Identification and Estimation of Nonparametric Hedonic Equilibrium Model with Unobserved Quality*

Ruoyao Shi†

December 24, 2016

JOB MARKET PAPER

Latest version can be found at `http://RuoyaoShi.github.io`

Abstract

This paper 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- 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.

Keywords: hedonic equilibrium, unobserved quality, distributional effects, nonparametric identification, series estimation

*I am deeply indebted to Jinyong Hahn, Zhipeng Liao and Rosa L. 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, Shuyang Sheng and Till Von Wachter for helpful discussions. Comments from the 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 and UCLA applied micro proseminars helped shape the paper as well.

†Department of Economics, UCLA. Email: shiruoyao@ucla.edu. Comments are very welcome.

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

$$z_i^m \equiv h_i^m \cdot e(x_i^m, a_i^m), \tag{1.1}$$

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 (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]

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]

[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 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 4.1 does not require explicitly estimating the price schedule functions.

[4] I also propose an algorithm to solve for the counterfactual equilibrium of the model in Appendix 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.

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 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 paper is organized as follows. Section 2.1 sets up the model and describes some important properties of the equilibrium; Section 2.2 discusses several applications to which my model and method can be applied for counterfactual policy analysis. Section 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 3.2. Section 4.1 describes the series estimators, and Section 4.2 derives their uniform rates of convergence. An illustration of the estimation procedure using the 2015 ATUS data is given in Section 5. Section 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 Appendices.

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 2.1 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 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.

[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.

[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).

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).

Assumption 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 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 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.

[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 C.

Assumption 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 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 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 2-4 (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 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 2-4, 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)

$$P_z^m(s^m(x, a)) \cdot e(x, a) - U_h\left(\frac{s^m(x, a)}{e(x, a)}, x, a\right) = 0. \tag{2.1}$$

Applying the Implicit Function Theorem (Hildebrandt and Graves, 1927) to equation (2.1) gives rise to

$$\frac{\partial z^s}{\partial a} = \frac{\partial s^m(x, a)}{\partial a} = \frac{eU_{ha} - P_z^m e e_a - U_{hh} h^m e_a}{P_{zz}^m e^2 - U_{hh}}, \tag{2.2}$$

where the arguments of the functions on the right-hand side of equation (2.2) are suppressed for simplicity. By Assumptions 2 and 3 and the FOC in equation (2.1), $P_z^m > 0$. Then Assumptions 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

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

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

[10] As discussed later, equations (2.6) and (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

$$I^m(x, a) = P^m(s^m(x, a)) = P^m(h^m(x, a) \cdot e(x, a)). \tag{2.3}$$

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

$$R_z(d^m(y, b), y, b) - P^m(d^m(y, b)) = 0, \tag{2.4}$$

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

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

and the range of equilibrium effective amount demanded

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

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

$$\begin{aligned} & \int_{\mathcal{X}} f^m_{x,a}\left(x, (s^m)^{-1}(x, z)\right) \frac{\partial (s^m)^{-1}(x, z)}{\partial z^s} dx \\ = & \int_{\mathcal{Y}} f^m_{y,b}\left(y, (d^m)^{-1}(y, z)\right) \frac{\partial (d^m)^{-1}(y, z)}{\partial z^d} dy, \end{aligned} \tag{2.5}$$

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 2.1 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 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 2-4 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 (2.5) and suppress the arguments of the functions, one gets

$$P^m_{zz}(z) = \frac{\int_{\mathcal{Y}} \frac{f^m_{y,b}}{R_{zb}} R_{zz} dy + \int_{\mathcal{X}} \frac{f^m_{x,a}}{-(U_{ha}e - P^m_z e e_a - U_{hh} h^m e_a)} U_{hh} dx}{\int_{\mathcal{Y}} \frac{f^m_{y,b}}{R_{zb}} dy + \int_{\mathcal{X}} \frac{e^2 f^m_{x,a}}{-(U_{ha}e - P^m_z e e_a - U_{hh} h^m e_a)} dx}. \tag{2.6}$$

Equation (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 and 3 imply that the second-order condition (SOC)[11]

$$R_{zz} \cdot e^2 - U_{hh} < 0 \tag{2.7}$$

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 (2.6)

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

and equation (2.7) imply $P^m_{zz}e^2 - U_{hh} < 0$.

Since the structural functions (U, e, R) remain invariant across markets, equation (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 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 A.

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.

2.2.1 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, work-

[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.

ers 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 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 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. 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).

[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 overtime.

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

[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).

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.

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 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 3.1, 3.2 and 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 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 4.2 account for the fact that they need to be estimated.

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 5. *Suppose that $x \perp a$ and $y \perp b$ 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 6. *Suppose that the sellers' unobserved characteristic a follows the uniform distribution $U[0,1]$ in all markets.*

Assumption 6 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 6 requires that the sellers' unobserved characteristic has the same distribution (probably unknown) across all markets.[18] In Appendix 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 6 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 1. *Under Assumptions 1-6, 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 (2.3), $I^m(x,a)$ 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 (2.1), Assumptions 2 and 3 guarantee that $P^m_z > 0$. On the other hand, by the SOC in equation (2.7) and the equilibrium condition in equation (2.6), we have $P^m_{zz} \cdot e^2 - U_{hh} < 0$. Then the expression of $\partial s^m(x,a)/\partial a$ in equation (2.2) is positive under Assumptions 3 and 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 6, the identification of $I^m(x,a)$ follows the same argument as in Matzkin (2003) (Specification I). In particular, by monotonicity, Assumptions 5 and 6, we have

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

Then

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

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.

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 1. *Under the conditions for Lemma 1, the partial derivatives of the payment function $I^m_{x_j}(x,a)$ $(j=1,\ldots,d_x)$ 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. *Under the conditions for Lemma 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 $e(x,a)$.

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 3.2.1 illustrates the intuition for scalar-valued x. The intuition applies to vector-valued x as well. Section 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 7. *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.

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 (2.3),

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

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

$$\frac{e(x_i^m, a_i^m)}{e(x_j^m, a_j^m)} = \frac{h^m(x_j^m, a_j^m)}{h^m(x_i^m, a_i^m)}. \tag{3.1}$$

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 3.2.1. The solid green line in Step 1 of Figure 3.2.1 represents the iso-payment curve in Market 1 that contains the normalization seller $(\bar{x}, \bar{a})$. By equation (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

$$\frac{e(x_2, a_2)}{e(\tilde{x}, \tilde{a})} = \frac{h^1(\tilde{x}, \tilde{a})}{h^1(x_2, a_2)}. \tag{3.2}$$

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

$$e(x_2, a_2) = \frac{h^2(\bar{x}, \bar{a})}{h^2(x_2, a_2)}. \tag{3.3}$$

Combining equation (3.2) and equation (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 3.2.1, 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 3.2.1, 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 3.2.1, 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.

3.2.2 General Results

It is not hard to generalize the intuition explained in Section 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

$$\left(\frac{I_{x_1}^m(x,a)}{I_a^m(x,a)}, \ldots, \frac{I_{x_{d_x}}^m(x,a)}{I_a^m(x,a)}, 1\right) \neq \left(\frac{I_{x_1}^{m'}(x,a)}{I_a^{m'}(x,a)}, \ldots, \frac{I_{x_{d_x}}^{m'}(x,a)}{I_a^{m'}(x,a)}, 1\right). \tag{3.4}$$

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 8. *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 9. *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 (3.4) holds for all $(x,a) \in \mathcal{X} \times \mathcal{A}$, then Assumption 9 is satisfied. Moreover, if Assumption 9 holds, there could be more than two markets satisfying equation (3.4).

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

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

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

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

[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.

where the superscript "−" indicates the generalized inverse of a matrix.

Theorem 1. *Suppose that Assumptions 7-9 and the conditions for Lemma 1 are satisfied. The quality function is then nonparametrically identified on $\mathcal{X} \times \mathcal{A}$ as*

$$e(x,a) = \exp\left(\sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} g_j(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})ds_j + \int_{\bar{a}}^{a} g_{d_x+1}(x,t)dt\right) \tag{3.5}$$

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 (2.3) yields

$$\begin{aligned} \nabla_x I^m(x,a) &= P_z^m\left(h^m(x,a)\cdot e(x,a)\right)\cdot\left[\nabla_x h^m(x,a)e(x,a) + h^m(x,a)\nabla_x e(x,a)\right], \\ I_a^m(x,a) &= P_z^m\left(h^m(x,a)\cdot e(x,a)\right)\cdot\left[h_a^m(x,a)e(x,a) + h^m(x,a)e_a(x,a)\right]. \end{aligned} \tag{3.6}$$

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:

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

$$\begin{aligned} \Longrightarrow \quad & I_a^m(x,a)\frac{\nabla_x e(x,a)}{e(x,a)} - \nabla_x I^m(x,a)\frac{e_a(x,a)}{e(x,a)} \\ = \; & \left[h_a^m(x,a)\nabla_x I^m(x,a) - I_a^m(x,a)\nabla_x h^m(x,a)\right]/h^m(x,a), \end{aligned} \tag{3.7}$$

for all $m \in \mathcal{M}$ and all $(x,a) \in \mathcal{X}\times\mathcal{A}$. Stack equation (3.7) for all markets, one gets a system of equations

$$B(x,a)\cdot\left(\frac{\nabla_x e(x,a)'}{e(x,a)}, \quad \frac{e_a(x,a)}{e(x,a)}\right)' = A(x,a), \tag{3.8}$$

for all $(x,a) \in \mathcal{X}\times\mathcal{A}$. Suppose that Assumption 9 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

$$\left(\frac{\nabla_x \epsilon(x,a)'}{\epsilon(x,a)}, \quad \frac{\epsilon_a(x,a)}{\epsilon(x,a)}\right)' = [B(x,a)'B(x,a)]^{-1}[B(x,a)'A(x,a)], \tag{3.9}$$

which depends only on the identified reduced form functions $I^m(x,a)$, $h^m(x,a)$ and their

[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 1.

derivatives. Then the identification of the quality function $e(x,a)$ amounts to a unique solution to the differential equations in (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 (3.9). Note that

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

Then,

$$\begin{aligned}\log\left(e(x_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})\right) &= \int_{\bar{x}_1}^{x_1} g_1(s_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})ds_1+\log\left(e(\bar{x},\bar{a})\right)\\ &= \int_{\bar{x}_1}^{x_1} g_1(s_1,\bar{x}_2,\ldots,\bar{x}_{d_x},\bar{a})ds_1, \qquad (3.10)\end{aligned}$$

for all $x_1\in\mathcal{X}_1$, where the second equality holds by Assumption 7. Then, consider the second equation in (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 (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

$$\mathcal{Z}_s^m = \{z\in\mathcal{Z}\colon \text{there exists some } (x,a)\in\mathcal{X}\times\mathcal{A} \text{ in market}$$

$m \in \mathcal{M}$ such that in equilibrium $z = h^m(x,a) \cdot e(x,a)\}$.

Corollary 2. *Under the conditions for Theorem 1, the unobserved effective amount* $z = h^m(x,a) \cdot e(x,a)$ *is identified.*

Corollary 3. *Under the conditions for Theorem 1, the price schedule function* $P^m(z)$ *for market* $m \in \mathcal{M}$ *is nonparametrically identified on* $\mathcal{Z}_s^m$.

Proof. Assumption 1, earnings equation (2.3), Lemmas 1 and 2, and Theorem 1 together imply the result. □

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:

$$\begin{aligned}\mathcal{HXA} \;=\; &\{(h,x,a) \in \mathcal{H} \times \mathcal{X} \times \mathcal{A}\text{: there exists a market } m \in \mathcal{M} \text{ and}\\ &\text{some } (x,a) \in \mathcal{X} \times \mathcal{A} \text{ such that in equilibrium } h = h^m(x,a)\}.\end{aligned}$$

If $|\mathcal{M}| = 1$, then $\mathcal{HXA}$ is degenerate since h is endogenous. As discussed in Section 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. *Under the conditions for Theorem 1, the sellers' marginal disutility function* $U_h(h,x,a)$ *is nonparametrically identified on* $\mathcal{HXA}$.

Proof. The result follows from Theorem 1, Corollary 3, 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 $m \in \mathcal{M}$. □

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:

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

Assumption 10. *Suppose that the buyers' unobserved characteristic b follows the uniform distribution $U[0, 1]$ in all markets.*

Lemma 3. ***(Heckman, Matzkin, and Nesheim 2010 Theorem 4.1)*** *Under Assumption 10 and the conditions for Theorem 1, the buyers' marginal utility function $R_z(z, y, b)$ is nonparametrically identified on $\mathcal{ZYB}$.*[23]

4 Estimation

This section provides an estimation procedure for the structural functions. Section 4.1 describes the estimation procedure step by step, and in Section 4.2 I derive the uniform rates of convergence for the estimators.

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 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] For the rest of the paper, I maintain the following sampling assumptions.

Assumption 11. *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 12. *For notational simplicity, suppose that the sample sizes from all the markets are equal, i.e., $N^1 = N^2 = \cdots = N^M = N$.*

[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.

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

In the rest of this paper, I maintain Assumptions 1-12. Assumption 12 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 paper, 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 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 1, the quality function $e(x,a)$ can be estimated by first solving an estimated version of the equations (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 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.

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 2.1, I showed that the payment function $I^m(x,a)$ is strictly increasing in a under Assumptions 1-4. 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

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

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

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

$$\begin{aligned} \hat{a}_i^m &\equiv \hat{F}_{I^m|x^m=x_i^m}(I_i^m) \\ &\equiv \Lambda_{k_{Q,N}}(x_i^m)' \left(\sum_{j=1}^{N} \Lambda_{k_{Q,N}}(x_j^m) \Lambda_{k_{Q,N}}(x_j^m)' \right)^{-} \left(\sum_{j=1}^{N} \Lambda_{k_{Q,N}}(x_j^m) \mathbb{I}(I_j^m \leq I_i^m) \right). \quad (4.1) \end{aligned}$$

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

$$\hat{I}^m(x,a) \equiv \Phi_{k_{I,N}}(x,a)' \hat{\xi}_{I,k_{I,N}}^m. \quad (4.2)$$

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

$$\hat{\xi}_{I,k_{I,N}}^m = \left(\sum_{i=1}^{N} \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m) \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m)' \right)^{-} \left(\sum_{i=1}^{N} \Phi_{k_{I,N}}(x_i^m, \hat{a}_i^m) I_i^m \right). \quad (4.3)$$

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

$$\hat{I}_{x_j}^m(x,a) \equiv \left(\frac{\partial \phi_1(x,a)}{\partial x_j}, \ldots, \frac{\partial \phi_{k_{I,N}}(x,a)}{\partial x_j} \right) \hat{\xi}_{I,k_{I,N}}^m, \quad (4.4)$$

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

$$\hat{I}_a^m(x,a) \equiv \left(\frac{\partial \phi_1(x,a)}{\partial a}, \ldots, \frac{\partial \phi_{k_{I,N}}(x,a)}{\partial a} \right) \hat{\xi}_{I,k_{I,N}}^m. \quad (4.5)$$

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

$$\hat{\xi}^m_{h,k_{h,N}} \equiv \left(\sum_{i=1}^{N} \Phi_{k_{h,N}}(x_i^m, \hat{a}_i^m)\Phi_{k_{h,N}}(x_i^m, \hat{a}_i^m)'\right)^{-} \left(\sum_{i=1}^{N} \Phi_{k_{h,N}}(x_i^m, \hat{a}_i^m)h_i^m\right). \tag{4.6}$$

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

$$\hat{h}^m(x,a) \equiv \Phi_{k_{h,N}}(x,a)'\hat{\xi}^m_{h,k_{h,N}}, \tag{4.7}$$

$$\hat{h}^m_a(x,a) \equiv \left(\frac{\partial \phi_1(x,a)}{\partial a}, \dots, \frac{\partial \phi_{k_{h,N}}(x,a)}{\partial a}\right)\hat{\xi}^m_{h,k_{h,N}}, \tag{4.8}$$

and

$$\hat{h}^m_{x_j}(x,a) \equiv \left(\frac{\partial \phi_1(x,a)}{\partial x_j}, \dots, \frac{\partial \phi_{k_{h,N}}(x,a)}{\partial x_j}\right)\hat{\xi}^m_{h,k_{h,N}}. \tag{4.9}$$

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

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 (3.8). Replace $I^m(x,a)$, $h^m(x,a)$ and their derivatives in equation (3.8) (i.e., in the expressions of $B(x,a)$ and $A(x,a)$) with their counterparts estimated in Section 4.1.1. Use the series of basis functions $\Phi_{k_{x_j,MN}}(x,a) = (\phi_1(x,a), \dots, \phi_{k_{x_j,MN}}(x,a))'$ to approximate $e_{x_j}(x,a)/e(x,a)$ and $\Phi_{k_{a,MN}}(x,a) = (\phi_1(x,a), \dots, \phi_{k_{a,MN}}(x,a))'$ to approximate $e_a(x,a)/e(x,a)$. Let the series coefficients be $\beta_{x_j,k_{x_j,MN}}$ ($j = 1, \dots, d_x$) and $\beta_{a,k_{a,MN}}$, respectively. And let $\beta_{MN} \equiv (\beta'_{x_1,k_{x_1,MN}}, \dots, \beta'_{x_{d_x},k_{x_{d_x},MN}}, \beta'_{a,k_{a,MN}})'$. Then, for each seller i and each market m, one obtains an estimated version of the equations (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

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

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

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

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:

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

where

$$\hat{S}_{\Phi\Phi} \quad \equiv \quad (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m), \tag{4.10}$$

$$\hat{S}_{\Phi A} \quad \equiv \quad (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Phi}(x_i^m, \hat{a}_i^m)' \hat{A}^m(x_i^m, \hat{a}_i^m). \tag{4.11}$$

In equations (4.10) and (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

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

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

Then the estimated ratios of the quality function are

$$\begin{cases} \widehat{\frac{e_{x_1}(x,a)}{e(x,a)}} & \equiv \hat{g}_1(x,a) = \Phi_{k_{x_1,MN}}(x,a)'\hat{\beta}_{x_1,k_{x_1,MN}}, \\ \vdots & \vdots \\ \widehat{\frac{e_{x_{d_x}}(x,a)}{e(x,a)}} & \equiv \hat{g}_{d_x}(x,a) = \Phi_{k_{x_{d_x},MN}}(x,a)'\hat{\beta}_{x_{d_x},k_{x_{d_x},MN}}, \\ \widehat{\frac{e_a(x,a)}{e(x,a)}} & \equiv \hat{g}_{d_x+1}(x,a) = \Phi_{k_{a,MN}}(x,a)'\hat{\beta}_{a,k_{a,MN}}. \end{cases} \tag{4.12}$$

By replacing the relevant ratios of the quality function in equation (3.5) with their estimators given in equation (4.12), one obtains the estimator of the quality function

$$\begin{aligned} \hat{e}(x,a) &= \exp\left(\sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} \hat{g}_j(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})ds_j + \int_{\bar{a}}^{a} \hat{g}_{d_x+1}(x,t)dt\right) \\ &= \exp\left(\sum_{j=1}^{d_x} \int_{\bar{x}_j}^{x_j} \left[\Phi_{k_{x_j,MN}}(x_1,\ldots,x_{j-1},s_j,\bar{x}_{j+1},\ldots,\bar{x}_{d_x},\bar{a})'\hat{\beta}_{x_j,k_{x_j,MN}}\right] ds_j \right. \\ &\quad \left. + \int_{\bar{a}}^{a} \left[\Phi_{k_{a,MN}}(x,t)'\hat{\beta}_{a,k_{a,MN}}\right] dt\right). \end{aligned} \tag{4.13}$$

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 (3.6). Combined with the sellers' FOC in equation (2.1), they imply that for $\forall (x,a) \in \mathcal{X} \times \mathcal{A}$,

$$\begin{cases} \nabla_x I^m(x,a) & = \left[\nabla_x h^m(x,a) + h^m(x,a)\frac{\nabla_x e(x,a)}{e(x,a)}\right] \cdot U_h(h^m(x,a),x,a), \\ I_a^m(x,a) & = \left[h_a^m(x,a) + h^m(x,a)\frac{e_a(x,a)}{e(x,a)}\right] \cdot U_h(h^m(x,a),x,a). \end{cases} \tag{4.14}$$

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 (4.14). Specifically, define

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

and

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

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

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

where

$$
\hat{S}_{\Psi\Psi} \quad \equiv \quad (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)' \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m), \tag{4.15}
$$

$$
\hat{S}_{\Psi I} \quad \equiv \quad (MN)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{N} \hat{S}_{\Psi}(h_i^m, x_i^m, \hat{a}_i^m)' \hat{S}_{I}(h_i^m, x_i^m, \hat{a}_i^m). \tag{4.16}
$$

In equations (4.15) and (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

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

The steps described in Sections 4.1.1-4.1.3 complete the estimation of the supply side structural functions (e, U_h).

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 (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 (2.4), 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 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.

4.2 Uniform Rates of Convergence of Structural Function Estimators

In this section and Appendix 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.

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

This subsection derives the convergence rates of the within market kernel estimators of the conditional quantile a_i^m given in equation (4.1).

Assumption 13. *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.*

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

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 3. *Suppose that Assumption 13 is satisfied. Then,*

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

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

Assumption 15. *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 15, 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 16. *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 5, 6 and 16 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.

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

Assumption 17. *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 $|I^m - \Phi'_{k_I}\xi^m_{0,I,k_I}|_{\delta_I} \leq Ck_I^{-\alpha_I}$ 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 $|h^m - \Phi'_{k_h}\xi^m_{0,h,k_h}|_{\delta_h} \leq Ck_h^{-\alpha_h}$ for all positive integers k_h.*[28]

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 (4.2) or equation (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 (4.4), equation (4.5), equation (4.8) or equation (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 paper. Moreover, note that $\nu_{l,N} = \mathcal{O}(\nu_{l_j,N})$, and $\nu_{l,N} = \mathcal{O}(\nu_{l_a,N})$.

Theorem 4. *Suppose that Assumptions 14-17 and the conditions of Theorem 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 $\sqrt{k_{l,N}}\nu_{a,N}\zeta_a(k_{l,N}) \to 0$ 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 5. *Suppose that the conditions for Theorem 4 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.

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

[29] Recall equations (4.3) and (4.6).

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 18. *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 $j = 1, \ldots, d_x$) and $\beta_{0,a,k_a} \in \mathbb{R}^{k_a}$ such that $|e_{x_j}/e - \Phi'_{k_{x_j}} \beta_{0,x_j,k_{x_j}}|_{\delta_e} \leq C k_{x_j}^{-\alpha_e}$ and $|e_a/e - \Phi'_{k_a} \beta_{0,a,k_a}|_{\delta_e} \leq C k_a^{-\alpha_e}$ for all positive integers k_{x_j} ($j = 1, \ldots, d_x$) and k_a.*

Define

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

where

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

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

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 19. *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 20. *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 paper.

Lemma 4. *Suppose that Assumptions 18-20 and the conditions of Theorems 3-5 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 6. *Suppose that the conditions for Lemma 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 7. *Suppose that the conditions for Theorem 6 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.

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

Assumption 22. *Suppose that there exists some positive constant B_U such that $|U_h|_1 \leq B_U$.*

Assumption 23. *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 24. *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

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

Assumption 25. *Suppose that $\lambda_{\min}(\mathbb{E}(S_{\Psi\Psi})) \geq c > 0$.*

Lemma 5. *Suppose that Assumptions 21-25 and the conditions for Theorem 6 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 paper.

Theorem 8. *Suppose that the conditions of Lemma 5 are satisfied. Then*

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

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.

5 Empirical Illustration in Labor Markets

In this section, I apply the estimation procedure provided in Sections 4.1 to estimate the efficiency (quality) function e in labor markets. Section 5.1 introduces the data set, and Section 5.2 estimates the workers' unobserved efficiency function.

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]

[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

Figure 5.1 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 within- and cross-market variation is crucial for the identification of the unobserved efficiency function.

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 2.1 and 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 5.2 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 5.2 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 9 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 9 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 9 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

to more observed variables poses no theoretical problem, but it may take more computing time.

[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.

defined as in equation (4.13). Figure 5.2 plots the estimated efficiency function $\hat{e}(x, a)$ on the support $\mathcal{X} \times \mathcal{A}$.

Figure 5.2 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]

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 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 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

[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.

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

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 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.

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 $i = 1, \ldots, N^m$, $m = 1, \ldots, M$). 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 4.1 and will remain constant under the intervention.[40] The equilibrium is solved for each market m separately.

A.1 Algorithm to Solve for Counterfactual Equilibrium

The algorithm consists of two steps:

1. Obtain the general solution to the ODE in equation (2.6) that characterizes the equilibrium.

2. Determine the initial value condition of the ODE in equation (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 (2.6). 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 (2.6) can be approximated numerically. In particular, the first term of the numerator can be approximated as

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

[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.

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' 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

$$\begin{aligned}
T_1(z,x) &\equiv \hat{U}_{ha}\left(\frac{z}{\hat{e}\left(x, a^*(z,x)\right)}, x, a^*(z,x)\right) \hat{e}\left(x, a^*(z,x)\right), \\
T_2(z,x) &\equiv P_z^m(z) \hat{e}\left(x, a^*(z,x)\right) \hat{e}_a\left(x, a^*(z,x)\right), \\
T_3(z,x) &\equiv \hat{U}_{hh}\left(\frac{z}{\hat{e}\left(x, a^*(z,x)\right)}, x, a^*(z,x)\right) \frac{z}{\hat{e}\left(x, a^*(z,x)\right)} \hat{e}_a\left(x, a^*(z,x)\right),
\end{aligned}$$

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 (2.6) 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 (2.6) 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 equilibrium outcomes, such as $\tilde{I}_i^m$ and $\tilde{h}_i^m$, can be constructed as a result.[43]

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 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 A.1 to solve for the equilibrium in a market with 1000 worker-firm pairs.[44] The first panel of Figure A.2 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

[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.

[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.

cases when the standard deviations of the perturbations are 0.05 and 0.1, respectively.

Figure A.2 has two important implications. First, even though I approximate the integrals in the equilibrium condition in equation (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 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 A.2 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.

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.

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 $h^m(x,a) \equiv (h^m_1(x,a), \ldots, h^m_L(x,a))'$ 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 1 and the payment equation (2.3) in market m, then we have

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

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

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

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

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

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

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

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 $B_{L+1}^m(x,a)$ is

$$B_{L+1}^m(x,a) \equiv \left(B_{L+1,2}^m,\ldots,B_{L+1,L}^m\right),$$

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

$$B_{L+1,l'}^m \;\equiv\; \frac{\partial I^m(x,a)}{\partial a}\nabla_x h_{l'}^m(x,a) - \frac{\partial h_{l'}^m(x,a)}{\partial a}\nabla_x I^m(x,a)$$

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

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

where

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

and

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

Therefore, there exists a unique solution of $(\nabla e_1(x,a)', \ldots, \nabla e_L(x,a)', \frac{e_2(x,a)}{e_1(x,a)}, \ldots, \frac{e_L(x,a)}{e_1(x,a)})'$ 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 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 $e_l(x,a)$ $(l = 1,\ldots,L)$ with the steps described in the proof of Theorem 1, one can recover all the quality functions $e_l(x,a)$ $(l = 1,\ldots,L)$.

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

D Proofs of the Theorems in Section 4.2

D.1 Proof of the Theorem in Section 4.2.1

This section provides the proof of Theorem 3. But some notation is needed first. Let $\Lambda_i^m \equiv \Lambda_{k_{Q,N}}(x_i^m)$, $\omega_{ij} \equiv \mathbb{I}(I_j^m \leq I_i^m) - F_{I^m|x^m}(I_i^m|x_j^m)$ $(i,j = 1,\ldots,N)$ and $\hat{W}^m \equiv \sum_{i=1}^N \Lambda_i^m \Lambda_i^{m\prime}/N$.

Lemma 6. *For $x^m \equiv (x_1^m,\ldots,x_N^m)$ and $k_{Q,N} \times 1$ vectors of functions $b_i(x^m)$ $(i = 1,\ldots,N)$, 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 7. *Suppose that Assumption 13 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. □

Proof of Theorem 3

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

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

D.2 Proofs of the Theorems in Section 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 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 (4.3) and equation (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 9 presents step (i), Lemma 10 presents step (ii), and step (iii) is given by Theorems 4 and 5.

Lemma 8. *Suppose that Assumptions 14 and 15 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 5 and 6, the joint density of (x,a) is bounded away from zero. Combine this with Assumption 15, then the results follow from equations (3.13)-(3.16) in Andrews (1991). □

Lemma 9. *Suppose that the conditions of Theorem 3 and Lemma 8 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 $i \in \{1, \ldots N\}$,

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

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 8, $\|\partial \Phi_{k_{l,N}}(x_i, \tilde{a}_i)/\partial a\| \leq C\zeta_a(k_{l,N})$. Then by Cauchy-Schwarz inequality, $\|\tilde{\Phi}_i - \Phi_i\| \leq C\zeta_a(k_{l,N})$ $|\hat{a}_i - a_i|$. Together with Theorem 3, 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

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

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

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

$$\begin{aligned}
\|\tilde{Q} - \bar{Q}\| &\leq \sum_{i=1}^{N} \|\tilde{\Phi}_i \tilde{\Phi}_i' - \Phi_i \Phi_i'\|/N \\
&\leq \sum_{i=1}^{N} \|\tilde{\Phi}_i - \Phi_i\|^2/N + 2\left(\sum_{i=1}^{N} \|\tilde{\Phi}_i - \Phi_i\|^2/N\right)^{1/2} \left(\sum_{i=1}^{N} \|\Phi_i\|^2/N\right)^{1/2}.
\end{aligned}$$

Moreover, by the Markov's inequality

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

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 10. *Suppose that Assumptions 16 and 17, and the conditions of Theorem 3 and Lemma 9 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 $i \in \{1, \ldots N\}$,

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

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 9. Since $\hat{a}_i$ and a_i are in $[0,1]$, so is $\tilde{a}_i$. Together with Assumptions 14 and 16, this implies that $|\partial l(x_i, \tilde{a}_i)/\partial a| \leq C$. Moreover, by Lemma 9, we have that $\lambda_{\min}(\tilde{Q}) \geq c$ with probability 1, so

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

Then (i) holds by Theorem 3 and Lemma 9 (iv).

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

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

where the last equality holds by Assumption 17. Therefore the result holds by Lemma 9 (iv). $\square$

Proof of Theorem 4

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,

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

The first equality holds by the Cauchy-Schwarz inequality, Assumption 17, and Lemmas 8 and 10. The second equality holds since $\zeta_0(k_{l,N}) \to \infty$ as $N \to \infty$. This completes the proof. $\square$

Proof of Theorem 5

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

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

The first equality holds by the Cauchy-Schwarz inequality, Assumption 17, and Lemmas 8 and 10. 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. $\square$

D.3 Proofs of the Theorems in Section 4.2.3

This subsection proceeds with the same steps as in Appendix 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 11. *Suppose that Assumption 16, and the conditions of Theorems 3 and 5 and Lemma 8 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)

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

(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 8, Assumption 16, the triangular inequality, and the Cauchy-Schwarz inequality, we get

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

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

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

Equation (D.2) implies that

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

And equation (D.3) implies that

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

As a result,

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

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

On the other hand, by Lemma 8, Assumption 16 and the Cauchy-Schwarz inequality, we have

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

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,

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

By similar argument, we also have

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

which implies that

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

As a result,

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

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 theorem 3, equation (D.4) and equation (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 5, equation (D.5) and equation (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 $j = 1, \ldots, d_x$, 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,

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

On the other hand, by the fact that $\hat{a}_i^m \in [0,1]$, Lemma 8, Assumption 16, 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

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

Now consider (ii),

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

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

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

So (ii) holds.

To prove (iii), note that

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

$$+2B_I^4 \sum_{j=1}^{d_x} \mathbb{E}\left[(MN)^{-1} \left\| \Phi_{k_{a,MN}}(x_i^m, a_i^m) \Phi_{k_{x_j,MN}}(x_i^m, a_i^m)' \right\|^2 \right]$$

$$+d_x^2 B_I^4 \mathbb{E}\left[\left\| \Phi_{k_{a,,N}}(x_i^m, a_i^m) \Phi_{k_{a,MN}}(x_i^m, a_i^m)' \right\|^2 \right], \tag{D.10}$$

where the first inequality holds by the definition of $S_{\Phi\Phi}$, Assumption 11, and that the second moment of a random variable is no less than its variance; the second inequality holds by Assumption 16. Recall that I assume the series basis functions are orthonormal (i.e. $Q = \mathbb{I}$), then by Lemma 8, we have that for $j = 1, \ldots, d_x$,

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

By the same token,

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

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

and

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

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

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

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 9(iv). This completes the proof of the lemma. $\square$

Define

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

where

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

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

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

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

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

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 12. *Suppose that Assumptions 18-20, and the conditions of Theorem 4 and Lemma 11 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 4 and 5, 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 $\Gamma_{x,a}^{(j)}(I, h)$ indexed by $(x, a) \in \mathcal{X} \times \mathcal{A}$ and $j \in \{1, \ldots, d_x\}$ as follows:

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

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

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

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

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

where the inequality holds by Assumptions 16 and 20(ii), and the triangular inequality. And in equation (D.15), the second terms of the summands are nonlinear functionals with

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

where the inequality holds by Assumptions 16 and 20(ii), the triangular inequality, and the Cauchy-Schwarz inequality.

By the consistency of $\hat{I}^m$, $\hat{h}^m$ and their derivatives, equation (D.16) and equation (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 (D.15), equation (D.16) and Assumptions 16 and 20(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 11(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 11(iii).

For (ii), consider

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

$$= \mathcal{O}_p\left(\sum_{j=1}^{d_x} k_{x_j,MN}^{-2\alpha_e} + k_{a,MN}^{-2\alpha_e}\right).$$

where the second equality holds by equation (3.7); the third equality holds by Assumptions 16 and 18, 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 11(iii).

For (iii), note that

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

where the third equality holds by Theorem 5, Assumptions 18 and 19, 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

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

So (iii) holds by Lemma 11(iii). This completes the proof of the lemma. □

Proof of Lemma 4

Proof. By the triangular inequality, we have

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

So the result follows by Lemma 12. □

Proof of Theorem 6

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

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

Then the result follows by Lemma 4, 8, Assumption 18, 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. □

Proof of Theorem 7

Proof. By Assumption 14, Theorem 6 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

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

and

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

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

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

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

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

where the inequality holds since $\mathcal{X} \times \mathcal{A}$ is compact by Assumption 14. 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,

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

And the result follows by equation (D.18), equation (D.19) and Theorem 6. □

D.4 Proofs of the Theorems in Section 4.2.4

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

Lemma 13. *Suppose that Assumption 23 and the conditions for Theorem 3 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 8. □

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 14. *Suppose that Assumptions 22, 23 and 25, and the conditions for Theorem 6 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

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

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 13, Assumptions 16, 19 and 20(ii), the triangular inequality, and the Cauchy-Schwarz inequality, we get

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

By the same token,

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

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

$$\begin{aligned}
& \left\|\bar{S}_\Psi(h^m_i,x^m_i,\hat{a}^m_i)-S_\Psi(h^m_i,x^m_i,a^m_i)\right\|^2 \\
\leq\ & C|\hat{a}^m_i-a^m_i|^2\zeta^2_a(k_{U,MN}).
\end{aligned} \tag{D.22}$$

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

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

This implies, together with Lemma 13, Assumptions 16, 19 and 20(ii), and the Cauchy-Schwarz inequality, that

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

Now consider (i),

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

where the inequality holds by the triangular inequality and the Cauchy-Schwarz inequality. Combine this result with Theorem theorem 3, equation (D.22) and equation (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:

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

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

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

Note that

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

and

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

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 $\Delta\left(\frac{e_{x_j}}{e}\right) \equiv \frac{\widehat{e_{x_j}}}{e} - \frac{e_{x_j}}{e}$ for $j = 1, \ldots, d_x$, and let $\Delta\left(\frac{e_a}{a}\right) \equiv \frac{\widehat{e_a}}{e} - \frac{e_a}{e}$. Then we have

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

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

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

Moreover,

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

$$+C\left|R\Upsilon^{(j)}\left(h^m,\frac{e_{x_j}}{e};\Delta h^m,\Delta\left(\frac{e_{x_j}}{e}\right)\right)\right|^2, \tag{D.29}$$

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

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

where the inequality holds by Assumptions 16, 19 and 20(ii), and the triangular inequality. And the second term in equation (D.29) is a nonlinear functional with

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

where the inequality holds by Assumptions 16, 19 and 20(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 (D.30) and equation (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 (D.29), equation (D.30), Assumptions 16 and 20, we get

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

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

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

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

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

On the other hand, by equation (D.28), Assumptions 22 and 25, and the Cauchy-Schwarz inequality, we have

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

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 (D.34) and equation (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

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

where the first inequality holds by the definition of $S_{\Psi\Psi}$, Assumption 11, and that the second moment of a random variable is no less than its variance; the second inequality holds by Assumptions 16, 19 and 20(ii). Recall that I assume the series basis functions are orthonormal, then by Lemma 13, we have

$$
\mathbb{E}\left[\left\|\Psi_{k_{U,MN}}(h_i^m,x_i^m,a_i^m)\Psi_{k_{U,MN}}(h_i^m,x_i^m,a_i^m)'\right\|^2\right]
$$

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

Plug the bounds in equation (D.37) into equation (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, $k_{U,MN}\nu_{ej,M,N} \to 0$ for $j = 1, \ldots, d_x$, and $k_{U,MN}\nu_{ea,M,N} \to 0$, we have $\|\bar{S}_{\Psi\Psi} - S_{\Psi\Psi}\| = o_p(1)$. Then (iii) follows by the same argument for the proof of Lemma 9(iv). This completes the proof of the lemma. □

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

$$
\bar{S}_I(h_i^m, x_i^m, \hat{a}_i^m) \equiv \left(\nabla_x I^m(x_i^m, \hat{a}_i^m)', \; I_a^m(x_i^m, \hat{a}_i^m) \right)',
$$

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

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

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

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

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

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 15. *Suppose that Assumption 21 and the conditions of Lemma 14 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

$$\left\|\hat{S}_{\Psi\Psi}^{1/2}\left(\hat{\gamma}_{k_{U,MN}} - \bar{\gamma}_{k_{U,MN}}\right)\right\|^2 = \left(\hat{S}_I - \bar{S}_I\right)'\hat{S}_\Psi\hat{S}_{\Psi\Psi}^{-}\hat{S}_\Psi'\left(\hat{S}_I - \bar{S}_I\right)/(MN)^2$$

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

where the second equality holds by Theorem 5. So (i) holds by Lemma 14(iii).

For (ii), note that equation (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 16, 19, 20(ii) and 21. 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 14(iii).

For (iii), consider the mean value expansion

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

Then we have

$$
\left|\Psi_{k_{U,MN}}(h^m(x_i^m,\hat{a}_i^m),x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}-\Psi_{k_{U,MN}}(h_i^m,x_i^m,\hat{a}_i^m)'\gamma_{0,k_{U,MN}}\right|^2
$$

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

where the first inequality holds by the triangular inequality and the Cauchy-Schwarz inequality; the second inequality holds by Assumptions 16, 20(ii), 21, 22, and equation (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 theorem 3, 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 14(iii).

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

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

In equation (D.40),

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

where the first inequality holds by the triangular inequality and the Cauchy-Schwarz inequality; the second inequality holds by Assumptions 21 and 22. Together, equation (D.32), equation (D.33), equation (D.40), equation (D.41) and Lemma 14(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 5

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 15. □

Proof of Theorem 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 5 and 13, Assumption 21, and the Cauchy-Schwarz inequality. So the result holds. □

References

Ackerberg, D., C. L. Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics 6*, 4171–4276.

Andrews, D. W. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica: Journal of the Econometric Society*, 307–345.

Bajari, P. and C. L. Benkard (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S., J. Levinsohn, and A. Pakes (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, S. and A. Pakes (2007). The pure characteristics demand model. *International Economic Review 48*(4), 1193–1225.

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

Brown, J. N. and H. S. Rosen (1982). On the estimation of structural hedonic price models. *Econometrica: Journal of the Econometric Society*, 765–768.

Chay, K. Y. and M. Greenstone (2005). Does air quality matter? evidence from the housing market. *Journal of political Economy 113*(2), 376–424.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Chernozhukov, V., A. Galichon, and M. Henry (2014). Identifying multi-attribute hedonic models. *Available at SSRN 2392549*.

Chiappori, P.-A., R. J. McCann, and L. P. Nesheim (2010). Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory 42*(2), 317–354.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Ekeland, I. (2010). Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types. *Economic Theory 42*(2), 275–315.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004). Identification and estimation of hedonic models. *Journal of Political Economy 112*(1 pt 2).

Epple, D. (1987). Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. *Journal of political economy 95*(1), 59–80.

Epple, D. and H. Sieg (1999). Estimating equilibrium models of local jurisdictions. *The Journal of Political Economy 107*(4), 645–681.

Fessler, P. and M. Kasy (2016). How to use economic theory to improve estimators, with an application to labor demand and wage inequality in europe. Technical report.

Fessler, P., M. Kasy, and P. Lindner (2013). Survey mode effects on income inequality measurement. *Economic Analysis 617*, 495–7730.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica*, 953–973.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of labor economics 4*, 1–102.

Harrison, D. and D. L. Rubinfeld (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management 5*(1), 81–102.

Heckman, J. J., R. L. Matzkin, and L. Nesheim (2010). Nonparametric identification and estimation of nonadditive hedonic models. *Econometrica 78*(5), 1569–1591.

Hildebrandt, T. H. and L. M. Graves (1927). Implicit functions and their differentials in general analysis. *Transactions of the American Mathematical Society 29*(1), 127–153.

Hofferth, S. L., S. M. Flood, and M. 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.

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

Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of political Economy*, 410–442.

Kahn, S. and K. Lang (1988). Efficient estimation of structural hedonic systems. *International Economic Review*, 157–166.

Lorentz, G. G. (1986). *Approximation of functions.* Chelsea, New York.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics 20*(4), 445–465.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 1339–1375.

Nesheim, L. (2015). Identification in multidimensional hedonic models. *Working Paper*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*(4), 705–729.

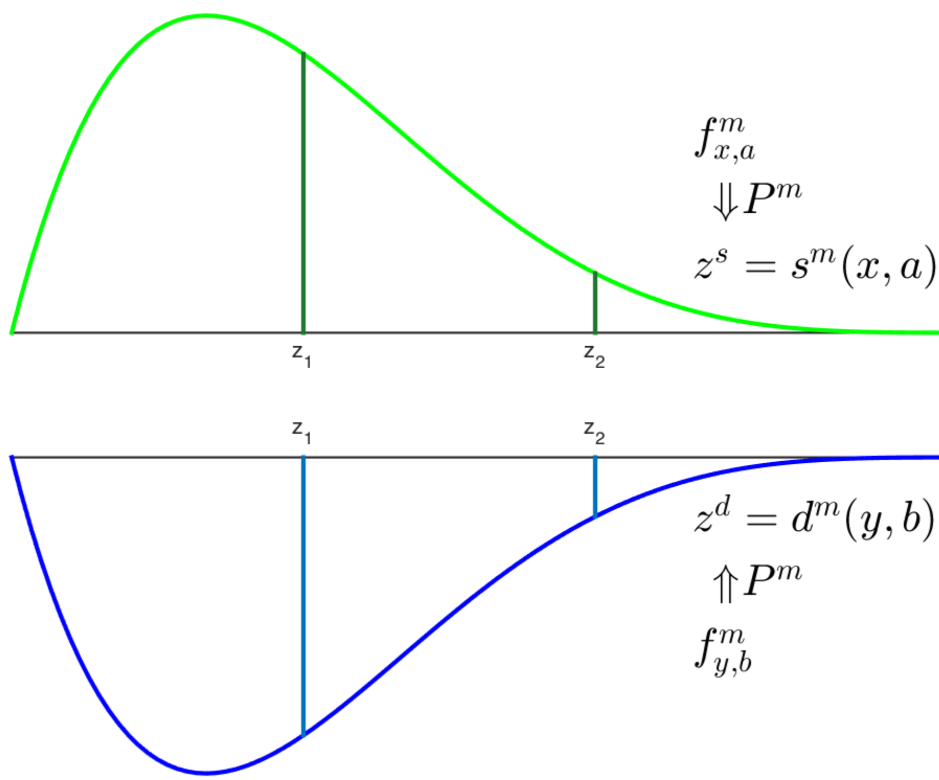
Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, 34–55.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15*(4), 321–340.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics 155*(1), 56–70.

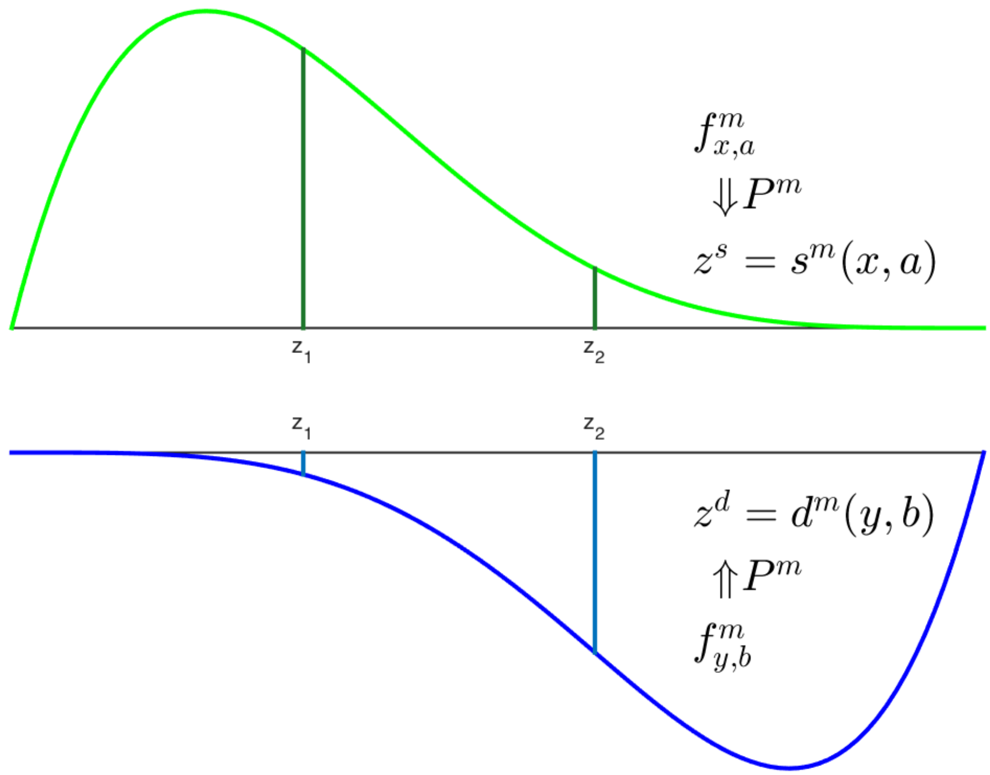
Sieg, H., V. K. Smith, H. S. Banzhaf, and R. Walsh (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review 45*(4), 1047–1077.

Stock, J. 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.



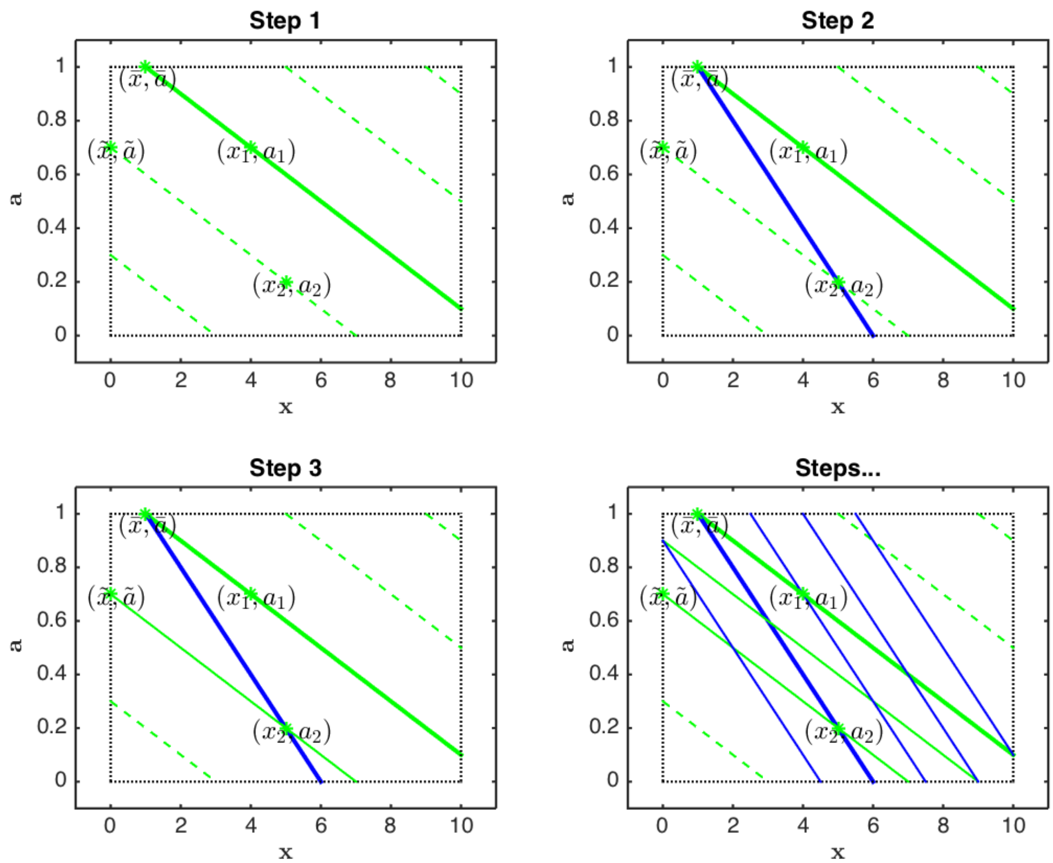
The green line illustrates the distribution of the optimal effective labor supply z^s under the price schedule function P^m in market m, as a function of sellers' observed characteristics x and unobserved characteristic a, which follow the distribution $f^m_{x,a}$. Similarly, the blue line illustrates 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.

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 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 2.2: Off Equilibrium



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

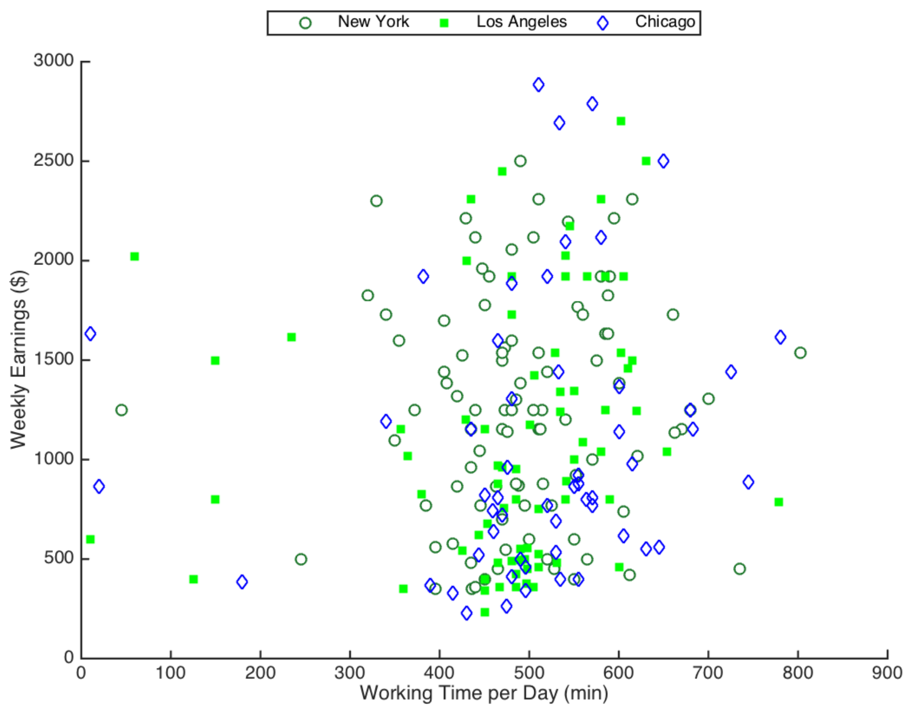
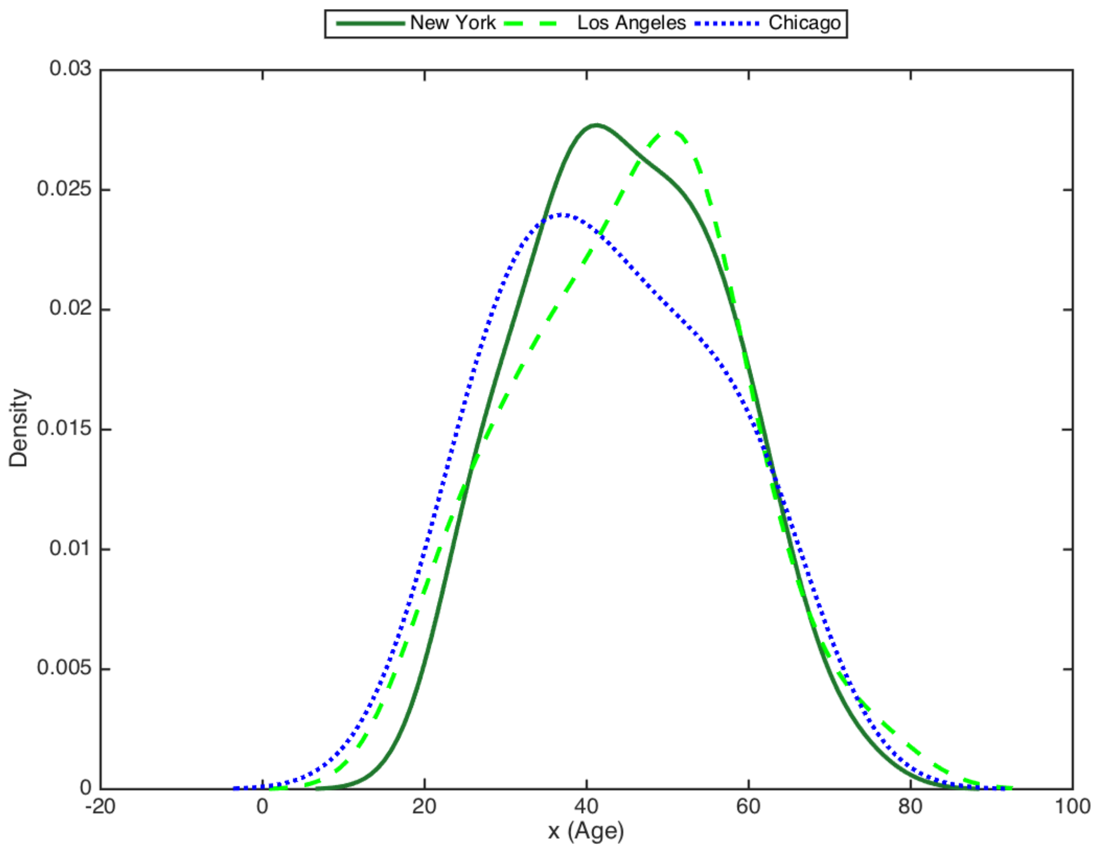
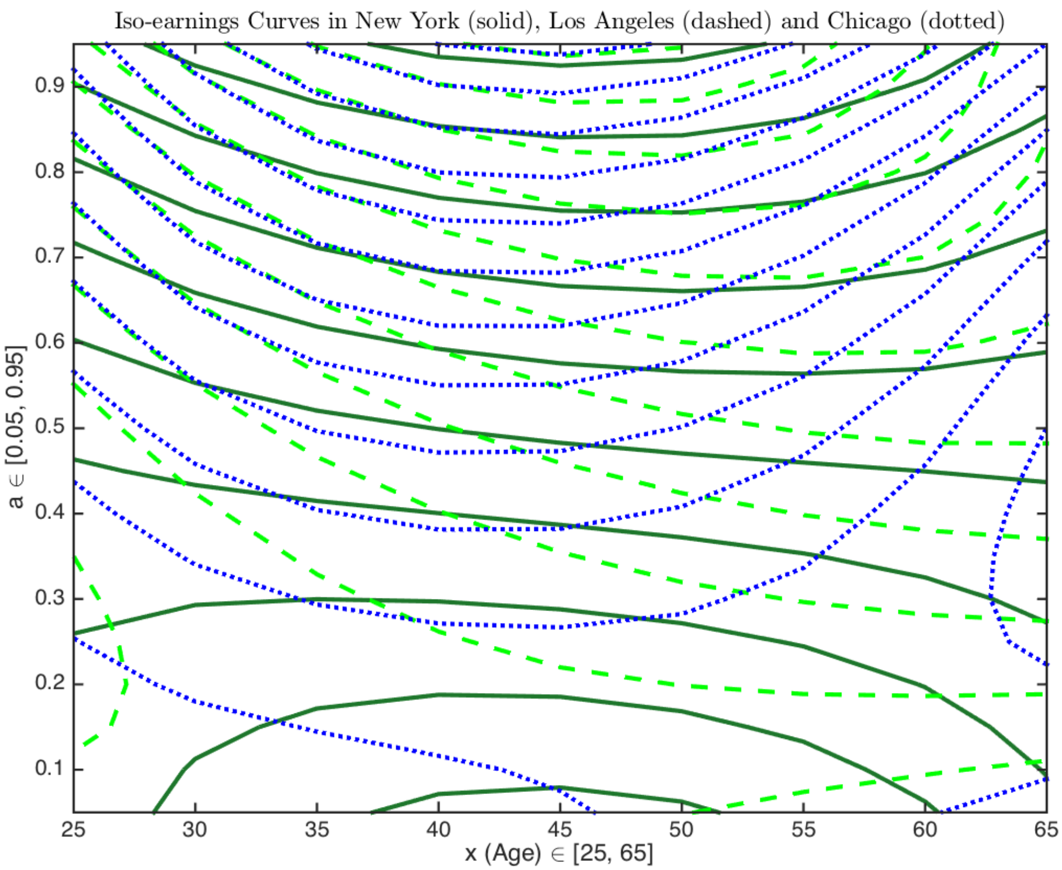


Figure 5.1: Scatter Plots of Weekly Earnings and Working Time in the Three Cities



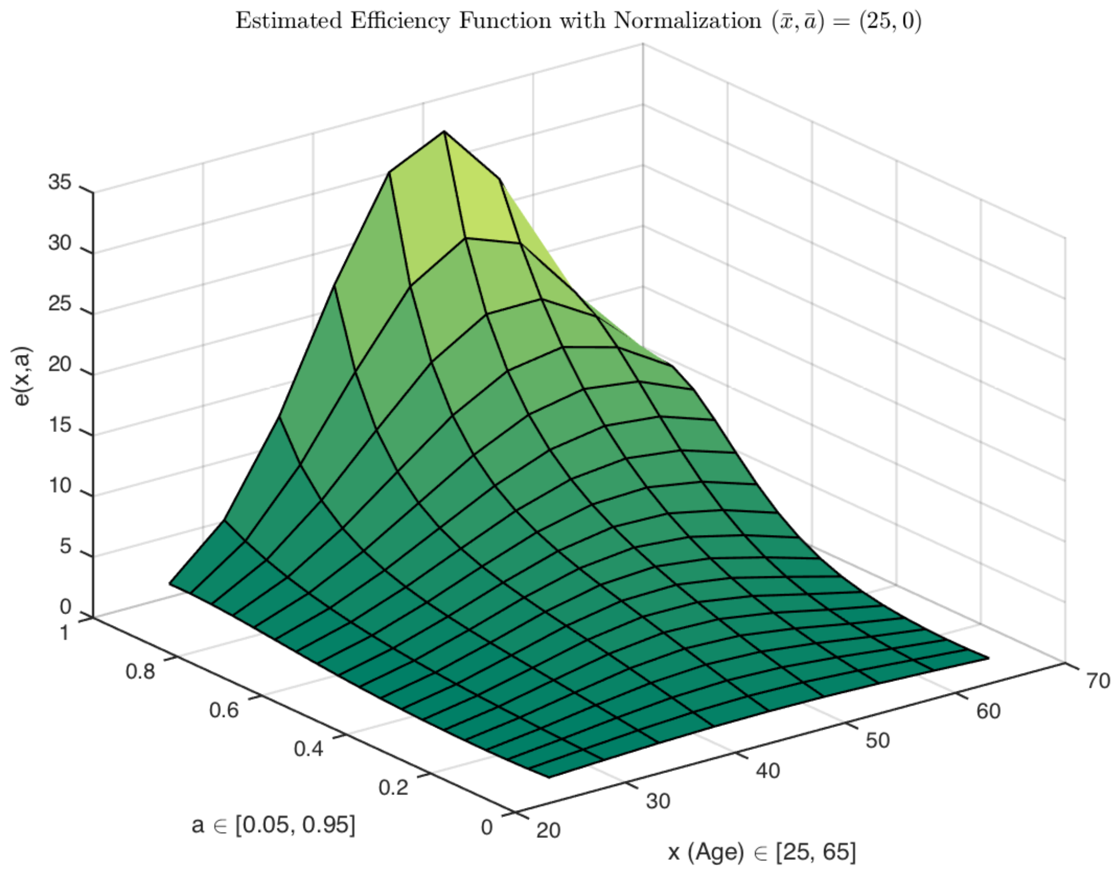
For ease of illustration, age is used as the single observed characteristic (x) of the workers. This figure shows that there is decent cross-market variation in the distributions of x, which drives (partially) the cross-market variation in the equilibrium payment functions.

Figure 5.2: Distributions of Age in the Three Cities



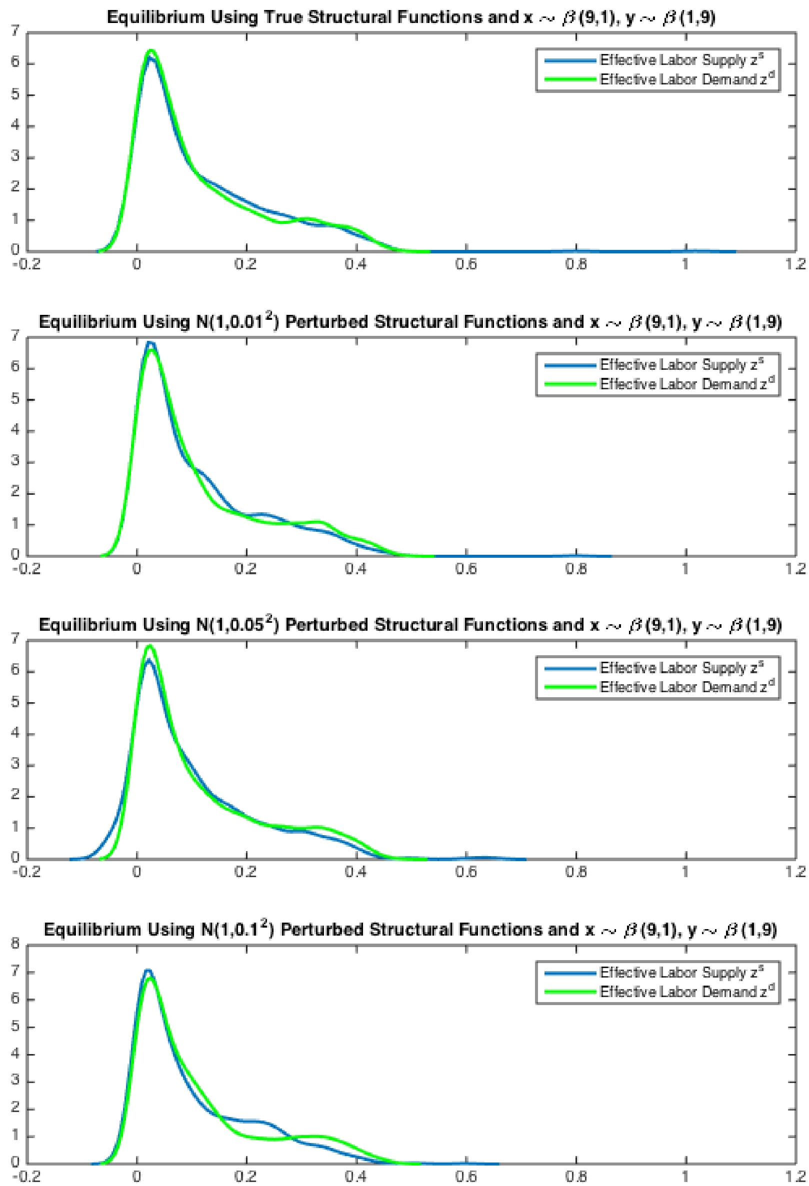
This figure shows representative iso-earnings curves for the three cities, on the support of age (x) and "ability" (a). For majority of the support, the iso-earnings curves from at least two markets cross. This suggests that the identification condition for the efficiency function $e(x, a)$ is satisfied.

Figure 5.3: Iso-Earnings Curves in the Three Cities



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

Figure 5.4: Estimated Efficiency Function $e(x, a)$



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.

Figure A.1: Numerically Solved Equilibrium Using True and Perturbed Structural Functions