical variances or not. To our knowledge, Ferman and Pinto (2016) were the first to recognize the usefulness of Andrews' test in the synthetic control context.

Andrews (2003)'s test utilizes time series variation seriously. When T_0 is relatively small, perhaps the researcher would like to have a procedure that is based on cross section variation.

[14] The results are available upon request.

If the factor structure is taken seriously and if the number of factors is a priori known, we can produce such a procedure by combining the ideas in Conley and Taber (2011) and Holtz-Eakin, Newey, and Rosen (1988). For simplicity, assume that the model is given by

$$\begin{aligned} Y_{j,t} =& Y_{j,t}(0) = \gamma_j \delta_t + x'_{j,t}\beta + \epsilon_{j,t} \quad t = 1, 2; j = 1, \ldots, J \\ & Y_{0,t}(0) = \gamma_0 \delta_t + x'_{0,t}\beta + \epsilon_{0,t} \end{aligned}$$

where we normalize $\delta_1 = 1$. Let $Y_{0,1} = Y_{0,1}(0)$ and $Y_{0,2} = Y_{0,2}(1) = Y_{0,2}(0) + \eta$. This is a case where $T_0 = 1$, and $T = 2$. We then have

$$\frac{Y_{j,2}}{\delta_2} - Y_{j,1} = \left(\frac{x_{j,2}}{\delta_2} - x_{j,1} \right)' \beta + \frac{\epsilon_{j,2}}{\delta_2} - \epsilon_{j,1} \quad t = 1, 2; j = 1, \ldots, J$$

Under strict exogeneity assumption on x's, we can consistently estimate (β, δ_2) as $J \to \infty$ by using the control group. Now, assume that $(\epsilon_{j,1}, \epsilon_{j,2})$ $j = 0, 1, 2, \ldots$ are i.i.d., which would imply

$$\frac{\epsilon_{j,2}}{\delta_2} - \epsilon_{j,1} = \frac{Y_{j,2} - x'_{j,2}\beta}{\delta_2} - \left(Y_{j,1} - x'_{j,1}\beta \right)$$

are i.i.d. A simple modification of Conley and Taber (2011)'s argument establishes that the distribution of $\frac{\epsilon_{j,2}}{\delta_2} - \epsilon_{j,1}$ can be consistently estimated by the empirical distribution of

$$\frac{Y_{j,2} - x'_{j,2}\widehat{\beta}}{\widehat{\delta}_2} - \left(Y_{j,1} - x'_{j,1}\widehat{\beta} \right) \quad j = 1, \ldots, J$$

where $\left(\widehat{\beta}, \widehat{\delta}_2\right)$ denote Holtz-Eakin, Newey, and Rosen (1988)'s estimator. Therefore, in order to test that $\eta = \overline{\eta}$, it suffices to consider a test that rejects whenever

$$\frac{Y_{0,2} - \overline{\eta} - x'_{0,2}\widehat{\beta}}{\widehat{\delta}_2} - \left(Y_{0,1} - x'_{0,1}\widehat{\beta}\right)$$

is in the extreme tails of such empirical distribution. When there are multiple factors, the idea can be implemented easily by using a generalization of Holtz-Eakin, Newey, and Rosen (1988), for example, Ahn, Lee, and Schmidt (2013). The idea of combining Holtz-Eakin, Newey, and Rosen (1988) with Conley and Taber (2011), although straightforward, does not seem to have been considered elsewhere.

We have considered two alternative methods of inference, one based on $T_0 \to \infty$ asymptotics, and the other one based on $J \to \infty$ asymptotics. In addition to these two methods, we can also entertain the possibility that if both T_0 and J are large, it may be possible to use the panel technique as in Bai (2009) as well.[15] The latter two procedures are based on the predication that the researcher takes the linear factor structure seriously, so it may be more powerful than the Andrews (2003)'s test. On the other hand, if a researcher views the linear factor model as just a toy model[16] to illustrate the potential problem of difference-

[15]If one were to assume that $Y_{0,t}(1) = Y_{0,t}(0) + \beta$, the factor model (4.2.3) becomes

$$\begin{aligned} Y_{j,t} &= \alpha_j + \theta_t + \gamma'_j \delta_t + \epsilon_{j,t}, \quad t = 1, \ldots, T_0 \\ Y_{j,T_0+1} &= \alpha_j + \theta_{T_0+1} + \gamma'_j \delta_{T_0+1} + \epsilon_{j,T_0+1}, \quad j = 1, \ldots, J \\ Y_{0,T_0+1} &= \beta + \alpha_0 + \theta_{T_0+1} + \gamma'_0 \delta_{T_0+1} + \epsilon_{0,T_0+1} \end{aligned}$$

Using the pre-treatment data, one can consistently estimate $\left(\alpha_j, \gamma'_j\right)$ and (θ_t, δ'_t) as long as $J, T_0 \to \infty$. Using the control outcome for the period $t = T_0 + 1$ along with $\left(\alpha_j, \gamma'_j\right)$ consistently estimated, one can consistently estimate $\left(\theta_{T_0+1}, \delta'_{T_0+1}\right)$, which is possible if $J \to \infty$. Combining (α_0, γ'_0) as well as $\left(\theta_{T_0+1}, \delta'_{T_0+1}\right)$, one can make an inference of β.

[16]Indeed Abadie, Diamond, and Hainmueller (2010) (Section 2.2) consider some other model (in addition

in-difference, then she would probably be hesitant to discard the synthetic control method, which can in theory accommodate potentially complicated statistical structure that may go beyond the linear factor model.

4.6 Summary

We considered the performance of the permutation test (placebo test) in the context of the synthetic control method. The symmetry assumption, one of the crucial condition for validity of the permutation test, is often violated in synthetic control studies. Using Monte Carlo simulations, we show that the size of the permutation tests can be distorted. Several possible alternatives were discussed.

to the factor model) for motivation of the synthetic control.

Appendix

4.A Adding-Up and Positivity Constraints

As mentioned in a footnote in Section 4.2, Abadie, Diamond, and Hainmueller (2010) in effect imposes three constraints, i.e., (i) $E\left[Y_{0,t}\left(0\right)\right] = \sum_{j=1}^{J} \omega_j E\left[Y_{j,t}\left(0\right)\right]$, (ii) $1 = \sum_{j=1}^{J} \omega_j$, and (iii) $\omega_j \geq 0$ for all j. We argue that the positivity restriction may be reasonable as a way of improving finite sample properties, but it is unnatural to adopt in placebo calculation.

We first argue that combition of positivity restriction may lead to bias in synthetic control. For this purpose, consider an example where $J = 2$ and $\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = 2$. In order for the synthetic control to have zero bias, we want to find $w \in [0, 1]$ such that

$$\alpha_0 = w \times \alpha_1 + (1 - w) \times \alpha_2$$

or

$$0 = w + 2\left(1 - w\right)$$

which is impossible. Abadie, Diamond, and Hainmueller (2010)'s algorithm amounts to minimizing

$$\left(\alpha_0 - \left(w \times \alpha_1 + (1 - w) \times \alpha_2\right)\right)^2$$

subject to $w \in [0, 1]$. Because

$$\left(\alpha_0 - \left(w \times \alpha_1 + (1 - w) \times \alpha_2\right)\right)^2 = \left(w + 2\,(1 - w)\right)^2 = (2 - w)^2$$

we find that the solution is given by $w = 1$. In other words, the synthetic control is given by $1 \times Y_{1,T_0+1} + 0 \times Y_{2,T_0+1} = Y_{1,T_0+1}$. Because

$$E\left[Y_{0,T_0+1} - Y_{1,T_0+1}\right] = -1$$

the synthetic control does not offer an unbiased estimator.

The lack of unbiasedness does not necessarily mean that the positivity restriction should not be used for estimation. If a researcher is confident that the population counterpart of the synthetic control does satisfy the positivity restriction, she may want to impose it for estimaton for the purpose of finite sample property of the synthetic control. We argue that even though the restriction may serve such a useful purpose, it should be avoided in calculation of placebo. It is because imposing positivity restriction in the placebo may artificially introduces bias to the permutation test. This can be seen by calculating the weights without imposing the positivity, and verifying whether all the weights are in fact positive. We will use two examples to illustrate this point.

For this purpose, we will consider a simple specification where

$$Y_{j,t} = Y_{j,t}(0) = \alpha_j + \theta_t + \epsilon_{j,t},$$

and consider the minimization problem in the population

$$\min_{\omega} \omega'\omega$$

subject to

$$\alpha'\omega = \alpha_0$$

$$\iota'\omega = 1,$$

where $\alpha = (\alpha_1, \ldots, \alpha_J)'$ and ι denotes a column vector consisting of ones. Note that the positivity restriction is not imposed. It is straightforward to show that the solution is

$$\omega = \frac{(\iota'\iota)\,\alpha_0 - \alpha'\iota}{\alpha'\alpha \cdot \iota'\iota - (\alpha'\iota)^2}\alpha + \frac{\alpha'\alpha - (\alpha'\iota)\,\alpha_0}{\alpha'\alpha \cdot \iota'\iota - (\alpha'\iota)^2}\iota$$

Now, consider DGP #3 in Table 4.1, where $\alpha_0 = \frac{5}{3}$ and $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 0, \ldots, \alpha_{20} = 0$. We see that

$$\begin{aligned}\omega &= \frac{(20) \times \frac{5}{3} - 3}{5 \times 20 - (3)^2}\alpha + \frac{5 - (3) \times \frac{5}{3}}{5 \times 20 - (3)^2}\iota \\ &= \frac{1}{3}\alpha\end{aligned}$$

On the other hand,

$$\begin{aligned}\omega_{-3} &= \frac{(19)\,(0)-3}{5\cdot 19-(3)^2}\alpha_{-3}+\frac{5-(3)\,(0)}{5\cdot 19-(3)^2}\iota \\ &= -\frac{3}{86}\alpha_{-3}+\frac{5}{86}\iota\end{aligned}$$

and the second element of ω_{-3} is

$$-\frac{3}{86}\times 2+\frac{5}{86}=-\frac{1}{86}<0$$

i.e., the positivity restriction is violated in the placebo calculation.

Let's now consider DGP #1 in Table 4.1, where $\alpha_0=\frac{41}{3}$ and $\alpha_j=j>0$ for $j=1,\ldots,20$. It is straightforward to show that

$$\begin{aligned}\omega &= \frac{20\times\frac{41}{3}-210}{2870\times 20-210^2}\alpha+\frac{2870-210\times\frac{41}{3}}{2870\times 20-210^2}\iota \\ &= \frac{1}{210}\alpha\end{aligned}$$

so every element is positive, i.e., the positivity restriction is naturally satisfied by the DGP itself. In placebo calculations, we need to compute the delete-1 version of ω. It is straightforward to show that for each α_j, we have

$$\begin{aligned}\omega_{-j} &= \frac{19\times j-(210-j)}{(2870-j^2)\cdot 19-(210-j)^2}\alpha_{-j}+\frac{(2870-j^2)-(210-j)\times j}{(2870-j^2)\cdot 19-(210-j)^2}\iota \\ &= \frac{20j-210}{-20j^2+420j+10\,430}\alpha_{-j}+\frac{2870-210j}{-20j^2+420j+10\,430}\iota.\end{aligned}$$

For example, we see that the last six components of $\omega_{-1}$are

$$\begin{bmatrix} -0.01754386 \\ -0.035087719 \\ -0.052631579 \\ -0.070175439 \\ -0.087719298 \\ -0.105263158 \end{bmatrix}$$

so the positivity restriction is violated in ω_{-1}. If the positivity restriction is imposed in the placebo calculation, it would lead to the bias of the synthetic control for this particular placebo.

In our Monte Carlo, we set $\theta_t = 0$ (at least half the times). Examination of (4.2.5) shows that the role of the adding-up condition is to avoid the bias $\left(1 - \sum_{j=1}^{J} \omega_j\right) \theta_t$, we can achieve the same purpose in Monte Carlo by setting $\theta_t = 0$, and not insisting on adding-up in our Monte Carlo. Equation (4.4.2) reveals that positivity restriction is naturally imposed by insisting that every element of α is nonnegative in the population. Although we are skeptical about the role of positivity in the population, it would be natural to make the Monte Carlo environment as close to the common practice as possible, which is achieved by choosing the values of α to be nonnegative. The benefit of avoiding the adding-up in our Monte Carlo is that the positivity is now maintained in the placebo as well. The bias in the placebo is avoided by the condition $\theta_t = 0$.

Table 4.2: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 0.1$ and $\theta_t \sim \mathcal{N}(0, 1)$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.039	0.0.36	0.040	0.039
	Permutation Test 2	0.085	0.090	0.103	0.082
	Permutation Test 3	0.041	0.030	0.036	0.040
	Permutation Test 4	0.010	0.003	0.005	0.005
2	Permutation Test 1	0.274	0.244	0.243	0.261
	Permutation Test 2	0.263	0.254	0.247	0.263
	Permutation Test 3	0.321	0.287	0.280	0.298
	Permutation Test 4	0.238	0.021	0.211	0.227
3	Permutation Test 1	0.120	0.138	0.174	0.171
	Permutation Test 2	0.106	0.115	0.096	0.107
	Permutation Test 3	0.106	0.067	0.041	0.048
	Permutation Test 4	0.038	0.029	0.035	0.045
4	Permutation Test 1	0.183	0.195	0.219	0.243
	Permutation Test 2	0.218	0.206	0.189	0.205
	Permutation Test 3	0.375	0.297	0.209	0.202
	Permutation Test 4	0.187	0.173	0.170	0.182

Table 4.3: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 0.1$ and $\theta_t \equiv 0$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.074	0.092	0.093	0.070
	Permutation Test 2	0.195	0.232	0.206	0.183
	Permutation Test 3	0.087	0.112	0.099	0.076
	Permutation Test 4	0.007	0.013	0.015	0.008
2	Permutation Test 1	0.244	0.255	0.257	0.256
	Permutation Test 2	0.274	0.288	0.281	0.270
	Permutation Test 3	0.366	0.371	0.370	0.377
	Permutation Test 4	0.248	0.269	0.270	0.251
3	Permutation Test 1	0.131	0.138	0.156	0.120
	Permutation Test 2	0.116	0.129	0.151	0.115
	Permutation Test 3	0.146	0.134	0.167	0.123
	Permutation Test 4	0.120	0.128	0.160	0.113
4	Permutation Test 1	0.098	0.087	0.132	0.115
	Permutation Test 2	0.108	0.107	0.121	0.091
	Permutation Test 3	0.268	0.232	0.197	0.157
	Permutation Test 4	0.137	0.128	0.155	0.119

Table 4.4: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 1$ and $\theta_t \sim \mathcal{N}(0,1)$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.065	0.059	0.065	0.084
	Permutation Test 2	0.094	0.104	0.091	0.114
	Permutation Test 3	0.069	0.062	0.071	0.087
	Permutation Test 4	0.062	0.058	0.064	0.079
2	Permutation Test 1	0.121	0.120	0.107	0.138
	Permutation Test 2	0.131	0.142	0.134	0.152
	Permutation Test 3	0.147	0.159	0.142	0.169
	Permutation Test 4	0.133	0.136	0.144	0.156
3	Permutation Test 1	0.122	0.128	0.133	0.163
	Permutation Test 2	0.171	0.159	0.146	0.187
	Permutation Test 3	0.116	0.098	0.094	0.114
	Permutation Test 4	0.090	0.083	0.092	0.111
4	Permutation Test 1	0.142	0.145	0.131	0.154
	Permutation Test 2	0.161	0.165	0.143	0.179
	Permutation Test 3	0.218	0.183	0.123	0.134
	Permutation Test 4	0.121	0.132	0.114	0.134

Table 4.5: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 1$ and $\theta_t \equiv 0$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.074	0.092	0.092	0.070
	Permutation Test 2	0.111	0.122	0.108	0.098
	Permutation Test 3	0.088	0.112	0.099	0.076
	Permutation Test 4	0.067	0.097	0.092	0.070
2	Permutation Test 1	0.125	0.129	0.144	0.104
	Permutation Test 2	0.139	0.168	0.150	0.130
	Permutation Test 3	0.162	0.174	0.180	0.158
	Permutation Test 4	0.145	0.173	0.167	0.143
3	Permutation Test 1	0.128	0.137	0.153	0.124
	Permutation Test 2	0.134	0.149	0.169	0.130
	Permutation Test 3	0.142	0.141	0.166	0.124
	Permutation Test 4	0.153	0.152	0.173	0.134
4	Permutation Test 1	0.120	0.112	0.141	0.117
	Permutation Test 2	0.129	0.137	0.150	0.119
	Permutation Test 3	0.171	0.151	0.158	0.133
	Permutation Test 4	0.139	0.135	0.162	0.118

Table 4.6: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 10$ and $\theta_t \sim \mathcal{N}(0, 1)$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.080	0.079	0.088	0.093
	Permutation Test 2	0.091	0.105	0.108	0.110
	Permutation Test 3	0.092	0.085	0.095	0.112
	Permutation Test 4	0.100	0.105	0.107	0.124
2	Permutation Test 1	0.079	0.084	0.089	0.100
	Permutation Test 2	0.093	0.110	0.107	0.119
	Permutation Test 3	0.106	0.108	0.098	0.124
	Permutation Test 4	0.111	0.114	0.120	0.137
3	Permutation Test 1	0.128	0.130	0.131	0.161
	Permutation Test 2	0.139	0.139	0.139	0.159
	Permutation Test 3	0.113	0.126	0.129	0.151
	Permutation Test 4	0.130	0.144	0.139	0.156
4	Permutation Test 1	0.141	0.135	0.129	0.148
	Permutation Test 2	0.138	0.141	0.137	0.154
	Permutation Test 3	0.150	0.145	0.127	0.153
	Permutation Test 4	0.126	0.148	0.127	0.163

Table 4.7: Null Rejection Rates of Permutation Tests For DGP's in Table 4.1, $\sigma_\epsilon^2 = 10$ and $\theta_t \equiv 0$

DGP No.	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
1	Permutation Test 1	0.075	0.092	0.091	0.070
	Permutation Test 2	0.087	0.106	0.107	0.077
	Permutation Test 3	0.087	0.111	0.097	0.077
	Permutation Test 4	0.092	0.125	0.115	0.088
2	Permutation Test 1	0.082	0.093	0.097	0.073
	Permutation Test 2	0.101	0.108	0.116	0.085
	Permutation Test 3	0.095	0.112	0.104	0.094
	Permutation Test 4	0.102	0.131	0.126	0.105
3	Permutation Test 1	0.114	0.145	0.151	0.122
	Permutation Test 2	0.134	0.150	0.169	0.130
	Permutation Test 3	0.126	0.142	0.172	0.131
	Permutation Test 4	0.153	0.152	0.173	0.134
4	Permutation Test 1	0.120	0.147	0.154	0.122
	Permutation Test 2	0.134	0.149	0.170	0.126
	Permutation Test 3	0.131	0.153	0.165	0.137
	Permutation Test 4	0.151	0.157	0.177	0.125

Table 4.8: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 0.1$ and $\theta_t \sim \mathcal{N}(0, 1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.972	0.978	0.979	0.986
	Permutation Test 2	0.979	0.977	0.979	0.974
	Permutation Test 3	0.973	0.983	0.979	0.985
	Permutation Test 4	0.984	0.982	0.979	0.976
1	Permutation Test 1	0.927	0.932	0.923	0.898
	Permutation Test 2	0.926	0.944	0.929	0.909
	Permutation Test 3	0.934	0.938	0.925	0.909
	Permutation Test 4	0.934	0.952	0.932	0.917
10	Permutation Test 1	0.765	0.785	0.815	0.797
	Permutation Test 2	0.806	0.815	0.839	0.810
	Permutation Test 3	0.784	0.803	0.830	0.810
	Permutation Test 4	0.820	0.824	0.856	0.818

Table 4.9: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 0.1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.908	0.924	0.912	0.923
	Permutation Test 2	0.909	0.927	0.915	0.929
	Permutation Test 3	0.910	0.927	0.919	0.927
	Permutation Test 4	0.916	0.936	0.923	0.932
1	Permutation Test 1	0.783	0.790	0.785	0.801
	Permutation Test 2	0.804	0.791	0.786	0.807
	Permutation Test 3	0.802	0.814	0.797	0.819
	Permutation Test 4	0.822	0.801	0.806	0.815
10	Permutation Test 1	0.656	0.684	0.679	0.687
	Permutation Test 2	0.692	0.720	0.711	0.694
	Permutation Test 3	0.679	0.703	0.708	0.710
	Permutation Test 4	0.717	0.736	0.731	0.712

Table 4.10: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 1$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.983	0.982	0.985	0.982
	Permutation Test 2	0.983	0.977	0.982	0.974
	Permutation Test 3	0.987	0.981	0.988	0.982
	Permutation Test 4	0.985	0.982	0.983	0.978
1	Permutation Test 1	0.948	0.927	0.945	0.935
	Permutation Test 2	0.946	0.938	0.951	0.940
	Permutation Test 3	0.952	0.937	0.948	0.939
	Permutation Test 4	0.948	0.944	0.953	0.946
10	Permutation Test 1	0.815	0.817	0.818	0.830
	Permutation Test 2	0.838	0.847	0.833	0.845
	Permutation Test 3	0.825	0.823	0.834	0.843
	Permutation Test 4	0.855	0.860	0.842	0.858

Table 4.11: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.976	0.973	0.974	0.969
	Permutation Test 2	0.967	0.971	0.0981	0.973
	Permutation Test 3	0.980	0.977	0.976	0.970
	Permutation Test 4	0.970	0.971	0.983	0.975
1	Permutation Test 1	0.903	0.918	0.912	0.923
	Permutation Test 2	0.909	0.927	0.915	0.929
	Permutation Test 3	0.911	0.923	0.919	0.928
	Permutation Test 4	0.916	0.936	0.923	0.932
10	Permutation Test 1	0.744	0.767	0.785	0.809
	Permutation Test 2	0.804	0.791	0.786	0.807
	Permutation Test 3	0.756	0.791	0.806	0.827
	Permutation Test 4	0.822	0.801	0.806	0.815

Table 4.12: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 10$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.996	0.993	0.995	0.994
	Permutation Test 2	0.995	0.993	0.995	0.993
	Permutation Test 3	0.997	0.993	0.995	0.994
	Permutation Test 4	0.995	0.994	0.995	0.992
1	Permutation Test 1	0.975	0.980	0.978	0.976
	Permutation Test 2	0.977	0.978	0.981	0.973
	Permutation Test 3	0.977	0.980	0.978	0.980
	Permutation Test 4	0.981	0.980	0.983	0.975
10	Permutation Test 1	0.904	0.897	0.916	0.917
	Permutation Test 2	0.923	0.915	0.918	0.922
	Permutation Test 3	0.907	0.896	0.921	0.924
	Permutation Test 4	0.927	0.926	0.922	0.925

Table 4.13: Severe Upward Size Distortion 1 ($\sigma_\delta^2 = 10$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.992	0.990	0.989	0.996
	Permutation Test 2	0.991	0.992	0.992	0.992
	Permutation Test 3	0.993	0.993	0.989	0.996
	Permutation Test 4	0.994	0.992	0.993	0.993
1	Permutation Test 1	0.973	0.971	0.973	0.967
	Permutation Test 2	0.967	0.971	0.981	0.973
	Permutation Test 3	0.977	0.973	0.974	0.968
	Permutation Test 4	0.970	0.971	0.983	0.975
10	Permutation Test 1	0.883	0.907	0.912	0.923
	Permutation Test 2	0.909	0.927	0.915	0.929
	Permutation Test 3	0.891	0.913	0.917	0.926
	Permutation Test 4	0.916	0.936	0.923	0.932

Table 4.14: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 0.1$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.206	0.192	0.186	0.200
	Permutation Test 2	0.203	0.207	0.203	0.216
	Permutation Test 3	0.237	0.227	0.212	0.234
	Permutation Test 4	0.244	0.246	0.235	0.250
1	Permutation Test 1	0.088	0.091	0.091	0.108
	Permutation Test 2	0.097	0.098	0.110	0.122
	Permutation Test 3	0.100	0.113	0.099	0.123
	Permutation Test 4	0.126	0.130	0.119	0.139
10	Permutation Test 1	0.085	0.080	0.081	0.089
	Permutation Test 2	0.085	0.097	0.095	0.108
	Permutation Test 3	0.096	0.098	0.092	0.108
	Permutation Test 4	0.111	0.120	0.108	0.128

Table 4.15: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 0.1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.195	0.199	0.200	0.183
	Permutation Test 2	0.219	0.225	0.218	0.213
	Permutation Test 3	0.217	0.226	0.231	0.218
	Permutation Test 4	0.246	0.248	0.252	0.237
1	Permutation Test 1	0.089	0.097	0.103	0.072
	Permutation Test 2	0.099	0.112	0.118	0.089
	Permutation Test 3	0.099	0.113	0.119	0.097
	Permutation Test 4	0.114	0.139	0.134	0.117
10	Permutation Test 1	0.080	0.090	0.089	0.069
	Permutation Test 2	0.086	0.098	0.094	0.079
	Permutation Test 3	0.090	0.106	0.102	0.080
	Permutation Test 4	0.096	0.124	0.119	0.098

Table 4.16: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 1$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.562	0.534	0.572	0.545
	Permutation Test 2	0.587	0.552	0.590	0.570
	Permutation Test 3	0.606	0.578	0.594	0.600
	Permutation Test 4	0.616	0.587	0.611	0.620
1	Permutation Test 1	0.207	0.188	0.184	0.202
	Permutation Test 2	0.203	0.207	0.203	0.216
	Permutation Test 3	0.242	0.219	0.204	0.231
	Permutation Test 4	0.244	0.246	0.235	0.250
10	Permutation Test 1	0.091	0.089	0.092	0.107
	Permutation Test 2	0.097	0.098	0.110	0.122
	Permutation Test 3	0.108	0.114	0.103	0.119
	Permutation Test 4	0.126	0.130	0.119	0.139

Table 4.17: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.568	0.580	0.566	0.583
	Permutation Test 2	0.582	0.588	0.582	0.597
	Permutation Test 3	0.587	0.612	0.588	0.614
	Permutation Test 4	0.602	0.620	0.606	0.626
1	Permutation Test 1	0.199	0.204	0.202	0.186
	Permutation Test 2	0.219	0.225	0.218	0.213
	Permutation Test 3	0.209	0.225	0.232	0.219
	Permutation Test 4	0.246	0.248	0.252	0.237
10	Permutation Test 1	0.097	0.103	0.102	0.075
	Permutation Test 2	0.099	0.112	0.118	0.089
	Permutation Test 3	0.101	0.121	0.116	0.097
	Permutation Test 4	0.114	0.139	0.134	0.117

Table 4.18: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 10$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.846	0.849	0.842	0.841
	Permutation Test 2	0.851	0.836	0.847	0.855
	Permutation Test 3	0.858	0.860	0.853	0.856
	Permutation Test 4	0.867	0.848	0.861	0.868
1	Permutation Test 1	0.550	0.543	0.561	0.543
	Permutation Test 2	0.587	0.552	0.590	0.570
	Permutation Test 3	0.592	0.578	0.587	0.598
	Permutation Test 4	0.616	0.587	0.611	0.620
10	Permutation Test 1	0.189	0.178	0.184	0.203
	Permutation Test 2	0.203	0.207	0.203	0.216
	Permutation Test 3	0.232	0.234	0.204	0.238
	Permutation Test 4	0.244	0.246	0.235	0.250

Table 4.19: Severe Upward Size Distortion 2 ($\sigma_\delta^2 = 10$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.830	0.854	0.841	0.861
	Permutation Test 2	0.848	0.847	0.850	0.867
	Permutation Test 3	0.846	0.866	0.854	0.877
	Permutation Test 4	0.860	0.857	0.859	0.879
1	Permutation Test 1	0.538	0.576	0.563	0.581
	Permutation Test 2	0.582	0.588	0.582	0.597
	Permutation Test 3	0.575	0.607	0.584	0.609
	Permutation Test 4	0.602	0.620	0.606	0.626
10	Permutation Test 1	0.195	0.207	0.205	0.196
	Permutation Test 2	0.219	0.225	0.218	0.213
	Permutation Test 3	0.220	0.236	0.229	0.221
	Permutation Test 4	0.246	0.248	0.252	0.237

Table 4.20: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 0.1$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.423	0.450	0.464	0.470
	Permutation Test 2	0.639	0.620	0.648	0.626
	Permutation Test 3	0.487	0.527	0.538	0.547
	Permutation Test 4	0.609	0.617	0.636	0.635
1	Permutation Test 1	0.245	0.237	0.254	0.268
	Permutation Test 2	0.342	0.321	0.351	0.346
	Permutation Test 3	0.300	0.311	0.304	0.327
	Permutation Test 4	0.358	0.327	0.348	0.353
10	Permutation Test 1	0.129	0.126	0.127	0.152
	Permutation Test 2	0.144	0.145	0.151	0.179
	Permutation Test 3	0.171	0.166	0.149	0.170
	Permutation Test 4	0.167	0.176	0.163	0.191

Table 4.21: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 0.1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.477	0.469	0.473	0.495
	Permutation Test 2	0.654	0.631	0.642	0.658
	Permutation Test 3	0.506	0.527	0.553	0.572
	Permutation Test 4	0.646	0.627	0.640	0.650
1	Permutation Test 1	0.254	0.263	0.263	0.256
	Permutation Test 2	0.342	0.348	0.339	0.351
	Permutation Test 3	0.296	0.325	0.319	0.319
	Permutation Test 4	0.339	0.362	0.342	0.354
10	Permutation Test 1	0.140	0.129	0.146	0.124
	Permutation Test 2	0.165	0.177	0.172	0.152
	Permutation Test 3	0.172	0.172	0.173	0.157
	Permutation Test 4	0.173	0.191	0.195	0.172

Table 4.22: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 1$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.474	0.532	0.654	0.667
	Permutation Test 2	0.865	0.840	0.861	0.858
	Permutation Test 3	0.282	0.384	0.604	0.688
	Permutation Test 4	0.866	0.841	0.850	0.867
1	Permutation Test 1	0.276	0.317	0.429	0.467
	Permutation Test 2	0.639	0.620	0.648	0.626
	Permutation Test 3	0.256	0.325	0.469	0.529
	Permutation Test 4	0.609	0.617	0.636	0.635
10	Permutation Test 1	0.159	0.167	0.250	0.273
	Permutation Test 2	0.342	0.321	0.351	0.346
	Permutation Test 3	0.186	0.219	0.304	0.346
	Permutation Test 4	0.358	0.327	0.348	0.353

Table 4.23: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 1$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.481	0.557	0.651	0.698
	Permutation Test 2	0.873	0.855	0.874	0.892
	Permutation Test 3	0.290	0.378	0.616	0.693
	Permutation Test 4	0.865	0.856	0.865	0.885
1	Permutation Test 1	0.305	0.349	0.445	0.470
	Permutation Test 2	0.654	0.631	0.642	0.658
	Permutation Test 3	0.268	0.327	0.493	0.535
	Permutation Test 4	0.646	0.627	0.640	0.650
10	Permutation Test 1	0.167	0.194	0.250	0.253
	Permutation Test 2	0.342	0.348	0.339	0.351
	Permutation Test 3	0.204	0.222	0.323	0.328
	Permutation Test 4	0.339	0.362	0.342	0.354

Table 4.24: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 10$ and $\theta_t \sim \mathcal{N}(0,1)$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.444	0.518	0.662	0.697
	Permutation Test 2	0.963	0.954	0.946	0.967
	Permutation Test 3	0.093	0.154	0.314	0.413
	Permutation Test 4	0.959	0.954	0.950	0.963
1	Permutation Test 1	0.271	0.336	0.497	0.545
	Permutation Test 2	0.865	0.840	0.861	0.858
	Permutation Test 3	0.131	0.179	0.304	0.402
	Permutation Test 4	0.866	0.841	0.850	0.867
10	Permutation Test 1	0.144	0.181	0.284	0.352
	Permutation Test 2	0.639	0.620	0.648	0.626
	Permutation Test 3	0.141	0.168	0.255	0.350
	Permutation Test 4	0.609	0.617	0.636	0.635

Table 4.25: Severe Upward Size Distortion 3 ($\sigma_\delta^2 = 10$ and $\theta_t \equiv 0$)

σ_ϵ^2	Test No.	$T_0 = 40$	$T_0 = 80$	$T_0 = 400$	$T_0 = 800$
0.1	Permutation Test 1	0.464	0.520	0.657	0.708
	Permutation Test 2	0.955	0.953	0.959	0.964
	Permutation Test 3	0.123	0.163	0.282	0.399
	Permutation Test 4	0.954	0.950	0.953	0.963
1	Permutation Test 1	0.305	0.361	0.500	0.545
	Permutation Test 2	0.873	0.855	0.874	0.892
	Permutation Test 3	0.150	0.195	0.281	0.375
	Permutation Test 4	0.865	0.856	0.865	0.885
10	Permutation Test 1	0.165	0.202	0.287	0.322
	Permutation Test 2	0.654	0.631	0.642	0.658
	Permutation Test 3	0.165	0.195	0.264	0.318
	Permutation Test 4	0.646	0.627	0.640	0.650

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association* .

Abadie, Alberto and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American economic review* :113–132.

Ackerberg, Daniel, C Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. "Econometric tools for analyzing market outcomes." *Handbook of econometrics* 6:4171–4276.

Andrews, Donald WK. 1991. "Asymptotic normality of series estimators for nonparametric and semiparametric regression models." *Econometrica: Journal of the Econometric Society* :307–345.

———. 1999. "Consistent moment selection procedures for generalized method of moments estimation." *Econometrica* 67 (3):543–563.

———. 2003. "End-of-Sample Instability Tests." *Econometrica* 71 (6):1661–1694.

Andrews, Donald WK and Xu Cheng. 2013. "Maximum likelihood estimation and uniform inference with sporadic identification failure." *Journal of Econometrics* 173 (1):36–56.

Andrews, Donald WK, Xu Cheng, and Patrik Guggenberger. 2011. "Generic results for establishing the asymptotic size of confidence sets and tests." *Cowles Foundation Discussion Paper, No.1813* .

Andrews, Donald WK and Biao Lu. 2001. "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models." *Journal of Econometrics* 101 (1):123–164.

Andrews, DWK and P Guggenberger. 2006. "On the asymptotic maximal risk of estimators and the Hausman pre-test estimator." *unpublished manuscript* .

Ashley, Richard. 2009. "Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis." *Journal of Applied Econometrics* 24 (2):325–337.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica* 77 (4):1229–1279.

Bajari, Patrick and C Lanier Benkard. 2005. "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach." *Journal of Political Economy* .

Berkowitz, Daniel, Mehmet Caner, and Ying Fang. 2012. "The validity of instruments revisited." *Journal of Econometrics* 166 (2):255–266.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society* :841–890.

———. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1):68–105.

Berry, Steven and Ariel Pakes. 2007. "The pure characteristics demand model." *International Economic Review* 48 (4):1193–1225.

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

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* :249–275.

Brown, James N and Harvey S Rosen. 1982. "On the Estimation of Structural Hedonic Price Models." *Econometrica: Journal of the Econometric Society* :765–768.

Buckland, Steven T, Kenneth P Burnham, and Nicole H Augustin. 1997. "Model selection: an integral part of inference." *Biometrics* :603–618.

Canay, Ivan A, Joseph P Romano, and Azeem M Shaikh. 2014. "Randomization tests under an approximate symmetry assumption." *manuscript, July* 20.

Chay, Kenneth Y and Michael Greenstone. 2005. "Does air quality matter? Evidence from the housing market." *Journal of political Economy* 113 (2):376–424.

Cheng, Xu and Bruce E Hansen. 2015. "Forecasting with factor-augmented regression: A frequentist model averaging approach." *Journal of Econometrics* 186 (2):280–293.

Cheng, Xu and Zhipeng Liao. 2015. "Select the valid and relevant moments: An information-based lasso for gmm with many moments." *Journal of Econometrics* 186 (2):443–464.

Cheng, Xu, Zhipeng Liao, and Ruoyao Shi. 2016. "Uniform asymptotic risk of averaging GMM estimator robust to misspecification." Revise and resubmit.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on counterfactual distributions." *Econometrica* 81 (6):2205–2268.

Chernozhukov, Victor, Alfred Galichon, and Marc Henry. 2014. “Identifying multi-attribute hedonic models.” *Available at SSRN 2392549* .

Chiappori, Pierre-André, Robert J McCann, and Lars P Nesheim. 2010. “Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness.” *Economic Theory* 42 (2):317–354.

Claeskens, Gerda and Raymond J Carroll. 2007. “An asymptotic theory for model selection inference in general semiparametric problems.” *Biometrika* :249–265.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. “Plausibly exogenous.” *Review of Economics and Statistics* 94 (1):260–272.

Conley, Timothy G and Christopher R Taber. 2011. “Inference with "difference in differences" with a small number of policy changes.” *The Review of Economics and Statistics* 93 (1):113–125.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux. 1996. “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach.” *Econometrica* 64 (5):1001–1044.

DiTraglia, Francis J. 2016. “Using invalid instruments on purpose: Focused moment selection and averaging for GMM.” *Journal of Econometrics* 195 (2):187–208.

Doko Tchatoka, Firmin and Jean-Marie Dufour. 2008. “Instrument endogeneity and identification-robust tests: some analytical results.” *Journal of statistical planning and inference* 138 (9):2649–2661.

———. 2014. “Identification-robust inference for endogeneity parameters in linear structural models.” *The Econometrics Journal* 17 (1):165–187.

Doudchenko, Nikolay and Guido Imbens. 2016. “Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis.”

Eichenbaum, Martin S, Lars Peter Hansen, and Kenneth J Singleton. 1988. “A time series analysis of representative agent models of consumption and leisure choice under uncertainty.” *The Quarterly Journal of Economics* 103 (1):51–78.

Ekeland, Ivar. 2010. “Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types.” *Economic Theory* 42 (2):275–315.

Ekeland, Ivar, James J Heckman, and Lars Nesheim. 2004. “Identification and Estimation of Hedonic Models.” *Journal of Political Economy* 112 (1 pt 2).

Epple, Dennis. 1987. “Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products.” *Journal of political economy* 95 (1):59–80.

Epple, Dennis and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *The Journal of Political Economy* 107 (4):645–681.

Ferman, Bruno and Cristine Pinto. 2016. "Revisiting the synthetic control estimator."

Fessler, Pirmin and Maximilian Kasy. 2016. "How to use economic theory to improve estimators, with an application to labor demand and wage inequality in Europe." Tech. rep.

Fessler, Pirmin, Maximilian Kasy, and Peter Lindner. 2013. "Survey mode effects on income inequality measurement." *Economic Analysis* 617:495–7730.

Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* :953–973.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition methods in economics." *Handbook of labor economics* 4:1–102.

Guggenberger, Patrik. 2012. "On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption." *Econometric Theory* 28 (02):387–421.

Hahn, Jinyong and Ruoyao Shi. 2016. "Synthetic control and inference." In preparation for publication.

Hansen, Bruce E. 2007. "Least squares model averaging." *Econometrica* 75 (4):1175–1189.

———. 2015. "A Stein-Like 2SLS Estimator." *Econometric Reviews* .

———. 2016. "Efficient shrinkage in parametric models." *Journal of Econometrics* 190 (1):115–132.

Hansen, Bruce E and Jeffrey S Racine. 2012. "Jackknife model averaging." *Journal of Econometrics* 167 (1):38–46.

Hansen, Lars Peter. 1982. "Large sample properties of generalized method of moments estimators." *Econometrica: Journal of the Econometric Society* :1029–1054.

Harrison, David and Daniel L Rubinfeld. 1978. "Hedonic housing prices and the demand for clean air." *Journal of environmental economics and management* 5 (1):81–102.

Heckman, James J, Rosa L Matzkin, and Lars Nesheim. 2010. "Nonparametric identification and estimation of nonadditive hedonic models." *Econometrica* 78 (5):1569–1591.

Hildebrandt, Theophil Henry and Lawrence M Graves. 1927. "Implicit functions and their differentials in general analysis." *Transactions of the American Mathematical Society* 29 (1):127–153.

Hjort, Nils Lid and Gerda Claeskens. 2003. "Frequentist model average estimators." *Journal of the American Statistical Association* 98 (464):879–899.

———. 2006. “Focused information criteria and model averaging for the Cox hazard regression model.” *Journal of the American Statistical Association* 101 (476):1449–1464.

Hofferth, Sandra L., Sarah M. Flood, and Matthew Sobek. 2013. “American Time Use Survey Data Extract System: Version 2.4 [Machine-readable database] (http://www.atusdata.org).” Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S Rosen. 1988. “Estimating vector autoregressions with panel data.” *Econometrica: Journal of the Econometric Society* :1371–1395.

Hong, Han, Bruce Preston, and Matthew Shum. 2003. “GENERALIZED EMPIRICAL LIKELIHOOD–BASED MODEL SELECTION CRITERIA FOR MOMENT CONDITION MODELS.” *Econometric Theory* 19 (06):923–943.

Imbens, Guido W and Whitney K Newey. 2009. “Identification and estimation of triangular simultaneous equations models without additivity.” *Econometrica* 77 (5):1481–1512.

James, William and Charles Stein. 1961. “Estimation with quadratic loss.” In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1. 361–379.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce. 1993. “Wage inequality and the rise in returns to skill.” *Journal of political Economy* :410–442.

Kabaila, Paul. 1995. “The effect of model selection on confidence regions and prediction regions.” *Econometric Theory* 11 (03):537–549.

———. 2009. “The coverage properties of confidence regions after model selection.” *International Statistical Review* 77 (3):405–414.

Kahn, Shulamit and Kevin Lang. 1988. “Efficient estimation of structural hedonic systems.” *International Economic Review* :157–166.

Kang, Hyunseung, Anru Zhang, T Tony Cai, and Dylan S Small. 2016. “Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization.” *Journal of the American Statistical Association* 111 (513):132–144.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W Imbens. 2015. “Identification and inference with many invalid instruments.” *Journal of Business & Economic Statistics* 33 (4):474–484.

Leeb, Hannes and Benedikt M Pötscher. 2005. “Model selection and inference: Facts and fiction.” *Econometric Theory* 21 (01):21–59.

———. 2008. “Sparse estimators and the oracle property, or the return of Hodges’ estimator.” *Journal of Econometrics* 142 (1):201–211.

Leung, Gilbert and Andrew R Barron. 2006. "Information theory and mixing least-squares regressions." *IEEE Transactions on Information Theory* 52 (8):3396–3410.

Liao, Zhipeng. 2013. "ADAPTIVE GMM SHRINKAGE ESTIMATION WITH CONSISTENT MOMENT SELECTION." *Econometric Theory* 29 (05):857–904.

Liu, Chu-An. 2015. "Distribution theory of the least squares averaging estimator." *Journal of Econometrics* 186 (1):142–159.

Lorentz, George G. 1986. *Approximation of functions*. Chelsea, New York.

Lu, Xun and Liangjun Su. 2015. "Jackknife model averaging for quantile regressions." *Journal of Econometrics* 188 (1):40–58.

Machado, José AF and José Mata. 2005. "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of applied Econometrics* 20 (4):445–465.

Matzkin, Rosa L. 2003. "Nonparametric estimation of nonadditive random functions." *Econometrica* :1339–1375.

Moon, Hyungsik Roger and Frank Schorfheide. 2009. "Estimation with overidentifying inequality moment conditions." *Journal of Econometrics* 153 (2):136–154.

Nesheim, Lars. 2015. "Identification in multidimensional hedonic models." *Working Paper* .

Nevo, Aviv. 2001. "Measuring market power in the ready-to-eat cereal industry." *Econometrica* 69 (2):307–342.

Nevo, Aviv and Adam M Rosen. 2012. "Identification with imperfect instruments." *Review of Economics and Statistics* 94 (3):659–671.

Newey, Whitney K. 1997. "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics* 79 (1):147–168.

Nolan, John. 2003. *Stable distributions: models for heavy-tailed data*. Birkhauser Boston.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110 (4):705–729.

Pötscher, Benedikt M. 1991. "Effects of model selection on inference." *Econometric Theory* :163–185.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *The journal of political economy* :34–55.

Rossi, Peter E, Robert E McCulloch, and Greg M Allenby. 1996. "The value of purchase history data in target marketing." *Marketing Science* 15 (4):321–340.

Rothe, Christoph. 2010. "Nonparametric estimation of distributional policy effects." *Journal of Econometrics* 155 (1):56–70.

Sargan, John D. 1958. "The estimation of economic relationships using instrumental variables." *Econometrica: Journal of the Econometric Society* :393–415.

Shibata, Ritei. 1986. "Consistency of model selection and parameter estimation." *Journal of Applied Probability* 23 (A):127–141.

Sieg, Holger, V Kerry Smith, H Spencer Banzhaf, and Randy Walsh. 2004. "Estimating the general equilibrium benefits of large changes in spatially delineated public goods." *International Economic Review* 45 (4):1047–1077.

Small, Dylan S. 2007. "Sensitivity analysis for instrumental variables regression with overidentifying restrictions." *Journal of the American Statistical Association* 102 (479):1049–1058.

Stein, Charles. 1956. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution." In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, vol. 1. 197–206.

Stock, James H. 1991. "Nonparametric policy analysis: an application to estimating hazardous waste cleanup benefits." *Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge* :77–98.

Yang, Yuhong. 2005. "Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation." *Biometrika* 92 (4):937–950.