Title
Essays on Identification and Estimation of Structural Economic Models

Author
Wu, Shaomin

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Identification and Estimation of Structural Economic Models

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Shaomin Wu

2023

ABSTRACT OF THE DISSERTATION

Essays on Identification and Estimation of Structural Economic Models

by

Shaomin Wu

Doctor of Philosophy in Economics

University of California, Los Angeles, 2023

Professor Rosa Liliana Matzkin, Chair

This dissertation consists of three chapters that study the identification and estimation of structural economic models. Chapter 1, "Identification and Estimation of Nonseparable Triangular Equations with Mismeasured Instruments" studies the nonparametric identification and estimation of the marginal effect of an endogenous variable X on the outcome variable Y, given a potentially mismeasured instrument variable W^*, without assuming linearity or separability of the functions governing the relationship between observables and unobservables. In order to address the challenges arising from the co-existence of measurement error and nonseparability, I first employ the deconvolution technique from the measurement error literature to identify the joint distribution of Y, X, W^* using two error-laden measurements of W^*. I then recover the structural derivative of the function of interest and the "Local Average Response" (LAR) from the joint distribution via the "unobserved instrument" approach in Matzkin (2016). I also propose nonparametric estimators for these parameters and derive their uniform rates of convergence. Monte Carlo exercises show evidence that the estimators I propose have good finite sample performance.

Chapter 2, "Two-step Estimation of Network Formation Models with Unobserved Heterogeneities and Strategic Interactions", characterizes the network formation process as

a static game of incomplete information, where the latent payoff of forming a link between two individuals depends on the structure of the network, as well as private information on agents' attributes. I allow agents' private unobserved attributes to be correlated with observed attributes through individual fixed effects. Using data from a single large network, I propose a two-step estimator for the model primitives. In the first step, I estimate agents' equilibrium beliefs of other people's choice probabilities. In the second step, I plug in the first-step estimator to the conditional choice probability expression and estimate the model parameters and the unobserved individual fixed effects together using Joint MLE. Assuming that the observed attributes are discrete, I showed that the first step estimator is uniformly consistent with rate $N^{-1/4}$, where N is the total number of linking proposals. I also show that the second-step estimator converges asymptotically to a normal distribution at the same rate.

Chapter 3, "Identification and Estimation in Differentiated Products Markets Where Firms Affect Consumers' Attention" studies the nonparametric identification and estimation of a demand and supply system where firms affect consumers' consideration sets via costly marketing inputs, when market-level data is available. On the demand side, I characterize preferences and considerations nonparametrically, allowing rich heterogeneities and correlations between them. On the supply side, I characterize firms' optimal choices by a set of first-order conditions without specifying the form of the oligopoly model. The demand and supply sides form a simultaneous system of equations in the spirit of Berry and Haile (2014). I then show the identification of the system using the method proposed by Matzkin (2015). Moreover, using the variations of exclusive regressors entering preferences and considerations respectively, I separately identify features of the utility functions and the attention functions. Based on the constructive identification results, I propose nonparametric estimators of the demand, utility, and attention functions and show their asymptotic properties.

The dissertation of Shaomin Wu is approved.

Elisabeth Honka

Shuyang Sheng

Zhipeng Liao

Jinyong Hahn

Rosa Liliana Matzkin, Committee Chair

University of California, Los Angeles

2023

DEDICATIONS

To my parents and my family

Contents

List of Tables

ACKNOWLEDGMENT

I am especially grateful to my Committee Chair Rosa Matzkin, for her continuous support and invaluable guidance throughout my PhD studies. I am extremely thankful for the support and advice from my other committee members Zhipeng Liao, Shuyang Sheng, Jinyong Hahn, and Elisabeth Honka. I thank Andres Santos, Denis Chetverikov, and other participants at the UCLA econometrics proseminar for helpful comments and discussions. All errors are mine.

VITA

Education

University of California, Los Angeles

C.Phil. in Economics	2019
M.A. in Economics	2018

Peking University

M.A. in Economics	2017
B.A. in Economics	2015

Honors and Awards

Honor Pass in Econometrics Comprehensive Exam, UCLA	2018
Honor Pass in Microeconomics Comprehensive Exam, UCLA	2018
UCLA Economic Departmental Graduate Fellowship	2017-2018
Distinguished Graduate, Peking University	2017
Award for Academic Excellence, Peking University	2016
Peking University Guanghua School of Management Graduate Fellowship	2015-2017

Chapter 1

Identification and Estimation of Nonseparable Triangular Equations with Mismeasured Instruments

1.1 Introduction

This paper studies the nonparametric identification and estimation of the marginal effect of an endogenous variable X on the outcome variable Y, given a potentially mismeasured instrument variable W^*, without assuming linearity or separability of the functions governing the relationship between observables and unobservables. Without measurement error, this type of model is referred to as "nonseparable triangular equations" and its identification was studied by Chesher (2003), Imbens and Newey (2009), and Shaikh and Vytlacil (2011). Measurement error on the instrument variable poses additional challenges to the identification and estimation of the model primitives. Because of nonseparability, simply using an error-laden measurement as the instrument leads to inconsistent results. Because of measurement error, the true value of W^* is unobserved, making it impossible to proceed using existing methods in Imbens and Newey (2009). In this paper, I propose a way to deal with the two difficulties mentioned above and show the identification of average and individual-level marginal effects. I also propose estimators for these parameters. To illustrate ideas, I study the following model:

$$Y = m(X, \epsilon)$$
$$X = h(W^*, \eta),$$

where X is endogenous in the sense that it's correlated with ϵ, and W^* is the instrument variable independent with both of the error terms ϵ, η but cannot be measured accurately. Following the measurement error literature, I assume that there are two error-laden measurements of W^*, denoted as W_1, W_2. Denote the corresponding measurement errors as $\Delta W_1, \Delta W_2$, i.e. $W^* = W_1 + \Delta W_1$, $W^* = W_2 + \Delta W_2$. I could allow a vector of observed exogenous control variables Z to enter both equations and all the arguments will carry over by conditioning on Z, so they are omitted here for simplicity.

To give an example of where this model can be used, consider the Engel curve estimation studied in Blundell et al. (2007). They used Sieve Minimum Distance to estimate a semi-nonparametric model, but the curve can be estimated under less restrictive modeling assumptions as done in Imbens and Newey (2009). Y here is the share of expenditure on a commodity of a household. X is the log of the household's total expenditure. Z is a vector capturing the household's demographic composition and ϵ captures the household's unobserved heterogeneity. Researchers might be interested in the average response of households' expenditure on food Y, to changes in household's total expenditure X, holding the distribution of unobserved household heterogeneity $F_{\epsilon|X=x}$ fixed. This parameter is called "Local Average Response" (LAR) by Altonji and Matzkin (2005) and if the function m is differentiable, it can be written as

$$\mathbf{LAR}(x) = \int \frac{\partial m(x, e)}{\partial x} f_{\epsilon|X=x}(e) de. \tag{1.1}$$

In addition, researchers might also be interested in the structural derivative of the function m. This is a more disaggregate level parameter. It stands for the marginal response of Y to changes in X for a specific household, say the household with log total expenditure equal $\bar{x}$, and share of food expenditure equal $\bar{y}$. Denote the structural derivative of this household as $\rho(\bar{y}, \bar{x})$. If m is differentiable and strictly monotone in its second argument for all values of X, $\rho(\bar{y}, \bar{x})$ can be written as

$$\rho(\bar{y}, \bar{x}) = \left. \frac{\partial m(x, \epsilon)}{\partial x} \right|_{x=\bar{x}, \epsilon=m^{-1}(\bar{y}, \bar{x})}, \tag{1.2}$$

where m^{-1} is the inverse of m with respect to its second argument, and $m^{-1}(\bar{y}, \bar{x})$ is the value of ϵ of the specific household one is interested in. To estimate these parameters, under the assumption that heterogeneity in earnings is not correlated with households' preferences over consumption, one can use the income of the head of the household as the IV W^*. The income variable is likely to suffer from measurement errors. Researchers could obtain multiple measurements of it from panel data, for example.

To see how nonseparability makes it harder to identify parameters like the LAR and the structural derivative when the IV W^* is mismeasured, consider the case when both equations are linear:

$$Y = \alpha_0 + \alpha_1 X + \epsilon$$
$$X = \beta_0 + \beta_1 W^* + \eta,$$

where $E[\epsilon W^*] = 0$ (exclusion restriction) and $E[XW^*] \neq E[X]E[W^*]$ (relevance condition). Suppose I have $W_2 = W^* + \Delta W_2$ as an error-laden measurement of W^*, it's not hard to verify that $E[\epsilon W_2] = 0$ and $X \not\perp W_2$ still hold under mild assumptions on ΔW_2 (e.g. $E[\Delta W_2 | Y, X] = 0$). This means in a linear model, even if the instrument variable suffers from measurement errors, one can still use the error-laden measurement W_2 as an IV and proceed as usual. However, this is not the case in nonseparable models. Plugging W_2 into the second equation yields $X = h(W_2 - \Delta W_2, \eta)$, where both ΔW_2 and η are unobervable and importantly, $W_2 \not\perp \Delta W_2$. This means if I were to use W_2 as an IV, both of the two equations in the triangular system would contain endogenous variables and nothing can be done without additional IVs outside of this system. This also means that if one simply uses W_2 as the instrument variable and proceeds with the standard techniques in Imbens and Newey (2009), they would get inconsistent results.

Given that nonseparability makes the problem much harder to solve, a natural question is why one wants to deal with nonseparable models instead of an additive separable or linear model. There are multiple reasons why researchers might prefer a nonseparable model. First, nonseparable models allow the observed variable X and the unobserved variable ϵ to interact in a flexible way. For example, a recent paper by Brancaccio et al. (2020a) estimated the matching function $m(s, e)$ between ships (of number s) and exporters (of number e which is unobserved) at a seaport. Not imposing functional form assumptions (including separability) is important. It allows the authors to remain agnostic about the nature of the meeting process. In addition, the flexibility of functional forms can be key when deriving welfare and policy implications (see Brancaccio et al. (2020b)). Second, nonseparability is also important when X and ϵ are correlated, since in many cases the source of endogeneity is the nonseparable nature of the model. For example, X could come from the optimization problem of maximizing the expected value of Y minus the production cost, given an exogenous variable W, and some noisy information about ϵ. Think of X as an individual's education level, or a firm's input level, and Y as the individual's lifetime earnings, or the firm's output. Then X could be the solution of $\max_x \{E[m(x, \epsilon)|\eta, W] - c(x, W)\}$, leading to $X = h(W, \eta)$, where η is some noisy proxy of ϵ, and $c(x, W)$ is the cost function. If the function m were additively separable in ϵ, the optimal choice of x would not even depend on η.

When there is no measurement error, Imbens and Newey (2009) proposed a way to identify the model primitives in a nonseparable triangular system, making use of the fact that $X \perp \epsilon|\eta$. They first estimate the control variable η (or a strictly monotone function of η, denoted as V in their paper) from the second equation, and then estimate the model primitives in the first equation by first conditioning on the estimated η, and then integrate it out. Their method cannot directly apply when the instrument variable W^* in the second equation is mismeasured, because the control variable (or control function) cannot be observed from error-laden measurements of W^*. A recent paper by Aradillas-Lopez (2022) studies inference in models where control functions are unobserved. The setup of his paper is different from this paper in many aspects. He requires the availability of observable or estimable bounds for the unobserved control functions, which is not required by the model in this paper. Instead, this paper requires the availability of error-laden measurements and builds upon the measurement error literature. Also, his focus is on constructing confidence sets for finite-dimensional parameters, while my focus is on point identification and estimation of infinite-dimensional parameters.

In this paper, I propose a method that makes use of the same intuition as in Imbens and Newey (2009), but can deal with mismeasured W^*. Same as Imbens and Newey (2009), I utilize the fact that the correlation between X and ϵ is merely coming from η, but instead of estimating and conditioning on η, I separate out this correlation by writing ϵ as a function of η and a uniformly distributed random variable which is independent with X and W^*. In this way, I am able to write the model primitives like the structural derivative and LAR as functionals of the joint distribution of Y, X, W^*, which can be recovered using the two error-laden measurements and the deconvolution technique developed in the measurement error literature (e.g. Fan (1991a), Fan and Truong (1993), Schennach (2004a), Schennach (2004b)). I can thus identify the model primitives in a constructive way and estimate them using plug-in estimators. I also derive uniform rates of convergence of the estimators.

This paper is most related to Schennach et al. (2012) (SWC, hereafter), where the authors also consider a triangular simultaneous equations model with a mismeasured exogenous instrument. This paper differs from their paper in the modeling assumptions, parameters that can be identified, and also theoretical methods. SWC shows that under separability assumption on the second equation, the instrument conditioned marginal response[1] $E[m_x(X,\epsilon)|W^*=w^*]$ can be written as a ratio of the derivative of the conditional mean of Y given W^* over the derivative of the conditional mean of X given W^*, both of which can be recovered from the data given two error-laden measurements W_1, W_2. Integrating out W^*, they can also recover the average response $E[m_x(X,\epsilon)]$. However, under the nonseparability of both of the equations, their method cannot recover either of the two parameters. Different from SWC, this paper shows that it's possible to identify not only the instrument-conditioned marginal response and the average marginal response but also the structural derivative and LAR, even when both of the equations are nonseparable. This conclusion, however, comes at the expense of more assumptions on the function m and unobservables compared with SWC. In particular, this paper assumes strict monotonicity of the function m on its second argument, and scalar unobservables, which are not required in SWC. Regarding estimation, both SWC and this paper employ plug-in estimators, but the asymptotic analysis in this paper is a bit more complex than in SWC, because of the observed variable components Y, X of the joint density f_{Y,X,W^*}. They have non-trivial implications on the asymptotic treatment (including the convergence rates) so that SWC's asymptotic analysis cannot directly apply here.

This paper is also closely related to Song et al. (2015). In their paper, the coauthors study a nonseparable model with mismeasured endogenous variable, assuming a correctly measured control variable is available. The setup of this paper is different from theirs. This paper also studies nonseparable models with endogeneity, but instead of mismeasured endogenous variable, this paper studies the case when the endogenous variable is correctly measured, and instead of assuming the existence of a correctly measured control variable, this paper assumes that a potentially mismeasured instrument variable is available. The two papers are complementary depending on the availability of data. Regarding the parameters of interest, in addition to the various parameters studied in Song et al. (2015),

[1] m_x denotes the partial derivative of the m function with respect to its first argument.

including the average marginal response conditional on the control variable, the average marginal response, the LAR, and the weighted average version of these variables, this paper additionally show identification of the structural derivative, which is more disaggregate and could be helpful when researchers are interested heterogeneous effects.

The rest of the paper consists of six parts. Section 2 introduces the model and assumptions. Section 3 talks about the identification of the model primitives. Section 4 proposes plug-in estimators of the model primitives identified in Section 3, and talks about their asymptotic properties. Section 5 conducts limited Monte Carlo studies to show the finite sample performance of the estimator. Section 6 concludes.

1.2 The Model

I consider the following triangular model in this paper:

$$Y = m(X, Z, \epsilon)$$
$$X = h(W^*, Z, \eta)$$

where X is an observed endogenous variable that is correlated with ϵ. In the returns to education example, X stands for years of education, which is correlated with ϵ since it's chosen by the agent as an equilibrium outcome. Z is a vector of observed exogenous variables which are independent with ϵ and η. W^* is an instrument for X and satisfies $W^* \perp (\epsilon, \eta)$. Researchers cannot measure W^* exactly but have two error-laden measurements of it:

$$W_1 = W^* + \Delta W_1$$
$$W_2 = W^* + \Delta W_2.$$

The measurement errors satisfy $E[\Delta W_1 \mid W^*, \Delta W_2] = 0$ and $\Delta W_2 \perp W^*, Y, X, Z$. As in the literature studying the identification of nonseparable models (Chesher (2003), Matzkin

(2003),Altonji and Matzkin (2005),Matzkin (2015)), I impose monotonicity assumptions on the structural functions. I assume that m is strictly increasing in ϵ and that h is strictly increasing in η. Since the vector of covariates Z is correctly measured and exogenous, all the analysis can be done conditional on Z. For brevity, from now on I omit the vector Z in my notations and work on the simplified model below, while other assumptions on the structural functions and distributions of variables remain unchanged.

$$Y = m(X, \epsilon) \tag{1.3}$$

$$X = h(W^*, \eta) \tag{1.4}$$

The assumptions mentioned above are stated formally below:

Assumption 1.1. *Function m and h are continuously differentiable with respect to both of their arguments and are strictly increasing in their respective second argument.*

Assumption 1.2. $W^* \perp (\epsilon, \eta)$.

Assumption 1.3. $W_1 = W^* + \Delta W_1$ *and* $W_2 = W^* + \Delta W_2$.

Assumption 1.4. $E[\Delta W_1 \mid W^*, \Delta W_2] = 0$, $\Delta W_2 \perp W^*, Y, X$.

Following the measurement error literature, I also impose:

Assumption 1.5. *For any finite* $t \in \mathbb{R}$, $|E[\exp(\mathbf{i}tW_2)]| > 0$.

Note that for the measurement error ΔW_1, only the mean independence assumption is imposed, which is weaker than the assumption on ΔW_2. This weaker assumption is sufficient to identify the distribution of W^* (Schennach (2004b)). On the other hand, although ΔW_2 satisfies strong independence assumptions, making W_2 a measurement with classical measurement error, one still cannot use W_2 as the instrument because of the nonseparability of the model. More specifically, plugging W_2 into the second equation in the triangular system yields

$$X = h(W_2 - \Delta W_2, \eta).$$

This is a nonseparable model with two unobservables ΔW_2 and η, and W_2 is not independent with ΔW_2. This means both of the two equations in the system contain endogenous variables and thus cannot be identified without further assumptions.

In addition, I also impose the following assumptions on the distribution of Y, X, W^*:

Assumption 1.6. *The distribution of* (Y, X, W^*) *has compact support, denoted as* $\mathbb{S}_{(Y,X,W^*)}$ *and has a joint density* f_{Y,X,W^*} *which is twice continuously differentiable on* R^3.

Assumption 1.7. *For any* x, w^* *belonging to the support of* X, W^*, $\left|\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*}\right| > 0$.

Assumption 1.7 is a sufficient condition to ensure $\left.\frac{\partial h(w^*,\eta)}{\partial w^*}\right|_{\eta=r(x,w^*)} \neq 0$ for all x, w^* on the support of X, W^*. This assumption is like the rank condition imposed in the linear instrument variable models. To illustrate this idea, suppose h is a linear function such that $X = \pi W^* + \eta$, then $F_{X|W^*=w^*}(x) = F_\eta(x - \pi w^*)$, and $\pi \neq 0$ is a neccessary condition for Assumption 1.7.

1.3 Identification

In this section, I show the identification of the model parameters in two steps. First, assuming that the joint distribution of Y, X, W^* is known, I show identification of the derivative of the structural function m with respect to x, when the value of ϵ is fixed, and the identification of other model parameters including the LAR and AR. Then I show how to identify the joint distribution of Y, X, W^* from the two measurements W_1 and W_2.

1.3.1 Identification of parameters assuming the joint distribution of Y, X, W^* is known

The independence between W^* and ϵ, η implies that conditional on η, W^* and ϵ are independent. The next lemma uses this fact to show that one can write ϵ as a function of η and another random variable which is independent with X, W^*. This lemma is similar to Proposition 5.1 in Matzkin (2016). Before stating the lemma, I first impose the following assumption on the conditional distribution of ϵ given η:

Assumption 1.8. *$F_{\epsilon|\eta=\bar{\eta}}(\bar{\epsilon})$ is strictly increasing in $\bar{\epsilon}$, given any values of $\bar{\eta}$.*

Lemma 1.1. *Under the model setup, suppose Assumption 1.8 is satisfied. There exists a function $s : \mathbb{R}^2 \to \mathbb{R}$ strictly increasing in its second argument and an unobservable random term δ such that*

$$\epsilon = s(\eta, \delta), \tag{1.5}$$

and δ is independent of (X, W^*) and is $U(0, 1)$.

Proof. See Appendix B. □

Next, I plug (1.5) into the first structural equation. The assumption that h is strictly increasing in η implies that one can write the inverse of $h(W^*, \eta)$ w.r.t. η as $r(X, W^*)$. Then I have

$$Y = m(X, \epsilon) = m(X, s(\eta, \delta)) = m(X, s(r(X, W^*), \delta)) \equiv v(X, W^*, \delta). \tag{1.6}$$

Equation (1.6) builds a bridge between the structural function m and the reduced form function v. Note that v is a function of observable variables X, W^* and an unobservable variable δ which is independent of the observables. The derivatives of v can be identified by applying identification techniques in standard nonseparable models (Matzkin (2003)). To identify the main parameter of interest, the derivative of the structural function m, one can utilize the last equality in 1.6: $m(X, s(r(X, W^*), \delta)) \equiv v(X, W^*, \delta)$. Taking derivative w.r.t X and W^* yields:

$$\frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} + \frac{\partial m(x,\epsilon)}{\partial \epsilon}\bigg|_{\epsilon=s(r(x,w^*),\delta)} \cdot \frac{\partial s(\eta,\delta)}{\partial \eta}\bigg|_{\eta=r(x,w^*)} \cdot \frac{\partial r(x,w^*)}{\partial x} = \frac{\partial v(x,w^*,\delta)}{\partial x} \tag{1.7}$$

$$\frac{\partial m(x,\epsilon)}{\partial \epsilon}\bigg|_{\epsilon=s(r(x,w^*),\delta)} \cdot \frac{\partial s(\eta,\delta)}{\partial \eta}\bigg|_{\eta=r(x,w^*)} \cdot \frac{\partial r(x,w^*)}{\partial w^*} = \frac{\partial v(x,w^*,\delta)}{\partial w^*} \tag{1.8}$$

Plug (1.8) into (1.7) to cancel $\frac{\partial m(x,\epsilon)}{\partial \epsilon}\big|_{\epsilon=s(r(x,w^*),\delta)} \cdot \frac{\partial s(\eta,\delta)}{\partial \eta}\big|_{\eta=r(x,w^*)}$ yields

$$\frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} = \frac{\partial v(x,w^*,\delta)}{\partial x} - \frac{\partial v(x,w^*,\delta)}{\partial w^*}\frac{\frac{\partial r(x,w^*)}{\partial x}}{\frac{\partial r(x,w^*)}{\partial w^*}}.$$

Taking derivative w.r.t. W^* on both sides of $X \equiv h(W^*, r(X, W^*))$ and cancelling out the unobserved $\frac{\partial h(w^*,\eta)}{\partial \eta}\big|_{\eta=r(x,w^*)}$, one can get $\frac{\partial h(w^*,\eta)}{\partial w^*}\big|_{\eta=r(x,w^*)} = -\frac{\frac{\partial r(x,w^*)}{\partial w^*}}{\frac{\partial r(x,w^*)}{\partial x}}$. Then one can write

$$\frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} = \frac{\partial v(x,w^*,\delta)}{\partial x} + \frac{\partial v(x,w^*,\delta)}{\partial w^*}\frac{1}{\frac{\partial h(w^*,\eta)}{\partial w^*}\big|_{\eta=r(x,w^*)}}. \tag{1.9}$$

The right-hand side of the equation (1.9) can be identified from the data. To show this, note that

$$\begin{aligned} &F_{Y|X=x,W^*=w^*}(v(x,w^*,\delta)) = F_\delta(\delta) = \delta \qquad (1.10)\\ \Rightarrow &\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial x} + \frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y}\frac{\partial v(x,w^*,\delta)}{\partial x} = 0\\ \Rightarrow &\frac{\partial v(x,w^*,\delta)}{\partial x} = -\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial x}\Bigg/\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y}, \end{aligned}$$

Similarly

$$\frac{\partial v(x,w^*,\delta)}{\partial w^*} = -\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial w^*}\Bigg/\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y}$$

$$\frac{\partial h(w^*,\eta)}{\partial w^*} = -\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*} \Bigg/ \frac{\partial F_{X|W^*=w^*}(x)}{\partial x},$$

where δ is the unique value such that $y = v(x, w^*, \delta)$ and η is the unique value such that $x = h(w^*, \eta)$. Plug into (1.9), one can get

$$\rho(\bar{y},\bar{x}) \equiv \frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{x=\bar{x},\epsilon=m^{-1}(\bar{x},\bar{y})} = -\frac{\frac{\partial F_{Y|X=\bar{x},W^*=w^*}(\bar{y})}{\partial x}}{\frac{\partial F_{Y|X=\bar{x},W^*=w^*}(\bar{y})}{\partial y}} + \frac{\frac{\partial F_{Y|X=\bar{x},W^*=w^*}(\bar{y})}{\partial w^*}}{\frac{\partial F_{Y|X=\bar{x},W^*=w^*}(\bar{y})}{\partial y}} \times \frac{\frac{\partial F_{X|W^*=w^*}(\bar{x})}{\partial x}}{\frac{\partial F_{X|W^*=w^*}(\bar{x})}{\partial w^*}} \tag{1.11}$$

for all $\bar{y}, \bar{x}$ belongs to support, and values of $\bar{\epsilon}$ such that $\bar{y} = m(\bar{x}, \bar{\epsilon})$. Note that as long as $\bar{y}$ and $\bar{x}$ are fixed, $\bar{\epsilon}$ is fixed, no matter which value is picked for w^*. This means there's actually overidentification for the structural derivative. To identify the **LAR** and **AR**, one needs independent variations of ϵ given X. Rewrite (1.10) with slightly different notations yields:

$$F_{Y|X=x,W^*=w^*}(v(x,w^*,\delta)) = F_\delta(\delta) = \delta$$
$$\Rightarrow v(x,w^*,\delta) = F^{-1}_{Y|X=x,W^*=w^*}(\delta)$$

so that one can write

$$\frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} = \frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial x} - \frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial w^*} \frac{\frac{\partial F_{X|W^*=w^*}(x)}{\partial x}}{\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*}}. \tag{1.12}$$

To show the identification of the **LAR** and **AR**, one needs to show that the support of $s(r(X, W^*), \delta)$ given $X = x$ is the same as the support of ϵ given $X = x$. This is stated formally in the lemma below.

Lemma 1.2. *Define random variable* $\mathbf{s} \equiv s(r(X, W^*), \delta)$. *Denote the support of* $\mathbf{s}$ *and* ϵ *conditional on* $X = x$ *as* $\mathbb{S}_{\mathbf{s}|X=x}$ *and* $\mathbb{S}_{\epsilon|X=x}$*, respectively. If Assumption 1.8 holds, then* $\mathbb{S}_{\mathbf{s}|X=x} = \mathbb{S}_{\epsilon|X=x}$*, for all* x *belonging to its support.*

Proof. See Appendix B. □

Lemma 1.1 and 1.2 ensures that $F_{\mathbf{s}|X}$ is the same as $F_{\epsilon|X}$, so that integrating $\frac{\partial m(x,\epsilon)}{\partial x}$ over ϵ given $X=x$ will be equivalent to integrating $\frac{\partial m(x,\mathbf{s})}{\partial x}$ over $\mathbf{s}$ given $X=x$. Then I have the following identification results:

$$\textbf{LAR}: \ E\left[\frac{\partial m(X,\epsilon)}{\partial x} \mid X=x\right] = \int\int_0^1 \frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} f_{W^*|X=x}(w^*)d\delta dw^* \quad (1.13)$$

$$\textbf{AR}: \ E\left[\frac{\partial m(X,\epsilon)}{\partial x}\right] = \int\int\int_0^1 \frac{\partial m(x,\epsilon)}{\partial x}\bigg|_{\epsilon=s(r(x,w^*),\delta)} f_{X,W^*}(x,w^*)d\delta dw^* dx. \quad (1.14)$$

1.3.2 Identification of the joint density of Y, X, W^*

First note that by Theorem 1 in Schennach (2004a)

$$\phi_{W^*}(t) = \exp\left(\int_0^t \frac{E\left[iW_1 e^{i\xi W_2}\right]}{E\left[e^{i\xi W_2}\right]} d\xi\right).$$

Then I can identify the density $f_{W^*}(w^*)$ by taking the Inverse Fourier Transform:

$$f_{W^*}(w^*) = \frac{1}{2\pi}\int e^{-itw^*}\phi_{W^*}(t)dt.$$

For the joint density $f_{Y,X,W^*}(y,x,w^*)$, by Assumption 1.4, I have the following convolution:

$$f_{Y,X,W_2}(x,w_2) = \int f_{Y,X,W^*}(x,\nu) f_{\Delta W_2}(w_2-\nu)d\nu$$

Applying Fourier transformation on both sides (holding x constant) yields

$$\phi_{f_{Y,X,W_2}(x,\cdot)}(t) = \phi_{f_{Y,X,W^*}(y,x,\cdot)}(t)\phi_{\Delta W_2}(t),$$

which, by Assumption 1.5 implies

$$\phi_{f_{Y,X,W^*}(y,x,\cdot)}(t) = \frac{\phi_{f_{Y,X,W_2}(y,x,\cdot)}(t)}{\phi_{\Delta W_2}(t)} = \frac{\phi_{f_{Y,X,W_2}(y,x,\cdot)}(t)\phi_{W^*}(t)}{\phi_{W_2}(t)}. \tag{1.15}$$

Then applying the inverse Fourier transform, one can get

$$f_{Y,X,W^*}(x,w^*) = \frac{1}{2\pi}\int e^{-itw^*}\frac{\phi_{f_{X,W_2}(y,x,\cdot)}(t)\phi_{W^*}(t)}{\phi_{W_2}(t)}dt,, \tag{1.16}$$

and similarly,

$$f_{Y,X,W^*}(y,x,w^*) = \frac{1}{2\pi}\int e^{-itw^*}\frac{\phi_{f_{Y,X,W_2}(y,x,\cdot)}(t)\phi_{W^*}(t)}{\phi_{W_2}(t)}dt. \tag{1.17}$$

Then $F_{X|W^*=w^*}(x)$ follows from

$$F_{X|W^*=w^*}(x) = \frac{\int_{-\infty}^{x} f_{X,W^*}(y,x,w^*)dx}{\int_{-\infty}^{\infty} f_{X,W^*}(y,x,w^*)dx} \tag{1.18}$$

and $F_{Y|X=x,W^*=w^*}(y)$ follows from

$$F_{Y|X=x,W^*=w^*}(y) = \frac{\int_{-\infty}^{y} f_{Y,X,W^*}(y,x,w^*)dy}{\int_{-\infty}^{\infty} f_{Y,X,W^*}(y,x,w^*)dy}.$$

From the equations above, one can write the derivatives of the conditional CDFs as functionals of the joint densities:

$$\begin{aligned}
\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial w^*} &= \frac{\left(\int_{-\infty}^{y}\frac{\partial f_{Y,X,W^*}(s,x,w^*)}{\partial w^*}ds\right)f_{X,W^*}(x,w^*) - \left(\int_{-\infty}^{y} f_{Y,X,W^*}(s,x,w^*)ds\right)\frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*}}{f_{X,W^*}^2(x,w^*)}\\
\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial x} &= \frac{\left(\int_{-\infty}^{y}\frac{\partial f_{Y,X,W^*}(s,x,w^*)}{\partial x}ds\right)f_{X,W^*}(x,w^*) - \left(\int_{-\infty}^{y} f_{Y,X,W^*}(s,x,w^*)ds\right)\frac{\partial f_{X,W^*}(x,w^*)}{\partial x}}{f_{X,W^*}^2(x,w^*)}\\
\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y} &= \frac{f_{Y,X,W^*}(y,x,w^*)}{f_{X,W^*}(x,w^*)}\\
\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*} &= \frac{\left(\int_{-\infty}^{x}\frac{\partial f_{X,W^*}(s,w^*)}{\partial w^*}ds\right)f_{W^*}(w^*) - \left(\int_{-\infty}^{x} f_{X,W^*}(s,w^*)ds\right)\frac{\partial f_{W^*}(w^*)}{\partial w^*}}{f_{W^*}^2(w^*)}\\
\frac{\partial F_{X|W^*=w^*}(x)}{\partial x} &= \frac{f_{X,W^*}(x,w^*)}{f_{W^*}(w^*)}.
\end{aligned}$$

This means that the right-hand side of the equation (1.11) can be written as a functional of the joint density f_{Y,X,W^*}, which can be identified from the data. To show identification of the left-hand side of the equation (1.12), one needs to build the relationship between the derivatives of conditional quantiles and the joint densities, which is shown in the following lemma:

Lemma 1.3. *Let $\delta \in [0,1]$ be a constant, then*

$$\frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial w^*} = \frac{\delta\frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*} - \int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} \frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial w^*}dy}{f_{Y,X,W^*}\left(F^{-1}_{Y|X=x,W^*=w^*}(\delta),x,w^*\right)}$$
$$\frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial x} = \frac{\delta\frac{\partial f_{X,W^*}(x,w^*)}{\partial x} - \int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} \frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial x}dy}{f_{Y,X,W^*}\left(F^{-1}_{Y|X=x,W^*=w^*}(\delta),x,w^*\right)}.$$

Proof. see Appendix B. □

The identification of the left-hand side of (1.13) and (1.14) is a straightforward implication of Lemma 1.3 and equation (1.12).

1.4 Estimation

1.4.1 The Estimator

From now on, I use the following function to denote the joint density of Y, X, W^*, or its derivative with respect to Y, X, W^*:

$$g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*) = \frac{\partial^\lambda f_{Y,X,W^*}(y,x,w^*)}{\partial w^{*\lambda_1}\partial y^{\lambda_{2,1}}\partial x^{\lambda_{2,2}}},$$

where $\lambda_1, \lambda_{2,1}, \lambda_{2,2} \in \{0,1\}$ and $\lambda \equiv \max\{\lambda_1, \lambda_{2,1}, \lambda_{2,2}\} \le 1$. The $0-th$ order derivative of a function is defined as the function itself. For convenience of notation, I also define $\lambda_2 \equiv \max\{\lambda_{2,1}, \lambda_{2,2}\}$. In this section, I first propose an estimator for $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$, and then propose plug-in estimators for the structural derivative and a weighted average version of the **LAR** since they can be written as known functionals of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$.

To make the expressions more transparent and clear, I show the explicit forms of the functionals by taking the structural derivative as an example. I write each component on the right-hand-side of equation 1.11 as

$$\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial w^*} = \frac{\left(\int_{-\infty}^{y} g_{1,0,0}(s,x,w^*)ds\right)}{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)} - \frac{\left(\int_{-\infty}^{y} g_{0,0,0}(s,x,w^*)ds\right)\left(\int_{-\infty}^{\infty} g_{1,0,0}(y,x,w^*)dy\right)}{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)^2}$$

and

$$\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial x} = \frac{\left(\int_{-\infty}^{y} g_{0,0,1}(s,x,w^*)ds\right)}{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)} - \frac{\left(\int_{-\infty}^{y} g_{0,0,0}(s,x,w^*)ds\right)\left(\int_{-\infty}^{\infty} g_{0,0,1}(y,x,w^*)dy\right)}{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)^2}$$

and

$$\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y} = \frac{g_{0,0,0}(y,x,w^*)}{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)}$$

and

$$\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*} = \frac{\left(\int_{-\infty}^{x}\int_{-\infty}^{\infty} g_{1,0,0}(y,s,w^*)dyds\right)}{\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dydx\right)} - \frac{\left(\int_{-\infty}^{x}\int_{-\infty}^{\infty} g_{0,0,0}(y,s,w^*)dyds\right)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{1,0,0}(y,x,w^*)dydx\right)}{\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dydx\right)^2}$$

and

$$\frac{\partial F_{X|W^*=w^*}(x)}{\partial x} = \frac{\left(\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dy\right)}{\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{0,0,0}(y,x,w^*)dydx\right)}.$$

For the **LAR**, I will introduce a weighted average version of it, and propose an estimator for this weighted average version of the **LAR**. The aim of introducing the weight is to ensure that the integration is taken on a set where the joint density f_{Y,X,W^*} is bounded away from zero. This has the benefit of guaranteeing that the relevant functional is Fréchet differentiable when f_{Y,X,W^*} appears in the denominator of it. The weighted average version of the **LAR** is defined as follows:

$$\begin{aligned}\mathbf{WLAR}(x) &\equiv E\left[\omega(X, W^*, \epsilon)\frac{\partial m(X, \epsilon)}{\partial x} \mid X = x\right] \\ &= \int\int_0^1 \omega(x, w^*, \epsilon)\left.\frac{\partial m(x, \epsilon)}{\partial x}\right|_{\epsilon = s(r(x,w^*),\delta)} f_{W^*|X=x}(w^*)d\delta dw^*, \end{aligned} \tag{1.19}$$

where $\omega(x, w^*, \epsilon)$ is a known or estimable weighting function taking value 0 outside a compact set $\mathbb{M}$. As a weighting function, $\omega(x, w^*, \epsilon)$ also satisfies that given any values of x such that $\omega(x, \cdot, \cdot)$ could obtain non-zero values, $\int_{w^*}\int_{\epsilon}\omega(x, w^*, \epsilon)f_{\epsilon,W^*|X=x}(\epsilon, w^*)d\epsilon dw^* = 1$. The specific form of ω is up to the researcher's choice, however, to ensure easy calculation of $\omega(x, w^*, s(r(x, w^*), \delta))$ on the right-hand side of (1.19), I require that function $\omega(x, w^*, \epsilon)$ satisfy the following restriction: $\omega(x, w^*, \epsilon) = \tilde{\omega}(x, w^*)$ if $\epsilon \in \left[q_{x,w^*}(\tau_l), q_{x,w^*}(\tau_u)\right]$ and $\omega(x, w^*, \epsilon) = 0$ otherwise, where $\tilde{\omega}(x, w^*)$ is a known function with compact support, and $q_{x,w^*}(\tau)$ denotes the conditional-τ quantile of ϵ given $X = x, W^* = w^*$. The specific form of function $\tilde{w}$ and the specific values of τ_l and τ_u are up to the researcher's choice. Under this restriction, $\omega(x, w^*, s(r(x, w^*), \delta))$ is equal to $\tilde{w}(x, w^*)$ if δ is between τ_l and τ_u, and is equal to 0 otherwise. An example of function $\omega(x, w^*, \epsilon)$ is when $\tilde{\omega}(x, w^*)$ is a constant equal to $\left[(\tau_u - \tau_l)\int_{\underline{w}^*}^{\bar{w}^*} f_{W^*|X=x}(w^*)dw^*\right]^{-1}$. The **WLAR** in this case can be interpreted as the Local Average Response of a subgroup of individuals whose unobserved ϵ is between its τ_l and τ_u conditional quantile given X, W^* and whose W^* is between $\underline{w}^*$ and $\bar{w}^*$. Note that even though X is correlated with ϵ, neither the distribution of $\epsilon|X$ nor the subgroup of individuals changes as I take the derivative with respect to x. This means that my objective of interest is the same group of people when studying the effect of the counterfactual changes of the endogenous X. The **WLAR** can also be written as a functional of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*)$. The explicit form can be derived from Lemma 1.3 and equation (1.12).

So far I have written the parameters of interest as functionals of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$ in explicit forms. Next I focus on the estimation of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$. To deal with the well-known ill-posed inverse problem when inverting a convolution operator, I base my estimator on a smoothed version of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$:

$$g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1) \equiv \int \frac{1}{h_1} K\left(\frac{\tilde{w}^* - w^*}{h_1}\right) g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,\tilde{w}^*) d\tilde{w}^*,$$

and use kernel functions satisfying the following assumption:

Assumption 1.9. *The kernel functions $K : \mathbb{R} \to \mathbb{R}$, $G_Y : \mathbb{R} \to \mathbb{R}$ and $G_X : \mathbb{R} \to \mathbb{R}$ are measurable, symmetric. $\int K(x)dx = 1$, $\int G_Y(y)dy = 1$, $\int G_X(x)dx = 1$. G_Y, G_X are differentiable. G_Y, G_X, and their derivatives are bounded, and denote the maximum of these bounds as $\bar{G}$. Their Fourier transforms $\xi \to \phi_K(\xi)$, $\xi \to \phi_{G_Y}(\xi)$ and $\xi \to \phi_{G_X}(\xi)$ obey: (i) ϕ_F is compactly supported (without loss of generality, the support is $[-1,1]$) for $F \in \{K, G_X, G_Y\}$; and (ii) there exists $\bar{\xi}_F$ such that $\phi_F(\xi) = 1$ for $|\xi| \leq \bar{\xi}_F$, where $F \in \{K, G_X, G_Y\}$.*

By assuming (ii), I use flat-top kernels proposed by Politis and Romano (1999). The benefit of using flat-top kernels is that the rate of decrease of the bias term will not be affected by the order of the kernel, and will only be affected by the smoothness of the function to be estimated. The restriction of compact support of the Fourier transform is without loss of generality. As discussed by Schennach (2004b), given any kernel K, one can always create a modified kernel $\tilde{K}$ that satisfies the assumption by using a "windowing" function. For example,

$$\phi_{\tilde{K}}(t) = W(t)\phi_K(t).$$

where

$$W(t) = \begin{cases} 1 & \text{if } |t| \leq \bar{t} \\ \left(1 + \exp\left((1-\bar{t})\left((1-|t|)^{-1} - (|t|-\bar{t})^{-1}\right)\right)\right)^{-1} & \text{if } 1 \geq |t| > \bar{t} \\ 0 & \text{if } |t| > 1 \end{cases}$$

for some $\bar{t} \in (0,1)$.

Assumption 1.10. *$E[|\Delta W_1|] < \infty$.*

The following lemma shows that the smoothed version $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_1)$ can be written as an expression involving the Fourier transform of the kernel function and the characteristic functions of different variables. This expression makes it more straightforward to introduce the estimator that will appear soon below.

Lemma 1.4. *For $(y, x, w^*) \in \mathbb{S}_{(Y,X,W^*)}$ and $h > 0$, let*

$$g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_1) \equiv \int \frac{1}{h_1} K\left(\frac{\tilde{w}^* - w^*}{h_1}\right) g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, \tilde{w}^*) d\tilde{w}^*.$$

where K satisfies Assumption 1.9. Then under Assumptions 1.5, 1.6, and 1.10,

$$g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_1) = \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_1 t) \frac{\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\phi_{W^*}(t)}{\phi_{W_2}(t)} dt$$

where $f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y, x, w_2) \equiv \frac{\partial^{\lambda_2} f_{Y,X,W_2}(y,x,w_2)}{\partial y^{\lambda_{2,1}} \partial x^{\lambda_{2,2}}}$.

Proof. See Appendix B. □

I now define my estimator for $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_1)$ by replacing $\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)$, $\phi_{W^*}(t)$, and $\phi_{W_2}(t)$ in Lemma 1.4 with their sample analogues. Formally,

Definition 1. *Let* $h_n \equiv (h_{1n}, h_{2n,1}, h_{2n,2}) \to 0$ *as* $n \to 0$. *The estimator for* $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*)$ *is defined as*

$$\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_n) \equiv \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_{1n}t) \frac{\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\hat{\phi}_{W^*}(t)}{\hat{\phi}_{W_2}(t)} dt,$$

where

$$\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) \equiv \hat{E}\left[e^{\mathbf{i}tW_2} \frac{1}{h_{2n,1}^{1+\lambda_{2,1}} h_{2n,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right],$$

$$\hat{\phi}_{W_2}(t) \equiv \hat{E}\left[e^{\mathbf{i}tW_2}\right],$$

$$\hat{\phi}_{W^*}(t) \equiv \exp\left(\int_0^t \frac{\hat{E}\left[iW_1 e^{i\xi W_2}\right]}{\hat{E}\left[e^{i\xi W_2}\right]} d\xi\right),$$

where $\hat{E}$ *denotes the sample average,* G_Y, G_X *denotes the kernel functions for* Y, X*, respectively, and* $G_X^{(\lambda_{2,2})}(x) \equiv \frac{\partial^{\lambda_{2,2}} G_X(x)}{\partial x^{\lambda_{2,2}}}$. G_Y, G_X *are not necessarily the same as* K.

Having defined $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*, h_n)$, I can now define the estimator for $\rho(\bar{y}, \bar{x})$ by replacing $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y, x, w^*)$ with its estimator on the right-hand-side of Equation (1.11). Note that by Fubini's Theorem

$$\int_{-\infty}^{y} \hat{g}_{\lambda_1,0,0,\lambda_{2,2}}(s, x, w^*, h_n) ds = \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_{1n}t) \frac{\int_{-\infty}^{y} \hat{\phi}_{f^{(0,\lambda_{2,2})}_{Y,X,W_2}(s,x,\cdot)}(t) ds \hat{\phi}_{W^*}(t)}{\hat{\phi}_{W_2}(t)} dt$$

$$\int_{-\infty}^{\infty} \hat{g}_{\lambda_1,0,0,\lambda_{2,2}}(y, x, w^*, h_n) dy = \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_{1n}t) \frac{\hat{\phi}_{f^{(0,\lambda_{2,2})}_{X,W_2}(x,\cdot)}(t)\hat{\phi}_{W^*}(t)}{\hat{\phi}_{W_2}(t)} dt$$

$$\int_{-\infty}^{x} \int_{-\infty}^{\infty} \hat{g}_{\lambda_1,0,0}(y, s, w^*, h_n) dy ds = \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_{1n}t) \frac{\int_{-\infty}^{x} \hat{\phi}_{f_{X,W_2}(s,\cdot)}(t) ds \hat{\phi}_{W^*}(t)}{\hat{\phi}_{W_2}(t)} dt$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{g}_{\lambda_1,0,0}(y, x, w^*, h_n) dy dx = \frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*} \phi_K(h_{1n}t) \hat{\phi}_{W^*}(t) dt,$$

where I've let $\tilde{G}_Y(y) \equiv \int_{-\infty}^{y} G_Y(u)\, du$ and $\tilde{G}_X(x) \equiv \int_{-\infty}^{x} G_X(u)\, du$ denote the kernel CDFs, and I've defined:

$$\int_{-\infty}^{y}\hat{\phi}_{f^{(0,\lambda_{2,2})}_{Y,X,W_2}(s,x,\cdot)}(t)ds = \hat{E}\left[e^{\mathbf{i}tW_2}\frac{1}{h_{2n,2}^{1+\lambda_{2,2}}}\tilde{G}_Y\left(\frac{y-Y}{h_{2n,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]$$

$$\hat{\phi}_{f^{(\lambda_{2,2})}_{X,W_2}(x,\cdot)}(t)ds = \hat{E}\left[e^{\mathbf{i}tW_2}\frac{1}{h_{2n,2}^{1+\lambda_{2,2}}}G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]$$

$$\int_{-\infty}^{x}\hat{\phi}_{f_{X,W_2}(s,\cdot)}(t)ds = \hat{E}\left[e^{\mathbf{i}tW_2}\tilde{G}_X\left(\frac{x-X}{h_{2n,2}}\right)\right].$$

1.4.2 The Asymptotic Properties of the Estimator

In this subsection, I analyze the asymptotic properties for $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)$. I begin by decomposing the difference between the estimator and the true value of the parameter as the sum of a bias term, a variance term, and a remainder term.

Lemma 1.5. *Suppose $\left\{Y_i, X_i, W_i^*, \Delta W_{1,i}, W_{2,i}\right\}$ is an IID sequence satisfying Assumptions 1.3, 1.4, 1.5, 1.6, and 1.10, and that Assumption 1.9 holds. Then for $(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}$, and $h>0$,*

$$\begin{aligned}&\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\\ =\ & B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\end{aligned}$$

where $B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is the bias term admitting the linear representation:

$$\begin{aligned}&B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\\ =\ & E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right] - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\\ =\ & g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\\ &+\frac{1}{2\pi}\int(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t)\frac{E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) - \phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right]\phi_{W^*}(t)}{\phi_{W_2}(t)}dt;\end{aligned}$$

$L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is the variance term admitting the linear representation:

$$\begin{aligned}
&L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\\
&= \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right]\\
&= \hat{E}\Bigg[\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(W_1 e^{\mathbf{i}\xi W_2} - E[W_1 e^{\mathbf{i}\xi W_2}]\right) d\xi\\
&\quad + \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(e^{\mathbf{i}\xi W_2} - E[e^{\mathbf{i}\xi W_2}]\right) d\xi\\
&\quad + \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\times\\
&\qquad\qquad \Bigg(e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\\
&\qquad\qquad\quad -E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\Bigg) d\xi\Bigg]\\
&= \hat{E}[l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h;Y,X,W_1,W_2)],
\end{aligned}$$

where I've let $\theta(\xi) \equiv E\left[W_1 e^{\mathbf{i}\xi W_2}\right]$ and defined

$$\begin{aligned}
\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) &= \frac{1}{2\pi}\frac{\mathbf{i}}{\phi_{W_2}(\xi)}\int_\xi^{\pm\infty}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_1 t)\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)dt\\
\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) &= -\frac{1}{2\pi}\frac{\mathbf{i}\theta(\xi)}{(\phi_{W_2}(\xi))^2}\int_\xi^{\pm\infty}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_1 t)\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)dt\\
&\quad -\frac{1}{2\pi}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}\xi w^*}\phi_K(h_1\xi)\frac{\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)}{\phi_{W_2}(\xi)}\\
\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) &= \frac{1}{2\pi}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}\xi w^*}\phi_K(h_1\xi)\frac{\phi_{W^*}(\xi)}{\phi_{W_2}(\xi)}.
\end{aligned}$$

where for a given function $\zeta \to f(\zeta)$, I've written $\int_\xi^{\pm\infty} f(\zeta)d\zeta \equiv \lim_{c\to+\infty}\int_\xi^{c\xi} f(\zeta)d\zeta$; and $R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is the remainder term:

$$R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) = \hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h).$$

Proof. See Appendix B. □

To derive the uniform rate of convergence, I impose some assumptions on the smoothness of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$, stated in terms of the rate of decay of the tail of its Fourier transform:

Assumption 1.11. *(i) There exist constants $C_\phi > 0$, $\alpha_\phi \leq 0$, $\beta_\phi \geq 0$ and $\gamma_\phi \in \mathbb{R}$ such that $\beta_\phi \gamma_\phi \geq 0$ and for $j = 1,2$*

$$\max_{\lambda_2 \in \{0,1\}} \sup_{(y,x)\in \mathbb{S}_{(Y,X)}} \left| \phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t) \right| \leq C_\phi (1+|t|)^{\gamma_\phi} \exp\left(\alpha_\phi |t|^{\beta_\phi}\right)$$

$$|\phi_{W^*}(t)| \leq C_\phi (1+|t|)^{\gamma_\phi} \exp\left(\alpha_\phi |t|^{\beta_\phi}\right),$$

where $f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,w_2) \equiv \frac{\partial^{\lambda_2} f_{Y,X,W_2}(y,x,w_2)}{\partial y^{\lambda_{2,1}} \partial x^{\lambda_{2,2}}}$. Moreover, if $\beta_\phi = 0$, then for given $\lambda \in \{0,1\}$, $\gamma_\phi < -\lambda_1 - 1$.

(ii) Denote the Fourier transform of $f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,w^*)$ as $\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}}(t_1,t_2,\xi)$. There exists constants $C_{f_1} > 0$, $C_{f_2} > 0$, $\alpha_{f_1} \leq 0$, $\alpha_{f_2} \leq 0$, $\beta_{f_1} \geq 0$, $\beta_{f_2} \geq 0$ and $\gamma_{f_1} \in \mathbb{R}$, $\gamma_{f_2} \in \mathbb{R}$ such that $\beta_{f_1}\gamma_{f_1} \geq 0$, $\beta_{f_2}\gamma_{f_2} \geq 0$, and

$$\left| \phi_{f^{(\lambda_{2,1},\lambda_{2,2})}}(t_1,t_2,\xi) \right| \leq C_{f_1} C_{f_2} (1+|t_1|)^{\gamma_{f_1}} (1+|t_2|)^{\gamma_{f_2}} \exp\left(\alpha_{f_1}|t_1|^{\beta_{f_1}} + \alpha_{f_2}|t_2|^{\beta_{f_2}}\right) |\phi_{W^*}(t)|.$$

Moreover, if $\beta_{f_1} = 0$, $\gamma_{f_1} < -1$, and if $\beta_{f_2} = 0$, $\gamma_{f_2} < -1$.

In Assumption 1.11(i) I impose the same bound for the tail behavior of $\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}$ and $\phi_{W^*}(t)$. This is without loss of generality since they have the same effect on the convergence rate. Assumption 1.11(ii) is similar to Assumption A7 in Li (2002), and can be viewed as a generalization of smoothness condition from the univariate case to the multivariate case.

I next state the first main result of this paper, the uniform asymptotic rate of the bias term:

Theorem 1.1. *Let the conditions of Lemma 1.5 hold with $\{Y_i, X_i, W_i^*, \Delta W_{1,i}, \Delta W_{2,i}\}$ IID,* *and suppose in addition that Assumption 1.11 holds. Then for $h > 0$,*

$$\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right| = O\left(\left(h_1^{-1}\right)^{1+\gamma_\phi+\lambda_1} \exp\left(\alpha_\phi \underline{\zeta}_K^{-\beta_\phi} \left(h_1^{-1}\right)^{\beta_\phi}\right)\right).$$

Proof. See Appendix B. □

I impose the following assumption to ensure finite variance:

Assumption 1.12. *For some $\delta > 0$, $E[|W_1|^{2+\delta}] < \infty$, $\sup_{w_2\in\mathbb{S}_{W_2}} E\left[|W_1|^{2+\delta} \mid W_2 = w_2\right] < \infty$.*

To derive the rate for $L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$, I impose the following assumption on the tail behavior of the Fourier transforms involved. These are common in the deconvolution literature (e.g. Fan (1991b), Fan and Truong (1993), and Schennach (2004b)).

Assumption 1.13. *(i) There exist constants $C_2 > 0$, $\alpha_2 \leq 0$, $\beta_2 \geq \beta_\phi \geq 0$ and $\gamma_2 \in \mathbb{R}$ such that $\beta_2\gamma_2 \geq 0$ and*

$$|\phi_{W_2}(t)| \geq C_2(1+|t|)^{\gamma_2} \exp\left(\alpha_2 |t|^{\beta_2}\right).$$

Moreover, if $\beta_2 = 0$, $\lambda_1 - \gamma_2 + \gamma_\phi > 0$.

(ii) There exist constants $C_* > 0$ and $\gamma_* \geq 0$ such that

$$\left|\frac{\phi'_{W^*}(t)}{\phi_{W^*}(t)}\right| \leq C_* \left(1+|t|\right)^{\gamma_*}.$$

I explicitly impose $\beta_2 \geq \beta_\phi$ because

$$\begin{aligned} C_\phi \left(1+|t|\right)^{\gamma_\phi} \exp\left(\alpha_{W^*}|t|^{\beta_\phi}\right) \geq |\phi_{W^*}(t)| &= \left|E[e^{\mathbf{i}tW^*}]\right| \\ &\geq \left|E[e^{\mathbf{i}tW^*}]\right| \left|E[e^{\mathbf{i}t\Delta W_2}]\right| \geq \left|E[e^{\mathbf{i}tW_2}]\right| \\ &= |\phi_{W_2}(t)| \geq C_2\left(1+|t|\right)^{\gamma_2} \exp\left(\alpha_2|t|^{\beta_2}\right). \end{aligned}$$

The following theorem states the asymptotic properties of the linear term $L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ and facilitates the analysis of various quantities of interest later.

Theorem 1.2. *Suppose the conditions of Lemma 1.5 hold. (i) Then for each* $(y,x,w^*) \in \mathbb{S}_{(y,x,w^*)}$, $E[L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)] = 0$, *and if Assumption 1.12 also holds, then* $E[L^2_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)] = n^{-1}\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$, *where* $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv E[\left(l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h;Y,X,W_1,W_2)\right)^2]$.

Further, if Assumption 1.11 and 1.13 also holds then

$$\sqrt{\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)}$$
$$= O\left(\max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi\mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right). \tag{1.20}$$

I also have

$$\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right| = \tag{1.21}$$
$$O\left(n^{-\frac{1}{2}}\max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi\mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right). \tag{1.22}$$

(ii) If Assumption 1.12 also holds, and if for each $(y,x,w^*) \in \mathbb{S}_{(y,x,w^*)}$, $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) > 0$ *for all n sufficiently large, then for each* $(y,x,w^*) \in \mathbb{S}_{(y,x,w^*)}$

$$n^{1/2}\left(\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right)^{-1/2} L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) \xrightarrow{d} N(0,1).$$

Proof. See Appendix B. □

Finally, I bound the remainder term. To do that, I first impose some restrictions on the moments of W_2:

Assumption 1.14. $E[|W_2|] < \infty, E[|W_1W_2|] < \infty$.

I then impose the following bounds for the bandwidths:

Assumption 1.15. *If* $\beta_2 = 0$ *in Assumption 1.13, then* $h_{1n}^{-1} = O\left(n^{(4+4\gamma_*-4\gamma_2)^{-1}-\eta}\right)$ *and* $h_{2n,j}^{-1} = O\left(n^{(16+8\lambda_2)^{-1}-\eta}\right)$*, for some* $\eta > 0$*,* $j = 1, 2$*; otherwise* $h_{1n}^{-1} = O\left((\ln n)^{\beta_2^{-1}-\eta}\right)$ *and* $h_{2n,j}^{-1} = O\left(n^{(8+4\lambda_2)^{-1}-\eta}\right)$*, for some* $\eta > 0$*,* $j = 1, 2$*.*

Theorem 1.3. *(i) Suppose the conditions of Lemma 1.5 and Assumption 1.11, 1.12, 1.13, and 1.14 hold. Then*

$$\begin{aligned}&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right| \\ &= o_p\left(h_{2,1}^{-2}(h_{2,2}^{-1})^{2+2\lambda_2} n^{-1+2\epsilon}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right) \\ &\quad\times O\left(\max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi \mathbf{1}\{\beta_\phi = \beta_2\} - \alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right)\end{aligned}$$

for arbitrarily small $\epsilon > 0$*. (ii) If Assumption 1.15 also holds, then*

$$\begin{aligned}&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right| = \\ &o_p\left(n^{-\frac{1}{2}}\max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi \mathbf{1}\{\beta_\phi = \beta_2\} - \alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right).\end{aligned}$$

Proof. See Appendix B. □

Collecting the results from Theorem 1.1,1.2 and 1.3 yields a straightforward corollary of the uniform rate of convergence of $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)$:

Corollary 1. *If the conditions of Theorem 1.3 (ii) hold, then*

$$\begin{aligned}&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\right| \\ &= O_p\left(\epsilon_{n,\lambda_1}\right) + O_p\left(\tilde{\epsilon}_{n,\lambda_1,\lambda_{2,1},\lambda_{2,2}}\right).\end{aligned}$$

where

$$\epsilon_{n,\lambda_1} \equiv \left(h_1^{-1}\right)^{1+\gamma_\phi+\lambda_1}\exp\left(\alpha_\phi \bar{\zeta}_K^{\beta_\phi}\left(h_1^{-1}\right)^{\beta_\phi}\right)$$

$\tilde{\epsilon}_{n,\lambda_1,\lambda_{2,1},\lambda_{2,2}}$

$$\equiv n^{-1/2} \max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi \mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right).$$

The uniform (over the whole support) rate of convergence of the kernel estimators of the density or its derivatives has been considered in Andrews (1995), Hansen (2008) and Schennach et al. (2012)[2]. Hansen (2008) obtained a faster rate than Andrews (1995) and Schennach et al. (2012) (Theorem 3.2, for the no-measurement error case), but requires more assumptions on the kernel functions (their Assumption 1 and 3), including bounded support or an integrable tail, which, unfortunately, are not satisfied by infinite order kernels. Andrews (1995)'s conclusion is stated assuming a kernel with finite order. Infinite order kernels are not essential when there is no measurement error (like in Andrews (1995) and Hansen (2008)), but are especially advantageous when there are measurement errors. If infinite order kernels are used, only the smoothness of the functions, but not the order of the kernels will affect the rate of convergence. Corollary 1 is similar to Corollary 4.7 in Schennach et al. (2012).

With the uniform rate of convergence of $\hat{g}$, I can then derive the uniform rate of convergence of the plug-in estimators of the structural derivative $\rho(y,x)$ and the **WLAR**, denoted as $\widehat{\rho(y,x)}$ and $\widehat{\mathbf{WLAR}}(x)$, respectively.

Theorem 1.4. *Suppose that* $\left\{Y_j, X_j, W_j^*, \Delta W_{1,j}, W_{2,j}\right\}$ *is an IID sequence satisfying the conditions of Corollary 1 with* $\lambda_1, \lambda_{2,1}, \lambda_{2,2} \in \{0,1\}$ *and* $\max\{\lambda_1, \lambda_{2,1}, \lambda_{2,2}\} \le 1$. *Suppose in addition, Assumption 1.7 and 1.8 hold.*
(i) Define $\mathbb{S}_{\tau_n} \equiv \left\{(y,x,w^*) \in \mathbb{S}_{(Y,X,W^*)} : \left|\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*}\right| > \tau_n \text{ and } f_{Y,X,W^*}(y,x,w^*) > \tau_n\right\}$. *Then*

$$\sup_{(y,x,w^*)\in\mathbb{S}_{\tau_n}} \left|\widehat{\rho(y,x)} - \rho(y,x)\right| \le O_p\left(\tilde{\epsilon}_{n,0,0,1}\right) + \frac{O_p\left(\epsilon_{n,1}\right) + O_p\left(\tilde{\epsilon}_{n,1,0,0}\right)}{\tau_n^2}.$$

[2] Schennach et al. (2012) studies the case both with and without measurement errors. Their Theorem 3.2 corresponds to the no-measurement error case, and Corollary 4.7 corresponds to the case with measurement errors.

and there exists $\{\tau_n\}$ *such that* $\tau_n > 0, \tau_n \to 0$ *as* $n \to \infty$*, and*

$$\sup_{(y,x,w^*)\in\mathbb{S}_{\tau_n}} \left|\widehat{\rho(y,x)} - \rho(y,x)\right| = o_p(1).$$

(ii)

$$\sup_{x\in\mathbb{S}_X} \left|\widehat{\mathbf{WLAR}}(x) - \mathbf{WLAR}(x)\right| \leq O_p\left(\epsilon_{n,1}\right) + O_p\left(\tilde{\epsilon}_{n,0,0,1}\right) + O_p\left(\tilde{\epsilon}_{n,1,0,0}\right).$$

Proof. See Appendix B. □

In Appendix C, I state an asymptotic normality result for $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)$ and $\widehat{\rho(y,x)}$. The proof of them requires a lower bound on $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$ relative to $B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$ and $R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$. The assumptions to ensure the lower bound are stated at a high level, and more primitive sufficient conditions need to be derived.

1.5 Monte Carlo Simulations

In this section, I conduct some Monte Carlo simulation exercises to study the finite sample performance of my proposed estimators. I consider two simulation designs. Design 1 is when both of the equations are linear, and Design 2 is when the equations are nonlinear.

Design 1

$$Y = 0.25X + 0.25\epsilon$$
$$X = W^* + \eta$$

Design 2

$$Y = \ln\left(\exp(X+\epsilon)+1\right)$$
$$X = \frac{3^3}{4^4}\frac{W^{*4}}{(-\eta)^3}$$

In both cases, ϵ and η are correlated through a common component θ: $\epsilon = \theta + \epsilon_1$, and $\eta = \theta + \eta_1$, where θ, ϵ_1 and η_1 are mutually independent, and are independent with W^*. In both designs, I have two error-laden measurements $W_1 = W^* + \Delta W_1, W_2 = W^* + \Delta W_2$, where ΔW_1 and ΔW_2 are independent and they are independent with all other variables. The distributions of variables in both designs are listed in Table 1.1 below:

Table 1.1: Distributions of Variables

Variables	Design 1 (linear)	Design 2 (nonlinear)
W^*	$N(0, \sqrt{0.5})$	$N(6, 1)$
θ	$N(0, 0.5)$	$N(-3, \sqrt{0.5})$
ϵ_1	$N(0, \sqrt{0.75})$	$N(3, \sqrt{0.5})$
η_1	$N(0, 0.5)$	$N(-3, \sqrt{0.5})$
ϵ	$N(0, 1)$	$N(0, 1)$
η	$N(0, \sqrt{0.5})$	$N(-6, 1)$
ΔW_1	$N(0, \sqrt{0.5})$	$N(0, \sqrt{0.5})$
ΔW_2	$\chi^2(2) - 2$	$\chi^2(2) - 2$

In the estimation, I use the following flat-top kernel[3] proposed by Politis and Romano (1999):

$$K(x) = \frac{\sin^2(2\pi u) - \sin^2(\pi u)}{\pi^2 u^2}$$

I use the sample size of 500 and replicate the estimation 500 times.

First I show the performance of my proposed method versus two other methods for estimating the structural derivative: (1) 2SLS using the error-laden measurement W_2 as IV; (2) a plug-in estimator replacing W^* with W_2 in my identification equation (1.11). Method (1) will be valid under the linear design (Design 1), since W_2, although error-laden, still satisfies the exclusion restriction and the rank condition for linear IV estimation. However, because of its misspecification of the model, it won't be valid under the nonlinear design (Design 2). Method (2) won't be valid under either the linear or nonlinear design, as the estimator will converge to a population value that is not equal to the right-hand side of

[3] I may use different flat-top kernels for G_X and G_Y. For simplicity I use the same flat-top kernel.

equation (1.11) in general. I estimate $\rho(y,x)$ evaluated at $y=0, x=0, w^*=0.7$ for Design 1, and $y=0.6, x=0.6, w^*=7$ for Design 2, yielding a true value of 0.25 and 0.4512, respectively. There are three bandwidths used in the estimation: h_1, h_{21}, h_{22}. The latter two correspond to estimating the unknown distribution of observed variables X and Y. I thus choose the values of h_{21} and h_{22} by cross-validation (i.e. minimizing the estimated MISE of $\hat{f}_{Y,X}(y,x)$). For the bandwidth h_1, I scan a set of values ranging from 0.5 to 3 for my proposed estimator and a set of values ranging from 0.5 to 6 for the method that uses the error-laden measurement. Table 1.2, 1.3, and 1.4 show the MSE, VAR, and the absolute value of BIAS of my proposed estimator, Method (2), and Method (1), respectively.

From Table 1.3, one can see that the bias from the error-contaminated estimator does not shrink toward zero as bandwidth decreases. Comparing results from Table 1.2 and Table 1.3, one can see that my estimator also gives smaller variances than the error-contaminated method. As a result, my estimator performs better than the error-contaminated method in terms of MSE under both the linear and the nonlinear designs. Comparing Table 1.2 with Table 1.4, one can see that under the linear design, both my estimator and 2SLS work well except that my estimator has a slightly larger variance. Under the nonlinear design, the bias and MSE of 2SLS are much larger than my estimator, which aligns with the theory.

Table 1.5 below shows the Monte Carlo simulation results of my proposed estimator as a function of the sample size. The value of h_{21} and h_{22} are selected by cross-validation for the respective sample sizes and the optimal values of h_1 are selected by minimizing the MSE in the corresponding set of values the same as when $N=500$. The MSE, VAR, and Bias decrease as the sample size increases, which is in accordance with the theory.

Table 1.2: Monte Carlo simulation results of my proposed estimator for $\rho(y, x)$

Design 1 (linear)						Design 2 (nonlinear)					
h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)	h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)
0.5	1.05	2.92	0.01072	0.01072	0.00009	0.5	2.09	0.88	0.31292	0.30489	0.08961
0.75	1.05	2.92	0.00766	0.00766	0.00050	0.75	2.09	0.88	0.17724	0.16339	0.11769
1	1.05	2.92	0.00728	0.00728	0.00138	1	2.09	0.88	0.15425	0.13973	0.12050
1.25	1.05	2.92	0.00738	0.00738	0.00253	1.25	2.09	0.88	0.14824	0.13360	0.12098
1.5	1.05	2.92	0.00755	0.00754	0.00328	1.5	2.09	0.88	0.14611	0.13129	0.12175
1.75	1.05	2.92	0.00772	0.00771	0.00376	1.75	2.09	0.88	0.14540	0.13034	0.12273
2	1.05	2.92	0.00789	0.00787	0.00408	2	2.09	0.88	0.14540	0.13008	0.12380
2.25	1.05	2.92	0.00806	0.00804	0.00431	2.25	2.09	0.88	0.14581	0.13023	0.12482
2.5	1.05	2.92	0.00822	0.00820	0.00447	2.5	2.09	0.88	0.14640	0.13060	0.12569
2.75	1.05	2.92	0.00836	0.00834	0.00461	2.75	2.09	0.88	0.14707	0.13109	0.12640
3	1.05	2.92	0.00851	0.00849	0.00474	3	2.09	0.88	0.14777	0.13163	0.12703
Optimal						Optimal					
h_1	h_{21}	h_{22}	MSE	VAR	BIAS	h_1	h_{21}	h_{22}	MSE	VAR	BIAS
1	1.05	2.92	0.00728	0.00728	0.00138	1.75	2.09	0.88	0.14540	0.13034	0.12273

Table 1.3: Monte Carlo simulation results of plugging in error-laden W_2 to (1.11) to estimate $\rho(y,x)$ (Method (2) above)

Design 1 (linear)						Design 2 (nonlinear)					
h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)	h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)
0.5	1.05	2.92	831.49239	827.49615	1.99906	0.5	2.09	0.88	342390.30868	342380.03811	3.20477
1	1.05	2.92	15.36480	15.18324	0.42611	1	2.09	0.88	3268.04002	3264.78862	1.80316
1.5	1.05	2.92	1609.24991	1605.98439	1.80708	1.5	2.09	0.88	795.12130	795.09105	0.17391
2	1.05	2.92	981.89810	980.63434	1.12417	2	2.09	0.88	1113.49474	1113.43014	0.25417
2.5	1.05	2.92	2739.03247	2729.13653	3.14578	2.5	2.09	0.88	7439.31131	7399.31171	6.32452
3	1.05	2.92	48.28533	48.27034	0.12243	3	2.09	0.88	2589.00861	2583.05235	2.44054
3.5	1.05	2.92	227.69069	227.32253	0.60677	3.5	2.09	0.88	7367.37049	7363.00243	2.08999
4	1.05	2.92	326.60891	325.63719	0.98576	4	2.09	0.88	4810.95117	4805.88266	2.25133
4.5	1.05	2.92	29.37044	29.30256	0.26055	4.5	2.09	0.88	12005.60726	11951.43663	7.36007
5	1.05	2.92	7904.23502	7887.80380	4.05354	5	2.09	0.88	863.16882	862.91399	0.50480
5.5	1.05	2.92	16.49962	16.48066	0.13769	5.5	2.09	0.88	6884.05497	6883.55213	0.70911
6	1.05	2.92	54.17185	54.16807	0.06147	6	2.09	0.88	844.96901	843.85308	1.05637
Optimal						Optimal					
h_1	h_{21}	h_{22}	MSE	VAR	BIAS	h_1	h_{21}	h_{22}	MSE	VAR	BIAS
1	1.05	2.92	15.36480	15.18324	0.42611	1.5	2.09	0.88	795.12130	795.09105	0.17391

Table 1.4: Monte Carlo simulation results of using 2SLS and error-laden W_2 to estimate $\rho(y,x)$ (Method (1) above)

Design 1 (linear)			Design 2 (nonlinear)		
MSE	VAR	abs(BIAS)	MSE	VAR	abs(BIAS)
0.00267	0.00267	0.00190	1.11718	0.00147	1.05627

Table 1.5: Monte Carlo simulation results of my $\hat{\rho}(y,x)$ as a function of the sample size

Design 1 (linear)							Design 2 (nonlinear)						
N	h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)	N	h_1	h_{21}	h_{22}	MSE	VAR	abs(BIAS)
500	1	1.05	2.92	0.00728	0.00728	0.00138	500	1.75	2.09	0.88	0.14540	0.13034	0.12273
1000	1	0.97	2.61	0.00354	0.00353	0.00359	1000	1.5	1.58	0.87	0.08341	0.06822	0.12323
5000	1.25	0.82	2.05	0.00102	0.00102	0.00026	5000	1.75	0.93	0.65	0.03053	0.03048	0.00708

Next, I study the performance of my estimator of the weighted LAR. For Design 1 (linear), I evaluate the WLAR at $x = 0$ and set the weighting function to be $\omega(0, w^*, \epsilon) = \omega_1$ if $\epsilon \in \left[q_{0,w^*}(0.25), q_{0,w^*}(0.35)\right]$ and $w^* \in [0.70, 0.90]$. The constant ω_1 is selected such that $\int_{0.7}^{0.9} \int_{q_{0,w^*}(0.25)}^{q_{0,w^*}(0.35)} \omega_1 f_{\epsilon,W^*|X=0}(\epsilon, w^*) d\epsilon dw^* = 1$. For Design 2 (nonlinear), I evaluate the WLAR at $x = 0.6$ and set the weighting function to be $\omega(0.6, w^*, \epsilon) = \omega_2$ if $\epsilon \in \left[q_{0.6,w^*}(0.25), q_{0.6,w^*}(0.35)\right]$ and $w^* \in [6, 6.23]$. The constant ω_2 is selected to ensure that $\int_{6}^{6.23} \int_{q_{0.6,w^*}(0.25)}^{q_{0.6,w^*}(0.35)} \omega_2 f_{\epsilon,W^*|X=0.6}(\epsilon, w^*) d\epsilon dw^* = 1$. These choices of the weighting functions yield a true value of the WLAR of 0.25 and 0.5032 for Design 1 and 2, respectively. In the estimation, I use the optimal values of h_1 coming from Table 1.2, and the same values of h_{21} and h_{22} as in Table 1.2. Table 1.6 below shows the results of my estimators of the WLAR under both designs. The performance of my estimator of the WLAR is comparable to the performance of my estimator for the point-wise structural derivative $\rho(y, x)$ (shown in Table 1.2). This shows evidence that my estimator could not only perform well at single points but also could maintain good performance in a global sense. The supports of the weighting functions I used here are not large. This is mainly due to computation constraints. One direction of future work is to look at the performance of my estimator for WLAR when the support of the weighting function is larger. This will give us a better understanding of the global performance of my estimator.

Table 1.6: Monte Carlo simulation results of my proposed estimator for the WLAR

	h_1	h_{21}	h_{22}	true value	MSE	VAR	abs(BIAS)
Design 1 (linear)	1	1.05	2.92	0.25	0.00779	0.00777	0.00405
Design 2 (nonlinear)	1.75	1.75	2.09	0.5032	0.15082	0.13829	0.11193

The Monte Carlo simulation exercise shown in this section is very limited. To have a comprehensive examination of the performance of my estimators, there are several future directions that I can work on. First, in my current exercise, the distribution of the measurement error ΔW_2 is set to be χ^2, which is ordinarily smooth. I can further explore the performance of my estimators under different distributions of the measurement error

ΔW_2 and other variables. Second, when estimating the WLAR, I set the support for the weighting function to be small because of the computation constraint. With more time and more computation power, I can look at the performance of my estimator when the support of the weighting function in the WLAR is larger, i.e. the WLAR is taking the average marginal effect over a larger population. Third, the way I select h_1 is by minimizing the actual MSE, i.e., the mean squared difference between my estimator and the true value of the parameter. This is infeasible in practice when one deals with real-world data. It will be helpful if I could propose a feasible way to select the bandwidth.

1.6 Conclusion

I study the nonparametric identification and estimation of the nonseparable triangular equations model when the instrument variable W^* is mismeasured. I don't assume linearity or separability of the functions governing the relationship between observables and unobservables. To deal with the challenges caused by the co-existence of the measurement error and nonseparability, I first employ the deconvolution technique developed in the measurement error literature (Schennach (2004a),Schennach (2004b)) to identify the joint distribution of Y, X, W^* using two error-laden measurements of W^*. I then recover the structural derivative of the function of interest and the "Local Average Response" (LAR) via the "unobserved instrument" approach in Matzkin (2016). Based on the constructive identification results, I propose plug-in nonparametric estimators for these parameters and derive their uniform rates of convergence. I also conducted limited Monte Carlo exercises to show the finite sample performance of my estimators.

I recognize that there are some important future directions for this paper. First, in the main text, I only demonstrated the uniform rate of convergence of the estimators. The appendix contains proofs of their asymptotic normality, which rely on high-level assumptions. To further enhance my results, it would be beneficial to derive more primitive sufficient conditions for these assumptions. Second, as discussed in Section 1.5, to conduct a comprehensive examination of the performance of my estimators, additional Monte Carlo studies are necessary. It will also be extremely useful if I could propose a feasible way to select the bandwidth h_1. Last, it will be an interesting exercise to apply my proposed method to real-world data.

1.7 Appendix

1.7.1 Appendix A: Useful Lemmas

Before stating the lemmas, I define some notations.

Let Y be a random variable, Z be a random vector of dimension 2. Denote $Z_1 \equiv X, Z_2 \equiv W$. Let $\mathbb{S}_{(Y,Z)}$ be the support of (Y, Z), and define

$$\mathbb{S}_\tau \equiv \left\{ (y, z) \in \mathbb{S}_{(Y,Z)} : \left| \frac{\partial F_{X|W=w}(x)}{\partial w} \right| \geq \tau, \text{ and } f_{Y,X,W}(y, x, w) > \tau \right\}$$

and $\mathcal{S}_\tau \equiv \left\{ (\delta, z) \in [0, 1] \times \mathbb{S}_Z : \left| \frac{\partial F_{X|W=w}(x)}{\partial w} \right| \geq \tau \text{ and } f_{Y,X,W}(v(x, w, \delta), x, w) > \tau \right\}$. Let $\bar{\mathbb{M}}$ be a compact subset of the support of (Y, X, W). Define the mapping **invm** as $\mathbf{invm}(Y, X, W) \equiv (m^{-1}(Y, X), X, W)$, where m^{-1} is as defined in (1.2). Define the mapping **invs** as $\mathbf{invs}(Y, X, W) \equiv (s^{-1}(r(X, W), m^{-1}(X, Y)), X, W)$, where s^{-1} is the inverse of the s function defined in Lemma 1.1 with respect to its second argument. Let $\mathbb{M} \equiv \mathbf{invm}(\bar{\mathbb{M}})$, and $\mathcal{M} = \mathbf{invs}(\bar{\mathbb{M}})$. By Assumption 1.1 and 1.6, $\mathbb{M}$ and $\mathcal{M}$ are also compact, and there exist $\tau > 0$ such that $\bar{\mathbb{M}} \subset \mathbb{S}_\tau$ and $\mathcal{M} \subset \mathcal{S}_\tau$.

For any function $g : \mathbb{R}^3 \to \mathbb{R}$, define $\tilde{g}(z) \equiv \int g(y, z)dy$, $\tilde{g}_{Z_2}(z_2) \equiv \int \tilde{g}(z)dx$, $G(y, z) \equiv \int_{-\infty}^{y} \int_{-\infty}^{z} g(s, t)dtds$, and if $\tilde{g}(z) \neq 0$, define $G_{Y|Z=z}(y) = \left(\int_{-\infty}^{y} g(s, z)ds \right) / \tilde{g}(z)$; if $\tilde{g}_{Z_2}(z_2) \neq 0$, define $G_{Z_1|Z_2=z_2}(z_1) = \left(\int_{-\infty}^{z_1} \tilde{g}(s, z_2)ds \right) / \tilde{g}_{Z_2}(z_2)$. Let F denote a set of twice continuously differentiable functions $g : \mathbb{R}^3 \to \mathbb{R}$ such that g vanishes outside of $\mathbb{S}_{(Y,Z)}$. Let f denote the joint density of (Y, Z). Assume that f belongs to F. For any value $(y, z) \in \mathbb{S}_{Y,Z}$, and any value $\delta \in [0, 1]$, define the functionals $\Lambda(\cdot)$, $\alpha(\cdot)$, $\Phi_{(j)}(\cdot)$, $\Psi_{(j)}(\cdot)$ and $\tilde{\Lambda}(\cdot)$, $\tilde{\Psi}(\cdot)$ on F by $\Lambda(g) \equiv G_{Y|Z=z}(y)$, $\alpha(g) \equiv G^{-1}_{Y|Z=z}(\delta)$, $\Phi_{(j)}(g) \equiv \frac{\partial G^{-1}_{Y|Z=z}(\delta)}{\partial z_j}$, $\Psi_{(j)}(g) \equiv - \frac{\partial G_{Y|Z=z}(y)}{\partial z_j} \Big/ \frac{\partial G_{Y|Z=z}(y)}{\partial y}$ and $\tilde{\Lambda}(g) \equiv G_{X|W=w}(x)$, $\tilde{\Psi}(g) \equiv - \frac{\partial G_{X|W=w}(x)}{\partial w} \Big/ \frac{\partial G_{X|W=w}(x)}{\partial x}$. For simplicity, I leave the argument (z, y, δ) implicit.

Lemma 1.6.

$$\Phi_{(j)}(f) = \frac{\delta \frac{\partial \tilde{f}(z)}{\partial z_j} - \int_{-\infty}^{\alpha(f)} \frac{\partial f(y,z)}{\partial z_j} dy}{f(\alpha(f), z)}$$

$$\Psi_{(j)}(f) = \frac{\Lambda(f) \frac{\partial \tilde{f}(z)}{\partial z_j} - \int_{-\infty}^{y} \frac{\partial f(s,z)}{\partial z_j} ds}{f(y, z)}$$

$$\tilde{\Psi}(f) = \frac{\tilde{\Lambda}(f) \frac{\partial \tilde{f}_W(w)}{\partial w} - \int_{-\infty}^{x} \frac{\partial \tilde{f}(s,w)}{\partial w} ds}{\tilde{f}(x, w)}.$$

Proof. By definition,

$$\begin{aligned} \delta &= F_{Y|Z=z}(\alpha(f)) \\ &= \frac{\int_{-\infty}^{\alpha(f)} f(y,z)dy}{\int_{-\infty}^{\infty} f(y,z)dy} \\ \implies \delta \tilde{f}(z) &= \int_{-\infty}^{\alpha(f)} f(y,z)dy \end{aligned}$$

Taking derivatives on both sides with respect to z_j yields

$$\begin{aligned} f(\alpha(f), z)\Phi_{(j)}(f) + \int_{-\infty}^{\alpha(f)} \frac{\partial f(y,z)}{\partial z_j} dy &= \delta \frac{\partial \tilde{f}(z)}{\partial z_j} \\ \implies \Phi_{(j)}(f) &= \frac{\delta \frac{\partial \tilde{f}(z)}{\partial z_j} - \int_{-\infty}^{\alpha(f)} \frac{\partial f(y,z)}{\partial z_j} dy}{f(\alpha(f), z)}. \end{aligned}$$

Then note that

$$\begin{aligned} \frac{\partial G_{Y|Z=z}(x)}{\partial z_j} &= \frac{\left(\int_{-\infty}^{y} \frac{\partial f(s,z)}{\partial z_j} ds \right) \tilde{f}(z) - \left(\int_{-\infty}^{y} f(s,z) ds \right) \frac{\partial \tilde{f}(z)}{\partial z_j}}{\tilde{f}^2(z)} \\ \frac{\partial G_{Y|Z=z}(x)}{\partial y} &= \frac{f(y,z)}{\tilde{f}(z)}. \end{aligned}$$

Then

$$\Psi_{(j)}(f) \equiv - \left. \frac{\partial G_{Y|Z=z}(x)}{\partial z_j} \middle/ \frac{\partial G_{Y|Z=z}(x)}{\partial y} \right.$$

$$= \frac{F_{Y|Z=z}(y)\frac{\partial \tilde{f}(z)}{\partial z_j} - \int_{-\infty}^{y} \frac{\partial f(s,z)}{\partial z_j} ds}{f(y,z)}.$$

The conclusion for $\tilde{\Psi}(f)$ follows by the same argument. □

Lemma 1.7. *For any h in* F *such that* $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$ *is small enough, I have that*

$$\Lambda(f+h) - \Lambda(f) = D\Lambda(f;h) + R\Lambda(f;h)$$
$$\tilde{\Lambda}(f+h) - \tilde{\Lambda}(f) = D\tilde{\Lambda}(f;h) + R\tilde{\Lambda}(f;h),$$

where

$$D\Lambda(f;h) = \frac{\int_{-\infty}^{y} h(s,z)ds - \tilde{h}(z)F_{Y|Z=z}(y)}{\tilde{f}(z)}$$
$$R\Lambda(f;h) = \left[\frac{\int_{-\infty}^{y} h(s,z)ds - \tilde{h}(z)F_{Y|Z=z}(y)}{\tilde{f}(z)}\right]\left[\frac{\tilde{h}(z)}{\tilde{f}(z)+\tilde{h}(z)}\right]$$
$$D\tilde{\Lambda}(f;h) = \frac{\int_{-\infty}^{x} \tilde{h}(s,w)ds - \tilde{h}_W(w)F_{X_1|W=w}(x)}{\tilde{f}_W(w)}$$
$$R\tilde{\Lambda}(f;h) = \left[\frac{\int_{-\infty}^{x} \tilde{h}(s,w)ds - \tilde{h}_W(w)F_{X_1|W=w}(x)}{\tilde{f}_W(w)}\right]\left[\frac{\tilde{h}_W(w)}{\tilde{f}_W(w)+\tilde{h}_W(w)}\right]$$

and for some $a_1 < \infty$,

$$\sup_{(y,z)\in\mathbb{S}_\tau} |D\Lambda(f;h)| \leq a_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(y,z)\in\mathbb{S}_\tau} |R\Lambda(f;h)| \leq a_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2;$$

for some $\tilde{a}_1 < \infty$,

$$\sup_{(y,z)\in\mathbb{S}_\tau} \left|D\tilde{\Lambda}(f;h)\right| \leq \tilde{a}_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(y,z)\in\mathbb{S}_\tau} \left|R\tilde{\Lambda}(f;h)\right| \leq \tilde{a}_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

Proof.

$$\begin{aligned}
\Lambda(f+h)-\Lambda(f) &= \frac{\int_{-\infty}^{y}(f(s,z)+h(s,z))ds}{\tilde{f}(z)+\tilde{h}(z)} - \frac{\int_{-\infty}^{y} f(s,z)ds}{\tilde{f}(z)} \\
&= \frac{\int_{-\infty}^{y} hds}{\tilde{f}+\tilde{h}} - \frac{\tilde{h}\int_{-\infty}^{y} fds}{\tilde{f}(\tilde{f}+\tilde{h})} \\
&= \frac{\int_{-\infty}^{y} hds - \tilde{h}F_{Y|Z=z}(y)}{\tilde{f}} + \left[\frac{\int_{-\infty}^{y} hds - \tilde{h}F_{Y|Z=z}(y)}{\tilde{f}}\right]\left[\frac{\tilde{h}}{\tilde{f}+\tilde{h}}\right]; \\
\tilde{\Lambda}(f+h)-\tilde{\Lambda}(f) &= \frac{\int_{-\infty}^{x} \tilde{h}ds - \tilde{h}_W F_{X_1|W=w}(x)}{\tilde{f}_W} + \left[\frac{\int_{-\infty}^{x} \tilde{h}ds - \tilde{h}_W F_{X_1|W=w}(x)}{\tilde{f}_W}\right]\left[\frac{\tilde{h}_W}{\tilde{f}_W+\tilde{h}_W}\right].
\end{aligned}$$

Define

$$\begin{aligned}
D\Lambda(f;h) &= \frac{\int_{-\infty}^{y} hds - \tilde{h}F_{Y|Z=z}(y)}{\tilde{f}} \\
R\Lambda(f;h) &= \left[\frac{\int_{-\infty}^{y} hds - \tilde{h}F_{Y|Z=z}(y)}{\tilde{f}}\right]\left[\frac{\tilde{h}}{\tilde{f}+\tilde{h}}\right] \\
D\tilde{\Lambda}(f;h) &= \frac{\int_{-\infty}^{x} \tilde{h}ds - \tilde{h}_W F_{X_1|W=w}(x)}{\tilde{f}_W} \\
R\tilde{\Lambda}(f;h) &= \left[\frac{\int_{-\infty}^{x} \tilde{h}ds - \tilde{h}_W F_{X_1|W=w}(x)}{\tilde{f}_W}\right]\left[\frac{\tilde{h}_W}{\tilde{f}_W+\tilde{h}_W}\right].
\end{aligned}$$

By the boundedness of $\mathbb{S}_{(Y,Z)}$, there exist finite constants C_X, C_Y such that for all $h \in \mathrm{F}$,

$$\begin{aligned}
\sup_{z\in\mathbb{S}_Z}\left|\tilde{h}(z)\right| &\le \sup_{z\in\mathbb{S}_Z}\int |h(y,z)|\,dy \le C_Y \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \\
\sup_{w\in\mathbb{S}_W}\left|\tilde{h}_W(w)\right| &= \sup_{w\in\mathbb{S}_W}\int \left|\tilde{h}(x,w)\right| dx \le C_X C_Y \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|
\end{aligned}$$

By the definition of set $\mathbb{S}_\tau$, I have that $\inf_{(y,z)\in\mathbb{S}_\tau}\tilde{f}(z) > \tau$ and that $\inf_{(y,z)\in\mathbb{S}_\tau}\tilde{f}_W(w) = \inf_{(y,z)\in\mathbb{S}_\tau}\int \tilde{f}(s,w)ds > \tau C_X$ Let $\epsilon_0 \equiv \min\{\frac{\tau}{2C_Y}, 1\}$, For any h in F that $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \le \epsilon_0$, I have that $\sup_{z\in\mathbb{S}_Z}\left|\tilde{h}(z)\right| \le \tau/2$, $\sup_{w\in\mathbb{S}_W}\left|\tilde{h}_W(w)\right| \le C_X\tau/2$, so that $\inf_{(y,z)\in\mathbb{S}_\tau}(\tilde{f}(z)+\tilde{h}(z)) \ge \tau/2$, and $\inf_{(y,z)\in\mathbb{S}_\tau}(\tilde{f}_W(w)+\tilde{h}_W(w)) \ge C_X\tau/2$. Then there exists a finite constant a_0 such that

$$\sup_{(y,z)\in\mathbb{S}_\tau} |D\Lambda(f;h)| \leq \frac{a_0}{\tau} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(y,z)\in\mathbb{S}_\tau} |R\Lambda(f;h)| \leq \frac{a_0}{\tau} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \times \frac{2\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|}{\tau} = \frac{2a_0}{\tau^2} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

Hence let $a_1 \equiv \max\{a_0/\tau, 2a_0/\tau^2\}$, I have that

$$\sup_{(y,z)\in\mathbb{S}_\tau} |D\Lambda(f;h)| \leq a_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(y,z)\in\mathbb{S}_\tau} |R\Lambda(f;h)| \leq a_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

Similarly, I can show that there exists a constant $\tilde{a}_1$ such that

$$\sup_{(y,z)\in\mathbb{S}_\tau} \left|D\tilde{\Lambda}(f;h)\right| \leq \tilde{a}_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(y,z)\in\mathbb{S}_\tau} \left|R\tilde{\Lambda}(f;h)\right| \leq \tilde{a}_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

□

Lemma 1.8. *For any h in* F *such that* $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$ *is small enough, I have that*

$$\alpha(f+h) - \alpha(f) = D\alpha(f;h) + R\alpha(f;h),$$

where

$$D\alpha(f;h) \equiv \frac{\tilde{h}(z)\int^{\alpha(f)} f(y,z)dy - \tilde{f}(z)\int^{\alpha(f)} h(y,z)dy}{\tilde{f}(z)f(\alpha(f),z)}$$
$$R\alpha(f;h) \equiv -\left[\frac{\frac{\partial f(r_f',z)}{\partial y}\left(r_f - \alpha(f)\right) + h(r_h,z)}{f(r_f,z) + h\left(r_h,z\right)}\right] D\alpha(f;h)$$

for some r_f and r_f' between $\alpha(f+h)$ and $\alpha(f)$ defined in the proof. Moreover, for some

$0 < a < \infty$,

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\alpha(f;h)| \le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$
$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\alpha(f;h)| \le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

Proof. First I show that for some $0 < a_2 < \infty$ and a_1 mentioned in Lemma 1.7,

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |\alpha(f+h) - \alpha(f)| \le \frac{2a_1a_2}{\tau} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|.$$

By the Mean Value Theorem, there exist r_1 between $\alpha(f+h)$ and $\alpha(f)$, such that

$$\begin{aligned} &(F+H)_{Y|Z=z}\left(\alpha(f+h)\right) - (F+H)_{Y|Z=z}\left(\alpha(f)\right) \\ =\ &(f+h)_{Y|Z=z}(r_1)\left(\alpha(f+h) - \alpha(f)\right) \end{aligned}$$

Hence since

$$\begin{aligned} &(F+H)_{Y|Z=z}\left(\alpha(f+h)\right) \\ =\ &(F+H)_{Y|Z=z}\left((F+H)^{-1}_{Y|Z=z}(\delta)\right) = \delta = F_{Y|Z=z}\left(F^{-1}_{Y|Z=z}(\delta)\right) \end{aligned}$$

it follows that

$$\alpha(f+h) - \alpha(f) = \frac{F_{Y|Z=z}\left(F^{-1}_{Y|Z=z}(\delta)\right) - (F+H)_{Y|Z=z}\left(F^{-1}_{Y|Z=z}(\delta)\right)}{(f+h)_{Y|Z=z}(r_1)}.$$

By the compactness of $\mathbb{S}_{Y,Z}$, for any h in F, there exist some finite constant a_2 such that $\sup_{z\in\mathbb{S}_Z}(\tilde{f}(z) + \tilde{h}(z)) \le a_2$. By the definition of $\mathbb{S}_\tau$, $\inf_{(y,z)\in\mathbb{S}_\tau} f(y,z) > \tau$. By similar argument as in Lemma 1.7, for any h in F such that $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \le \min\{\tau/2, 1\}$, I have $\inf_{(y,z)\in\mathbb{S}_\tau}(f(y,z) + h(y,z)) > \tau/2$. Then $\inf_{(y,z)\in\mathbb{S}_\tau}(f+h)_{Y|Z=z}(r_1) = \frac{\inf_{(y,z)\in\mathbb{S}_\tau} f(r_1,z)+h(r_1,z)}{\sup_{(y,z)\in\mathbb{S}_\tau} \tilde{f}(z)+\tilde{h}(z)} > \frac{\tau}{2a_2}$.

By Lemma 1.7,

$$\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|F_{Y|Z=z}\left(F^{-1}_{Y|Z=z}(\delta)\right)-(F+H)_{Y|Z=z}\left(F^{-1}_{Y|Z=z}(\delta)\right)\right|\leq a_1\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|$$

for some $0<a_1<\infty$. Then I have that

$$\sup_{(\delta,z)\in\mathcal{S}_\tau}|\alpha(f+h)-\alpha(f)|\leq\frac{2a_1a_2}{\tau}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|. \tag{1.23}$$

Next, I will obtain a first-order expansion for $\alpha(f+h)$. By the fact that $(F+H)_{Y|Z=z}(\alpha(f+h))=\delta=F_{Y|Z=z}(\alpha(f))$, I have

$$\begin{aligned}
&\frac{\int_{-\infty}^{\alpha(f)}f(y,z)dy}{\tilde{f}(z)}=\frac{\int^{\alpha(f+h)}(f(y,z)+h(y,z))dy}{\tilde{f}(z)+\tilde{h}(z)}\\
\Longrightarrow\ &(\tilde{f}(z)+\tilde{h}(z))\int^{\alpha(f)}f(y,z)dy=\tilde{f}(z)\int^{\alpha(f+h)}(f(y,z)+h(y,z))\,dy\\
\Longrightarrow\ &\tilde{h}(z)\int^{\alpha(f)}f(y,z)dy-\tilde{f}(z)\int^{\alpha(f)}h(y,z)dy\\
&\quad=\tilde{f}(z)\int_{\alpha(f)}^{\alpha(f+h)}h(y,z)dy+\tilde{f}(z)\int_{\alpha(f)}^{\alpha(f+h)}f(y,z)dy
\end{aligned}$$

By the Mean Value Theorem, there exist r_f and r_h, between $\alpha(f)$ and $\alpha(f+h)$, such that

$$\begin{aligned}
&\int_{\alpha(f)}^{\alpha(f+h)}f(y,z)dy=f(r_f,z)\left(\alpha(f+h)-\alpha(f)\right)\ \text{and}\\
&\int_{\alpha(f)}^{\alpha(f+h)}h(y,z)dy=h(r_h,z)\left(\alpha(f+h)-\alpha(f)\right).
\end{aligned}$$

Denote $Az\equiv\tilde{h}(z)\int^{\alpha(f)}f(y,z)dy-\tilde{f}(z)\int^{\alpha(f)}h(y,z)dy$. Then

$$\begin{aligned}
&Az=\tilde{f}(z)\left[f(r_f,z)+h(r_h,z)\right](\alpha(f+h)-\alpha(f))\\
\Longrightarrow\ &\alpha(f+h)-\alpha(f)=\frac{Az}{\tilde{f}(z)\left[f(r_f,z)+h(r_h,z)\right]}.
\end{aligned}$$

By the Mean Value Theorem, there exist r'_f between $\alpha(f)$ and r_f such that $f(r_f,z)-$

$f(\alpha(f), z) = [\partial f(r'_f, z)/\partial y](r_f - \alpha(f))$. Hence

$$\alpha(f+h) - \alpha(f) = \frac{Az}{\tilde{f}(z)\left(f(\alpha(f), z) + \frac{\partial f\left(r'_f, z\right)}{\partial y}\left(r_f - \alpha(f)\right) + h\left(r_h, z\right)\right)}$$

$$= \frac{Az}{\tilde{f}(z) f(\alpha(f), z)} - \left[\frac{\frac{\partial f(r'_f, z)}{\partial y}\left(r_f - \alpha(f)\right) + h(r_h, z)}{f(\alpha(f), z) + \frac{\partial f\left(r'_f, z\right)}{\partial y}\left(r_f - \alpha(f)\right) + h\left(r_h, z\right)}\right] \frac{Az}{\tilde{f}(z) f(\alpha(f), z)}$$

Denote

$$D\alpha(f; h) \equiv \frac{Az}{\tilde{f}(z) f(\alpha(f), z)} \text{ and}$$

$$R\alpha(f; h) \equiv -\left[\frac{\frac{\partial f(r'_f, z)}{\partial y}\left(r_f - \alpha(f)\right) + h(r_h, z)}{f(r_f, z) + h\left(r_h, z\right)}\right] \frac{Az}{\tilde{f}(z) f(\alpha(f), z)}.$$

Then

$$\alpha(f+h) - \alpha(f) = D\alpha(f; h) + R\alpha(f; h).$$

By the definition of r_f and by (1.23), $\sup_{(\delta,z)\in\mathcal{S}_\tau} |r_f - \alpha(f)| \le \sup_{(\delta,z)\in\mathcal{S}_\tau} |\alpha(f+h) - \alpha(f)| \le \frac{2a_1a_2}{\tau} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$. It follows by the continuity of $\partial f/\partial y$ and the compactness of $\mathbb{S}_{(Y,Z)}$ that

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} \left| \frac{\partial f(r'_f, z)}{\partial y} \left(r_f - \alpha(f)\right) + h(r_h, z) \right| \le d_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$$

for some finite constant d_1. Then for all $h \in \mathrm{F}$ such that $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \le \min\{\tau/2, 1\}$, I have $\inf_{(\delta,z)\in\mathcal{S}_\tau} f(r_f, z) + h\left(r_h, z\right) > \tau/2$. Then there exists a finite constant a such that

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\alpha(f; h)| \le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \text{ and}$$

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\alpha(f; h)| \le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.$$

$\square$

Lemma 1.9. *For any value* $z \in \mathbb{S}_Z$*, and any value* $\delta \in (0,1)$*, define functional* $\Phi_{1,(j)}(\cdot)$ *as* $\Phi_{1,(j)}(g) \equiv \int_{-\infty}^{\alpha(g)} \frac{\partial g(y,z)}{\partial z_j} dy$*. For simplicity, I leave the argument* (z, δ) *implicit. For any* h *in* F *that* $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$ *is sufficiently small, I have that*

$$\Phi_{1,(j)}(f+h) - \Phi_{1,(j)}(f) = D\Phi_{1,(j)}(f;h) + R\Phi_{1,(j)}(f;h),$$

where

$$\begin{aligned}
D\Phi_{1,(j)}(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial z_j} D\alpha(f;h) + \int_{-\infty}^{\alpha(f)} \frac{\partial h(y,z)}{\partial z_j} dy \\
R\Phi_{1,(j)}(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial z_j} R\alpha(f;h) + \frac{\partial^2 f(\bar{r}'_f, z)}{\partial y \partial z_j} \left(\bar{r}_f - \alpha(f)\right) \left(\alpha(f+h) - \alpha(f)\right) \\
&\quad + \frac{\partial h(\bar{r}_h, z)}{\partial z_j} \left(\alpha(f+h) - \alpha(f)\right),
\end{aligned}$$

for some $\bar{r}_f$ *and* $\bar{r}_h$*, and* $\bar{r}'_f$ *between* $\alpha(f+h)$ *and* $\alpha(f)$ *defined in the proof. Moreover, for some* $0 < b_1 < \infty$*,*

$$\begin{aligned}
\sup_{(\delta,z)\in\mathcal{S}_\tau} \left| D\Phi_{1,(j)}(f;h) \right| &\leq b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| + b_1 \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left| \frac{\partial h(y,z)}{\partial z_j} \right| \\
\sup_{(\delta,z)\in\mathcal{S}_\tau} \left| R\Phi_{1,(j)}(f;h) \right| &\leq b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left| \frac{\partial h(y,z)}{\partial z_j} \right|.
\end{aligned}$$

Proof.

$$\begin{aligned}
\Phi_{1,(j)}(f+h) - \Phi_{1,(j)}(f) &= \int_{-\infty}^{\alpha(f+h)} \left(\frac{\partial f(y,z)}{\partial z_j} + \frac{\partial h(y,z)}{\partial z_j} \right) dy - \int_{-\infty}^{\alpha(f)} \frac{\partial f(y,z)}{\partial z_j} dy \\
&= \int_{-\infty}^{\alpha(f)} \frac{\partial h(y,z)}{\partial z_j} dy + \int_{\alpha(f)}^{\alpha(f+h)} \frac{\partial h(y,z)}{\partial z_j} dy + \int_{\alpha(f)}^{\alpha(f+h)} \frac{\partial f(y,z)}{\partial z_j} dy
\end{aligned}$$

By the Mean Value Theorem, there exist $\bar{r}_f$ and $\bar{r}_h$ between $\alpha(f)$ and $\alpha(f+h)$ such that

$$\int_{\alpha(f)}^{\alpha(f+h)} \frac{\partial f(y,z)}{\partial z_j} dy = \frac{\partial f(\bar{r}_f, z)}{\partial z_j} \left(\alpha(f+h) - \alpha(f)\right)$$

$$\int_{\alpha(f)}^{\alpha(f+h)} \frac{\partial h(y,z)}{\partial z_j} dy = \frac{\partial h(\bar{r}_h, z)}{\partial z_j} \left(\alpha(f+h) - \alpha(f)\right)$$

Apply the Mean Value Theorem again. I have that there exist $\bar{r}'_f$ between $\alpha(f)$ and $\alpha(f+h)$ such that

$$\int_{\alpha(f)}^{\alpha(f+h)} \frac{\partial f(y,z)}{\partial z_j} dy = \frac{\partial f(\alpha(f),z)}{\partial z_j} \left(\alpha(f+h) - \alpha(f)\right) + \frac{\partial^2 f(\bar{r}'_f, z)}{\partial y \partial z_j} \left(\bar{r}_f - \alpha(f)\right) \left(\alpha(f+h) - \alpha(f)\right).$$

Let

$$\begin{aligned} D\Phi_{1,(j)}(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial z_j} D\alpha(f;h) + \int_{-\infty}^{\alpha(f)} \frac{\partial h(y,z)}{\partial z_j} dy \\ R\Phi_{1,(j)}(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial z_j} R\alpha(f;h) + \frac{\partial^2 f(\bar{r}'_f, z)}{\partial y \partial z_j} \Big(\bar{r}_f - \alpha(f)\Big) \left(\alpha(f+h) - \alpha(f)\right) \\ &\quad + \frac{\partial h(\bar{r}_h, z)}{\partial z_j} \left(\alpha(f+h) - \alpha(f)\right). \end{aligned}$$

By the boundedness of $\mathbb{S}_{Y,Z}$, continuity of $\partial f/\partial z_j$, $\partial^2 f/\partial y \partial z_j$, $\partial h/\partial z_j$ and Lemma 1.8 there exist finite constants d_2, d_3, d_4, d_5 such that

$$\begin{aligned} \sup_{(\delta,z)\in\mathcal{S}_\tau} \left|D\Phi_{1,(j)}(f;h)\right| &\leq d_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| + d_3 \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial z_j}\right| \\ \sup_{(\delta,z)\in\mathcal{S}_\tau} \left|R\Phi_{1,(j)}(f;h)\right| &\leq d_4 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + d_5 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial z_j}\right| \end{aligned}$$

Let $b_1 \equiv \max\{d_2, d_3, d_4, d_5\}$, then

$$\begin{aligned} \sup_{(\delta,z)\in\mathcal{S}_\tau} \left|D\Phi_{1,(j)}(f;h)\right| &\leq b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| + b_1 \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial z_j}\right| \\ \sup_{(\delta,z)\in\mathcal{S}_\tau} \left|R\Phi_{1,(j)}(f;h)\right| &\leq b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + b_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial z_j}\right|. \end{aligned}$$

□

Lemma 1.10. *Define functional $\Phi_2()$ on F by $\Phi_2(g) = g(\alpha(g), z)$. For any h in F that*

$\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$ *is small enough, I have that*

$$\Phi_2(f+h) - \Phi_2(f) = D\Phi_2(f;h) + R\Phi_2(f;h),$$

where

$$\begin{aligned} D\Phi_2(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial y} D\alpha(f;h) + h(\alpha(f),z) \\ R\Phi_2(f;h) &\equiv \frac{\partial f(\alpha(f),z)}{\partial y} R\alpha(f;h) + \frac{\partial^2 f(\tilde{r}'_f,z)}{\partial y^2}\left(\tilde{r}_f - \alpha(f)\right)\left(\alpha(f+h) - \alpha(f)\right) + \\ &\quad \frac{\partial h(\tilde{r}_h,z)}{\partial y}\left(\alpha(f+h) - \alpha(f)\right). \end{aligned}$$

for some $\tilde{r}_f$ *and* $\tilde{r}_h$*, and* $\tilde{r}'_f$ *between* $\alpha(f+h)$ *and* $\alpha(f)$ *defined in the proof. Moreover, for some* $0 < b_2 < \infty$*,*

$$\begin{aligned} &\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\Phi_2(f;h)| \le b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|. \\ &\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\Phi_2(f;h)| \le b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,X}} \left|\frac{\partial h(y,z)}{\partial y}\right|. \end{aligned}$$

Proof.

$$\begin{aligned} &\Phi_2(f+h) - \Phi_2(f) \\ &= (f+h)\left(\alpha(f+h),z\right) - f\left(\alpha(f),z\right) \\ &= f\left(\alpha(f+h),z\right) - f\left(\alpha(f),z\right) + h\left(\alpha(f+h),z\right) - h\left(\alpha(f),z\right) + h\left(\alpha(f),z\right) \end{aligned}$$

By the Mean Value Theorem, there exist $\tilde{r}_f$ and $\tilde{r}_h$ between $\alpha(f+h)$ and $\alpha(f)$, such that

$$\begin{aligned} f\left(\alpha(f+h),z\right) - f\left(\alpha(f),z\right) &= \frac{\partial f(\tilde{r}_f,z)}{\partial y}\left(\alpha(f+h) - \alpha(f)\right) \\ h\left(\alpha(f+h),z\right) - h\left(\alpha(f),z\right) &= \frac{\partial h(\tilde{r}_h,z)}{\partial y}\left(\alpha(f+h) - \alpha(f)\right). \end{aligned}$$

Apply the Mean Value Theorem again. There exist $\tilde{r}'_f$ between $\tilde{r}_f$ and $\alpha(f)$ such that

$$
\begin{aligned}
& f\left(\alpha(f+h), z\right) - f\left(\alpha(f), z\right) \\
= & \frac{\partial f(\alpha(f), z)}{\partial y} \left(\alpha(f+h) - \alpha(f)\right) + \frac{\partial^2 f(\tilde{r}'_f, z)}{\partial y^2} \left(\tilde{r}_f - \alpha(f)\right) \left(\alpha(f+h) - \alpha(f)\right)
\end{aligned}
$$

Then by Lemma 1.8 I have that

$$
\begin{aligned}
& \Phi_2(f+h) - \Phi_2(f) \\
= & \frac{\partial f(\alpha(f), z)}{\partial y} \left(D\alpha(f;h) + R\alpha(f;h)\right) + \frac{\partial^2 f(\tilde{r}'_f, z)}{\partial y^2} \left(\tilde{r}_f - \alpha(f)\right) \left(\alpha(f+h) - \alpha(f)\right) + \\
& \quad \frac{\partial h(\tilde{r}_h, z)}{\partial y} \left(\alpha(f+h) - \alpha(f)\right) + h(\alpha(f), z),
\end{aligned}
$$

where for some $0 < a < \infty$ and $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \le \epsilon_0 \equiv \min\{\tau/2, 1\}$,

$$
\begin{aligned}
\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\alpha(f;h)| &\le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \\
\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\alpha(f;h)| &\le a \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2.
\end{aligned}
$$

Define

$$
\begin{aligned}
D\Phi_2(f;h) &\equiv \frac{\partial f(\alpha(f), z)}{\partial y} D\alpha(f;h) + h(\alpha(f), z) \\
R\Phi_2(f;h) &\equiv \frac{\partial f(\alpha(f), z)}{\partial y} R\alpha(f;h) + \frac{\partial^2 f(\tilde{r}'_f, z)}{\partial y^2} \left(\tilde{r}_f - \alpha(f)\right) \left(\alpha(f+h) - \alpha(f)\right) + \\
& \quad \frac{\partial h(\tilde{r}_h, z)}{\partial y} \left(\alpha(f+h) - \alpha(f)\right).
\end{aligned}
$$

Then

$$
\Phi_2(f+h) - \Phi_2(f) = D\Phi_2(f;h) + R\Phi_2(f;h).
$$

By the compactness of $\mathbb{S}_{(Y,X)}$ and continuity of $\partial f/\partial y$, $\partial^2 f/\partial y^2$, and $\partial h/\partial y$, I have that there exist some finite constant d_6 such that

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\Phi_2(f;h)| \leq d_6 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|.$$

holds for all $h \in \mathrm{F}$. On the other hand, by Lemma 1.8, there exist some finite constant a_1, a_2 such that $\sup_{(\delta,z)\in\mathcal{S}_\tau} |\tilde{r}_f - \alpha(f)| \leq \sup_{(\delta,z)\in\mathcal{S}_\tau} |\alpha(f+h) - \alpha(f)| \leq \frac{2a_1a_2}{\tau} \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$. Then there exist some finite constants d_7, d_8 such that

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\Phi_2(f;h)| \leq d_7 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + d_8 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial y}\right|$$

Let $b_2 \equiv \max\{d_6, d_7, d_8\}$. Then

$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |D\Phi_2(f;h)| \leq b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|.$$
$$\sup_{(\delta,z)\in\mathcal{S}_\tau} |R\Phi_2(f;h)| \leq b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|^2 + b_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}} \left|\frac{\partial h(y,z)}{\partial y}\right|.$$

□

Lemma 1.11. *For any value $z \in \mathbb{S}_Z$, and any value $\delta \in (0,1)$, define functionals $\Psi_{1,(j)}(\cdot)$ and $\tilde{\Psi}_1(\cdot)$ as $\Psi_{1,(j)}(g) \equiv \int_{-\infty}^{y} \frac{\partial g(s,z)}{\partial z_j} ds$ and $\tilde{\Psi}_1(g) \equiv \int_{-\infty}^{x} \frac{\partial \tilde{g}(s,w)}{\partial w} ds$. For simplicity, I leave the argument z implicit. For any h in* F *that $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}} |h|$ is sufficiently small, I have that*

$$\Phi_{(j)}(f+h) - \Phi_{(j)}(f) = D\Phi_{(j)}(f;h) + R\Phi_{(j)}(f;h),$$
$$\Psi_{(j)}(f+h) - \Psi_{(j)}(f) = D\Psi_{(j)}(f;h) + R\Psi_{(j)}(f;h),$$
$$\tilde{\Psi}(f+h) - \tilde{\Psi}(f) = D\tilde{\Psi}(f;h) + R\tilde{\Psi}(f;h),$$

where

$$D\Phi_{(j)}(f;h) \equiv \frac{\left(\delta \frac{\partial \tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h)\right)\Phi_2(f) - \left(\delta \frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right) D\Phi_2(f;h)}{\Phi_2^2(f)}$$

$$R\Phi_{(j)}(f;h) \equiv -\frac{\Phi_2(f)R\Phi_{1,(j)}(f;h) + \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right) R\Phi_2(f;h)}{\Phi_2^2(f)}$$

$$- \frac{\Delta\Phi_2(f;h)\left[\left(\delta\frac{\partial \tilde{h}(z)}{\partial z_j} - \Delta\Phi_{1,(j)}(f;h)\right)\Phi_2(f) - \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)\Delta\Phi_2(f;h)\right]}{\Phi_2(f+h)\Phi_2^2(f)}$$

$$D\Psi_{(j)}(f;h) \equiv \frac{\left(D\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j} - \int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds\right) \times f(y,z) - \left(\Lambda(f)\frac{\partial \tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f)\right) \times h(y,z)}{f^2(y,z)}$$

$$R\Psi_{(j)}(f;h) \equiv \frac{\left(R\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\left(D\Lambda(f;h) + R\Lambda(f;h)\right)\right)}{f(y,z)}$$

$$- h(y,z) \times \left[\frac{\Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j}f(y,z) + \Delta\Lambda(f;h)\left(\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\right)f(y,z)}{\left(f(y,z) + h(y,z)\right)f^2(y,z)}\right.$$

$$\left. - \frac{\Lambda(f)\frac{\partial \tilde{f}(z)}{\partial z_j}h(y,z) + \Delta\Psi_{1,(j)}(f;h)f(y,z) - \Psi_{1,(j)}(f;h)h(y,z)}{\left(f(y,z) + h(y,z)\right)f^2(y,z)}\right]$$

$$D\tilde{\Psi}(f;h) \equiv \frac{\left(D\tilde{\Lambda}(f;h)\frac{\partial \tilde{f}_W(w)}{\partial w} + \tilde{\Lambda}(f)\frac{\partial \tilde{h}_W(w)}{\partial w} - \int_{-\infty}^{x}\frac{\partial \tilde{h}(s,w)}{\partial w}ds\right) \times \tilde{f}(x,w) - \left(\tilde{\Lambda}(f)\frac{\partial \tilde{f}_W(w)}{\partial w} - \tilde{\Psi}_1(f)\right) \times \tilde{h}(x,w)}{\tilde{f}^2(x,w)}$$

$$R\tilde{\Psi}(f;h) \equiv \frac{\left(R\tilde{\Lambda}(f;h)\frac{\partial \tilde{f}_W(w)}{\partial w} + \frac{\partial \tilde{h}_W(w)}{\partial w}\left(D\tilde{\Lambda}(f;h) + R\tilde{\Lambda}(f;h)\right)\right)}{\tilde{f}(x,w)}$$

$$- \tilde{h}(x,w) \times \left[\frac{\tilde{\Lambda}(f)\frac{\partial \tilde{h}_W(w)}{\partial w}\tilde{f}(x,w) + \Delta\tilde{\Lambda}(f;h)\left(\frac{\partial \tilde{f}_W(w)}{\partial w} + \frac{\partial \tilde{h}_W(w)}{\partial w}\right)\tilde{f}(x,w)}{\left(\tilde{f}(x,w) + \tilde{h}(x,w)\right)\tilde{f}^2(x,w)}\right.$$

$$\left. - \frac{\tilde{\Lambda}(f)\frac{\partial \tilde{f}_W(w)}{\partial w}\tilde{h}(x,w) + \Delta\tilde{\Psi}_{1,(j)}(f;h)\tilde{f}(x,w) - \tilde{\Psi}_1(f;h)\tilde{h}(x,w)}{\left(\tilde{f}(x,w) + \tilde{h}(x,w)\right)\tilde{f}^2(x,w)}\right],$$

with $\Delta\Phi_l(f;h) \equiv D\Phi_l(f;h) + R\Phi_l(f;h)$ *for* $l = 1,(j)$ *and* $l = 2$, $\Delta\Psi_{1,(j)}(f;h) \equiv D\Psi_{1,(j)}(f;h) + R\Psi_{1,(j)}(f;h)$ *and* $\Delta\tilde{\Psi}_1(f;h) \equiv D\tilde{\Psi}_1(f;h) + R\tilde{\Psi}_1(f;h)$. *Moreover, for some* $c_1, c_2, c_3 < \infty$,

$$\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|D\Phi_{(j)}(f;h)\right| \leq c_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + c_1 \sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|$$

$$\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|R\Phi_{(j)}(f;h)\right| \leq c_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + c_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial y}\right|$$

$$+ c_1 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|$$

$$\sup_{(y,z)\in\mathbb{S}_\tau}\left|D\Psi_{(j)}(f;h)\right| \leq c_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + c_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|$$

$$\sup_{(y,z)\in\mathbb{S}_\tau}\left|R\Psi_{(j)}(f;h)\right| \leq c_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + c_2 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|$$

$$\sup_{(y,z)\in\mathbb{S}_\tau}\left|D\tilde{\Psi}(f;h)\right| \le c_3 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + c_3 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_{j+1}}\right|$$

$$\sup_{(y,z)\in\mathbb{S}_\tau}\left|R\tilde{\Psi}(f;h)\right| \le c_3 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + c_3 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_{j+1}}\right|.$$

Proof. By Lemma 1.6,

$$\Phi_{(j)}(f) = \frac{\delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)}{\Phi_2(f)}$$
$$\Psi_{(j)}(f) = \frac{\Lambda(f)\frac{\partial\tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f)}{f(y,z)} \text{ and}$$
$$\tilde{\Psi}(f) = \frac{\tilde{\Lambda}(f)\frac{\partial\tilde{f}_W(w)}{\partial w} - \tilde{\Psi}_1(f)}{\tilde{f}(x,w)}$$

For all $h \in \mathrm{F}$,

$$\begin{aligned}
\Phi_{(j)}(f+h) - \Phi_{(j)}(f) &= \frac{\delta\frac{\partial\tilde{f}(z)}{\partial z_j} + \delta\frac{\partial\tilde{h}(z)}{\partial z_j} - \Phi_{1,(j)}(f+h)}{\Phi_2(f+h)} - \frac{\delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)}{\Phi_2(f)}\\
&\equiv \frac{N_1'}{D_1'} - \frac{N_1}{D_1}\\
\Psi_{(j)}(f+h) - \Psi_{(j)}(f) &= \frac{\Lambda(f+h)\left(\frac{\partial\tilde{f}(z)}{\partial z_j} + \frac{\partial\tilde{h}(z)}{\partial z_j}\right) - \Psi_{1,(j)}(f+h)}{f(y,z)+h(y,z)} - \frac{\Lambda(f)\frac{\partial\tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f)}{f(y,z)}\\
&\equiv \frac{N_2'}{D_2'} - \frac{N_2}{D_2}\\
\tilde{\Psi}(f+h) - \tilde{\Psi}(f) &= \frac{\tilde{\Lambda}(f+h)\left(\tilde{f}_W(w) + \tilde{h}_W(w)\right) - \tilde{\Psi}_1(f+h)}{\tilde{f}(x,w)+\tilde{h}(x,w)} - \frac{\tilde{\Lambda}(f)\tilde{f}_W(w) - \tilde{\Psi}_1(f)}{\tilde{f}(x,w)}\\
&\equiv \frac{N_3'}{D_3'} - \frac{N_3}{D_3},
\end{aligned}$$

where I denote

$$N_1' = \delta\frac{\partial\tilde{f}(z)}{\partial z_j} + \delta\frac{\partial\tilde{h}(z)}{\partial z_j} - \Phi_{1,(j)}(f+h)$$
$$N_1 = \delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)$$

$$
\begin{aligned}
N_2' &= \Lambda(f+h)\left(\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\right) - \Psi_{1,(j)}(f+h) \\
N_2 &= \Lambda(f)\frac{\partial \tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f) \\
D_1' &= \Phi_2(f+h) \\
D_1 &= \Phi_2(f) \\
D_2' &= f(y,z) + h(y,z) \\
D_2 &= f(y,z).
\end{aligned}
$$

Note that

$$
\begin{aligned}
N_1' - N_1 &= \delta\frac{\partial \tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h) - R\Phi_{1,(j)}(f;h) \equiv DN_1 + RN_1 \\
N_2' - N_2 &= (\Lambda(f+h) - \Lambda(f))\left(\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\right) + \Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j} - \int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds \\
&= (D\Lambda(f;h) + R\Lambda(f;h))\left(\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\right) + \Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j} - \int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds \\
&\equiv DN_2 + RN_2 \\
D_1' - D_1 &= D\Phi_2(f;h) + R\Phi_2(f;h) \\
D_2' - D_2 &= h(y,z),
\end{aligned}
$$

where

$$
\begin{aligned}
DN_1 &\equiv \delta\frac{\partial \tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h) \\
RN_1 &\equiv -R\Phi_{1,(j)}(f;h) \\
DN_2 &\equiv D\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j} - \int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds \\
RN_2 &\equiv R\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\left(D\Lambda(f;h) + R\Lambda(f;h)\right).
\end{aligned}
$$

I will make use of the equation:

$$\begin{aligned}\frac{N_j'}{D_j'} - \frac{N_j}{D_j} &= \frac{N_j' D_j - N_j D_j'}{D_j^2} - \frac{(D_j' - D_j)(N_j' D_j - N_j D_j')}{D_j' D_j^2} \\ &= \frac{(N_j' - RN_j)D_j - N_j\left(D_j' - RD_j\right)}{D_j^2} \\ &\quad + \frac{D_j \times RN_j - N_j \times RD_j}{D_j^2} - \frac{(D_j' - D_j)(N_j' D_j - N_j D_j')}{D_j' D_j^2}.\end{aligned}$$

Then

$$\begin{aligned}&\Phi_{(j)}(f+h) - \Phi_{(j)}(f) \\ &= \frac{\left(\delta\frac{\partial \tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h)\right)\Phi_2(f) - \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)D\Phi_2(f;h)}{\Phi_2^2(f)} \\ &\quad - \frac{\Phi_2(f)R\Phi_{1,(j)}(f;h) + \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)R\Phi_2(f;h)}{\Phi_2^2(f)} \\ &\quad - \frac{\big(D\Phi_2(f;h) + R\Phi_2(f;h)\big)\left(\left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} + \delta\frac{\partial \tilde{h}(z)}{\partial z_j} - \Phi_{1,(j)}(f+h)\right)\Phi_2(f) - \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)\Phi_2(f+h)\right)}{\Phi_2(f+h)\Phi_2^2(f)}.\end{aligned}$$

$$\begin{aligned}&\Psi_{(j)}(f+h) - \Psi_{(j)}(f) \\ &= \frac{DN_2 \times D_2 - N_2 \times h(y,z)}{D_2^2} + \frac{D_2 \times RN_2}{D_2^2} - \frac{h(y,z)(N_2' D_2 - N_2 D_2')}{D_2' D_2^2} \\ &= \frac{\left(D\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \Lambda(f)\frac{\partial \tilde{h}(z)}{\partial z_j} - \int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds\right) \times f(y,z) - \left(\Lambda(f)\frac{\partial \tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f)\right) \times h(y,z)}{f^2(y,z)} \\ &\quad + \frac{\left(R\Lambda(f;h)\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\left(D\Lambda(f;h) + R\Lambda(f;h)\right)\right)}{f(y,z)} \\ &\quad - \frac{h(y,z)\left(\left(\Lambda(f+h)\left(\frac{\partial \tilde{f}(z)}{\partial z_j} + \frac{\partial \tilde{h}(z)}{\partial z_j}\right) - \Psi_{1,(j)}(f+h)\right)f(y,z) - \left(\Lambda(f)\frac{\partial \tilde{f}(z)}{\partial z_j} - \Psi_{1,(j)}(f)\right)\left(f(y,z) + h(y,z)\right)\right)}{\left(f(y,z) + h(y,z)\right)f^2(y,z)}.\end{aligned}$$

Denote $D\Phi_{(j)}(f;h)$, $D\Psi_{(j)}(f;h)$, $R\Phi_{(j)}(f;h)$, and $R\Psi_{(j)}(f;h)$ by

$$\begin{aligned}D\Phi_{(j)}(f;h) &\equiv \frac{\left(\delta\frac{\partial \tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h)\right)\Phi_2(f) - \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)D\Phi_2(f;h)}{\Phi_2^2(f)} \\ R\Phi_{(j)}(f;h) &\equiv -\frac{\Phi_2(f)R\Phi_{1,(j)}(f;h) + \left(\delta\frac{\partial \tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)R\Phi_2(f;h)}{\Phi_2^2(f)}\end{aligned}$$

$$
-\frac{\left(D\Phi_2(f;h)+R\Phi_2(f;h)\right)\left(\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j}+\delta\frac{\partial\tilde{h}(z)}{\partial z_j}-\Phi_{1,(j)}(f+h)\right)\Phi_2(f)-\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j}-\Phi_{1,(j)}(f)\right)\Phi_2(f+h)\right)}{\Phi_2(f+h)\Phi_2^2(f)}
$$

$$
=-\frac{\Phi_2(f)R\Phi_{1,(j)}(f;h)+\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j}-\Phi_{1,(j)}(f)\right)R\Phi_2(f;h)}{\Phi_2^2(f)}
$$

$$
-\frac{\Delta\Phi_2(f;h)\left[\left(\delta\frac{\partial\tilde{h}(z)}{\partial z_j}-\Delta\Phi_{1,(j)}(f;h)\right)\Phi_2(f)-\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j}-\Phi_{1,(j)}(f)\right)\Delta\Phi_2(f;h)\right]}{\Phi_2(f+h)\Phi_2^2(f)}
$$

$$
D\Psi_{(j)}(f;h)\equiv\frac{\left(D\Lambda(f;h)\frac{\partial\tilde{f}(z)}{\partial z_j}+\Lambda(f)\frac{\partial\tilde{h}(z)}{\partial z_j}-\int_{-\infty}^{y}\frac{\partial h(s,z)}{\partial z_j}ds\right)\times f(y,z)-\left(\Lambda(f)\frac{\partial\tilde{f}(z)}{\partial z_j}-\Psi_{1,(j)}(f)\right)\times h(y,z)}{f^2(y,z)}
$$

$$
R\Psi_{(j)}(f;h)\equiv\frac{\left(R\Lambda(f;h)\frac{\partial\tilde{f}(z)}{\partial z_j}+\frac{\partial\tilde{h}(z)}{\partial z_j}\left(D\Lambda(f;h)+R\Lambda(f;h)\right)\right)}{f(y,z)}
$$

$$
-\frac{h(y,z)\left(\left(\Lambda(f+h)\left(\frac{\partial\tilde{f}(z)}{\partial z_j}+\frac{\partial\tilde{h}(z)}{\partial z_j}\right)-\Psi_{1,(j)}(f+h)\right)f(y,z)-\left(\Lambda(f)\frac{\partial\tilde{f}(z)}{\partial z_j}-\Psi_{1,(j)}(f)\right)\left(f(y,z)+h(y,z)\right)\right)}{\left(f(y,z)+h(y,z)\right)f^2(y,z)}
$$

$$
=\frac{\left(R\Lambda(f;h)\frac{\partial\tilde{f}(z)}{\partial z_j}+\frac{\partial\tilde{h}(z)}{\partial z_j}\left(D\Lambda(f;h)+R\Lambda(f;h)\right)\right)}{f(y,z)}
$$

$$
-h(y,z)\times\left[\frac{\Lambda(f)\frac{\partial\tilde{h}(z)}{\partial z_j}f(y,z)+\Delta\Lambda(f;h)\left(\frac{\partial\tilde{f}(z)}{\partial z_j}+\frac{\partial\tilde{h}(z)}{\partial z_j}\right)f(y,z)}{\left(f(y,z)+h(y,z)\right)f^2(y,z)}\right.
$$

$$
\left.-\frac{\Lambda(f)\frac{\partial\tilde{f}(z)}{\partial z_j}h(y,z)+\Delta\Psi_{1,(j)}(f;h)f(y,z)-\Psi_{1,(j)}(f;h)h(y,z)}{\left(f(y,z)+h(y,z)\right)f^2(y,z)}\right],
$$

where $\Delta\Phi_l(f;h)\equiv D\Phi_l(f;h)+R\Phi_l(f;h)$ for $l=1,(j)$ and $l=2$, and $\Delta\Psi_{1,(j)}(f;h)\equiv D\Psi_{1,(j)}(f;h)+R\Psi_{1,(j)}(f;h)$. Then

$$
\Phi_{(j)}(f+h)-\Phi_{(j)}(f)=D\Phi_{(j)}(f;h)+R\Phi_{(j)}(f;h)
$$

$$
\Psi_{(j)}(f+h)-\Psi_{(j)}(f)=D\Psi_{(j)}(f;h)+R\Psi_{(j)}(f;h).
$$

By the same logic, I can write

$$
\tilde{\Psi}(f+h)-\tilde{\Psi}(f)=D\tilde{\Psi}(f;h)+R\tilde{\Psi}(f;h),
$$

where $D\tilde{\Psi}(f;h)$ and $R\tilde{\Psi}(f;h)$ are as defined in the Lemma.

By compactness of $\mathbb{S}_{(Y,Z)}$ and continuity of f, $\partial f/\partial z$, I have that there exist finite constant d_9 such that

$$\sup_{(\delta,z)\in[0,1]\times\mathbb{S}_\tau}\left|\Phi_{1,(j)}(f)\right| = \sup_{(\delta,z)\in[0,1]\times\mathbb{S}_\tau}\left|\int_{-\infty}^{\alpha(f)}\frac{\partial f(y,z)}{\partial z_j}dy\right| \leq d_9 \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial f(y,z)}{\partial z_j}\right| < \infty$$

$$\sup_{(\delta,z)\in[0,1]\times\mathbb{S}_\tau}|\Phi_2(f)| \leq \sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|f(y,z)| < \infty \tag{1.24}$$

Also, by the definition of $\mathbb{S}_\tau$,

$$\inf_{(\delta,z)\in\mathcal{S}_\tau}\Phi_2^2(f) \geq \inf_{(y,z)\in\mathbb{S}_\tau} f^2(y,z) > \tau^2. \tag{1.25}$$

For $\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \leq \min\{\tau/2, 1\}$, I have

$$\begin{aligned}\inf_{(y,z)\in\mathbb{S}_\tau}\Phi_2(f+h) &= \inf_{(y,z)\in\mathbb{S}_\tau} f(\alpha(f+h),z) + h(\alpha(f+h),z)\\ &> \tau - \frac{\tau}{2} = \frac{\tau}{2}.\end{aligned} \tag{1.26}$$

By Lemma 1.9, 1.10 there exist finite constants $d_{10}, d_{11}, d_{12}, d_{13}, d_{14}, d_{15}$ such that

$$\begin{aligned}
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\delta\frac{\partial\tilde{h}(z)}{\partial z_j} - D\Phi_{1,(j)}(f;h)\right| \leq d_{10}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + d_{10}\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|\\
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)D\Phi_2(f;h)\right| \leq d_{11}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\\
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_{1,(j)}(f)\right)R\Phi_2(f;h)\right| \leq d_{12}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + d_{12}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial y}\right|\\
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\Delta\Phi_2(f;h)\right| \leq d_{13}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\\
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\delta\frac{\partial\tilde{h}(z)}{\partial z_j} - \Delta\Phi_{1,(j)}(f;h)\right| \leq d_{14}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + d_{14}\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|\\
&\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|\left(\delta\frac{\partial\tilde{f}(z)}{\partial z_j} - \Phi_1(f)\right)\Delta\Phi_2(f;h)\right| \leq d_{15}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|.
\end{aligned} \tag{1.27}$$

Combining results from (1.25), (1.26), and (1.27), I have that for

$\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| \leq \min\{\tau/2, 1\}$ and some finite constants d_{16}, d_{17}

$$\begin{aligned}
\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|D\Phi_{(j)}(f;h)\right| &\leq d_{16}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + d_{16}\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right| \\
\sup_{(\delta,z)\in\mathcal{S}_\tau}\left|R\Phi_{(j)}(f;h)\right| &\leq d_{17}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + d_{17}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial y}\right| \\
&\quad + d_{17}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\sup_{(y,z)\in\mathbb{S}_{Y,Z}}\left|\frac{\partial h(y,z)}{\partial z_j}\right|.
\end{aligned}$$

Let $c_1 \equiv \max\{d_{16}, d_{17}\}$ gives the desired result. By the same logic, it can be shown that

$$\begin{aligned}
\sup_{(y,z)\in\mathbb{S}_\tau}\left|D\Psi_{(j)}(f;h)\right| &\leq d_{18}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + d_{18}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_j}\right| \\
\sup_{(y,z)\in\mathbb{S}_\tau}\left|R\Psi_{(j)}(f;h)\right| &\leq d_{19}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + d_{19}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_j}\right| \\
\sup_{(y,z)\in\mathbb{S}_\tau}\left|D\tilde{\Psi}(f;h)\right| &\leq d_{20}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h| + d_{20}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_{j+1}}\right| \\
\sup_{(y,z)\in\mathbb{S}_\tau}\left|R\tilde{\Psi}(f;h)\right| &\leq d_{21}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|^2 + d_{21}\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}|h|\sup_{(y,z)\in\mathbb{S}_{(Y,Z)}}\left|\frac{\partial h(y,z)}{\partial z_{j+1}}\right|.
\end{aligned}$$

Let $c_2 \equiv \max\{d_{18}, d_{19}\}$, $c_3 \equiv \max\{d_{20}, d_{21}\}$ gives the desired result. □

Lemma 1.12. *For any value $(y,z) \in \mathbb{S}_{(Y,Z)}$, and any value $\delta \in [0,1]$, define functional Ξ and $\tilde{\Xi}$ on* F *by $\Xi(g) \equiv \Phi_{(1)}(g) + \frac{\Phi_{(2)}(g)}{\tilde{\Psi}(g)}$ and $\tilde{\Xi}(g) \equiv \Psi_{(1)}(g) + \frac{\Psi_{(2)}(g)}{\tilde{\Psi}(g)}$. Then I have that*

$$\begin{aligned}
\Xi(f+h) - \Xi(f) &= D\Xi(f;h) + R\Xi(f;h) \\
\tilde{\Xi}(f+h) - \tilde{\Xi}(f) &= D\tilde{\Xi}(f;h) + R\tilde{\Xi}(f;h),
\end{aligned}$$

and that there exist constants e_1, e_2 such that

$$\begin{aligned}
\sup_{(\delta,x,w)\in\mathcal{S}_\tau}\left|D\Xi(f;h)\right| &\leq \frac{e_1}{\tau^2}\|h\| + e_1\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + \frac{e_1}{\tau^2}\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right| \\
\sup_{(\delta,x,w)\in\mathcal{S}_\tau}\left|R\Xi(f;h)\right| &\leq \frac{e_1}{\tau^3}\|h\|^2 + \frac{e_1}{\tau^3}\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right| + \frac{e_1}{\tau}\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial y}\right| \\
&\quad + e_1\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + \frac{e_1}{\tau^3}\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|^2
\end{aligned}$$

$$\sup_{(y,x,w)\in\mathbb{S}_\tau}\left|D\tilde{\Xi}(f;h)\right| \le \frac{e_2}{\tau^2}\|h\| + e_2 \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + \frac{e_2}{\tau^2}\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|$$

$$\sup_{(y,x,w)\in\mathbb{S}_\tau}\left|R\tilde{\Xi}(f;h)\right| \le \frac{e_2}{\tau^3}\|h\|^2 + \frac{e_2}{\tau^3}\|h\| \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|$$

$$+ e_2\|h\| \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + \frac{e_2}{\tau^3}\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|^2.$$

Proof.

$$\begin{aligned}
&\Xi(f+h)-\Xi(f)\\
&=\left(\Phi_{(1)}(f+h)+\frac{\Phi_{(2)}(f+h)}{\tilde{\Psi}_{(1)}(f+h)}\right)-\left(\Phi_{(1)}(f)+\frac{\Phi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f)}\right)\\
&=\Delta\Phi_{(1)}(f;h)+\frac{D\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-D\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}\\
&\quad-\frac{D\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)-D\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)}{\tilde{\Psi}^2_{(1)}(f)\tilde{\Psi}_{(1)}(f+h)}\\
&\quad+\frac{R\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-R\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f+h)\tilde{\Psi}_{(1)}(f)}
\end{aligned}$$

$$\begin{aligned}
&\tilde{\Xi}(f+h)-\tilde{\Xi}(f)\\
&=\left(\Psi_{(1)}(f+h)+\frac{\Psi_{(2)}(f+h)}{\tilde{\Psi}_{(1)}(f+h)}\right)-\left(\Psi_{(1)}(f)+\frac{\Psi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f)}\right)\\
&=\Delta\Psi_{(1)}(f;h)+\frac{D\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-D\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}\\
&\quad-\frac{D\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)-D\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)}{\tilde{\Psi}^2_{(1)}(f)\tilde{\Psi}_{(1)}(f+h)}\\
&\quad+\frac{R\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-R\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f+h)\tilde{\Psi}_{(1)}(f)}
\end{aligned}$$

Define

$$\begin{aligned}
D\Xi(f;h)&\equiv D\Phi_{(1)}(f;h)+\frac{D\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-D\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}\\
R\Xi(f;h)&\equiv R\Phi_{(1)}(f;h)-\frac{D\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)-D\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)}{\tilde{\Psi}^2_{(1)}(f)\tilde{\Psi}_{(1)}(f+h)}\\
&\quad+\frac{R\Phi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-R\tilde{\Psi}_{(1)}(f;h)\Phi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f+h)\tilde{\Psi}_{(1)}(f)}\\
D\tilde{\Xi}(f;h)&\equiv D\Psi_{(1)}(f;h)+\frac{D\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)-D\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}
\end{aligned}$$

$$R\tilde{\Xi}(f;h) \equiv R\Psi_{(1)}(f;h) - \frac{D\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f)\Delta\tilde{\Psi}_{(1)}(f;h) - D\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)\Delta\tilde{\Psi}_{(1)}(f;h)}{\tilde{\Psi}^2_{(1)}(f)\tilde{\Psi}_{(1)}(f+h)}$$
$$+ \frac{R\Psi_{(2)}(f;h)\tilde{\Psi}_{(1)}(f) - R\tilde{\Psi}_{(1)}(f;h)\Psi_{(2)}(f)}{\tilde{\Psi}_{(1)}(f+h)\tilde{\Psi}_{(1)}(f)}.$$

Then

$$\Xi(f+h) - \Xi(f) = D\Xi(f;h) + R\Xi(f;h)$$
$$\tilde{\Xi}(f+h) - \tilde{\Xi}(f) = D\tilde{\Xi}(f;h) + R\tilde{\Xi}(f;h).$$

Note that there exist some constant d_{14} such that

$$\sup_{(\delta,x,w)\in\mathcal{S}_\tau} \left|\Phi_{(2)}(f)\right| = \sup_{(\delta,x,w)\in\mathcal{S}_\tau} \left| \frac{\delta\frac{\partial\tilde{f}(x,w)}{\partial w} - \Phi_{1,(2)}(f)}{\Phi_2(f)} \right| \leq \frac{\sup_{(\delta,x,w)\in\mathcal{S}_\tau}\left|\delta\frac{\partial\tilde{f}(x,w)}{\partial w} - \Phi_{1,(2)}(f)\right|}{\inf_{(\delta,x,w)\in\mathcal{S}_\tau}\left|f(\alpha(f),x,w)\right|} \leq d_{14}$$
$$\sup_{(y,x,w)\in\mathbb{S}_\tau} \left|\Psi_{(2)}(f)\right| = \sup_{(y,x,w)\in\mathbb{S}_\tau} \left| \frac{\Lambda(f)\frac{\partial\tilde{f}(x,w)}{\partial w} - \Psi_{1,(2)}(f)}{f(y,x,w)} \right| \leq \frac{\sup_{(y,x,w)\in\mathbb{S}_\tau}\left|\Lambda(f)\frac{\partial\tilde{f}(x,w)}{\partial w} - \Psi_{1,(2)}(f)\right|}{\inf_{(y,x,w)\in\mathbb{S}_\tau}\left|f(y,x,w)\right|} \leq d_{14}.$$

Application of Lemma 1.11 yields the desired result. □

Lemma 1.13. *In Lemma 1.6, let the dimension of vector $L = 2$. For any value $(y,z) \in \mathbb{S}_{(Y,Z)}$, and any value $\delta \in [0,1]$, define functional $\tilde{\Gamma}$ on* F *by*

$$\tilde{\Gamma}(g) \equiv \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))\Xi(g)\frac{\int g(y,x.w)dy}{\int\int g(y,x,w)dydw}d\delta dw,$$

where $\omega()$ is a known[4] *compactly supported weighting function as defined in (1.19), and $s()$ and $r()$ are as defined in (1.5) and (1.6). Then I have that*

$$\tilde{\Gamma}(f+h) - \tilde{\Gamma}(f) = D\tilde{\Gamma}(f;h) + R\tilde{\Gamma}(f;h),$$

[4] The same logic carries over if $\omega()$ is estimated. Take the example mentioned in the main text for instance. When $\tilde{\omega}(x,w^*)$ is a constant equal to $\left[(\tau_u - \tau_l)\int_{\underline{w}^*}^{\bar{w}^*} f_{W^*|X=x}(w^*)dw^*\right]^{-1}$, and $f_{W^*|X=x}(w^*)$ is estimated by a plug-in estimator, the same argument in this proof will follow by replacing $\int\int f(y,x,w)dydw$ with $\int_{\underline{w}^*}^{\bar{w}^*}\int f(y,x,w)dydw$ in the definition of Γ.

and there exists a constant e_4 such that

$$\sup_{x\in\mathbb{S}_X}\left|D\tilde{\Gamma}(f;h)\right| \leq e_4\|h\| + e_4 \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + e_4 \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|$$

$$\sup_{x\in\mathbb{S}_X}\left|R\tilde{\Gamma}(f;h)\right| \leq e_4\|h\|^2 + e_4\|h\| \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right| + e_4\|h\| \sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial y}\right|.$$

Proof. Define functional Γ on F by $\Gamma(g) \equiv \Xi(g)\frac{\int g(y,x.w)dy}{\int\int g(y,x,w)dydw}$, then

$$\begin{aligned}\Gamma(f+h) - \Gamma(f) &= \frac{\Xi(f+h)\int f(y,x.w) + h(y,x.w)dy}{\int\int f(y,x,w) + h(y,x.w)dydw} - \frac{\Xi(f)\int f(y,x.w)dy}{\int\int f(y,x,w)dydw} \\ &\equiv \frac{N'}{D'} - \frac{N}{D}\end{aligned}$$

where I've defined

$$\begin{aligned}N' &\equiv \Xi(f+h)\int f(y,x.w) + h(y,x.w)dy \\ N &\equiv \Xi(f)\int f(y,x.w)dy \\ D' &\equiv \int\int f(y,x,w) + h(y,x.w)dydw \\ D &\equiv \int\int f(y,x,w)dydw\end{aligned}$$

Note that

$$\begin{aligned}N' - N &= D\Xi(f;h)\int f(y,x,w)dy + \Xi(f)\int h(y,x,w)dy \\ &\quad + D\Xi(f;h)\int h(y,x,w)dy + R\Xi(f;h)\int (f(y,x,w) + h(y,x,w))dy \\ &\equiv DN + RN \\ D' - D &= \int\int h(y,x,w)dydw\end{aligned}$$

where $DN \equiv D\Xi(f;h)\int f(y,x,w)dy + \Xi(f)\int h(y,x,w)dy$ and $RN \equiv D\Xi(f;h)\int h(y,x,w)dy + R\Xi(f;h)\int (f(y,x,w) + h(y,x,w))dy$, and $D\Xi(f;h)$ and

$R\Xi(f;h)$ are as defined in Lemma 1.12. Then by the same logic as Lemma 1.11,

$$\begin{aligned}
&\Gamma(f+h)-\Gamma(f)\\
&=\frac{\left(D\Xi(f;h)\int f(y,x,w)dy+\Xi(f)\int h(y,x,w)dy\right)\int\int f(y,x,w)dydw-\Xi(f)\int f(y,x.w)dy\left(\int\int h(y,x.w)dydw\right)}{\left(\int\int f(y,x,w)dydw\right)^2}\\
&+\frac{\int\int f(y,x,w)dydw\times\left(D\Xi(f;h)\int h(y,x,w)dy+R\Xi(f;h)\int(f(y,x,w)+h(y,x,w))dy\right)}{\left(\int\int f(y,x,w)dydw\right)^2}\\
&-\int\int h(y,x,w)dydw\times\\
&\frac{\left(\left(\Xi(f+h)\int f(y,x.w)+h(y,x.w)dy\right)\int\int f(y,x,w)dydw-\Xi(f)\int f(y,x.w)dy\left(\int\int f(y,x,w)+h(y,x.w)dydw\right)\right)}{\left(\int\int f(y,x,w)+h(y,x.w)dydw\right)\left(\int\int f(y,x,w)dydw\right)^2}\\
&=D\Gamma(f;h)+R\Gamma(f;h).
\end{aligned}$$

where I've defined

$$\begin{aligned}
&D\Gamma(f;h)\equiv\frac{\left(D\Xi(f;h)\int f(y,x,w)dy+\Xi(f)\int h(y,x,w)dy\right)\int\int f(y,x,w)dydw-\Xi(f)\int f(y,x.w)dy\left(\int\int h(y,x.w)dydw\right)}{\left(\int\int f(y,x,w)dydw\right)^2}\\
&R\Gamma(f;h)\\
&\equiv+\frac{\int\int f(y,x,w)dydw\times\left(D\Xi(f;h)\int h(y,x,w)dy+R\Xi(f;h)\int(f(y,x,w)+h(y,x,w))dy\right)}{\left(\int\int f(y,x,w)dydw\right)^2}\\
&-\int\int h(y,x,w)dydw\times\\
&\frac{\left(\left(\Xi(f+h)\int f(y,x.w)+h(y,x.w)dy\right)\int\int f(y,x,w)dydw-\Xi(f)\int f(y,x.w)dy\left(\int\int f(y,x,w)+h(y,x.w)dydw\right)\right)}{\left(\int\int f(y,x,w)+h(y,x.w)dydw\right)\left(\int\int f(y,x,w)dydw\right)^2}.
\end{aligned}$$

By the boundedness of support, the definition of $\mathcal{S}_\tau$ and Lemma 1.12, there exists a constant e_3 such that

$$\begin{aligned}
\sup_{(\delta,x,w)\in\mathcal{S}_\tau}\left|D\Gamma(f;h)\right|&\le e_3\|h\|+e_3\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right|+e_3\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|\\
\sup_{(\delta,x,w)\in\mathcal{S}_\tau}\left|R\Gamma(f;h)\right|&\le e_3\|h\|^2+e_3\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|+e_3\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial y}\right|
\end{aligned}$$

Denote the support of $\omega(x,w^*,s(r(x,w^*),\delta))$ as $\mathcal{M}$, then by the argument stated at the beginning of Appendix A, there exists $\tau>0$ such that $\mathcal{M}\subset\mathcal{S}_\tau$ so that I have

$$\begin{aligned}
\sup_{(\delta,x,w)\in\mathcal{M}}\left|D\Gamma(f;h)\right|&\le e_3\|h\|+e_3\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right|+e_3\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|\\
\sup_{(\delta,x,w)\in\mathcal{M}}\left|R\Gamma(f;h)\right|&\le e_3\|h\|^2+e_3\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right|+e_3\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial y}\right|.
\end{aligned}$$

By definition, I have

$$
\begin{aligned}
&\int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))\Gamma(f+h)d\delta dw - \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))\Gamma(f)d\delta dw \\
&= \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))D\Gamma(f;h)d\delta dw + \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))R\Gamma(f;h)d\delta dw \\
&\equiv D\tilde{\Gamma}(f;h) + R\tilde{\Gamma}(f;h)
\end{aligned}
$$

where I've defined $D\tilde{\Gamma}(f;h) \equiv \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))D\Gamma(f;h)d\delta dw$ and $R\tilde{\Gamma}(f;h) \equiv \int\int_0^1 \omega(x,w^*,s(r(x,w^*),\delta))R\Gamma(f;h)d\delta dw$. By the compactness of $\mathcal{M}$, there exist a constant e_4 such that

$$
\begin{aligned}
&\sup_{x\in\mathbb{S}_X}\left|D\tilde{\Gamma}(f;h)\right| \leq e_4\|h\| + e_4\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial x}\right| + e_4\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right| \\
&\sup_{x\in\mathbb{S}_X}\left|R\tilde{\Gamma}(f;h)\right| \leq e_4\|h\|^2 + e_4\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial w}\right| + e_4\|h\|\sup_{(y,x,w)\in\mathbb{S}_{Y,X,W}}\left|\frac{\partial h(y,x,w)}{\partial y}\right|.
\end{aligned}
$$

□

1.7.2 Appendix B: Proofs of Theorems and Lemmas in the Main Text

Proof of Lemma 1.1. Since $W^* \perp (\epsilon,\eta)$, then $\epsilon \perp W^* \mid \eta$. Define $\delta = F_{\epsilon|\eta}(\epsilon)$. Then I can write $\epsilon = F^{-1}_{\epsilon|\eta}(\delta) =: s(\eta,\delta)$. By Assumption 1.8, s is strictly increasing in δ. To see that $\delta \perp (X,W^*)$ and is uniform (0,1), note that

$$
\begin{aligned}
F_{\delta|W^*=w^*,\eta=\bar{\eta}}(t) &= Pr\left(F_{\epsilon|\eta}(\epsilon) \leq t \mid W^*=w^*,\eta=\bar{\eta}\right) \\
&= Pr\left(\epsilon \leq F^{-1}_{\epsilon|\eta=\bar{\eta}}(t) \mid W^*=w^*,\eta=\bar{\eta}\right) \\
&= Pr\left(\epsilon \leq F^{-1}_{\epsilon|\eta=\bar{\eta}}(t) \mid \eta=\bar{\eta}\right) \\
&= F_{\epsilon|\eta=\bar{\eta}}(F^{-1}_{\epsilon|\eta=\bar{\eta}}(t)) = t,
\end{aligned}
$$

which says that the conditional distribution of δ given W^*,η is $U(0,1)$ regardless of the values of W^* and η. In other words, $\delta \sim U(0,1)$ and is independent of (W^*,η). Since X is a function of (W^*, η), δ is independent of X as well. □

Proof of Lemma 1.2. (i)First I show that $\mathbb{S}_{\epsilon|X=x} \subseteq \mathbb{S}_{\mathbf{s}|X=x}$. For any $\bar{\epsilon} \in \mathbb{S}_{\epsilon|X=x}$, let $\bar{\delta} \equiv F_{\epsilon|\eta=r(x,\bar{w}^*)}(\bar{\epsilon})$ for some $\bar{w}^* \in \mathbb{S}_{W^*|X=x}$. Then $\bar{\delta} \in [0,1]$ and thus $\bar{\epsilon} \equiv s(r(x,\bar{w}^*),\bar{\delta}) \in \mathbb{S}_{\mathbf{s}|X=x}$.

(ii) Then I show that $\mathbb{S}_{\mathbf{s}|X=x} \subseteq \mathbb{S}_{\epsilon|X=x}$. By definition of $\mathbb{S}_{\mathbf{s}|X=x}$, for each $\tilde{s} \in \mathbb{S}_{\mathbf{s}|X=x}$, there is some $(\tilde{w}^*, \tilde{\delta}) \in \mathbb{S}_{W^*|X=x} \times [0,1]$ such that $\tilde{s} = s(r(x,\tilde{w}^*),\tilde{\delta})$. Denote the upper and lower bound (potentially infinity) of $\mathbb{S}_{\epsilon|X=x}$ as $\bar{e}_x$ and $\underline{e}_x$, respectively. From Lemma 1.1, I know that $\epsilon \perp W^* \mid \eta$. Since X is a function of W^*, η, I have that $\epsilon \perp X \mid \eta$. Then if $\bar{e}_x \in \mathbb{S}_{\epsilon|X=x}$, then $Pr\left(\epsilon \leq \bar{e}_x \mid \eta = r(x,\tilde{w}^*)\right) = Pr\left(\epsilon \leq \bar{e}_x \mid \eta = r(x,\tilde{w}^*), X = x\right) = 1$, which implies $F^{-1}_{\epsilon|\eta=r(x,\tilde{w}^*)}(1) \leq \bar{e}_x$. If $\bar{e}_x \notin \mathbb{S}_{\epsilon|X=x}$, then $Pr\left(\epsilon < \bar{e}_x \mid \eta = r(x,\tilde{w}^*)\right) = Pr\left(\epsilon < \bar{e}_x \mid \eta = r(x,\tilde{w}^*), X = x\right) = 1$, which implies $F^{-1}_{\epsilon|\eta=r(x,\tilde{w}^*)}(1) < \bar{e}_x$. Similarly, $F^{-1}_{\epsilon|\eta=r(x,\tilde{w}^*)}(0) \geq \bar{e}_x$, and the inequality is strict if $\underline{e}_x \notin \mathbb{S}_{\epsilon|X=x}$ Since $\tilde{\delta} \in [0,1]$, by Assumption 1.8, $\underline{e}_x \leq F^{-1}_{\epsilon|\eta=r(x,\tilde{w}^*)}(\tilde{\delta}) \leq \bar{e}_x$, or equivalently, $\underline{e}_x \leq s(r(x,\tilde{w}^*),\tilde{\delta}) \leq \bar{e}_x$, and the inequalities are strict if $\underline{e}_x \notin \mathbb{S}_{\epsilon|X=x}$ or $\bar{e}_x \notin \mathbb{S}_{\epsilon|X=x}$, respectively. Thus $\tilde{s} \in \mathbb{S}_{\epsilon|X=x}$. Combining results from (i) and (ii) yields the desired conclusion. □

Proof of Lemma 1.3.

$$
\begin{aligned}
\delta &= F_{Y|X=x,W^*=w^*}\left(F^{-1}_{Y|X=x,W^*=w^*}(\delta)\right) \\
&= \frac{\int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} f_{Y,X,W^*}(y,x,w^*)dy}{\int_{-\infty}^{\infty} f_{Y,X,W^*}(y,x,w^*)dy} \\
\implies \delta f_{X,W^*}(x,w^*) &= \int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} f_{Y,X,W^*}(y,x,w^*)dy
\end{aligned}
$$

Taking derivatives on both sides with respect to w^* yields

$$
\begin{aligned}
&f_{Y,X,W^*}\left(F^{-1}_{Y|X=x,W^*=w^*}(\delta), x, w^*\right)\frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial w^*} + \int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} \frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial w^*}dy \\
&= \delta\frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*}
\end{aligned}
$$

This implies

$$\frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial w^*} = \frac{\delta \frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*} - \int_{-\infty}^{F^{-1}_{Y|X=x,W^*=w^*}(\delta)} \frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial w^*} dy}{f_{Y,X,W^*}\left(F^{-1}_{Y|X=x,W^*=w^*}(\delta), x, w^*\right)}.$$

The conclusion for $\frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial x}$ follows the same logic. □

Proof of Lemma 1.4. Assumption 1.5, 1.6, and 1.10 ensure the existence of

$$\begin{aligned}
&\phi_{f_{Y,X,W_2}(y,x,\cdot)}(t) = E\left[e^{\mathbf{i}tW_2} \mid Y=y, X=x\right] f_{Y,X}(y,x) \\
&\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) = \int e^{\mathbf{i}tw_2} \frac{\partial^{\lambda_2} f_{Y,X,W_2}(y,x,w_2)}{\partial y^{\lambda}_{2,1} \partial x^{\lambda}_{2,2}} dw_2 \quad \text{and thus} \\
&g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1) \equiv \int \frac{1}{h_1} K\left(\frac{\tilde{w}^* - w^*}{h_1}\right) g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,\tilde{w}^*) d\tilde{w}^* \\
&= \int \frac{1}{h_1} K\left(\frac{\tilde{w}^* - w^*}{h_1}\right) \left(\frac{1}{2\pi} \int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}t\tilde{w}^*} \frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) \phi_{W^*}(t)}{\phi_{W_2}(t)} dt\right) d\tilde{w}^*.
\end{aligned}$$

Denote $k_{h_1}(v) \equiv \frac{1}{h_1} K\left(\frac{v}{h_1}\right)$, then for all $x \in \mathbb{R}$, $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1)$ is the convolution between $k_{h_1}(\cdot)$ and $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,\cdot)$. By the convolution theorem, it's equal to the inverse Fourier transform of the product of the Fourier transformation of $k_{h_1}(\cdot)$ and the Fourier transformation of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,\cdot)$. The Fourier transformation of $k_{h_1}(\cdot)$ is $\phi_K(ht)$, and the Fourier transformation of $g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,\cdot)$ is $(-\mathbf{i}t)^{\lambda_1} \frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\phi_{W^*}(t)}{\phi_{W_2}(t)}$. □

Proof of Lemma 1.5. Write $\hat{\theta}(\zeta) \equiv \hat{E}\left[W_1 e^{\mathrm{i}\zeta W_2}\right]$, $\delta\hat{\theta}(t) \equiv \hat{\theta}(t) - \theta(t)$, $\delta\hat{\phi}_Z(t) = \hat{\phi}_Z(t) - \phi_Z(t)$ for random variable Z, and $\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) = \hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) - \phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)$.

Denote $q_{\lambda_2}(t) \equiv \frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)}{\phi_{W_2}(t)}$, $\hat{q}_{\lambda_2}(t) \equiv \frac{\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)}{\hat{\phi}_{W_2}(t)}$ and $\delta\hat{q}_{\lambda_2}(t) \equiv \hat{q}_{\lambda_2}(t) - q_{\lambda_2}(t)$, then $\delta\hat{q}_{\lambda_2}(t)$ can be written as

$$\delta\hat{q}_{\lambda_2}(t) = \left(\frac{\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)}{\phi_{W_2}(t)} - \frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\delta\hat{\phi}_{W_2}(t)}{\left(\phi_{W_2}(t)\right)^2}\right)\left(1 + \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\right)^{-1} \quad \text{or}$$

$$\delta\hat{q}_{\lambda_2}(t) = \delta_1\hat{q}_{\lambda_2}(t) + \delta_2\hat{q}_{\lambda_2}(t), \quad \text{with}$$

$$\delta_1 \hat{q}_{\lambda_2}(t) \equiv \frac{\delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)}{\phi_{W_2}(t)} - \frac{\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\delta\hat{\phi}_{W_2}(t)}{\left(\phi_{W_2}(t)\right)^2}$$

$$\delta_2 \hat{q}_{\lambda_2}(t) \equiv \frac{\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)}{\phi_{W_2}(t)} \left(\frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\right)^2 \left(1 + \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\right)^{-1} - \frac{\delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)}{\phi_{W_2}(t)} \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)} \left(1 + \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\right)^{-1}. \tag{1.28}$$

Similarly, I denote $q_{W_1}(\xi) \equiv \frac{\theta(\xi)}{\phi_{W_2}(\xi)}$, $\hat{q}_{W_1}(\xi) \equiv \frac{\hat{\theta}(\xi)}{\hat{\phi}_{W_2}(\xi)}$, and $\delta\hat{q}_{W_1}(\xi) \equiv \hat{q}_{W_1}(\xi) - q_{W_1}(\xi)$. Then $\delta\hat{q}_{W_1}(\xi)$ can be written as

$$\delta\hat{q}_{W_1}(\xi) = \left(\frac{\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)} - \frac{\theta(\xi)\delta\hat{\phi}_{W_2}(\xi)}{\left(\phi_{W_2}(\xi)\right)^2}\right)\left(1 + \frac{\delta\hat{\phi}_{W_2}(\xi)}{\phi_{W_2}(\xi)}\right)^{-1} \quad \text{or}$$

$$\delta\hat{q}_{W_1}(\xi) = \delta_1\hat{q}_{W_1}(\xi) + \delta_2\hat{q}_{W_1}(\xi), \quad \text{with}$$

$$\delta_1\hat{q}_{W_1}(\xi) \equiv \frac{\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)} - \frac{\theta(\xi)\delta\hat{\phi}_{W_2}(\xi)}{\left(\phi_{W_2}(\xi)\right)^2}$$

$$\delta_2\hat{q}_{W_1}(\xi) \equiv \frac{\theta(\xi)}{\phi_{W_2}(\xi)}\left(\frac{\delta\hat{\phi}_{W_2}(\xi)}{\phi_{W_2}(\xi)}\right)^2\left(1 + \frac{\delta\hat{\phi}_{W_2}(\xi)}{\phi_{W_2}(\xi)}\right)^{-1} - \frac{\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)}\frac{\delta\hat{\phi}_{W_2}(\xi)}{\phi_{W_2}(\xi)}\left(1 + \frac{\delta\hat{\phi}_{W_2}(\xi)}{\phi_{W_2}(\xi)}\right)^{-1}. \tag{1.29}$$

For $Q(t) = \int_0^t \frac{\mathbf{i}\theta(\xi)}{\phi_{W_2}(\xi)}d\xi$, $\delta\hat{Q}(t) = \int_0^t \frac{\mathbf{i}\hat{\theta}(\xi)}{\hat{\phi}_{W_2}(\xi)}d\xi - Q(t)$ and some random function $\delta\bar{Q}(t)$ such that $|\delta\bar{Q}(t)| \leq |\delta\hat{Q}(t)|$ for all t,

$$\exp\left(Q(t) + \delta\hat{Q}(t)\right) = \exp(Q(t))\left(1 + \delta Q(t) + \frac{1}{2}[\exp(\delta\bar{Q}(t))]\left(\delta\hat{Q}(t)\right)^2\right) \tag{1.30}$$

Then I can state a useful representation for

$$\begin{aligned}
&\frac{\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\hat{\phi}_{W^*}(t)}{\hat{\phi}_{W_2}(t)} \\
&= (q_{\lambda_2}(t) + \delta\hat{q}_{\lambda_2}(t))\exp(Q(t))\left(1 + \delta\hat{Q}(t) + \frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right) \\
&= q_{\lambda_2}(t)\exp(Q(t)) + q_{\lambda_2}(t)\exp(Q(t))\left(\delta\hat{Q}(t) + \frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \delta\hat{q}_{\lambda_2}(t)\exp(Q(t))\left(1 + \delta\hat{Q}(t) + \frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right) \\
= {} & q_{\lambda_2}(t)\exp(Q(t)) + q_{\lambda_2}(t)\exp(Q(t))\left[\int_0^t \mathbf{i}\delta_1\hat{q}_{W_1}(\xi)d\xi\right] + \delta_1\hat{q}_{\lambda_2}(t)\exp(Q(t)) \\
&+ q_{\lambda_2}(t)\exp(Q(t))\left[\int_0^t \mathbf{i}\delta_2\hat{q}_{W_1}(\xi)d\xi\right] + q_{\lambda_2}(t)\exp(Q(t))\left[\frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right] \\
&+ \delta_2\hat{q}_{\lambda_2}(t)\exp(Q(t)) + \delta\hat{q}_{\lambda_2}(t)\exp(Q(t))\left(\delta\hat{Q}(t) + \frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right). \qquad (1.31)
\end{aligned}
$$

Plug (1.31) into $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$, I get

$$
\begin{aligned}
&\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \\
= {} & Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + Rg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)
\end{aligned}
$$

where $Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is the term linear in $\delta\hat{\theta}(t)$, $\delta\hat{\phi}_{W_2}$, and $\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)$, and $Rg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is the higher-order remainder term. More specifically,

$$
\begin{aligned}
Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv {} & \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t) \\
& \left(q_{\lambda_2}(t)\exp(Q(t))\left[\int_0^t \mathbf{i}\delta_1\hat{q}_{W_1}(\xi)d\xi\right] + \delta_1\hat{q}_{\lambda_2}(t)\exp(Q(t))\right)dt \\
Rg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv {} & \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t) \\
& \left\{q_{\lambda_2}(t)\exp(Q(t))\left[\int_0^t \mathbf{i}\delta_2\hat{q}_{W_1}(\xi)d\xi\right]\right. \\
& + q_{\lambda_2}(t)\exp(Q(t))\left[\frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right] \\
& + \delta_2\hat{q}_{\lambda_2}(t)\exp(Q(t)) \\
& \left. + \delta\hat{q}_{\lambda_2}(t)\exp(Q(t))\left(\delta\hat{Q}(t) + \frac{1}{2}\exp\left(\delta\bar{Q}(t)\right)\left(\delta\hat{Q}(t)\right)^2\right)\right\}dt.
\end{aligned}
$$

Define $\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ Then I have the following expression:

$$
\begin{aligned}
&\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*) \\
= {} & B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) + R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)
\end{aligned}
$$

where

$$B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right] - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)$$
$$L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right]$$
$$R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv \hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h).$$

To see the explicit form of $B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$, note that

$$\begin{aligned}
&B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right] - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\\
&= g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*) + E\left[Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right]\\
&= g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\\
&\quad + \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t)\frac{E\left[\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right]\phi_{W^*}(t)}{\phi_{W_2}(t)}dt
\end{aligned}$$

To see the explicit form of $L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$, note that

$$\begin{aligned}
&L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \equiv \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - E\left[\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right]\\
&= Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - E\left[Dg_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right]\\
&= \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t)\\
&\quad\left(\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\left[\int_0^t\left(\frac{\mathbf{i}\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)} - \frac{\mathbf{i}\theta(\xi)\delta\hat{\phi}_{W_2}(\xi)}{\left(\phi_{W_2}(\xi)\right)^2}\right)d\xi\right] - \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right)dt\\
&\quad + \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t)\frac{\left(\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) - E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right]\right)\phi_{W^*}(t)}{\phi_{W_2}(t)}dt\\
&= \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_1t)\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\left[\int_0^t\left(\frac{\mathbf{i}\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)} - \frac{\mathbf{i}\theta(\xi)\delta\hat{\phi}_{W_2}(\xi)}{\left(\phi_{W_2}(\xi)\right)^2}\right)d\xi\right]dt\\
&+ \quad \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_1t)\\
&\times\left(\frac{\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) - E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right]}{\phi_{W_2}(t)}\phi_{W^*}(t) - \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right)dt
\end{aligned}$$

Then using the identity

$$\int_{-\infty}^{\infty}\int_{0}^{\xi} f(\xi,\zeta)d\zeta d\xi = \int_{0}^{\infty}\int_{\zeta}^{\infty} f(\xi,\zeta)d\xi d\zeta + \int_{-\infty}^{0}\int_{\zeta}^{-\infty} f(\xi,\zeta)d\xi d\zeta$$
$$:= \iint_{\zeta}^{\pm\infty} f(\xi,\zeta)d\xi d\zeta$$

for any absolutely integrable function f, I obtain

$$L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \tag{1.32}$$
$$= \bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$$
$$= \frac{1}{2\pi}\int\int_{\xi}^{\pm\infty} (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_1 t)\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)dt \left(\frac{\mathbf{i}\delta\hat{\theta}(\xi)}{\phi_{W_2}(\xi)} - \frac{\mathbf{i}\theta(\xi)\delta\hat{\phi}_{W_2}(\xi)}{\left(\phi_{W_2}(\xi)\right)^2}\right)d\xi$$
$$+ \frac{1}{2\pi}\int (-\mathbf{i}t)^{\lambda_1} e^{-\mathbf{i}tw^*}\phi_K(h_1 t)\times \tag{1.33}$$
$$\left(\frac{\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t) - E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right]}{\phi_{W_2}(t)}\phi_{W^*}(t) - \frac{\delta\hat{\phi}_{W_2}(t)}{\phi_{W_2}(t)}\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right)dt$$
$$= \int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(\hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}]\right)d\xi$$
$$+ \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(\hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}]\right)d\xi$$
$$+ \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(\hat{E}\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\right.$$
$$\left. - E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\right)d\xi$$
$$= \hat{E}\left[\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(W_1 e^{\mathbf{i}\xi W_2} - E[W_1 e^{\mathbf{i}\xi W_2}]\right)d\xi\right.$$
$$+ \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(e^{\mathbf{i}\xi W_2} - E[e^{\mathbf{i}\xi W_2}]\right)d\xi$$
$$+ \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\times\left(e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right.$$
$$\left.\left. - E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\right)d\xi\right]$$
$$= \hat{E}[l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h;Y,X,W_1,W_2)], \tag{1.34}$$

where

$$\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) = \frac{1}{2\pi}\frac{\mathbf{i}}{\phi_{W_2}(\xi)}\int_{\xi}^{\pm\infty}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_1t)\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)dt$$

$$\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) = -\frac{1}{2\pi}\frac{\mathbf{i}\theta(\xi)}{(\phi_{W_2}(\xi))^2}\int_{\xi}^{\pm\infty}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_1t)\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)dt$$
$$-\frac{1}{2\pi}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}\xi w^*}\phi_K(h_1\xi)\frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)}{\phi_{W_2}(\xi)}$$

$$\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) = \frac{1}{2\pi}(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}\xi w^*}\phi_K(h_1\xi)\frac{\phi_{W^*}(\xi)}{\phi_{W_2}(\xi)}.$$

□

Definition 2. *Write $f(t) \preceq g(t)$ for $f,g:\mathbb{R}\to\mathbb{R}$ when there is a constant $C>0$, independent of t, such that $f(t)\le Cg(t)$ for all $t\in\mathbb{R}$ (and similarly for $\succeq$). Write $a_n \preceq b_n$ for two sequences a_n, b_n if there exists a constant C independent of n such that $a_n \le Cb_n$ for all $n\in\mathbb{N}$*

Proof of Theorem 1.1. By Parseval's identity, I have

$$\begin{aligned}
&|g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)| \\
&= \left|\frac{1}{2\pi}\int(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\left(\phi_K(h_1t)-1\right)\frac{\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)}{\phi_{W_2}(t)}\phi_{W^*}(t)dt\right| \\
&\le \frac{1}{2\pi}\int|t|^{\lambda_1}\left|\phi_K(h_1t)-1\right|\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|dt \\
&= \frac{1}{2\pi}\int_{|h_1t|>\bar{\zeta}_K}|t|^{\lambda_1}\left|\phi_K(h_1t)-1\right|\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|dt \\
&\preceq \int_{|h_1t|>\bar{\zeta}_K}|t|^{\lambda_1}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|dt
\end{aligned}$$

where I have used Assumption 1.9 to ensure $\phi_K(t)=1$ for $|t|\le\bar{\zeta}_K$ and $\sup_\zeta|\phi_K(\zeta)|<\infty$. Then by Assumption 1.11 and Lemma 7 in Schennach (2004b),

$$|g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)| \preceq \int_{|h_1t|>\bar{\zeta}_K}|t|^{\lambda_1}(1+|t|)^{\gamma_\phi}\exp\left(\alpha_\phi|t|^{\beta_\phi}\right)dt$$

$$= O\left(\left(\frac{\bar{\zeta}_K}{h_1}\right)^{1+\gamma_\phi+\lambda_1} \exp\left(\alpha_\phi \left(\frac{\bar{\zeta}_K}{h_1}\right)^{\beta_\phi}\right)\right)$$
$$= O\left(\left(h_1^{-1}\right)^{1+\gamma_\phi+\lambda_1} \exp\left(\alpha_\phi \bar{\zeta}_K^{\beta_\phi} \left(h_1^{-1}\right)^{\beta_\phi}\right)\right).$$

Next I show that the second term in $B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)$ is of a smaller order. To do this, I first state a useful representation of $E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right]$

$$\begin{aligned}
&E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\xi)\right]\\
&= E\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]\\
&= \int\int e^{\mathbf{i}\xi\tilde{w}} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-\tilde{y}}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-\tilde{x}}{h_{2,2}}\right) f_{Y,X,W_2}(\tilde{y},\tilde{x},\tilde{w})d\tilde{y}d\tilde{x}d\tilde{w}\\
&= \int \phi_{f^{(0,0)}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi) \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-\tilde{y}}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-\tilde{x}}{h_{2,2}}\right) d\tilde{y}d\tilde{x}
\end{aligned}$$

If $\lambda_{2,1} = 1, \lambda_{2,2} = 0$,

$$\begin{aligned}
&E\left[\hat{\phi}_{f^{(1,0)}_{Y,X,W_2}(y,x,\cdot)}(\xi)\right]\\
&= -\int\left(\int \phi_{f^{(0,0)}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi)\frac{1}{h_{2,1}} dG_Y\left(\frac{y-\tilde{y}}{h_{2,1}}\right)\right)\frac{1}{h_{2,2}} G_X\left(\frac{x-\tilde{x}}{h_{2,2}}\right)d\tilde{x}\\
&= \int\left(-\phi_{f^{(0,0)}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi)\frac{1}{h_{2,1}} G_Y\left(\frac{\tilde{y}-y}{h_{2,1}}\right)\Big|_{-\infty}^{\infty}\right.\\
&\qquad\left. + \int \frac{1}{h_{2,1}} G_Y\left(\frac{y-\tilde{y}}{h_{2,1}}\right)\phi_{f^{(1,0)}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi)d\tilde{y}\right)\frac{1}{h_{2,2}} G_X\left(\frac{x-\tilde{x}}{h_{2,2}}\right)d\tilde{x}\\
&= \int\int \phi_{f^{(1,0)}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi)\frac{1}{h_{2,1}} G_Y\left(\frac{y-\tilde{y}}{h_{2,1}}\right)\frac{1}{h_{2,2}} G_X\left(\frac{x-\tilde{x}}{h_{2,2}}\right)d\tilde{y}d\tilde{x}
\end{aligned}$$

For $\lambda_{2,1} = 0, \lambda_{2,2} = 1$, I have similar results, so

$$E\left[\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\xi)\right] = \int\int \phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(\tilde{y},\tilde{x},\cdot)}(\xi)\frac{1}{h_{2,1}} G_Y\left(\frac{y-\tilde{y}}{h_{2,1}}\right)\frac{1}{h_{2,2}} G_X\left(\frac{x-\tilde{x}}{h_{2,2}}\right)d\tilde{y}d\tilde{x}.$$

Then note that

$$\begin{aligned}
&\left|E\left[\delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)\right]\frac{\phi_{W^*}(\xi)}{\phi_{W_2}(\xi)}\right| \\
&=\left|\int\int \phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(\tilde{y},\tilde{x},\cdot)}(\xi)\frac{1}{h_{2,1}}G_Y\left(\frac{y-\tilde{y}}{h_{2,1}}\right)\frac{1}{h_{2,2}}G_X\left(\frac{x-\tilde{x}}{h_{2,2}}\right)d\tilde{y}d\tilde{x}-\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)\right| \\
&=\left|\int\int \phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(\tilde{y},\tilde{x},\cdot)}(\xi)\left(\frac{1}{(2\pi)^2}\int\int e^{-\mathbf{i}t_1(y-\tilde{y})-\mathbf{i}t_2(x-\tilde{x})}\phi_{G_Y}(t_1h_{2,1})\phi_{G_X}(t_2h_{2,2})dt_1dt_2\right)d\tilde{y}d\tilde{x}\right. \\
&\left.-\frac{1}{(2\pi)^2}\int\int e^{-\mathbf{i}t_1y-\mathbf{i}t_2x}\left(\int\int e^{\mathbf{i}t_1y+\mathbf{i}t_2x}\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)dydx\right)dt_1dt_2\right| \\
&=\left|\frac{1}{(2\pi)^2}\int\int e^{-\mathbf{i}t_1y-\mathbf{i}t_2x}\left(\phi_{G_Y}(t_1h_{2,1})\phi_{G_X}(t_2h_{2,2})-1\right)\left(\int\int e^{\mathbf{i}t_1\tilde{y}+\mathbf{i}t_2\tilde{x}}\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(\tilde{y},\tilde{x},\cdot)}(\xi)d\tilde{y}d\tilde{x}\right)dt_1dt_2\right| \\
&=\left|\frac{1}{(2\pi)^2}\int_{\frac{\bar{\xi}_{G_X}}{h_{2,2}}}\int_{\frac{\bar{\xi}_{G_Y}}{h_{2,1}}} e^{-\mathbf{i}t_1y-\mathbf{i}t_2x}\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}}(t_1,t_2,\xi)dt_1dt_2\right| \\
&\le \frac{1}{(2\pi)^2}\int_{\frac{\bar{\xi}_{G_X}}{h_{2,2}}}\int_{\frac{\bar{\xi}_{G_Y}}{h_{2,1}}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}}(t_1,t_2,\xi)\right|dt_1dt_2 \\
&\preceq \left|\phi_{W^*}(\xi)\right|\int_{\frac{\bar{\xi}_{G_X}}{h_{2,2}}}\int_{\frac{\bar{\xi}_{G_Y}}{h_{2,1}}}(1+|t_1|)^{\gamma_{f_1}}(1+|t_2|)^{\gamma_{f_2}}\exp\left(\alpha_{f_1}|t_1|^{\beta_{f_1}}+\alpha_{f_2}|t_2|^{\beta_{f_2}}\right)dt_1dt_2 \\
&\preceq \left|\phi_{W^*}(\xi)\right|O\left(\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{1+\gamma_{f_1}}\left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{1+\gamma_{f_2}}\exp\left(\alpha_{f_1}\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{\beta_{f_1}}+\alpha_{f_2}\left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{\beta_{f_2}}\right)\right) \\
&\preceq \left|\phi_{W^*}(\xi)\right|o(1)., \qquad (1.35)
\end{aligned}$$

so

$$\begin{aligned}
&\left|\frac{1}{2\pi}\int(-\mathbf{i}t)^{\lambda_1}e^{-\mathbf{i}tw^*}\phi_K(h_{1n}t)\frac{E\left[\delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(v,\cdot)}(t)\right]\phi_{W^*}(t)}{\phi_{W_2}(t)}dt\right| \\
&\preceq o(1)\int_0^{h_{1n}^{-1}}|t|^{\lambda_1}\frac{\left|\phi_{W^*}(t)\right|}{\left|\phi_{W_2}(t)\right|}\left|\phi_{W_2}(t)\right|dt \\
&\preceq o(1)\int_0^{h_{1n}^{-1}}(1+|t|)^{\lambda_1+\gamma_\phi-\gamma_2}\exp\left(\alpha_\phi|t|^{\beta_\phi}\right)\exp\left(-\alpha_2|t|^{\beta_2}\right)dt \\
&\preceq o(1)O\left(\left(h_1^{-1}\right)^{1+\gamma_\phi+\lambda_1-\gamma_2}\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)\exp\left(-\alpha_2\left(h_1^{-1}\right)^{\beta_2}\right)\right) \\
&= o\left(\left(h_1^{-1}\right)^{1+\gamma_\phi+\lambda_1}\exp\left(\alpha_\phi\bar{\zeta}_K^{\beta_\phi}\left(h_1^{-1}\right)^{\beta_\phi}\right)\right).
\end{aligned}$$

This completes the proof. □

Lemma 1.14. *Suppose the conditions of Lemma 1.5 hold. For each* ξ *and* h_1*, let* $\Psi^+_{l,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, h_1) \equiv \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} |\Psi_{l,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, y, x, w^*, h_1)|$ *for* $l = 1, 2, 3$*. Define*

$$\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h) \equiv \sum_{l=1}^{2} \int \Psi^+_{l,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, h_1)d\xi + (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}} \int \Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, h_1)d\xi.$$

If Assumption 1.11 and 1.13 also hold, then for $h_1 > 0$

$$\begin{aligned}
&\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h)\\
&= O\left(\max\left\{(h_1^{-1})^{1+\gamma_*}, (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi \mathbf{1}\{\beta_\phi = \beta_2\} - \alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right).
\end{aligned}$$

Proof.

$$\begin{aligned}
&\Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, h_1)\\
&= \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} |\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, y, x, w^*, h_1)|\\
&\preceq \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \frac{1}{|\phi_{W_2}(\xi)|}\int_\xi^{\pm\infty} |t|^{\lambda_1}\left|e^{-\mathbf{i}tw^*}\right| |\phi_K(h_1 t)| \left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}} \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right) dt\\
&\preceq \frac{1}{|\phi_{W_2}(\xi)|}\int_\xi^{\pm\infty} |t|^{\lambda_1} |\phi_K(h_1 t)| \left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}} \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right) dt
\end{aligned}$$

so that

$$\begin{aligned}
\int \Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi, h_1)d\xi &\preceq \int \left[\frac{1}{|\phi_{W_2}(\xi)|}\mathbf{1}\{|\xi| \le h_1^{-1}\}\int_\xi^{h_1^{-1}} |t|^{\lambda_1}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right| dt\right] d\xi\\
&\preceq \int \left[\frac{1}{|\phi_{W_2}(\xi)|}\mathbf{1}\{|\xi| \le h_1^{-1}\}\int_\xi^{h_1^{-1}} |t|^{\lambda_1}(1+|t|)^{\gamma_\phi}\exp\left(\alpha_\phi|t|^{\beta_\phi}\right) dt\right] d\xi\\
&\preceq (h_1^{-1})^{2-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(-\alpha_2\left(h_1^{-1}\right)^{\beta_2}\right)\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)
\end{aligned}$$

where in the third $\preceq$, I used the assumption $1-\gamma_2+\gamma_\phi+\lambda_1>0$ when $\beta_2=0$ and invoked Lemma 7 and Lemma 8 in Schennach (2004b).

$$
\begin{aligned}
&\Psi^{+}_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)\\
=&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}}\left|\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right|\\
\preceq&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}}\frac{|\theta(\xi)|}{|\phi_{W_2}(\xi)|^2}\int_\xi^{\pm\infty}|t|^{\lambda_1}|e^{-\mathbf{i}tw^*}||\phi_K(h_1t)|\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|dt\\
&+\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}}|\xi|^{\lambda_1}|e^{-\mathbf{i}\xi w^*}||\phi_K(h_1\xi)|\frac{\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right|}{|\phi_{W_2}(\xi)|}\\
\preceq&\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left\{\frac{|\theta(\xi)|}{|\phi_{W_2}(\xi)|^2}\int_\xi^{\pm\infty}|t|^{\lambda_1}|\phi_K(h_1t)|\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|dt\right.\\
&\qquad\left.+|\xi|^{\lambda_1}|\phi_K(h_1\xi)|\frac{\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|}{|\phi_{W_2}(\xi)|}\right\}\\
\preceq&\frac{1}{|\phi_{W_2}(\xi)|}\left(\frac{|\theta(\xi)|}{|\phi_{W_2}(\xi)|}\int_\xi^{\pm\infty}|t|^{\lambda_1}|\phi_K(h_1t)|\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)dt\right.\\
&\left.+|\xi|^{\lambda_1}|\phi_K(h_1\xi)|\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right|\right)\right)\\
\preceq&\frac{1}{|\phi_{W_2}(\xi)|}\left(\frac{|\phi'_{W^*}(\xi)|}{|\phi_{W^*}(\xi)|}\int_\xi^{\pm\infty}|t|^{\lambda_1}|\phi_K(h_1t)|\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)dt\right.\\
&\left.+|\xi|^{\lambda_1}|\phi_K(h_1\xi)|\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right|\right)\right)
\end{aligned}
$$

where I used the fact that

$$
\begin{aligned}
\frac{\theta(\xi)}{\phi_{W_2}(\xi)}&=\frac{E\left[W_1e^{\mathrm{i}\xi W_2}\right]}{E\left[e^{\mathrm{i}\xi W_2}\right]}=\frac{E\left[\left(W^*+\Delta W_1\right)e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]}{E\left[e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]}\\
&=\frac{E\left[W^*e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]+E\left[E\left[\Delta W_1\mid W^*,\Delta W_2\right]e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]}{E\left[e^{\mathrm{i}\xi(W^*+\Delta_2)}\right]}\\
&=\frac{E\left[W^*e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]}{E\left[e^{\mathrm{i}\xi(W^*+\Delta W_2)}\right]}=\frac{E\left[W^*e^{\mathrm{i}\xi W^*}\right]}{E\left[e^{\mathrm{i}\xi W^*}\right]}
\end{aligned}
$$

$$= \frac{-\mathbf{i}(d/d\xi)E\left[e^{\mathrm{i}\xi W^*}\right]}{E\left[e^{\mathrm{i}\xi W^*}\right]} = -\mathbf{i}\frac{(d/d\xi)\phi_{W^*}(\xi)}{\phi_{W^*}(\xi)}.$$

Integrating $\Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$ with respect to ξ and using Assumption 1.11 and 1.13 gives

$$\begin{aligned}
&\int \Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi \\
&\preceq \int \frac{1}{|\phi_{W_2}(\xi)|}\left(\frac{|\phi'_{W^*}(\xi)|}{|\phi_{W^*}(\xi)|}\int_{\xi}^{\pm\infty}\mathbf{1}\left\{|t|\leq h_1^{-1}\right\}|t|^{\lambda_1}\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)dt\right. \\
&\quad\left.+\mathbf{1}\left\{|\xi|\leq h_1^{-1}\right\}|\xi|^{\lambda_1}\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right|\right)\right)d\xi \\
&\preceq \int \frac{1}{|\phi_{W_2}(\xi)|}\left(\frac{|\phi'_{W^*}(\xi)|}{|\phi_{W^*}(\xi)|}\mathbf{1}\left\{|\xi|\leq h_1^{-1}\right\}\int_{|\xi|}^{h_1^{-1}}|t|^{\lambda_1}\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)dt\right. \\
&\quad\left.+\mathbf{1}\left\{|\xi|\leq h_1^{-1}\right\}|\xi|^{\lambda_1}\left(\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(\xi)\right|\right)\right)d\xi \\
&\preceq \int (1+|\xi|)^{-\gamma_2}\exp(-\alpha_2|\xi|^{\beta_2})\mathbf{1}\left\{|\xi|\leq h_1^{-1}\right\} \\
&\quad\times\left((1+|\xi|)^{\gamma_*}\int_0^{h_1^{-1}}|t|^{\lambda_1}(1+|t|)^{\gamma_\phi}\exp\left(\alpha_\phi|t|^{\beta_\phi}\right)dt+|\xi|^{\lambda_1}(1+|\xi|)^{\gamma_\phi}\exp\left(\alpha_\phi|\xi|^{\beta_\phi}\right)\right)d\xi \\
&\preceq (1+h_1^{-1})^{1-\gamma_2}\exp\left(-\alpha_2\left(h_1^{-1}\right)^{\beta_2}\right) \\
&\quad\times\left((1+h_1^{-1})^{\gamma_*}(1+h_1^{-1})^{1+\gamma_\phi+\lambda_1}\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)+(1+h_1^{-1})^{\gamma_\phi+\lambda_1}\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)\right) \\
&\preceq (h_1^{-1})^{2-\gamma_2+\gamma_\phi+\gamma_*+\lambda_1}\exp\left(-\alpha_2\left(h_1^{-1}\right)^{\beta_2}\right)\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)
\end{aligned}$$

$$\begin{aligned}
\Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) &= \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}}\left|\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \\
&\preceq \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}}\frac{|\phi_{W^*}(\xi)|}{|\phi_{W_2}(\xi)|}|\xi|^{\lambda_1}|e^{-\mathbf{i}\xi w^*}||\phi_K(h_1\xi)| \\
&= \frac{|\phi_{W^*}(\xi)|}{|\phi_{W_2}(\xi)|}|\xi|^{\lambda_1}|\phi_K(h_1\xi)|.
\end{aligned}$$

so that

$$
\begin{aligned}
&(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\int \Psi^{+}_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi\\
&\preceq (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\int \frac{|\phi_{W^*}(\xi)|}{|\phi_{W_2}(\xi)|}|\xi|^{\lambda_1}\mathbf{1}\left\{|\xi|\leq h_1^{-1}\right\}d\xi\\
&\preceq (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\int_0^{h_1^{-1}} \frac{|\phi_{W^*}(\xi)|}{|\phi_{W_2}(\xi)|}|\xi|^{\lambda_1}d\xi\\
&\preceq (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(-\alpha_2\left(h_1^{-1}\right)^{\beta_2}\right)\exp\left(\alpha_\phi\left(h_1^{-1}\right)^{\beta_\phi}\right)
\end{aligned}
$$

Collecting together these rates delivers the conclusion of the lemma. □

Lemma 1.15. *Suppose the conditions for Lemma 1.5 hold, then*

$$
\begin{aligned}
\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\Bigg|&\hat{E}\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\\
&-E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\Bigg|\preceq\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\sqrt{n}}.
\end{aligned}
$$

Proof. Denote the Fourier transform of $f_{Y,X,W_2}(y,x,w^*)$ as $\phi_{f_2}(t_1,t_2,\xi)$.

$$
\begin{aligned}
&E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\\
&=\int\int\int e^{\mathbf{i}\xi\tilde{w}}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-\tilde{y}}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-\tilde{x}}{h_{2,2}}\right)f_{Y,X,W_2}(\tilde{y},\tilde{x},\tilde{w})d\tilde{w}d\tilde{y}d\tilde{x}\\
&=\int\int\int e^{\mathbf{i}\xi\tilde{w}}\left(\frac{1}{(2\pi)^2}\int\int e^{-\mathbf{i}t_1(x-\tilde{x})-\mathbf{i}t_2(y-\tilde{y})}(\mathbf{i}t_1)^{\lambda_{2,1}}(\mathbf{i}t_2)^{\lambda_{2,2}}\phi_{G_X}(t_1h_{2,1})\phi_{G_Y}(t_2h_{2,2})dt_1dt_2\right)\times\\
&\qquad f_{Y,X,W_2}(\tilde{y},\tilde{x},\tilde{w})d\tilde{w}d\tilde{y}d\tilde{x}\\
&=\frac{1}{(2\pi)^2}\int\int(\mathbf{i}t_1)^{\lambda_{2,1}}(\mathbf{i}t_2)^{\lambda_{2,2}}e^{-\mathbf{i}t_1y-\mathbf{i}t_2x}\phi_{G_Y}(t_1h_{2,1})\phi_{G_X}(t_2h_{2,2})\phi_{f_2}(t_1,t_2,\xi)dt_1dt_2
\end{aligned}
$$

so

$$
\begin{aligned}
\sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\Bigg|&\hat{E}\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\\
&-E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\Bigg|
\end{aligned}
$$

$$= \sup_{(y,x)\in\mathbb{S}_{(Y,X)}} \left| \frac{1}{(2\pi)^2} \int\int (\mathbf{i}t_1)^{\lambda_{2,1}} (\mathbf{i}t_2)^{\lambda_{2,2}} e^{-\mathbf{i}t_1 y - \mathbf{i}t_2 x} \phi_{G_Y}(t_1 h_{2,1}) \phi_{G_X}(t_2 h_{2,2}) \times \right.$$
$$\left. \left(\frac{1}{n} \sum_{j=1}^{n} e^{\mathbf{i}t_1 Y_j + \mathbf{i}t_2 X_j + \mathbf{i}\xi W_{2,j}} - \phi_{f_2}(t_1,t_2,\xi) \right) dt_1 dt_2 \right|$$
$$\leq \frac{1}{(2\pi)^2} \int\int |t_1|^{\lambda_{2,1}} |t_2|^{\lambda_{2,2}} \left| \phi_{G_Y}(t_1 h_{2,1}) \phi_{G_X}(t_2 h_{2,2}) \right| \left| \frac{1}{n} \sum_{j=1}^{n} e^{\mathbf{i}t_1 Y_j + \mathbf{i}t_2 X_j + \mathbf{i}\xi W_{2,j}} - \phi_{f_2}(t_1,t_2,\xi) \right| dt_2.$$

Then since

$$E\left[\left(\frac{1}{n} \sum_{j=1}^{n} e^{\mathbf{i}t_1 Y_j + \mathbf{i}t_2 X_j + \mathbf{i}\xi W_{2,j}} - \phi_{f_2}(t_1,t_2,\xi) \right)^2 \right]$$
$$= \frac{1}{n} E\left[\left(e^{\mathbf{i}t_1 Y_j + \mathbf{i}t_2 X_j + \mathbf{i}\xi W_{2,j}} - \phi_{f_2}(t_1,t_2,\xi) \right)^2 \right]$$
$$\leq \frac{1}{n} E\left[\left(e^{\mathbf{i}t_1 Y_j + \mathbf{i}t_2 X_j + \mathbf{i}\xi W_{2,j}} \right)^2 \right] = O\left(\frac{1}{n}\right),$$

I have

$$\sup_{(y,x)\in\mathbb{S}_{(Y,X)}} \left| \hat{E}\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right) \right] \right.$$
$$\left. - E\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right) \right] \right|$$
$$\preceq \frac{1}{\sqrt{n}} \int\int |t_1|^{\lambda_{2,1}} |t_2|^{\lambda_{2,2}} \left| \phi_{G_Y}(t_1 h_{2,1}) \phi_{G_X}(t_2 h_{2,2}) \right| dt_1 dt_2$$
$$\preceq \frac{1}{\sqrt{n}} \int_0^{h_{2,2}^{-1}} \int_0^{h_{2,1}^{-1}} |t_1|^{\lambda_{2,1}} |t_2|^{\lambda_{2,2}} dt_1 dt_2 \preceq \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \sqrt{n}}.$$

□

Lemma 1.16. *For a finite integer J, let $\left\{P_{n,j}(a)\right\}$ be a sequence of nonrandom real-valued continuously differentiable functions of random vectors a, for $j = 1,...,J$ respectively. Let A be a random vector. If for each n, $\sigma_n = \sqrt{var\left(\sum_{j=1}^{J} P_{n,j}(A)\right)}$ exist, and $\sigma_n > 0$ for n sufficiently large, then*

$$\sigma_n^{-1} n^{1/2} \left(\hat{E}\left[\sum_{j=1}^{J} P_{n,j}(A) \right] \right) \xrightarrow{d} N(0,1).$$

Proof. Denote the centered version of the summands as $Z_{n,j}(A) \equiv P_{n,j}(A) - E\left[P_{n,j}(A)\right]$. Then $\sigma_n^2 \equiv E\left[\left(\sum_{j=1}^J Z_{n,j}(A)\right)^2\right]$ It's sufficient to verify that the Lindeberg condition holds:

$$\sum_{i=1}^n E\left[\mathbf{1}\left\{\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right| > \epsilon\right\}\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right|^2\right] \to 0, \text{ as } n \to \infty,$$

for all $\epsilon > 0$. Note that as $n \to \infty$

$$E\left[\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right|^2\right] = \frac{1}{n}E\left[\frac{\left|\sum_{j=1}^J Z_{n,j}(A)\right|^2}{E\left[\left(\sum_{j=1}^J Z_{n,j}(A)\right)^2\right]}\right] \to 0$$

By Markov Inequality, for any $\epsilon > 0$, as $n \to \infty$, $Pr\left(\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right| > \epsilon\right) \to 0$, which implies $\mathbf{1}\left\{\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right| > \epsilon\right\} = o_p(1)$. Note that $\mathbf{1}\left\{\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right| > \epsilon\right\}\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sigma_n}\right|^2 \leq \left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sigma_n}\right|^2$ and that $E\left[\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sigma_n}\right|^2\right] = 1 < \infty$. By Dominated Convergence Theorem

$$E\left[\mathbf{1}\left\{\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sqrt{n}\sigma_n}\right| > \epsilon\right\}\left|\frac{\sum_{j=1}^J Z_{n,j}(A)}{\sigma_n}\right|^2\right] \to 0, \text{ as } n \to \infty.$$

The conclusion follows. □

Proof of Theorem 1.2. (i) The fact that $E[L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)] = 0$ follows directly from Eq. (1.34). Next, with fixed value of h, Assumption 1.9 and 1.12 ensures the existence and finiteness of

$$\begin{aligned}E[L^2_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)] &= E\left[\left(\hat{E}\left[l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h;Y,X,W_1,W_2)\right]\right)^2\right] \\ &= n^{-1}E\left[\left(l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h;Y,X,W_1,W_2)\right)^2\right] \\ &= n^{-1}\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h).\end{aligned}$$

From Eq. (1.34), I have

$$\begin{aligned}
&\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \\
&\equiv E\left[n\left(\bar{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right)^2\right] \\
&= E\Bigg[\Bigg(\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)n^{1/2}\delta\hat{\theta}(\xi)d\xi \\
&\qquad + \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)n^{1/2}\delta\hat{\phi}_{W_2}(\xi)d\xi \\
&\qquad + \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)n^{1/2}\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\xi)d\xi\Bigg)^2\Bigg] \\
&= \int\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\theta}(\xi)\delta\hat{\theta}^\dagger(\zeta)\right]\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\phi}_{W_2}(\xi)\delta\hat{\phi}^\dagger_{W_2}(\zeta)\right]\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int (h_{2,1}^{-2})^{1+\lambda_{2,1}}(h_{2,2}^{-2})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\times \\
&\qquad\qquad E\left[nh_{2,1}^{2+2\lambda_{2,1}}h_{2,2}^{2+2\lambda_{2,2}}\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\xi)\delta\hat{\phi}^\dagger_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right]\times \\
&\qquad\qquad \left(\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\theta}(\xi)\delta\hat{\phi}^\dagger_{W_2}(\zeta)\right]\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\theta}(\xi)h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\hat{\phi}^\dagger_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right]\times \\
&\qquad\qquad\qquad (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\left(\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\phi}_{W_2}(\xi)h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\hat{\phi}^\dagger_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right]\times \\
&\qquad\qquad\qquad (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\left(\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)E\left[n\delta\hat{\phi}_{W_2}(\zeta)\delta\hat{\theta}^\dagger(\xi)\right]\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\times \\
&\qquad E\left[nh_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\hat{\phi}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\delta\hat{\theta}^\dagger(\xi)\right]\times\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \\
&\quad + \int\int (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\times \\
&\qquad E\left[nh_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\delta\hat{\phi}^\dagger_{W_2}(\xi)\right]\times\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right)^\dagger d\xi d\zeta \qquad (1.36)
\end{aligned}$$

Note that I have

$$E\left[n\delta\hat{\theta}(\xi)\delta\hat{\theta}^{\dagger}(\zeta)\right] = E\left[n\left(\hat{\theta}(\xi)-\theta(\xi)\right)\left(\hat{\theta}^{\dagger}(\zeta)-\theta^{\dagger}(\zeta)\right)\right]$$
$$= E\left[\left(W_1e^{\mathbf{i}\xi W_2}-\theta(\xi)\right)\left(W_1e^{-\mathbf{i}\zeta W_2}-\theta(-\zeta)\right)\right]$$
$$= E\left[W_1^2e^{\mathbf{i}(\xi-\zeta)W_2}\right]-\theta(\xi)E\left[W_1e^{-\mathbf{i}\zeta W_2}\right]-E\left[W_1e^{\mathbf{i}\xi W_2}\right]\theta^{\dagger}(\zeta)+\theta(\xi)\theta^{\dagger}(\zeta)$$
$$= E\left[W_1^2e^{\mathbf{i}(\xi-\zeta)W_2}\right]-\theta(\xi)\theta(-\zeta),$$

$$E\left[n\delta\hat{\theta}(\xi)\delta\hat{\phi}^{\dagger}_{W_2}(\zeta)\right] = E\left[n\left(\hat{\theta}(\xi)-\theta(\xi)\right)\left(\hat{\phi}^{\dagger}_{W_2}(\zeta)-\phi^{\dagger}_{W_2}(\zeta)\right)\right]$$
$$= E\left[\left(W_1e^{\mathbf{i}\xi W_2}-\theta(\xi)\right)\left(e^{-\mathbf{i}\zeta W_2}-\phi^{\dagger}_{W_2}(\zeta)\right)\right]$$
$$= E\left[W_1e^{\mathbf{i}\xi W_2}e^{-\mathbf{i}\zeta W_2}\right]-\theta(\xi)E\left[e^{-\mathbf{i}\zeta W_2}\right]-E\left[W_1e^{\mathbf{i}\xi W_2}\right]\phi^{\dagger}_{W_2}(\zeta)+\theta(\xi)\phi^{\dagger}_{W_2}(\zeta)$$
$$= E\left[W_1e^{\mathbf{i}(\xi-\zeta)W_2}\right]-\theta(\xi)\phi_{W_2}(-\zeta),$$

$$E\left[n\delta\hat{\theta}(\xi)h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\hat{\phi}^{\dagger}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right]$$
$$= E\Bigg[\left(W_1e^{\mathbf{i}\xi W_2}-\theta(\xi)\right)\times$$
$$\left(e^{-\mathbf{i}\zeta W_2}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)-h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\phi^{\dagger}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right)\Bigg]$$
$$= E\left[W_1e^{\mathbf{i}(\xi-\zeta)W_2}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]-\theta(\xi)h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(-\zeta),$$

$$E\left[n\delta\hat{\phi}_{W_2}(\xi)\delta\hat{\phi}^{\dagger}_{W_2}(\zeta)\right] = E\left[\left(e^{\mathbf{i}\xi W_2}-\phi_{W_2}(\xi)\right)\left(e^{-\mathbf{i}\zeta W_2}-\phi^{\dagger}_{W_2}(\zeta)\right)\right]$$
$$= E\left[e^{\mathbf{i}(\xi-\zeta)W_2}\right]-\phi_{W_2}(\xi)\phi_{W_2}(-\zeta),$$

$$E\left[n\delta\hat{\phi}_{W_2}(\xi)h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}\delta\hat{\phi}^{\dagger}_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(\zeta)\right]$$
$$= E\Bigg[\left(e^{\mathbf{i}\xi W_2}-\phi_{W_2}(\xi)\right)\times$$

$$\left. \left(e^{-\mathbf{i}\zeta W_2} G_Y \left(\frac{y-Y}{h_{2,1}} \right) G_X^{(\lambda_{2,1},\lambda_{2,2})} \left(\frac{x-X}{h_{2,2}} \right) - h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \phi^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta) \right) \right]$$

$$= E\left[e^{\mathbf{i}(\xi-\zeta)W_2} G_Y \left(\frac{y-Y}{h_{2,1}} \right) G_X^{(\lambda_{2,1},\lambda_{2,2})} \left(\frac{x-X}{h_{2,2}} \right) \right] - \phi_{W_2}(\xi) h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(-\zeta),$$

$$E\left[n h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi) h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta) \right]$$

$$= E\left[\left(e^{\mathbf{i}\xi W_2} G_Y \left(\frac{y-Y}{h_{2,1}} \right) G_X^{(\lambda_{2,1},\lambda_{2,2})} \left(\frac{x-X}{h_{2,2}} \right) - h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi) \right) \times \right.$$

$$\left. \left(e^{-\mathbf{i}\zeta W_2} G_Y \left(\frac{y-Y}{h_{2,1}} \right) G_X^{(\lambda_{2,1},\lambda_{2,2})} \left(\frac{x-X}{h_{2,2}} \right) - h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \phi^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta) \right) \right]$$

$$= E\left[e^{\mathbf{i}(\xi-\zeta)W_2} \left(G_Y \left(\frac{y-Y}{h_{2,1}} \right) G_X^{(\lambda_{2,1},\lambda_{2,2})} \left(\frac{x-X}{h_{2,2}} \right) \right)^2 \right]$$

$$- h_{2,1}^{2+2\lambda_{2,1}} h_{2,2}^{2+2\lambda_{2,2}} \phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi) \phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(-\zeta).$$

By Assumption 1.12,

$$\begin{aligned} \left| E\left[n\delta\hat{\theta}(\xi)\delta\hat{\theta}^{\dagger}(\zeta) \right] \right| &= \left| E\left[W_1^2 e^{\mathbf{i}(\xi-\zeta)W_2} \right] - \theta(\xi)\theta(-\zeta) \right| \\ &\leq E\left[W_1^2 \left| e^{\mathbf{i}(\xi-\zeta)W_2} \right| \right] + E\left[|W_1| \left| e^{\mathbf{i}\xi W_2} \right| \right] E\left[|W_1| \left| e^{-\mathbf{i}\zeta W_2} \right| \right] \\ &\leq E\left[W_1^2 \right] + E\left[|W_1| \right] E\left[|W_1| \right] \preceq 1, \end{aligned}$$

$$\begin{aligned} \left| E\left[n\delta\hat{\theta}(\xi)\delta\hat{\phi}^{\dagger}_{W_2}(\zeta) \right] \right| &= \left| \theta(\xi-\zeta) - \theta(\xi)\phi_{W_2}(-\zeta) \right| \\ &\leq E\left[|W_1| \left| e^{\mathbf{i}(\xi-\zeta)W_2} \right| \right] + E\left[|W_1| \left| e^{\mathbf{i}\xi W_2} \right| \right] E\left[\left| e^{-\mathbf{i}\zeta W_2} \right| \right] \\ &\leq 2E[|W_1|] \preceq 1, \end{aligned}$$

$$\sup_{x\in\mathbb{S}_X} \left| E\left[n\delta\hat{\theta}(\xi) h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta) \right] \right|$$

$$
\begin{aligned}
&= \sup_{x\in\mathbb{S}_X}\left|E\left[W_1 e^{\mathbf{i}(\xi-\zeta)W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] - \theta(\xi) h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(-\zeta)\right| \\
&\leq E\left[|W_1|\left|e^{\mathbf{i}(\xi-\zeta)W_2}\right| \sup_{x\in\mathbb{S}_X}\left|G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right|\right] \\
&\qquad + E\left[|W_1|\left|e^{\mathbf{i}\xi W_2}\right|\right] E\left[\left|e^{-\mathbf{i}\zeta W_2}\right| \sup_{x\in\mathbb{S}_X}\left|G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right|\right] \\
&\leq 2E[|W_1|] \preceq 1,
\end{aligned}
$$

where the last line follows by Assumption 1.9. Following the same logic, I have that

$$
\begin{aligned}
&\left|E\left[n\delta\hat{\phi}_{W_2}(\xi)\delta\hat{\phi}_{W_2}^{\dagger}(\zeta)\right]\right| \preceq 1, \\
&\sup_{x\in\mathbb{S}_X}\left|E\left[n\delta\hat{\phi}_{W_2}(\xi) h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta)\right]\right| \preceq 1, \\
&\sup_{x\in\mathbb{S}_X}\left|E\left[n h_{2,1}^{2+2\lambda_{2,1}} h_{2,2}^{2+2\lambda_{2,2}} \delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi) \delta\hat{\phi}^{\dagger}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta)\right]\right| \preceq 1, \\
&\left|E\left[n\delta\hat{\phi}_{W_2}(\zeta)\delta\hat{\theta}^{\dagger}(\xi)\right]\right| \preceq 1, \\
&\sup_{x\in\mathbb{S}_X}\left|E\left[n h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta)\delta\hat{\theta}^{\dagger}(\xi)\right]\right| \preceq 1, \\
&\sup_{x\in\mathbb{S}_X}\left|E\left[n h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}} \delta\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\zeta)\delta\hat{\phi}^{\dagger}_{W_2}(\xi)\right]\right| \preceq 1.
\end{aligned}
$$

It follows from Equation (1.36) that

$$
\begin{aligned}
&\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) \\
&\leq \int\int \left|\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^{\dagger}\right| d\xi d\zeta \\
&\quad + \int\int \left|\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^{\dagger}\right| d\xi d\zeta \\
&\quad + \int\int \left|(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \times \\
&\qquad\qquad \left|\left((h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^{\dagger}\right| d\xi d\zeta \\
&\quad + \int\int \left|\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^{\dagger}\right| d\xi d\zeta
\end{aligned}
$$

$$
\begin{aligned}
&+\int\int \left|\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left((h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger\right| d\xi d\zeta \\
&+\int\int \left|\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left((h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger\right| d\xi d\zeta \\
&+\int\int \left|\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| \left|\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right)^\dagger\right| d\xi d\zeta \\
&+\int\int \left|(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right| \left|\left(\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right)^\dagger\right| d\xi d\zeta \\
&+\int\int \left|(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\zeta,y,x,w^*,h_1)\right| \left|\left(\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right)^\dagger\right| d\xi d\zeta \\
&= \Bigg(\int \left|\Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| d\xi + \int \left|\Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| d\xi \\
&\qquad + \int \left|(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\right| d\xi\Bigg)^2 \\
&\le \left(\int \Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi + \int \Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi + \int (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h)d\xi\right)^2 \\
&= \left(\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h)\right)^2,
\end{aligned}
$$

where $\Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$, $\Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$, $\Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h)$, and $\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h)$ are as defined in Lemma 1.14 and

$$
\begin{aligned}
&\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h) \\
&= O\left(\max\left\{(h_1^{-1})^{1+\gamma_*},(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi\mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right).
\end{aligned}
$$

Hence I have proved Eq. (1.20)

Next I show Eq. (1.22). From Eq. (1.34) I have

$$
\begin{aligned}
&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right| \\
&= \sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \Bigg|\int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(\hat{E}[W_1e^{\mathbf{i}\xi W_2}]-E[W_1e^{\mathbf{i}\xi W_2}]\right)d\xi \\
&\qquad + \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(\hat{E}[e^{\mathbf{i}\xi W_2}]-E[e^{\mathbf{i}\xi W_2}]\right)d\xi \\
&\qquad + \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\times \\
&\qquad\qquad \Bigg(\hat{E}\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] \\
&\qquad\qquad\quad -E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\Bigg)d\xi\Bigg|
\end{aligned}
$$

$$
\begin{aligned}
&\le \int \left(\sup_{(y,x,w^*)\in \mathbb{S}_{(Y,X,W^*)}} \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) \right) \left| \hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}] \right| d\xi \\
&\quad + \int \left(\sup_{(y,x,w^*)\in \mathbb{S}_{(Y,X,W^*)}} \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) \right) \left| \hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}] \right| d\xi \\
&\quad + \int \left(\sup_{(y,x,w^*)\in \mathbb{S}_{(Y,X,W^*)}} (h_{2,1}^{-1})^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1) \right) \times \\
&\quad \sup_{x\in\mathbb{S}_X} \left| \hat{E}\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] - E\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] \right| d\xi \\
&= \int \Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) \left| \hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}] \right| d\xi \\
&\quad + \int \Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) \left| \hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}] \right| d\xi \\
&\quad + \int (h_{2,1}^{-1})^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} \Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) \times \\
&\quad \sup_{x\in\mathbb{S}_X} \left| \hat{E}\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] - E\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] \right| d\xi
\end{aligned}
$$

where $\Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$, $\Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$ and $\Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)$ are defined in Lemma 1.14. The integrals are finite because of two reasons. First, with probability approaching 1, I have $\left| \hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}] \right| \le 2E[|W_1|] + \epsilon \le \infty$, for some positive real number ϵ. This follows from Assumption 1.12, the fact that

$$
\begin{aligned}
\left| \hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}] \right| &\le \left| \hat{E}[W_1 e^{\mathbf{i}\xi W_2}] \right| + \left| E[W_1 e^{\mathbf{i}\xi W_2}] \right| \\
&\le \hat{E}[\left| W_1 e^{\mathbf{i}\xi W_2} \right|] + E[\left| W_1 e^{\mathbf{i}\xi W_2} \right|] \\
&\le \hat{E}[|W_1|] + E[|W_1|],
\end{aligned}
$$

and that $\hat{E}[|W_1|] \xrightarrow{p} E[|W_1|]$. Also, it's easy to show that $\left| \hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}] \right| \le 2 + \epsilon < \infty$ and that $\sup_{x\in\mathbb{S}_X} \left| \hat{E}\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] - E\left[e^{\mathbf{i}\xi W_2} G_Y\left(\frac{y-Y}{h_{2,1}}\right) G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right] \right| \preceq 1$. Second, it follows by Lemma 1.14 that $\int \Psi^+_{j,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) d\xi < \infty$ for $j = 1,2$ and that $\int (h_{2,1}^{-1})^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} \Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1) d\xi < \infty$.

By Assumption 1.12,

$$
\begin{aligned}
E\left[\left(\hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}] \right)^2 \right] &\le \frac{1}{n} E\left[\left(W_1 e^{\mathbf{i}\xi W_2} - E[W_1 e^{\mathbf{i}\xi W_2}] \right)^2 \right] \le \frac{1}{n} E\left[\left(W_1 e^{\mathbf{i}\xi W_2} \right)^2 \right] \\
&\le \frac{1}{n} E\left[W_1^2 \right] = O\left(\frac{1}{n}\right),
\end{aligned}
$$

which implies $\left|\hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}]\right| = O_p(n^{-1/2})$. Similarly, I also have $\left|\hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}]\right| = O_p(n^{-1/2})$.

By the conclusions above and Lemma 1.15, I have

$$
\begin{aligned}
&\sup_{(y,x,w^*)\in\mathbb{S}_{(Y,X,W^*)}} \left|L_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h)\right| \\
&\leq \int \Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)\left|\hat{E}[W_1 e^{\mathbf{i}\xi W_2}] - E[W_1 e^{\mathbf{i}\xi W_2}]\right| d\xi \\
&\quad + \int \Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)\left|\hat{E}[e^{\mathbf{i}\xi W_2}] - E[e^{\mathbf{i}\xi W_2}]\right| d\xi \\
&\quad + \int \Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)\times \\
&\qquad \sup_{(y,x)\in\mathbb{S}_{(Y,X)}}\left|\hat{E}\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\right. \\
&\qquad\qquad \left. -E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]\right| d\xi \\
&\preceq n^{-1/2}\times \\
&\quad \left(\int \Psi^+_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi + \int \Psi^+_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi + \int (h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\Psi^+_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,h_1)d\xi\right) \\
&= n^{-1/2}\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h),
\end{aligned}
$$

where

$$
\begin{aligned}
&\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h) \\
&= O\left(\max\left\{(h_1^{-1})^{1+\gamma_*},(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi\mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right),
\end{aligned}
$$

as shown in Lemma 1.14.

(ii) Next I show asymptotic normality. I apply Lemma 1.16 to

$$
\begin{aligned}
&l_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n;Y,X,W_1,W_2) \\
&= \int \Psi_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(W_1 e^{\mathbf{i}\xi W_2} - E[W_1 e^{\mathbf{i}\xi W_2}]\right) d\xi \\
&\quad + \int \Psi_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\left(e^{\mathbf{i}\xi W_2} - E[e^{\mathbf{i}\xi W_2}]\right) d\xi \\
&\quad + \int \Psi_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}}(\xi,y,x,w^*,h_1)\times \\
&\qquad \left(\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)e^{\mathbf{i}\xi W_2}\right.
\end{aligned}
$$

$$-E\left[\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,1},\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)e^{\mathbf{i}\xi W_2}\right]\right)d\xi,$$

where $A = Y, X, W_1, W_2$ Previous argument ensures that for fixed h, $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) < \infty$. The desired conclusion follows. □

Lemma 1.17. *(Schennach (2004a) Lemma 6) Let A and W_2 be random variables satisfying $E[|A|^2] < \infty$ and $E[|A||W_2|] < \infty$ and let $(A_j, W_{2,j})_{j=1,\ldots,n}$ be a corresponding IID sample. Then for any $u, U \geq 0$ and $\epsilon > 0$,*

$$\sup_{\zeta\in[-Un^u,Un^u]}\left|\hat{E}[A\exp(\mathbf{i}\zeta W_2)] - E[A\exp(\mathbf{i}\zeta W_2)]\right| = o_p\left(n^{-1/2+\epsilon}\right).$$

Proof of Theorem 1.3. Plug (1.28) (1.29) and (1.30) into

$$\begin{aligned}
&\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1)\\
&= \frac{1}{2\pi}\int(-\mathbf{i}t)^{\lambda_2}e^{-\mathbf{i}tw^*}\phi_K(h_1t)\\
&\qquad\times\left[\frac{\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)}{\hat{\phi}_{W_2}(t)}\exp\left(\int_0^t\frac{\mathbf{i}\hat{\theta}(\xi)}{\hat{\phi}_{W_2}(\xi)}d\xi\right) - \frac{\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)}{\phi_{W_2}(t)}\exp\left(\int_0^t\frac{\mathbf{i}\theta(\xi)}{\phi_{W_2}(\xi)}d\xi\right)\right]dt.
\end{aligned}$$

and remove terms linear in $\delta\hat{\theta}(t)$, $\delta\hat{\phi}_{W_2}(\xi)$, and $\delta\hat{\phi}_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}$. For notation simplicity, I write h instead of h_n here. I can then find that $\left|\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_1)\right| \preceq \frac{1}{2\pi}\sum_{l=1}^7 R_{l,\lambda_1,\lambda_{2,1},\lambda_{2,2}}$, where

$$\begin{aligned}
R_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}} &= \int_0^\infty|t|^{\lambda_1}|\phi_K(h_1t)||\delta_1\hat{q}_{\lambda_2}(t)||\phi_{W^*}(t)|\left(\int_0^t|\delta_1\hat{q}_{W_1}(\xi)|d\xi\right)dt\\
R_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}} &= \int_0^\infty|t|^{\lambda_1}|\phi_K(h_1t)||\delta_2\hat{q}_{\lambda_2}(t)||\phi_{W^*}(t)|dt\\
R_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}} &= \int_0^\infty|t|^{\lambda_1}|\phi_K(h_1t)||\delta_2\hat{q}_{\lambda_2}(t)||\phi_{W^*}(t)|\left(\int_0^t|\delta_1\hat{q}_{W_1}(\xi)|d\xi\right)dt\\
R_{4,\lambda_1,\lambda_{2,1},\lambda_{2,2}} &= \int_0^\infty|t|^{\lambda_1}|\phi_K(h_1t)||q_{\lambda_2}(t)||\phi_{W^*}(t)|\left(\int_0^t|\delta_2\hat{q}_{W_1}(\xi)|d\xi\right)dt\\
R_{5,\lambda_1,\lambda_{2,1},\lambda_{2,2}} &= \int_0^\infty|t|^{\lambda_1}|\phi_K(h_1t)||\delta\hat{q}_{\lambda_2}(t)||\phi_{W^*}(t)|\left(\int_0^t|\delta_2\hat{q}_{W_1}(\xi)|d\xi\right)dt
\end{aligned}$$

$$R_{6,\lambda_1,\lambda_{2,1},\lambda_{2,2}} = \int_0^\infty |t|^{\lambda_1} |\phi_K(h_1 t)| |q_{\lambda_2}(t)| |\phi_{W^*}(t)| \frac{1}{2} \exp\left(\left|\delta\bar{Q}(t)\right|\right) \left(\int_0^t |\delta\hat{q}_{W_1}(\xi)| \, d\xi\right)^2 dt$$
$$R_{7,\lambda_1,\lambda_{2,1},\lambda_{2,2}} = \int_0^\infty |t|^{\lambda_1} |\phi_K(h_1 t)| |\delta\hat{q}_{\lambda_2}(t)| |\phi_{W^*}(t)| \frac{1}{2} \exp\left(\left|\delta\bar{Q}(t)\right|\right) \left(\int_0^t |\delta\hat{q}_{W_1}(\xi)| \, d\xi\right)^2 dt,$$

where $q_{\lambda_2}(t)$, $\delta\hat{q}_{\lambda_2}(t)$, $\delta_1\hat{q}_{\lambda_2}(t)$, $\delta_2\hat{q}_{\lambda_2}(t)$, $q_{W_1}(t)$, $\delta\hat{q}_{W_1}(t)$, $\delta_1\hat{q}_{W_1}(t)$, $\delta_2\hat{q}_{W_1}(t)$, and $\delta\bar{Q}(t)$ were defined in the proof of Lemma 1.5.

Lemma 1.17 gives that for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{(y,x)\in\mathbb{S}_(Y,X)} \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \Bigg| \hat{E}\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right] \\
&\qquad - E\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]\Bigg| \\
&= \left(h_{2,1}^{-1}\right)^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} \sup_{x\in\mathbb{S}_X} \left| G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right| \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \left|\hat{E}\left[e^{\mathbf{i}\xi W_2}\right] - E\left[e^{\mathbf{i}\xi W_2}\right]\right| \\
&= o_p\left(\left(h_{2,1}^{-1}\right)^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} n^{-1/2+\epsilon}\right).
\end{aligned}$$

By (1.35) I have that

$$\begin{aligned}
&\sup_{(y,x)\in\mathbb{S}_(Y,X)} \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \left| E\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right] - \phi_{f^{(\lambda_2)}_{Y,X,W_2}(y,x,\cdot)}(\xi)\right| \\
&\preceq \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \left|\phi_{W_2}(\xi)\right| O\left(\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{1+\gamma_{f_1}} \left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{1+\gamma_{f_2}} \exp\left(\alpha_{f_1}\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{\beta_{f_1}} + \alpha_{f_2}\left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{\beta_{f_2}}\right)\right) \\
&\preceq O\left(\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{1+\gamma_{f_1}} \left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{1+\gamma_{f_2}} \exp\left(\alpha_{f_1}\left(\frac{\bar{\xi}_{G_Y}}{h_{2,1}}\right)^{\beta_{f_1}} + \alpha_{f_2}\left(\frac{\bar{\xi}_{G_X}}{h_{2,2}}\right)^{\beta_{f_2}}\right)\right) \\
&= o\left(\left(h_{2,1}^{-1}\right)^{1+\lambda_{2,1}} (h_{2,2}^{-1})^{1+\lambda_{2,2}} n^{-1/2+\epsilon}\right),
\end{aligned}$$

where the last inequality holds under Assumption 1.15. Combining results from above, I have that for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{(y,x)\in\mathbb{S}_(Y,X)} \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \left|\hat{\phi}_{f^{(\lambda_2)}_{Y,X,W_2}(y,x,\cdot)}(\xi) - \phi_{f^{(\lambda_2)}_{Y,X,W_2}(y,x,\cdot)}(\xi)\right| \\
&= \sup_{(y,x)\in\mathbb{S}_(Y,X)} \sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]} \left| \hat{E}\left[e^{\mathbf{i}\xi W_2} \frac{1}{h_{2,1}^{1+\lambda_{2,1}} h_{2,2}^{1+\lambda_{2,2}}} G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right) G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]\right.
\end{aligned}$$

$$
-E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]\Bigg|
$$

$$
+\sup_{(y,x)\in\mathbb{S}_(Y,X)}\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\Bigg|E\left[e^{\mathbf{i}\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y^{(\lambda_{2,1})}\left(\frac{y-Y}{h_{2n,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2n,2}}\right)\right]
$$

$$
-\phi_{f_{Y,X,W_2}^{(\lambda_2)}(y,x,\cdot)}(\xi)\Bigg|
$$

$$
=o_p\left(\left(h_{2,1}^{-1}\right)^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}n^{-1/2+\epsilon}\right).
$$

I define $\Upsilon(h_1)$ and $\hat{\Phi}_n$ as

$$
\begin{aligned}
\Upsilon(h_{1n}) &\equiv (1+h_{1n}^{-1})\left(\sup_{\xi\in[-h^{-1},h^{-1}]}\frac{|\phi_{W^*}^{'}(\xi)|}{|\phi_{W^*}(\xi)|}\right)\left(\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\phi_{W_2}(\xi)\right|^{-1}\right)\\
&= O\left((1+h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)\\
\hat{\Phi}_{n,\lambda_2} &\equiv \max\Bigg\{\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{\theta}(\xi)-\theta(\xi)\right|,\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{\phi}_{W_2}(\xi)-\phi_{W_2}(\xi)\right|,\\
&\qquad \sup_{(y,x)\in\mathbb{S}_(Y,X)}\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{\phi}_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)-\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(\xi)\right|\Bigg\}\\
&= o_p\left((h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}n^{-1/2+\epsilon}\right)
\end{aligned}
$$

for any $\epsilon>0$. The latter order of magnitude follows from Lemma 1.17 and Assumption 1.15 and 1.14. Then $R_{1,\lambda_1,\lambda_{2,1},\lambda_{2,2}}-R_{7,\lambda_1,\lambda_{2,1},\lambda_{2,2}}$ can be bounded in terms of $\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h_{1n})$, $\Upsilon(h_{1n})$, and $\hat{\Phi}_{n,\lambda_2}$. Note that under Assumption 1.15,

$$
\begin{aligned}
\sup_{\xi\in[-h_{1n}^{-1},h_{1n}^{-1}]}\frac{\hat{\Phi}_{n,\lambda_2}}{|\phi_{W_2}(\xi)|} &\preceq \hat{\Phi}_{n,\lambda_2}\Upsilon(h_{1n})\\
&= o_p\left((h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}n^{-1/2+\epsilon}\right)O\left((1+h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)\\
&= o_p(1).
\end{aligned}
$$

Now I have

$$
R_1\le 2\int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1t)\right|\left(\frac{1}{\left|\phi_{W_2}(t)\right|}+\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|^2}\right)\hat{\Phi}_{n,\lambda_2}\left|\phi_{W^*}(t)\right|\left(\int_0^t\left|\delta_1\hat{q}_{W_1}(\xi)\right|d\xi\right)dt
$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}\int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left(1+\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|}\right)\left|\phi_{W^*}(t)\right|\left(\int_0^t \left|\delta_1\hat{q}_{W_1}(\xi)\right|d\xi\right)d\tau dt$$

$$= \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}\int_0^\infty \left[\int_\xi^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left(1+\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|}\right)\left|\phi_{W^*}(t)\right|dt\right]\left|\delta_1\hat{q}_{W_1}(\xi)\right|d\xi$$

$$= \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}\int_0^\infty \left[\int_\xi^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left(\left|\phi_{W^*}(t)\right|+\left|\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|\right)dt\right]\left|\delta_1\hat{q}_{W_1}(\xi)\right|d\xi$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\int_0^\infty \left[\int_\xi^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left(\left|\phi_{W^*}(t)\right|+\left|\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|\right)dt\right]$$

$$\times\left(1+\frac{\left|\theta(\xi)\right|}{\left|\phi_{W_2}(\xi)\right|}\right)\frac{1}{\left|\phi_{W_2}(\xi)\right|}d\xi$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h)$$

$$= o_p\left(h_{2,1}^{-2}(h_{2,2}^{-1})^{2+2\lambda_2}n^{-1+2\epsilon}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)\times$$

$$O\left(\max\left\{(h_1^{-1})^{1+\gamma_*},(h_{2,1}^{-1})^{1+\lambda_{2,1}}(h_{2,2}^{-1})^{1+\lambda_{2,2}}\right\}(h_1^{-1})^{1-\gamma_2+\gamma_\phi+\lambda_1}\exp\left(\left(\alpha_\phi\mathbf{1}\{\beta_\phi=\beta_2\}-\alpha_2\right)\left(h_1^{-1}\right)^{\beta_2}\right)\right).$$

If further Assumption 1.15 holds, I just need to show that

$$o_p\left(h_{2,1}^{-2}(h_{2,2}^{-1})^{2+2\lambda_2}n^{-1/2+2\epsilon}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)=o_p(1).$$

If $\beta_2\neq 0$, $h_{1n}^{-1}=O\left((\ln n)^{1/\beta_2-\eta}\right)$, $h_{2n,2}^{-1}=O\left(n^{(8+4\lambda_2)^{-1}-\eta}\right)$, so that

$$o_p\left(n^{-1/2+2\epsilon}h_{2,1}^{-2}(h_{2,2}^{-1})^{2+2\lambda_2}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)$$

$$= o_p\left(n^{-1/2+2\epsilon}n^{1/2-2\eta(2+\lambda_2)}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(h_{1n}^{-1}\right)^{\beta_2}\right)\right)$$

$$= o_p\left(n^{2\epsilon-2\eta(2+\lambda_2)}((\ln n)^{1/\beta_2-\eta})^{1+\gamma_*-\gamma_2}\exp\left(-\alpha_2\left(\ln n\right)^{1-\eta\beta_2}\right)\right)$$

$$= o_p\left(\exp\left[-\alpha_2\left(\ln n\right)^{1-\eta\beta_2}+(2\epsilon-2\eta(2+\lambda_2))\ln n+(1+\gamma_*-\gamma_2)\left(1/\beta_2-\eta\right)\ln\left(\ln n\right)\right]\right)$$

$$= o_p(1),$$

where the equality follows since $(\ln n)^{1-\eta\beta_2}$ and $\ln(\ln n)$ are dominated by $\ln n$ and by picking

$\epsilon < \eta(2+\lambda_2)$.

If $\beta_2 = 0$, $h_{1n}^{-1} = O\left(n^{(4+4\gamma_*-4\gamma_2)^{-1}-\eta}\right)$, $h_{2n,2}^{-1} = O\left(n^{(16+8\lambda_2)^{-1}-\eta}\right)$ and with $\epsilon < \eta$,

$$\begin{aligned}
&o_p\left(n^{-1/2+2\epsilon}(h_{1n}^{-1})^{1+\gamma_*-\gamma_2}(h_{2,1}^{-2}(h_{2,2}^{-1})^{2+2\lambda_2}\right)\\
&= o_p\left(n^{-1/2+2\epsilon}(n^{(4+4\gamma_*-4\gamma_2)^{-1}-\eta})^{1+\gamma_*-\gamma_2}n^{1/4-2\eta(2+\lambda_2)}\right)\\
&= o_p\left(n^{2\epsilon-\eta(5+\gamma_*-\gamma_2+2\lambda_2)}\right)\\
&= o_p(1).
\end{aligned}$$

The remaining terms are similarly bounded

$$\begin{aligned}
R_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}} \leq& \int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right| \left|\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|^2}\frac{1}{\left|\phi_{W_2}(t)\right|}\hat{\Phi}_{n,\lambda_2}^2\left|1+o_p(1)\right|^{-1}\right.\\
&\left.+\frac{1}{|\phi_{W_2}(t)|^2}\hat{\Phi}_{n,\lambda_2}^2\left|1+o_p(1)\right|^{-1}\right|\left|\phi_{W^*}(t)\right|dt\\
&\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\left|1+o_p(1)\right|^{-1}\int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\frac{1}{\left|\phi_{W_2}(t)\right|}\left|\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|}+1\right|\left|\phi_{W^*}(t)\right|dt\\
&\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\left|1+o_p(1)\right|^{-1}\left(\int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\frac{\left|\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|}dt\right.\\
&\left.+\int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\frac{\left|\phi_{W^*}(t)\right|}{\left|\phi_{W_2}(t)\right|}dt\right)\\
&\preceq \Upsilon(h_1)\hat{\Phi}_{n,,\lambda_2}^2\Psi_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}^+(h)\left(1+o_p(1)\right)^{-1};\\
R_{3,\lambda_1,\lambda_{2,1},\lambda_{2,2}} \preceq& \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}R_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}} = o_p(1)R_{2,\lambda_1,\lambda_{2,1},\lambda_{2,2}};\\
R_{4,\lambda_1,\lambda_{2,1},\lambda_{2,2}} =& \int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left|\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|\left(\int_0^t\left|\delta_2\hat{q}_{W_1}(\xi)\right|d\xi\right)dt\\
&\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\left|1+o_p(1)\right|^{-1}\int_0^\infty\frac{\int_\xi^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left|\phi_{f_{Y,X,W^*}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|dt}{\left|\phi_{W_2}(\xi)\right|}d\xi\\
&= \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2}^2\Psi_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}^+(h)(1+o_p(1));\\
R_{5,\lambda_1,\lambda_{2,1},\lambda_{2,2}} \leq& \int_0^\infty |t|^{\lambda_1}\left|\phi_K(h_1 t)\right|\left(\frac{1}{\left|\phi_{W_2}(t)\right|}+\frac{\left|\phi_{f_{Y,X,W_2}^{(\lambda_{2,1},\lambda_{2,2})}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|^2}\right)
\end{aligned}$$

$$\times \hat{\Phi}_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \left|\phi_{W^*}(t)\right| \left(\int_0^t \left|\delta_2 \hat{q}_{W_1}(\xi)\right| d\xi\right) dt$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \int_0^\infty \left|\phi_K(h_1 t)\right| \left(\left|\phi_{W^*}(t)\right| + \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)$$

$$\times \left(\int_0^t \left|\delta_2 \hat{q}_{W_1}(\xi)\right| d\xi\right) d\tau dt$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2} \left(1+o_p(1)\right) R_{4,\lambda_1,\lambda_{2,1},\lambda_{2,2}} = o_p(1) R_{4,\lambda_1,\lambda_{2,1},\lambda_{2,2}};$$

$$R_{6,\lambda_1,\lambda_{2,1},\lambda_{2,2}} \le \int_0^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right| \frac{1}{2} \exp\left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right) \left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right)^2 dt$$

$$\le \frac{1}{2} \exp\left(o_p(1)\right) \int_0^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right| \left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right)^2 dt$$

$$\preceq \frac{1}{2} \exp\left(o_p(1)\right) \Upsilon(h_1)\hat{\Phi}^2_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \int_0^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|$$

$$\times \left(\int_0^t \frac{1}{\left|\phi_{W_2}(\xi)\right|} + \frac{\left|\theta(\xi)\right|}{\left|\phi_{W_2}(\xi)\right|^2} d\xi\right) dt$$

$$\preceq \frac{1}{2} \exp\left(o_p(1)\right) \Upsilon(h_1)\hat{\Phi}^2_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \int_0^\infty \left(\int_\xi^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right| dt\right)$$

$$\times \left(\frac{1}{\left|\phi_{W_2}(\xi)\right|} + \frac{\left|\theta(\xi)\right|}{\left|\phi_{W_2}(\xi)\right|^2}\right) d\xi$$

$$\preceq O_p(1)\Upsilon(h_1)\hat{\Phi}^2_{n,\lambda_2} \Psi^+_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(h)$$

$$R_{7,\lambda_1,\lambda_{2,1},\lambda_{2,2}} \le \int_0^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left(1 + \frac{\left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W_2}(y,x,\cdot)}(t)\right|}{\left|\phi_{W_2}(t)\right|}\right) \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \left|\phi_{W^*}(t)\right|$$

$$\times \frac{1}{2} \exp\left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right) \left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right)^2 dt$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} \int_0^\infty |t|^{\lambda_1} \left|\phi_K(h_1 t)\right| \left(\left|\phi_{W^*}(t)\right| + \left|\phi_{f^{(\lambda_{2,1},\lambda_{2,2})}_{Y,X,W^*}(y,x,\cdot)}(t)\right|\right)$$

$$\times \frac{1}{2} \exp\left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right) \left(\int_0^t \left|\delta \hat{q}_{W_1}(\xi)\right| d\xi\right)^2 dt$$

$$\preceq \Upsilon(h_1)\hat{\Phi}_{n,\lambda_2} \left|1+o_p(1)\right|^{-1} R_{6,\lambda_1,\lambda_{2,1},\lambda_{2,2}} = o_p(1) R_{6,\lambda_1,\lambda_{2,1},\lambda_{2,2}}.$$

□

Proof of Theorem 1.4. In Lemma 1.12, let $Z = (X, W^*)$. By (1.9), for any value $(y, x, w^*) \in \mathbb{S}_{Y,X,W^*}$, and any value $\delta \in [0,1]$, I have that $\mathbf{WLAR}(x) = \tilde{\Gamma}(f)$ and that

$\rho(y,x) = \tilde{\Xi}(f)$. (i) By Lemma 1.12,

$$\begin{aligned}
&\sup_{(y,x,w^*)\in\mathbb{S}_\tau}\left|\widehat{\rho(y,x)} - \rho(y,x)\right| \\
&\leq \frac{O_p\left(\epsilon_{n,0}\right) + O_p\left(\tilde{\epsilon}_{n,0,0,0}\right)}{\tau^2} + O_p\left(\epsilon_{n,0}\right) + O_p\left(\tilde{\epsilon}_{n,0,0,1}\right) + \frac{O_p\left(\epsilon_{n,1}\right) + O_p\left(\tilde{\epsilon}_{n,1,0,0}\right)}{\tau^2} \\
&\leq O_p\left(\tilde{\epsilon}_{n,0,0,1}\right) + \frac{O_p\left(\epsilon_{n,1}\right) + O_p\left(\tilde{\epsilon}_{n,1,0,0}\right)}{\tau^2}.
\end{aligned}$$

The last inequality follows since $\epsilon_{n,0}$ is of smaller order than $\epsilon_{n,1}$, $\tilde{\epsilon}_{n,0,0,0}$ is of smaller order than $\tilde{\epsilon}_{n,1,0,0}$. Choose τ_n such that $\tau_n > 0, \tau_n \to 0$ as $n \to \infty$, and that $\frac{\epsilon_{n,1}}{\tau^2} \to 0$ and $\frac{\tilde{\epsilon}_{n,1,0,0}}{\tau^2} \to 0$, I can get the desired result.

(ii) A direct application of Lemma 1.13 gives the desired result. □

1.7.3 Appendix C: Asymptotic Normality

For asymptotic normality, I need to place a lower bound on $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$ relative to $B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$ and $R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*.h_n)$. The following assumption is stated at a high level. More primitive sufficient conditions can be derived using techniques of Schennach (2004b). Combining this assumption with the results from Theorem 1.1,1.2 and 1.3 yields a corollary establishing the asymptotic normality of $\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)$.

Assumption 1.16. *(High Level) For given $\lambda_1, \lambda_{2,1}, \lambda_{2,2} \in \{0,1\}$ and given $j \in \{1,2\}$, $h_n \to 0$ at a rate such that for each $(y,x,w^*) \in \mathbb{S}_{(Y,X,W^*)}$ such that $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) > 0$ for all n sufficiently large, I have $n^{1/2}\left(\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right)^{-1/2}\left|B_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right| \xrightarrow{p} 0$, and $n^{1/2}\left(\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right)^{-1/2}\left|R_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right| \xrightarrow{p} 0$.*

Corollary 2. *If the conditions of Theorem 1.2 and Assumption 1.16 hold, then for each $(y,x,w^*) \in \mathbb{S}_\tau$ such that $\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) > 0$ for all n sufficiently large I have*

$$n^{1/2}\left(\Omega_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n)\right)^{-1/2}\left(\hat{g}_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*,h_n) - g_{\lambda_1,\lambda_{2,1},\lambda_{2,2}}(y,x,w^*)\right) \xrightarrow{d} N(0,1).$$

In the next few lemmas, I show the asymptotic normality of $\widehat{\rho(y,x)}$ under some high-level assumptions. The asymptotic normality of $\widehat{\mathbf{WLAR}}(x)$ can be shown following the same logic. I first state a Lemma in a similar fashion as Lemma 1.5

Lemma 1.18. *For any value $(y,x,w^*) \in \mathbb{S}_\tau$, I have that*

$$\widehat{\rho(y,x)} - \rho(y,x) = BD\tilde{\Xi}_n + LD\tilde{\Xi}_n + RD\tilde{\Xi}_n + R\tilde{\Xi}_n,$$

where $R\tilde{\Xi}_n$ corresponds to the nonlinear part of functional $\tilde{\Xi}$; $BD\tilde{\Xi}_n$ and $RD\tilde{\Xi}_n$ are the bias term contained in the linear part of functional $\tilde{\Xi}$; $LD\tilde{\Xi}_n \equiv \hat{E}\left[lD\tilde{\Xi}_n\right]$ is the linear term[5] *contained in the linear part of functional $\tilde{\Xi}$, and is equal to $\hat{E}\left[\sum_{k=1}^{11} \tilde{s}_k(y,x,w^*)lD\tilde{\Xi}_{n,k}\right]$ in which*

$$\begin{aligned}
lD\tilde{\Xi}_{n,1} &= \int l_{0,0,0}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,2} &= \int_{-\infty}^{y} l_{0,0,0}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,3} &= \int l_{0,0,1}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,4} &= \int_{-\infty}^{y} l_{0,0,1}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,5} &= l_{0,0,0}(y,x,w^*,h_n) \\
lD\tilde{\Xi}_{n,6} &= \int l_{1,0,0}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,7} &= \int_{-\infty}^{y} l_{1,0,0}(y,x,w^*,h_n)dy \\
lD\tilde{\Xi}_{n,8} &= \int \tilde{l}_{0,0}(s,w^*,h_n)ds \\
lD\tilde{\Xi}_{n,9} &= \int_{-\infty}^{x} \tilde{l}_{0,0}(s,w^*,h_n)ds \\
lD\tilde{\Xi}_{n,10} &= \int \tilde{l}_{1,0}(s,w^*,h_n)ds \\
lD\tilde{\Xi}_{n,11} &= \int_{-\infty}^{x} \tilde{l}_{1,0}(s,w^*,h_n)ds.
\end{aligned}$$

[5] Linear in terms of $\hat{E}\left[W_1 e^{i\xi W_2}\right]$, $\hat{E}\left[e^{i\xi W_2}\right]$ and $\hat{E}\left[e^{i\xi W_2}\frac{1}{h_{2,1}^{1+\lambda_{2,1}}h_{2,2}^{1+\lambda_{2,2}}}G_Y\left(\frac{y-Y}{h_{2,1}}\right)G_X^{(\lambda_{2,2})}\left(\frac{x-X}{h_{2,2}}\right)\right]$

$$\tilde{s}_1(y,x,w^*) = -\frac{F_{Y|X=x,W^*=w^*}(y)}{f_{X,W^*}(x,w^*)f_{Y,X,W^*}(y,x,w^*)}\left(\frac{\partial f_{X,W^*}(x,w^*)}{\partial x} + \frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*}\frac{1}{\tilde{\Psi}_{(1)}(f)}\right)$$
$$- \frac{F_{X|W^*=w^*}(x)\frac{\partial f_{W^*}(w^*)}{\partial w^*} - \int_{-\infty}^{x}\frac{\partial f_{X,W^*}(s,w^*)}{\partial w^*}ds}{\tilde{f}^2(x,w^*)} \times \frac{\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}$$
$$\tilde{s}_2(y,x,w^*) = \frac{1}{f_{X,W^*}(x,w^*)f_{Y,X,W^*}(y,x,w^*)}\left(\frac{\partial f_{X,W^*}(x,w^*)}{\partial x} + \frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*}\frac{1}{\tilde{\Psi}_{(1)}(f)}\right)$$
$$\tilde{s}_3(y,x,w^*) = \frac{F_{Y|X=x,W^*=w^*}(y)}{f_{Y,X,W^*}(y,x,w^*)}$$
$$\tilde{s}_4(y,x,w^*) = -\frac{1}{f_{Y,X,W^*}(y,x,w^*)}$$
$$\tilde{s}_5(y,x,w^*) = -\frac{1}{f^2(y,x,w^*)}\left[F_{Y|X=x,W^*=w^*}(y)\frac{\partial f_{X,W^*}(x,w^*)}{\partial x} - \int_{-\infty}^{y}\frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial x}dy\right.$$
$$\left. + \left(F_{Y|X=x,W^*=w^*}(y)\frac{\partial f_{X,W^*}(x,w^*)}{\partial w^*} - \int_{-\infty}^{y}\frac{\partial f_{Y,X,W^*}(y,x,w^*)}{\partial w^*}dy\right)\frac{1}{\tilde{\Psi}_{(1)}(f)}\right]$$
$$\tilde{s}_6(y,x,w^*) = \frac{F_{Y|X=x,W^*=w^*}(y)}{f_{Y,X,W^*}(y,x,w^*)}\frac{1}{\tilde{\Psi}_{(1)}(f)}$$
$$\tilde{s}_7(y,x,w^*) = -\frac{1}{f_{Y,X,W^*}(y,x,w^*)}\frac{1}{\tilde{\Psi}_{(1)}(f)}$$
$$\tilde{s}_8(y,x,w^*) = -\frac{F_{X|W^*=w^*}(x)\frac{\partial f_{W^*}(w^*)}{\partial w^*}}{f_{W^*}(w^*)f_{X,W^*}(x,w^*)} \times \frac{\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}$$
$$\tilde{s}_9(y,x,w^*) = \frac{\frac{\partial f_{W^*}(w^*)}{\partial w^*}}{f_{W^*}(w^*)f_{X,W^*}(x,w^*)} \times \frac{\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}$$
$$\tilde{s}_{10}(y,x,w^*) = \frac{F_{X|W^*=w^*}(x)}{f_{X,W^*}(x,w^*)} \times \frac{\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)}$$
$$\tilde{s}_{11}(y,x,w^*) = -\frac{1}{f_{X,W^*}(x,w^*)} \times \frac{\Psi_{(2)}(f)}{\tilde{\Psi}^2_{(1)}(f)},$$

where $\alpha(f) \equiv F^{-1}_{Y|X=x,W^*=w^*}(\delta)$, $\Phi_{(2)}(f) \equiv \frac{\partial F^{-1}_{Y|X=x,W^*=w^*}(\delta)}{\partial w^*}$, $\Psi_{(2)}(f) \equiv -\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial w^*}\Big/\frac{\partial F_{Y|X=x,W^*=w^*}(y)}{\partial y}$, *and* $\tilde{\Psi}_{(1)}(f) = -\frac{\partial F_{X|W^*=w^*}(x)}{\partial w^*}\Big/\frac{\partial F_{X|W^*=w^*}(x)}{\partial x}$.

Proof of Lemma 1.18. By (1.11), for any value $(y,x,w^*) \in \mathbb{S}_\tau$, I have that

$$\widehat{\rho(y,x)} - \rho(y,x) = D\tilde{\Xi}_n + R\tilde{\Xi}_n.$$

By Lemma 1.12,

$$D\tilde{\Xi}_n = \sum_{k=1}^{11}\tilde{s}_k(x,w^*,\delta)D\tilde{\Xi}_{n,k},$$

where $\tilde{s}_k(x, w^*, \delta)$ is as stated in the Lemma. and

$$\begin{aligned}
D\tilde{\Xi}_{n,1} &= \int \Big(\hat{g}_{0,0,0}(y, x, w^*, h_n) - g_{0,0,0}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,2} &= \int_{-\infty}^{y} \Big(\hat{g}_{0,0,0}(y, x, w^*, h_n) - g_{0,0,0}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,3} &= \int \Big(\hat{g}_{0,1,1}(y, x, w^*, h_n) - g_{0,1,1}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,4} &= \int_{-\infty}^{y} \Big(\hat{g}_{0,1,1}(y, x, w^*, h_n) - g_{0,1,1}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,5} &= \hat{g}_{0,0,0}(y, x, w^*, h_n) - g_{0,0,0}(y, x, w^*, h_n) \\
D\tilde{\Xi}_{n,6} &= \int \Big(\hat{g}_{1,0,0}(y, x, w^*, h_n) - g_{1,0,0}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,7} &= \int_{-\infty}^{y} \Big(\hat{g}_{1,0,0}(y, x, w^*, h_n) - g_{1,0,0}(y, x, w^*, h_n)\Big)\, dy \\
D\tilde{\Xi}_{n,8} &= \int \Big(\hat{\tilde{g}}_{0,0}(s, w^*, h_n) - \tilde{g}_{0,0}(s, w^*, h_n)\Big)\, ds \\
D\tilde{\Xi}_{n,9} &= \int_{-\infty}^{x} \Big(\hat{\tilde{g}}_{0,0}(s, w^*, h_n) - \tilde{g}_{0,0}(s, w^*, h_n)\Big)\, ds \\
D\tilde{\Xi}_{n,10} &= \int \Big(\hat{\tilde{g}}_{1,0}(s, w^*, h_n) - \tilde{g}_{1,0}(s, w^*, h_n)\Big)\, ds \\
D\tilde{\Xi}_{n,11} &= \int_{-\infty}^{x} \Big(\hat{\tilde{g}}_{1,0}(s, w^*, h_n) - \tilde{g}_{1,0}(s, w^*, h_n)\Big)\, ds.
\end{aligned}$$

By Lemma 1.5, a first order expansion of $D\tilde{\Xi}_n$ can be written as

$$\begin{aligned}
D\tilde{\Xi}_n &= BD\tilde{\Xi}_n + LD\tilde{\Xi}_n + RD\tilde{\Xi}_n \\
&= BD\tilde{\Xi}_n + \sum_{k=1}^{11} \tilde{s}_k(x, w^*, \delta) LD\tilde{\Xi}_{n,k} + RD\tilde{\Xi}_n
\end{aligned}$$

where

$$\begin{aligned}
LD\tilde{\Xi}_{n,1} &= \int L_{0,0,0}(y, x, w^*, h_n) dy = \int \hat{E}\Big[l_{0,0,0}(y, x, w^*, h_n)\Big]\, dy = \hat{E}\left[\int l_{0,0,0}(y, x, w^*, h_n) dy\right] \\
LD\tilde{\Xi}_{n,2} &= \int_{-\infty}^{y} L_{0,0,0}(y, x, w^*, h_n) dy = \int_{-\infty}^{y} \hat{E}\Big[l_{0,0,0}(y, x, w^*, h_n)\Big]\, dy \\
&= \hat{E}\left[\int_{-\infty}^{y} l_{0,0,0}(y, x, w^*, h_n) dy\right] \\
LD\tilde{\Xi}_{n,3} &= \int L_{0,0,1}(y, x, w^*, h_n) dy = \int \hat{E}\Big[l_{0,0,1}(y, x, w^*, h_n)\Big]\, dy
\end{aligned}$$

$$
\begin{aligned}
&= \hat{E}\left[\int l_{0,0,1}(y,x,w^*,h_n)dy\right] \\
LD\tilde{\Xi}_{n,4} &= \int_{-\infty}^{y} L_{0,0,1}(y,x,w^*,h_n)dy = \int_{-\infty}^{y} \hat{E}\left[l_{0,0,1}(y,x,w^*,h_n)\right] dy \\
&= \hat{E}\left[\int_{-\infty}^{y} l_{0,0,1}(y,x,w^*,h_n)dy\right] \\
LD\tilde{\Xi}_{n,5} &= L_{0,0,0}(y,x,w^*,h_n) = \hat{E}\left[l_{0,0,0}(y,x,w^*,h_n)\right] \\
LD\tilde{\Xi}_{n,6} &= \int L_{1,0,0}(y,x,w^*,h_n)dy = \int \hat{E}\left[l_{1,0,0}(y,x,w^*,h_n)\right] dy \\
&= \hat{E}\left[\int l_{1,0,0}(y,x,w^*,h_n)dy\right] \\
LD\tilde{\Xi}_{n,7} &= \int_{-\infty}^{y} L_{1,0,0}(y,x,w^*,h_n)dy = \int_{-\infty}^{y} \hat{E}\left[l_{1,0,0}(y,x,w^*,h_n)\right] dy \\
&= \hat{E}\left[\int_{-\infty}^{y} l_{1,0,0}(y,x,w^*,h_n)dy\right] \\
LD\tilde{\Xi}_{n,8} &= \int \tilde{L}_{0,0}(s,w^*,h_n)ds = \int \hat{E}\left[\tilde{l}_{0,0}(s,w^*,h_n)\right] ds = \hat{E}\left[\int \tilde{l}_{0}(s,w^*,h_n)ds\right] \\
LD\tilde{\Xi}_{n,9} &= \int_{-\infty}^{x} \tilde{L}_{0,0}(s,w^*,h_n)ds = \int_{-\infty}^{x} \hat{E}\left[\tilde{l}_{0,0}(s,w^*,h_n)\right] ds = \hat{E}\left[\int_{-\infty}^{x} \tilde{l}_{0,0}(s,w^*,h_n)ds\right] \\
LD\tilde{\Xi}_{n,10} &= \int \tilde{L}_{1,0}(s,w^*,h_n)ds = \int \hat{E}\left[\tilde{l}_{1,0}(s,w^*,h_n)\right] ds = \hat{E}\left[\int \tilde{l}_{1,0}(s,w^*,h_n)ds\right] \\
LD\tilde{\Xi}_{n,11} &= \int_{-\infty}^{x} \tilde{L}_{1,0}(s,w^*,h_n)ds = \int_{-\infty}^{x} \hat{E}\left[\tilde{l}_{1,0}(s,w^*,h_n)\right] ds = \hat{E}\left[\int_{-\infty}^{x} \tilde{l}_{1,0}(s,w^*,h_n)ds\right].
\end{aligned}
$$

Then I can write

$$
\begin{aligned}
LD\tilde{\Xi}_n &= \hat{E}\left[lD\tilde{\Xi}_n\right] \\
&= \hat{E}\left[\sum_{k=1}^{11} \tilde{s}_k(x,w^*,\delta) lD\tilde{\Xi}_{n,k}\right],
\end{aligned}
$$

where $lD\tilde{\Xi}_n$ are as stated in the Lemma. □

Lemma 1.19. *Suppose the conditions of Lemma 1.5 hold. (i) Then for each* $(y,x,w^*) \in \mathbb{S}_\tau$, $E[LD\tilde{\Xi}_n] = 0$, *and if Assumption 1.12 also holds, then* $E[LD\tilde{\Xi}_n^2] = n^{-1}\tilde{\Omega}(y,x,w^*,h)$, *where* $\tilde{\Omega}(y,x,w^*,h) \equiv E\left[lD\tilde{\Xi}_n^2\right] < \infty$.

(ii) If Assumption 1.12 also holds, and if for each $(y,x,w^*) \in \mathbb{S}_\tau$, $\tilde{\Omega}(y,x,w^*,h_n) > 0$

for all n sufficiently large, then for each $(y,x,w^*) \in \mathbb{S}_\tau$

$$n^{1/2}\left(\tilde{\Omega}(y,x,w^*,h_n)\right)^{-1/2} LD\tilde{\Xi}_n \xrightarrow{d} N(0,1).$$

Proof of Lemma 1.19. (i) The fact that $E\left[LD\tilde{\Xi}_n\right] = 0$ follows directly from Theorem 1.2. Next I show that for all $(y,x,w^*) \in \mathbb{S}_\tau$, $\tilde{\Omega}(y,x,w^*,h) < \infty$. It follows by Cauchy Schwartz that

$$\begin{aligned}\tilde{\Omega}(y,x,w^*,h_n) &\equiv E\left[lD\tilde{\Xi}_n^2\right] \\ &\preceq \sum_{k=1}^{11} \tilde{s}_k^2(y,x,w^*)E\left[lD\tilde{\Xi}_{n,k}^2\right].\end{aligned}$$

Note that

$$\begin{aligned}E\left[lD\tilde{\Xi}_{n,1}^2\right] &= E\left[\left(\int l_{0,0,0}(y,x,w^*,h_n)dy\right)^2\right] \\ &\leq E\left[\int \left(l_{0,0,0}(y,x,w^*,h_n)\right)^2 dy\right] \\ &= \int E\left[\left(l_{0,0,0}(y,x,w^*,h_n)\right)^2\right] dy\end{aligned}$$

where the second line follows from Jensen's Inequality and the third line follows from Tonelli's Theorem. By Theorem 1.2, $\int E\left[\left(l_{0,0,0}(y,x,w^*,h)\right)^2\right] dy < \infty$ for each h. Conclusions for other values of $k \in \{1,2,3,...,11\}$ follows similarly. Then by Assumption 1.6, I have that $\tilde{\Omega}(y,x,w^*,h) < \infty$ for all values $(y,x,w^*) \in \mathbb{S}_\tau$.

(ii) To show asymptotic normality, I apply Lemma 1.16 to $lD\tilde{\Xi}_n = \sum_{k=1}^{11} \tilde{s}_k(x,w^*,\delta)lD\tilde{\Xi}_{n,k}$. By previous argument, $\tilde{\Omega}(y,x,w^*,h) < \infty$ for all values $(y,x,w^*) \in \mathbb{S}_\tau$. I've assumed that for n sufficiently large, $\tilde{\Omega}(y,x,w^*,h_n) > 0$. The conclusion follows. □

Finally, I state the asymptotic normality result for the estimator of $\rho(y,x)$ under a high-level assumption:

Assumption 1.17. *(High Level) For given $\lambda_1, \lambda_{2,1}, \lambda_{2,2} \in \{0,1\}$ and given $j \in \{1,2\}$, $h_n \to 0$ at a rate such that for each $(y, x, w^*) \in \mathbb{S}_\tau$ such that $\tilde{\Omega}(y, x, w^*, h_n) > 0$ for all n sufficiently large, I have $n^{1/2}\left(\tilde{\Omega}(y, x, w^*, h_n)\right)^{-1/2}\left|BD\tilde{\Xi}_n\right| \xrightarrow{p} 0$ and $n^{1/2}\left(\tilde{\Omega}(y, x, w^*, h_n)\right)^{-1/2}\left|RD\tilde{\Xi}_n\right| \xrightarrow{p} 0$.*

Theorem 1.5. *If the conditions of Theorem 1.2 and Assumption 1.16, 1.17 hold, then for each $(y, x, w^*) \in \mathbb{S}_\tau$ such that $\tilde{\Omega}(y, x, w^*, h_n) > 0$ for all n sufficiently large I have*

$$n^{1/2}\left(\tilde{\Omega}(y, x, w^*, h_n)\right)^{-1/2}\left(\widehat{\rho(y,x)} - \rho(y, x)\right) \xrightarrow{d} N(0, 1).$$

Chapter 2

Two-step Estimation of Network Formation Models with Unobserved Heterogeneities and Strategic Interactions

2.1 Introduction

The social network to which an individual belongs can be an important element when studying many economic behaviors of such an individual. Peer effects in education and crime, the dynamics of product adoption, and financial contagions are just a few examples. However, most network studies of these behaviors are challenged by the endogeneity of the network. This highlights the importance of developing econometric models of network formation. Moreover, the network formation process is itself an interesting subject to study, since it helps us better understand people's real-life activities such as interactions on social apps.

There are two features that are important in a network formation model. First, the incentives of forming a link in a network not only include the two agents' characteristics but also the linking decisions of other agents, such as the "popularity effect", meaning that an agent i is more likely to link to j if j has many other friends. It's important to include such strategic interaction effects in a network formation model. Second, some individual-specific characteristics affecting the utilities and thus linking choices are private information (like people's personalities when using a dating app). They are correlated with observed characteristics but are unobserved to other agents or researchers. It is thus important to incorporate these unobserved individual-specific characteristics in the network formation process. Motivated by the two features, I study a directed network formation model with individual-specific unobserved heterogeneities and strategic interactions. I allow the incremental utility of a link from person i to j to depend on the linking choices of the person j to capture the popularity effect. I also allow the unobserved individual-specific heterogeneities to be correlated with observables by introducing individual fixed effects. I don't require that the conditional distribution of the individual fixed effects be known to the researchers.

There's growing literature on the estimation of network formation models. Among them, this paper is most related to Leung (2015) and Ridder and Sheng (2022). Both of them study the estimation of network formation games with incomplete information and strategic interactions and assume that the private information is independent of observed characteristics. In Leung (2015), the payoff depends on network structure in a separable way, through the sum of incremental utilities from each link. Then the optimal link choices are myopic, in the sense that an agent chooses to form a link with another member if the expected utility of forming that link is greater than 0. To be specific, let G_{ij} denote the linking proposal from individual i to j, and let X_i, X_j denote observed characteristics of the two individuals; let ϵ_{ij} denote unobserved link-specific characteristics that are independent with X. Leung (2015)'s model yields the following optimal linking decision:

$$G_{ij} = \mathbf{1}\left\{w(X_i, X_j)\beta_0 + \mathbb{E}[G_{-ij}|X, \sigma] + \varepsilon_{ij} \geqslant 0\right\}, \tag{2.1}$$

where w is a known function capturing the homophily effect, and σ denotes the equilibrium. Ridder and Sheng (2022) considers a more general case in which the utility function depends on the choice of potential partners in a non-separable way, for example, allowing the utility to depend on links-in-common. Using the Legendre transform, they show that even under this general case, the optimal linking choice is still equivalent to a sequence of myopic link choices. For estimation, both of the two papers assume that the data observed comes from a symmetric equilibrium, whereby agents with the same observable characteristics have the same equilibrium linking probabilities, i.e. $P(G_{ij}|X_{ij} = x, X) = P(G_{kl}|X_{kl} = x, X)$. Then the conditional linking probabilities can be estimated in the first step, by taking the empirical frequency with which agents with the same observable characteristics link to each other. In terms of strategic interactions, this paper adopts the same framework as Leung (2015), including only the popularity effect and keeping the dependence on the network structure to be separable, which is simpler than Ridder and Sheng (2022)'s framework. Different from the two papers, this paper studies the case when private information is correlated with observables by including individual fixed effects in the utility. For estimation, this paper also adopts a two-step procedure and estimates the realized equilibrium beliefs in the first step. This allows us to circumvent the difficulty to specify the equilibrium selection mechanism when there might be multiple equilibria.

This paper is also closely related to Graham (2017), which studies a network formation model with dyadic link formation. In their model, the linking decision between individual i and j only depends on the characteristics of i and j and there are no strategic interactions. Let A_i, A_j denote individual fixed effects unobserved to researchers. The linking decision in Graham (2017) is

$$G_{ij} = \mathbf{1}\left\{w(X_i, X_j)\beta_0 + A_i + A_j + \varepsilon_{ij} \geqslant 0\right\}. \tag{2.2}$$

Same as Graham (2017), this paper also incorporates unobserved individual fixed effects. The difference is that my model contains strategic interactions, so the information structure matters. I assume that individual fixed effects A_i are private information that is i.i.d. across individuals. The agents know the distribution of the individual fixed effects so that they can form beliefs of the expected "type"[1] of other people. From the modeling point of view, this paper studies a model which is a combination of (2.1) and (2.2). Note that a special case of (2.1) is when ϵ_{ij} can be written as the sum of an individual "random effect" A_i and an idiosyncratic error ν_{ij}. This is different from this paper's setting since A_i is assumed to be independent with X in Leung (2015).

Another strand of literature on estimating strategic network formation models assumes complete information, such as Miyauchi (2016) and Sheng (2020). These models are the hardest to deal with because they generally admit multiple equilibria and thus achieve set but not point identification of the model parameters. This paper shies away from these cases by assuming incomplete information.

The rest of the paper is organized as follows. In section 2, I develop the model and derive the optimal link choices. In section 3, I propose a two-step estimation procedure and show the consistency of the first-step estimator. In section 4, I show the asymptotic distribution of the estimators. In section 5, I conduct some Monte Carlo exercises to study the finite sample performance of my estimators. In the last section, I conclude.

2.2 The Model

I consider the directed network formation model in this paper. The formation process is a static game of incomplete information. An agent's payoff of forming a link depends on idiosyncratic private information. Given the belief of other people's linking decisions, agents form their own links simultaneously. Formally, the network formation game is set up as follows:

There are n agents indexed by $i \in \mathcal{I} = \{1, 2, ..., n\}$. Each agent chooses whether or not to link with the other $n-1$ agents. Player i's action vector $G_i = (G_{i1}, G_{i2}, ..., G_{ij}, ..., G_{in})'$ where $j \neq i$ is chosen from the action profile A which has 2^{n-1} components. The payoff

[1] I don't assume A_i to have discrete distribution, though.

function of individual i is

$$U_i(G, X, A_i, \varepsilon_i) = \sum_{j=1}^{n} G_{ij} \left(u_{ij}(G_{-i}, X, A_i; \beta) + \varepsilon_{ij} \right). \tag{2.3}$$

The deterministic part of incremental utility from link ij is specified as

$$u_{ij}(G_{-i}, X, A_i; \beta) = w(X_i, X_j)\beta_1 + A_i + G_{ji}\beta_2 + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk}\beta_3, \tag{2.4}$$

where the first term captures the homophily effect. w is a known function. $X = (X_1', ..., X_n')'$ is public information for all agents and is observable to researchers. For simplicity, write $w(X_i, X_j) = W_{ij}$ from now on. The second term A_i is individual-specific heterogeneity, which is unobserved both to other agents and researchers. Let $F_{A|X}$ be the distribution of A_i conditional on observables, which is assumed to be independent and identical across i, and known to all agents, but not necessarily known to researchers. A_i can be correlated with X. The third and the last term capture the popularity effect. The realization of $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{in})'$ is agent i's private information which is also unobserved to researchers. The model is therefore a static game with incomplete information, and the solution concept is Bayesian Nash Equilibrium. Different from Leung (2015), my model allows private information to be correlated with common information while doesn't require the conditional distribution of private information to be known to researchers. Also, I consider the "symmetric" equilibrium where pairs of agents with the same observable attributes and the same type A_i have the same conditional linking probabilities. This means that pairs of agents with the same observable attributes may have different choice probabilities if their (unobserved) types are different, which is different from Leung (2015).

For the above model, I impose the following assumptions:

Assumption 2.1. *(a)* $X_i \perp X_j$ *for* $i \neq j$. X_i *is discrete distributed with finite support* $\mathbb{X} = \{x_1, ..., x_{T_x}\}$. *(b)* A_i *are independently and identically distributed. The conditional CDF* $F_{A|X}$ *is known to all agents but unknown to researchers. (c)* ε_{ij} *are i.i.d. with logit* *distribution F_ε, which is known to both agents and researchers. (d) $\varepsilon_i \perp (X', A)'$ for all i.*

Let $\delta_j(X, A_j, \varepsilon_j)$ denote agent j's (pure) strategy. Let $\sigma_j(a|X, A_j) = Pr\left(\delta_j(X, A_j, \varepsilon_j) = a|X, A_j\right)$ denote the agent i's belief that agent j of type A_j chooses action a, given commonly known information X and agent i's private information. By Assumption 2.1 (b) and (c), actions G_i and G_j, $i \neq j$ are independent given commonly known attributes X. This fact simplifies the proof of consistency by weakening the correlation between links. Since agent i actually doesn't know the realization of A_j, his/her expected utility from choosing action $g_i \in S$ is $\sum_{g_{-i}} U_i(g_i, g_{-i}, X, A_i, \varepsilon_i)\mathbb{E}_{A_{-j}}\left[\sigma_{-i}(g_{-i}|X, A_{-j})\right]$. Therefore,

$$\begin{aligned} &Pr(G_i = g_i|X, A_i, \sigma) \\ &\quad = Pr\Bigg(\sum_{g_{-i}}\Big[U_i(g_i, g_{-i}, X, A_i, \varepsilon_i) - U_i(\tilde{g}_i, g_{-i}, X, A_i, \varepsilon_i)\Big]\mathbb{E}_{A_{-j}}\left[\sigma_{-i}(g_{-i}|X, A_{-j})\right] > 0, \\ &\qquad\qquad \forall \tilde{g}_i \in S \Bigg| X, A_i, \sigma\Bigg). \end{aligned}$$

A (Bayesian) equilibrium $\sigma^*(X, A_i)$ is a belief function that solves the fixed point equation:

$$\sigma_i^*(a|X, A_i) = Pr(G_i = a|X, A_i, \sigma^*)$$

for all $X \in \mathbf{X}$, agents $i \in \mathcal{I}$ and actions $a \in S$.

I consider "symmetric" equilibria in which pairs of agents with the same observable attributes and the same type (A_i) have the same conditional linking probabilities. For any $(X, A_i, \varepsilon_{ij})$ and "symmetric" belief profile σ in a neighborhood of an "symmetric" equilibrium σ^*, player i's optimal strategy $G_i(X, A_i, \varepsilon_i, \sigma) = \left(G_{ij}(X, A_i, \varepsilon_i, \sigma)\right)_{j\neq i}$ is given by:

$$G_{ij}(X, A_i, \varepsilon_i, \sigma) = \mathbf{1}\Big\{\mathbb{E}\Big[u_{ij}(G_{-i}, X, A_i; \beta)\Big|X, A_i, \sigma\Big] + \varepsilon_{ij} \geqslant 0\Big\}. \tag{2.5}$$

Assuming a symmetric equilibrium exists, the model is incomplete because there could be multiple equilibria for any realization of (X, A, ε). For completeness of the model, I specify the equilibrium selection mechanism in the following assumption. The mechanism, however, is not explicitly used in writing the likelihood function in part 3, because by using two-step estimation, I can avoid specifying the equilibrium theoretically. For the convenience of defining equilibrium selection mechanisms, I add subscript n to G, X, A, and ε. The equilibrium selection mechanism is a measurable function $\lambda_n : (X_n, \nu_n, \beta_0) \mapsto \sigma_n \in \mathcal{G}(X_n, A_n, \beta_0)$, where $\mathcal{G}(X_n, A_n, \beta_0)$ is the set of symmetric equilibria.

Assumption 2.2. ***(Equilibrium Selection)*** *There exist sequences of equilibrium selection mechanisms $\{\lambda_n(\cdot); n \in \mathbb{N}\}$ and public signals $\{\nu_n; n \in \mathbb{N}\}$ such that for n sufficiently large, $\mathcal{G}(X_n, \beta_0)$ is nonempty, and for any $g_n \in S^n$,*

$$Pr(G_n = \mathbf{g}_n | X_n, A_n) = \sum_{\sigma_n \in \mathcal{G}(X_n, A_n, \beta_0) = \sigma_n | X_n, A_n} Pr(\lambda(X_n, \nu_n; \beta_0) = \sigma_n | X_n, A_n) \prod_{i=1}^{n} \sigma_i(g_i | X_n, A_i).$$

2.3 Estimation

Define $P_{ij}(X, A_i, \sigma)$ to be the probability that individual i proposes to form a link with j conditional on X A_i, and σ. According to (2.5) and Assumption 2.1 (c),

$$P_{ij}(X, A_{i,n}, \sigma) = Pr(G_{ij}(X, A_i, \varepsilon_i, \sigma) = 1 | X, A_i, \sigma) =$$
$$\frac{\exp\left(W_{ij}\beta_0 + A_i + \mathbb{E}_{A_j}\left[\sigma_{ji}(G_{ji} = 1 \middle| X, A_j)\right]\beta_1 + \frac{1}{n-2}\sum_{k \neq i,j} \mathbb{E}_{A_j}\left[\sigma_{jk}(G_{jk} = 1 \middle| X, A_j)\right]\beta_2\right)}{1 + \exp\left(W_{ij}\beta_0 + A_i + \mathbb{E}_{A_j}\left[\sigma_{ji}(G_{ji} = 1 \middle| X, A_j)\right]\beta_1 + \frac{1}{n-2}\sum_{k \neq i,j} \mathbb{E}_{A_j}\left[\sigma_{jk}(G_{jk} = 1 \middle| X, A_j)\right]\beta_2\right)}. \tag{2.6}$$

Define $p_{ij}(X, A_i)$ to be the equilibrium probability that agent i proposes a link to agent j. which is realized in the data. Equilibrium condition requires that

$$\begin{aligned} & p_{ij}(X, A_i) = P_{ij}\left(X, A_i, p(X, A_i)\right) \\ =& \frac{\exp\left(W_{ij}\beta_0 + A_i + \mathbb{E}_{A_j|X}\left[p_{ji}(X, A_j)\right]\beta_1 + \frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}_{A_j|X}\left[p_{jk}(X, A_j)\right]\beta_2\right)}{1 + \exp\left(W_{ij}\beta_0 + A_i + \mathbb{E}_{A_j|X}\left[p_{ji}(X, A_j)\right]\beta_1 + \frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}_{A_j|X}\left[p_{jk}(X, A_j)\right]\beta_2\right)}. \end{aligned} \tag{2.7}$$

For notation simplicity, denote $q_{jk}(X, \sigma^*) := \mathbb{E}_{A_j}\left[Pr(G_{jk}(X, A_j, \varepsilon_j, \sigma) = 1|X, A_j, \sigma^*)\right]$, which is the probability that agent j proposes a link to k conditional on X and the realized equilibrium σ^*. Then (2.7) can be rewritten as

$$\begin{aligned} & P_{ij}\left(X, A_i, p(X, A_i)\right) \\ =& \frac{\exp\left(W_{ij}\beta_0 + A_i + q_{ji}(X, \sigma^*)\beta_1 + \frac{1}{n-2}\sum_{k\neq i,j} q_{jk}(X, \sigma^*)\beta_2\right)}{1 + \exp\left(W_{ij}\beta_0 + A_i + q_{ji}(X, \sigma^*)\beta_1 + \frac{1}{n-2}\sum_{k\neq i,j} q_{jk}(X, \sigma^*)\beta_2\right)} \\ :=& Q_{ij}\left(X, A_i, q(X, \sigma^*)\right). \end{aligned} \tag{2.8}$$

Although $p_{jk}(X, A_j)$ is not identified from data, $q_{ij}(X)$ is identified. With abuse of notations, let $q_{st}(X) = \mathbb{E}_{A_j}\left[p_{jk}(X_j = x_s, X_k = x_t, X, A_j)\right]$.

Consider the empirical frequency of pairs with the same observable characteristics proposing to form a link:

$$\hat{q}_{n,st} = \frac{\sum_i \sum_{j\neq i} G_{ij}\mathbf{1}\left\{X_i = x_s, X_j = x_t\right\}}{\sum_i \sum_{j\neq i} \mathbf{1}\left\{X_i = x_s, X_j = x_t\right\}}.$$

The following lemma shows that $q_{st}(X, \sigma^*)$ can be consistently estimated by $\hat{q}_{n,st}$ under the payoff function specified in 2.3.

Lemma 2.1. *For any X and realized symmetric equilibrium σ^*,*

$$\sup_{s,t} \Big| \hat{q}_{n,st} - q_{st}(X, \sigma^*) \Big| = O_p\Big(n^{-1/2}\Big).$$

Proof. See the Appendix. □

For the convenience of the following analysis, I introduce a change of notation:

$$Z_{ij} := (W'_{ij}, q_{ji}, \frac{1}{n-2} \sum_{k \neq i,j} q_{jk})'$$

and

$$\hat{Z}_{ij} := (W'_{ij}, \hat{q}_{ji}, \frac{1}{n-2} \sum_{k \neq i,j} \hat{q}_{jk})'.$$

Then by Lemma 2.1, $\sup_{s,t} |\hat{Z}_{s,t} - Z_{s,t}| = O_p\Big(n^{-1/2}\Big)$.

With the estimated $\hat{q}_n = \{\hat{q}_{st}\}_{\forall s,t}$, I propose to estimate the parameter β and individual fixed effects $\{A_i\}_{i=1}^n$ jointly by MLE. By Assumption 2.1 (c), the conditional likelihood of the network is

$$P(G = \mathbf{g}|X, A) = \prod_{i \neq j} Pr(G_{ij}(X, A_i, \varepsilon_i, \sigma) = g|X, A_i, \sigma).$$

By (2.6) and (2.8),

$$\begin{aligned} &Pr(G_{ij}(X, A_i, \varepsilon_i, \sigma) = g|X, A_i, \sigma) \\ =&Q_{ij}(X, A, q(X))^g \Big[1 - Q_{ij}(X, A, q(X))\Big]^{1-g}. \end{aligned}$$

Then one can construct the log-likelihood function as follows

$$\mathcal{L}_n(\beta, A, q) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} G_{ij} \ln Q_{ij}(\beta, A_i, q) + (1 - G_{ij}) \ln(1 - Q_{ij}(\beta, A_i, q)). \tag{2.9}$$

Let $\hat{\beta}$ and $\hat{A}$ be the maximizer of the log-likelihood with q replaced by $\hat{q}_n$.

$$\max_{\beta,A} \mathcal{L}_n(\beta, A, \hat{q}_n).$$

By first concentrating out A, the estimators are given by:

$$\hat{\beta} = \arg\max_{\beta} \mathcal{L}_n^c(\beta, \hat{A}(\beta), \hat{q}_n) \tag{2.10}$$

where

$$\begin{aligned} &\hat{A}(\beta) = \arg\max_{A} \mathcal{L}_n(\beta, A, \hat{q}_n) \\ \Longrightarrow &\hat{A}_i(\beta) = \arg\max_{A_i} \frac{1}{n-1} \sum_{j \neq i} G_{ij} \ln Q_{ij}(\beta, A_i, \hat{q}_n) + (1 - G_{ij}) \ln(1 - Q_{ij}(\beta, A_i, \hat{q}_n)). \end{aligned}$$

By rearranging the sample score of (2.9), it can be shown that $\hat{A}(\beta)$, when it exists, is the unique solution to the fixed point problem:

$$\hat{A}(\beta) = \varphi(\hat{A}(\beta)) \tag{2.11}$$

where

$$\varphi(A) = \begin{pmatrix} \ln \sum_{j \neq 1} G_{1j} - \ln \sum_{j \neq 1} \frac{\exp(\hat{Z}'_{1j}\beta)}{1+\exp(\hat{Z}'_{1j}\beta + A_1)} \\ \vdots \\ \ln \sum_{j \neq n} G_{nj} - \ln \sum_{j \neq n} \frac{\exp(\hat{Z}'_{nj}\beta)}{1+\exp(\hat{Z}'_{nj}\beta + A_n)} \end{pmatrix}. \tag{2.12}$$

2.4 Asymptotic Analysis

In this part, I first show the consistency of $\hat{\beta}$ and $\hat{A}$ and then prove the asymptotic normality of $\hat{\beta}$. Because link proposals from the same individual are correlated, the first step estimator has a slow convergence rate $\sqrt{n}$, which is equivalent to the usual convergence rate of $N^{1/4}$, since the number of summands in the likelihood function is $N = n(n-1)$. As is well discussed in the nonlinear panel literature, there is an estimation bias of $\hat{\beta}$ caused by the incidental parameters problem (e.g. Hahn and Newey (2004), Arellano and Hahn (2007)). However, as I will show in this part, the second-step bias has a higher order than the slow convergence rate of the first step, so a bias term won't show up in the asymptotic distribution.

Assumption 2.3. ***(Compact Support)*** *$\beta_0 \in int(\mathbb{B})$, with $\mathbb{B}$ a compact subset of $\mathbb{R}^K$.*

Assumption 2.4. ***(Joint FE Identification)*** *$\mathbb{E}[\mathcal{L}_n(\beta, A, q)|X, A_0]$ is uniquely maximized at $\beta = \beta_0$ and , $A = A_0$, for large enough n.*

Compactness of the support (Assumption 2.1 (a)(b) and Assumption 2.3) implies that

$$Q_{ij}(\beta, A_i, q) \in (\kappa, 1-\kappa) \tag{2.13}$$

for some $0 < \kappa < 1$ and for all $A_i \in \mathbb{A}$, $\beta \in \mathbb{B}$ and $\forall q \in (k, 1-k)$.

Theorem 2.1. ***(Consistency)*** *Under Assumptions 2.1, 2.2, 2.3, and 2.4*

$$\hat{\beta} \xrightarrow{p} \beta_0;$$
$$\hat{A} \xrightarrow{p} A_0.$$

Proof. See the Appendix. □

With a more involved argument, I can actually show the uniform convergence rate of $\hat{A}$, as shown in the Theorem below.

Theorem 2.2. *With probability $1 - O(n^{-2})$,*

$$\sup_{1 \leqslant i \leqslant n} |\hat{A}_i - A_{i0}| < O\left(\sqrt{\frac{\ln n}{n}}\right).$$

Proof. See the Appendix. □

To state the form of the asymptotic distribution, define

$$\mathcal{I}_0 = \lim_{n\to\infty} -\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} Z_{ij}Z_{ij}'Q_{ij}(1-Q_{ij}) + \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\frac{1}{n-1}\sum\limits_{j\neq i} Q_{ij}(1-Q_{ij})Z_{ij}\right)\left(\frac{1}{n-1}\sum\limits_{j\neq i} Q_{ij}(1-Q_{ij})Z_{ij}'\right)}{\frac{1}{n-1}\sum_{j\neq i} Q_{ij}(1-Q_{ij})}. \tag{2.14}$$

The asymptotic normality of $\hat{\beta}$ is formally stated in the theorem below.

Theorem 2.3. *Under Assumptions 2.1, 2.2, 2.3, and 2.4,*

$$\frac{\sqrt{n}a'(\hat{\beta}-\beta_0)}{\|a\|^{-1/2}(a'\mathcal{I}_0^{-1}\Omega_n\mathcal{I}_0^{-1}a)^{1/2}} \xrightarrow{d} N(0,1)$$

for any $d \times 1$ vector of real constants a and Ω_n as defined in the Appendix.

Proof. See the Appendix. □

2.5 Monte Carlo Simulation

In this section, I implement the proposed method in some simulation studies. Assume the following utility specification:

$$U_i(G, X, A_i, \varepsilon_i) = \sum_{j=1}^{n} G_{ij}\left(|X_i - X_j|\beta_1 + A_i + G_{ji}\beta_2 + \frac{1}{n-2}\sum_{k\neq i,j} G_{jk}\beta_3 + \varepsilon_{ij}\right)$$

where X_i is a random variable taking values in $\{1, -1\}$ with equal probability, and ϵ_{ij} follows the Logistic distribution. The distribution of A_i is generated according to

$$A_i = (\alpha_L + \gamma a_i)\mathbf{1}\{X_i = -1\} + \alpha_H\mathbf{1}\{X_i = 1\} + V_i,$$

with $\alpha_L < \alpha_H$ and $a_i \sim N(0, 0.1), V_i \sim N(0, \sqrt{0.1})$, and they are independent. In the simulation exercise, I consider three scenarios. In the first two scenarios, A_i is correlated with

X. In Scenario 1, I let $\alpha_L = -2/3, \alpha_H = -1/6$, and $\gamma = 0$, so that the correlation between A_i and X_i is only through the value of X_i. In Scenario 2, I let $\alpha_L = -2/3, \alpha_H = -1/6$, and $\gamma = 1$, so that the correlation between A_i and X_i is determined not only by the value of X_i but also by the identity of i (captured by the random variable a_i). In Scenario 3, I let $\alpha_L = -1/2, \alpha_H = -1/2$ and $\gamma = 0$, so that A_i is independent with X. The true values of the parameters are $(\beta_1, \beta_2, \beta_3) = (-2, 1, 1)$. The network is generated according to the $n-$player incomplete information game described in Section 2.2, with n taking values of $50, 100, 250$, and 500. For each value of n, I generate a single network and use the method proposed in this paper and Leung (2015) to estimate the parameters. When using Leung (2015)'s estimator, the private information η_{ij} is the sum of A_i and ϵ_{ij} with $A_i \perp \epsilon_{ij}$. Each experiment is repeated 1000 times. I report the means and standard errors of the estimated parameters in the tables below.

Table 2.1: Scenario 1 Correlated Private Information ($\alpha_L = -2/3, \alpha_H = -1/6, \gamma = 0$)

	This paper's estimator			Leung (2015)'s estimator		
n	β_1	β_2	β_3	β_1	β_2	β_3
50	-1.922	0.966	0.936	-2.092	0.827	1.390
	(0.049)	(0.101)	(0.069)	(0.249)	(0.484)	(1.024)
100	-1.930	0.951	0.926	-2.085	0.807	1.435
	(0.035)	(0.058)	(0.033)	(0.198)	(0.478)	(0.985)
250	-1.967	1.050	0.973	-2.065	0.864	1.334
	(0.042)	(0.101)	(0.055)	(0.170)	(0.476)	(0.961)
500	-2.015	1.065	0.975	-2.047	0.919	1.237
	(0.036)	(0.057)	(0.035)	(0.160)	(0.476)	(0.950)

This table gives the mean of each estimator across the 1000 Monte Carlo estimates. The standard deviation of the Monte Carlo estimates is reported below the mean value of the point estimates in parentheses (this is a quantile-based estimate which uses the 0.05 and 0.95 quantiles of the Monte Carlo distribution of point estimates and the assumption of Normality).

As can be seen in Table 2.1 and 2.2, when the private information is correlated with observed individual characteristics X, this paper's approach yields good estimates for the parameters, while Leung (2015)'s estimator doesn't perform well, both in terms of the mean and variance of the estimators. This is not surprising since Leung (2015) assumes that

Table 2.2: Scenario 2 Correlated Private Information ($\alpha_L = -2/3, \alpha_H = -1/6, \gamma = 1$)

	This paper's estimator			Leung (2015)'s estimator		
n	β_1	β_2	β_3	β_1	β_2	β_3
50	-1.921	0.952	0.928	-2.072	0.866	1.294
	(0.043)	(0.108)	(0.069)	(0.280)	(0.698)	(1.521)
100	-1.932	0.909	0.902	-2.079	0.820	1.408
	(0.048)	(0.034)	(0.021)	(0.258)	(0.732)	(1.539)
250	-1.955	0.956	0.925	-2.056	0.888	1.287
	(0.045)	(0.055)	(0.029)	(0.242)	(0.738)	(1.509)
500	-2.015	1.023	0.953	-2.035	0.956	1.162
	(0.047)	(0.069)	(0.038)	(0.233)	(0.724)	(1.460)

This table gives the mean of each estimator across the 1000 Monte Carlo estimates. The standard deviation of the Monte Carlo estimates is reported below the mean value of the point estimates in parentheses (this is a quantile-based estimate which uses the 0.05 and 0.95 quantiles of the Monte Carlo distribution of point estimates and the assumption of Normality).

private information and observable individual characteristics are independent. Under the correlated scenario, Leung (2015)'s estimator will not be consistent. Table 2.3 shows the simulation results when the individual private information A is independent with observed characteristics X. Not surprisingly, both this paper's estimator and Leung (2015)'s estimator perform reasonably well, except that Leung (2015)'s estimator has larger variances.

Table 2.3: Scenario 3 Independent Private Information ($\alpha_L = -1/2, \alpha_H = -1/2, \gamma = 0$)

	This paper's estimator			Leung (2015)'s estimator		
n	β_1	β_2	β_3	β_1	β_2	β_3
50	-1.913	0.951	0.925	-2.046	0.909	1.181
	(0.040)	(0.050)	(0.025)	(0.205)	(0.482)	(0.936)
100	-1.921	0.925	0.911	-2.014	0.956	1.085
	(0.027)	(0.030)	(0.014)	(0.154)	(0.470)	(0.901)
250	-1.945	0.946	0.919	-2.016	0.944	1.109
	(0.030)	(0.030)	(0.014)	(0.140)	(0.464)	(0.901)
500	-1.994	1.043	0.965	-2.010	0.967	1.064
	(0.029)	(0.042)	(0.020)	(0.134)	(0.463)	(0.892)

This table gives the mean of each estimator across the 1000 Monte Carlo estimates. The standard deviation of the Monte Carlo estimates is reported below the mean value of the point estimates in parentheses (this is a quantile-based estimate which uses the 0.05 and 0.95 quantiles of the Monte Carlo distribution of point estimates and the assumption of Normality).

2.6 Conclusion

In this paper, I characterize the network formation process as a static game of incomplete information, where the latent payoff of forming a link between two individuals depends on the structure of the network, as well as private information on agents' attributes. I allow agents' private unobserved attributes to be correlated with observables through individual fixed effects. Using data from a single large network, I propose a two-step estimator for the model primitives. In the first step, I estimate agents' equilibrium beliefs of other people's choice probabilities. In the second step, I plug in the first-step estimator to the conditional choice probability expression and estimate the model parameters and the unobserved individual fixed effects together using ioint MLE. Assuming that the observed attributes are discrete, I showed that the first step estimator is uniformly consistent with the rate $n^{-1/2}$, where n is the number of individuals in the network. This rate corresponds to the usual $N^{-1/4}$ rate where N stands for the total number of linking proposals and is the effective sample size. The slow convergence rate is translated to the second step so that the usual asymptotic bias of order $N^{-1/2}$ caused by the "incidental parameter problem" won't show up in the asymptotic distribution. The second-step estimator $\hat{\beta}$ subtracted by its mean converges asymptotically to a normal distribution at the rate $N^{-1/4}$. Monte Carlo Simulation shows that the estimator proposed in this paper performs well in finite samples.

2.7 Appendix

2.7.1 Useful Lemmas

The next two lemmas are to be used in the proofs of the asymptotics.

Lemma 2.2. *Under Assumptions 1,2 and 3,*

$$\sup_{1\leqslant i\leqslant n}\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-Q_{ij})\right| < O\left(\sqrt{\frac{\ln n}{n}}\right)$$

with probability $1-O(n^{-2})$*, and*

$$\sup_{1\leqslant i\leqslant n}\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-\hat{Q}_{ij})\right| < O\left(\sqrt{\frac{\ln n}{n}}\right)$$

with probability $1-O(n^{-2})$*, where*

$$Q_{ij} := Q_{ij}(\beta_0, A_{i0}, Z_{ij})$$
$$\hat{Q}_{ij} := Q_{ij}(\beta_0, A_{i0}, \hat{Z}_{ij}).$$

Proof. The first conclusion comes by applying Hoeffding's inequality

$$Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-Q_{ij})\right|\geqslant\epsilon\right)\leqslant 2\exp\left(-\frac{2(n-1)\epsilon^2}{(1-2\kappa)^2}\right)$$

for κ as defined by (2.13). Setting $\epsilon=\sqrt{\frac{3(1-2\kappa)^2}{2}\frac{\ln n}{n}}$ gives

$$Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-Q_{ij})\right|\geqslant\sqrt{\frac{3(1-2\kappa)^2}{2}\frac{\ln n}{n}}\right)$$
$$\leqslant 2\exp\left(-\frac{2(n-1)}{(1-2\kappa)^2}\frac{3(1-2\kappa)^2}{2}\frac{\ln n}{n}\right)$$

$$
\begin{aligned}
&= 2\exp\left(\ln\left(\frac{1}{n^3}\right)\frac{(n-1)}{n}\right) \\
&= 2\left(\frac{1}{n^3}\right)^{\frac{(n-1)}{n}} \\
&= O(n^{-3}).
\end{aligned}
$$

Applying Boole's inequality then gives

$$
\begin{aligned}
&Pr\left(\max_{1\leqslant i\leqslant n}\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-Q_{ij})\right| \geqslant \sqrt{\frac{3(1-2\kappa)^2}{2}\frac{\ln n}{n}}\right) \\
&\leqslant n * O(n^{-3}) \\
&= O(n^{-2}),
\end{aligned}
$$

from which the first conclusion follows.

To prove the second conclusion, first, observe that for any i, j

$$
\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-\hat{Q}_{ij})\right| \leqslant \left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij}-Q_{ij})\right| + \left|\frac{1}{n-1}\sum_{j\neq i}(Q_{ij}-\hat{Q}_{ij})\right|.
$$

By the triangle inequality,

$$
\left|\frac{1}{n-1}\sum_{j\neq i}(Q_{ij}-\hat{Q}_{ij})\right| \leqslant \sup_{i,j}\left|Q_{ij}-\hat{Q}_{ij}\right|.
$$

Applying mean value expansion gives that for any i, j

$$
\begin{aligned}
\left|Q_{ij}-\hat{Q}_{ij}\right| &= \left|\frac{\exp(\bar{Z}_{ij}\beta_0 + A_{i,0})\beta_0'}{(1+\exp(\bar{Z}_{ij}\beta_0 + A_{i,0}))^2}(\hat{Z}_{ij}-Z_{ij})\right| \\
&= O_p(1)O_p(n^{-1/2}) \\
&= O_p(n^{-1/2})
\end{aligned}
$$

where the second equality comes from condition (2.13), Assumption 2.3 and Lemma 2.1.

The second conclusion follows from the first conclusion. □

Lemma 2.3. *Under Assumptions 2.1,2.2 and 2.3, $\hat{A}_i(\beta_0) - A_i(\beta_0)$ has the asymptotically linear representation*

$$\hat{A}_i(\beta_0) - A_i(\beta_0) = \frac{\sum_{j\neq i}(G_{ij} - Q_{ij})}{\sum_{j\neq i} Q_{ij}(1-Q_{ij})} + \frac{\sum_{j\neq i}\hat{Q}_{ij} - Q_{ij}}{\sum_{j\neq i} Q_{ij}(1-Q_{ij})} + O_P\left(\frac{\ln n}{n}\right).$$

Proof. Consider the first order condition with respect to A

$$\left.\frac{\partial \mathcal{L}_n(\beta_0, A, \hat{q})}{\partial A}\right|_{A=\hat{A}(\beta_0)} = 0;$$

a mean value expansion gives that for all i

$$\begin{aligned}
0 =& \sum_{j\neq i}\left(G_{ij} - Q_{ij}(\beta_0, \hat{A}_i(\beta_0), \hat{q}_{ij})\right) \\
=& \sum_{j\neq i}\left(G_{ij} - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})\right) \\
&- \sum_{j\neq i}(\hat{A}_i(\beta_0) - A_i(\beta_0))Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})\left[1 - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})\right] \\
&+ \frac{1}{2}\sum_{j\neq i}(\hat{A}_i(\beta_0) - A_i(\beta_0))^2 Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\times \\
&\qquad \left[1 - Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\right]\left[1 - 2Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\right]. \qquad (2.15)
\end{aligned}$$

Denote the last term by R_i. The Triangle Inequality and Condition (2.13) then implies

$$\begin{aligned}
|R_i| \leqslant& \frac{1}{2}\left|\hat{A}_i(\beta_0) - A_i(\beta_0)\right|^2 \sum_{j\neq i}\Big|Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\times \\
&\qquad \left[1 - Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\right]\left[1 - 2Q_{ij}(\beta_0, \bar{A}_i(\beta_0), \hat{q}_{ij})\right]\Big| \qquad (2.16) \\
\leqslant& \lambda_n^2 O_p(n-1), \qquad (2.17)
\end{aligned}$$

where $\lambda_n = \sup_{1\leqslant i\leqslant n}\left|\hat{A}_i - A_{i0}\right| \leqslant O_p(\sqrt{\frac{\ln n}{n}})$ according to Theorem 2.2. From (2.15) I have

$$\begin{aligned}
&\hat{A}_i(\beta_0) - A_i(\beta_0)\\
=&\frac{\sum_{j\neq i}[G_{ij} - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})]}{\sum_{j\neq i} Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})[1 - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})]}\\
&+ \frac{R_i}{\sum_{j\neq i} Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})[1 - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})]}\\
=&\frac{\sum_{j\neq i}[G_{ij} - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})]}{\sum_{j\neq i} Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})[1 - Q_{ij}(\beta_0, A_i(\beta_0), \hat{q}_{ij})]} + O_p\left(\frac{\sqrt{\ln n}}{n}\right) + O_p\left(\frac{\ln n}{n}\right)\\
=&\frac{\sum_{j\neq i}(G_{ij} - Q_{ij})}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})} + \frac{\sum_{j\neq i}\hat{Q}_{ij} - Q_{ij}}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})} + O_p\left(\frac{\sqrt{\ln n}}{n}\right)
\end{aligned}$$

where the second equality follows from (2.17) and Condition (2.13) and the third equality come from a similar argument as in the proof of Lemma 2.2. More specifically, from the proof of Lemma 2.2, I know that $\hat{Q}_{ij} = Q_{ij} + O_p\left(\frac{1}{\sqrt{n}}\right)$, then applying Condition (2.13) yields

$$\begin{aligned}
&\frac{\sum_{j\neq i}(G_{ij} - \hat{Q}_{ij})}{\sum_{j\neq i}\hat{Q}_{ij}(1 - \hat{Q}_{ij})}\\
=&\frac{\sum_{j\neq i}\left(G_{ij} - Q_{ij}\right)}{\sum_{j\neq i}\hat{Q}_{ij}(1 - \hat{Q}_{ij})} + \frac{\sum_{j\neq i}\hat{Q}_{ij} - Q_{ij}}{\sum_{j\neq i}\hat{Q}_{ij}(1 - \hat{Q}_{ij})}\\
=&\frac{\sum_{j\neq i}\left(G_{ij} - Q_{ij}\right)}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})} + \frac{\left(\left[\sum_{j\neq i} Q_{ij}(1 - Q_{ij})\right] - \left[\sum_{j\neq i}\hat{Q}_{ij}(1 - \hat{Q}_{ij})\right]\right)\sum_{j\neq i}\left(G_{ij} - Q_{ij}\right)}{\left[\sum_{j\neq i} Q_{ij}(1 - Q_{ij})\right]\left[\sum_{j\neq i}\hat{Q}_{ij}(1 - \hat{Q}_{ij})\right]}\\
&+ \frac{\sum_{j\neq i}\hat{Q}_{ij} - Q_{ij}}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})}\\
=&\frac{\sum_{j\neq i}\left(G_{ij} - Q_{ij}\right)}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})} + \frac{\sum_{j\neq i}\hat{Q}_{ij} - Q_{ij}}{\sum_{j\neq i} Q_{ij}(1 - Q_{ij})} + O_p\left(\frac{1}{n}\right),
\end{aligned}$$

the conclusion thus follows. ☐

2.7.2 Proofs of the Theorems and Lemmas in the Main Text

Proof of Lemma 2.1. As specified in (2.5), the optimal linking decision of agent i with agent j is $G_{ij}(X, A_i, \varepsilon_i, \sigma^*)$.

$$\begin{aligned}
&\sup_{s,t}\left|\hat{q}_{n,st} - q_{st}(X,\sigma^*)\right| \\
=&\sup_{s,t}\left|\frac{\sum_i\sum_{j\neq i}\left(G_{ij} - P(G_{ij}=1|X_i=x_s, X_j=x_t, X,\sigma^*)\right)\mathbf{1}\left\{X_i=x_s, X_j=x_t\right\}}{\sum_i\sum_{j\neq i}\mathbf{1}\left\{X_i=x_s, X_j=x_t\right\}}\right|.
\end{aligned}$$

Denote the fraction term by $\Delta_{n,st}$. It suffices to show that

$$\lim_{\eta\to\infty}\lim_{n\to\infty} P\left(\sup_{s,t}\left|\Delta_{n,st}\right| > \eta n^{-1/2}\right) = 0.$$

By the law of iterated expectations and dominated convergence theorem, it suffices to show

$$P\left(\sup_{s,t}\left|\Delta_{n,st}\right| > \eta n^{-1/2}\middle| X,\sigma^*\right) \xrightarrow{p} 0 \text{ as } \eta, n\to\infty.$$

Note that

$$\begin{aligned}
P\left(\sup_{s,t}\left|\Delta_{n,st}\right| > \eta n^{-1/2}\middle| X,\sigma^*\right) &\leqslant \sum_{s,t} P\left(\left|\Delta_{n,st}\right| > \eta n^{-1/2}\middle| X,\sigma^*\right) \\
&\leqslant \sum_{s,t}\frac{n\mathbb{E}\left(\Delta_{st}^2\middle| X,\sigma^*\right)}{\eta^2} \\
&\leqslant \frac{nT_x^2}{\eta^2}\max_{s,t}\mathbb{E}\left(\Delta_{st}^2\middle| X,\sigma^*\right).
\end{aligned}$$

Then it suffices to show $E\left(\Delta_{st}^2\middle| X,\sigma^*\right) = O\left(n^{-1}\right)$ for all s,t.

$$\begin{aligned}
&\mathbb{E}\left(\Delta_{st}^2\middle| X,\sigma^*\right) \\
=&\frac{\sum_i\sum_{i\neq j} Var\left(G_{ij}|X_i=x_s, X_j=x_t, X,\sigma^*\right)\mathbf{1}\left\{X_i=x_s, X_j=x_t\right\}}{\left(\sum_i\sum_{j\neq i}\mathbf{1}\left\{X_i=x_s, X_j=x_t\right\}\right)^2}
\end{aligned}$$

$$+\frac{\sum_i \sum_{i\neq j} \sum_{k\neq i,j} Cov\left(G_{ij}, G_{ik}|X_i = x_s, X_j = x_t, X, \sigma^*\right) \mathbf{1}\left\{X_i = x_s, X_j = x_t\right\}}{\left(\sum_i \sum_{j\neq i} \mathbf{1}\left\{X_i = x_s, X_j = x_t\right\}\right)^2} \tag{2.18}$$

where I used the fact that link proposals from different agents are independent, i.e. $G_{ij}(X, A_i, \varepsilon_i, \sigma) \perp G_{i'j'}(X, A_{i'}, \varepsilon_{i'}, \sigma^*)\Big|X, \sigma,$ so $Cov\Big(G_{ij}, G_{i'j'}|X_i = X_{i'} = x_s, X_j = X_{j'} = x_t, X, \sigma^*\Big) = 0$ for all $i \neq i'$.

Since G_{ij} is a binary random variable, $Var\left(G_{ij}|X_i = x_s, X_j = x_t, X, \sigma^*\right) \leqslant \frac{1}{4}$. The first term is bounded by

$$\frac{1}{4}\left(\sum_i \sum_{j\neq i} \mathbf{1}\left\{X_i = x_s, X_j = x_t\right\}\right)^{-1}.$$

Then for the second term, by Cauchy-Schwarz Inequality,

$$\begin{aligned}
&Cov\left(G_{ij}, G_{ik}|X_i = x_s, X_j = X_k = x_t, X, \sigma^*\right) \\
\leqslant\, &Var\left(G_{ij}\Big|X_i = x_s, X_j = x_t, X, \sigma^*\right)^{1/2} Var\left(G_{ik}\Big|X_i = x_s, X_k = x_t, X, \sigma^*\right)^{1/2} \\
\leqslant\, &\frac{1}{4}
\end{aligned}$$

so the second term is bounded by

$$\begin{aligned}
&\frac{\frac{1}{4n}\frac{1}{n(n-1)(n-2)}\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \mathbf{1}\left\{X_i = x_s, X_j = X_k = x_t\right\}}{\frac{n-1}{n-2}\left(\frac{1}{n(n-1)}\sum_i \sum_{j\neq i}\left\{X_i = x_s, X_j = x_t\right\}\right)^2} \\
\leqslant &\frac{1}{4n}\frac{\frac{1}{n(n-1)(n-2)}\sum_i \sum_{j\neq i} \sum_{k\neq i,j} \mathbf{1}\left\{X_i = x_s, X_j = X_k = x_t\right\}}{\left(\frac{1}{n(n-1)}\sum_i \sum_{j\neq i}\left\{X_i = x_s, X_j = x_t\right\}\right)^2}.
\end{aligned}$$

Both the numerator and denominator are U-statistics. It's straightforward to show that they converge to their expectations. Therefore, the sum of the first and second terms are $O(\frac{1}{n})$, and the proof is complete. □

Proof of Theorem 2.1. According to Assumption 2.4, β_0, A_0 uniquely maximizes

$\mathbb{E}[\mathcal{L}_n(\beta, A, q)|X, A_0]$. Since $(\hat{\beta}, \hat{A})$ solves $\max_{\beta\in\mathbb{B}, A\in\mathbb{A}} \mathcal{L}_n(\beta, A, \hat{q})$, it suffices to show that

$$\sup_{\beta,A}\left|\mathcal{L}_n(\beta, A, \hat{q}) - \mathbb{E}[\mathcal{L}_n(\beta, A, q)|X, A_0]\right| \xrightarrow{p} 0. \tag{2.19}$$

By the triangle inequality, the left-hand side is less than or equal to

$$\underbrace{\sup_{\beta,A}\left|\mathcal{L}_n(\beta, A, \hat{q}) - \mathbb{E}[\mathcal{L}_n(\beta, A, \hat{q})|X, A_0]\right|}_{I} + \underbrace{\sup_{\beta,A}\left|\mathbb{E}[\mathcal{L}_n(\beta, A, \hat{q})|X, A_0] - \mathbb{E}[\mathcal{L}_n(\beta, A, q)|X, A_0]\right|}_{II}.$$

By Continuous Mapping Theorem and Lemma 2.1, $II = o_p(1)$. By the Logit formalization of $Q_{ij}(\beta, A_i, \hat{q})$,

$$\begin{aligned} I &= \sup_{\beta,A}\left|\mathcal{L}_n(\beta, A, \hat{q}) - \mathbb{E}[\mathcal{L}_n(\beta, A, \hat{q})|X, A_0]\right| \\ &= \sup_{\beta,A}\left|\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}(G_{ij} - Q_{ij})\ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})}\right)\right| \end{aligned}$$

where $Q_{ij} := Q_{ij}(\beta_0, A_0, \hat{q})$. According to the Triangle Inequality,

$$\begin{aligned} &\left|\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}(G_{ij} - Q_{ij})\ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})}\right)\right| \\ &\leqslant \frac{1}{n}\sum_i\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij} - Q_{ij})\ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})}\right)\right|. \end{aligned}$$

Condition (2.13) implies that $\ln(\frac{\kappa}{1-\kappa}) \leqslant \ln\left(\frac{Q_{ij}(\beta,A_i,\hat{q})}{1-Q_{ij}(\beta,A_i,\hat{q})}\right) \leqslant \ln(\frac{1-\kappa}{\kappa})$, thus $(\kappa - 1)\ln\frac{1-\kappa}{\kappa} \leqslant (G_{ij} - Q_{ij})\ln\left(\frac{Q_{ij}(\beta,A_i,\hat{q})}{1-Q_{ij}(\beta,A_i,\hat{q})}\right) \leqslant (1-\kappa)\ln\frac{1-\kappa}{\kappa}$. According to Hoeffding's inequality,

$$Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}(G_{ij} - Q_{ij})\ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})}\right)\right| \geqslant \epsilon\right) \leqslant 2\exp\left(-\frac{(n-1)\epsilon^2}{2(1-\kappa)^2(\ln\frac{1-\kappa}{\kappa})^2}\right).$$

Take $\epsilon = \sqrt{\frac{3 \ln n}{n}}$ and apply Boole's inequality, for any $\beta \in \mathbb{B}$, $A \in \mathbb{A}^n$,

$$\begin{aligned}
&Pr\left(\max_{1 \leqslant i \leqslant n} \left| \frac{1}{n-1} \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \\
\leqslant & nPr\left(\left| \frac{1}{n-1} \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \\
\leqslant & \left(\frac{2}{n^2} \right)^{-\frac{(n-1)}{2n(1-\kappa)^2 (\ln \frac{1-\kappa}{\kappa})^2}} \\
= & O\Big(\frac{1}{n^2}\Big) \\
\Longrightarrow & Pr\left(\left| \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \\
\leqslant & Pr\left(\frac{1}{n} \sum_i \left| \frac{1}{n-1} \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \\
\leqslant & Pr\left(\max_{1 \leqslant i \leqslant n} \left| \frac{1}{n-1} \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \leqslant O\Big(\frac{1}{n^2}\Big),
\end{aligned}$$

which implies the uniform convergence result:

$$Pr\left(\sup_{\beta, A} \left| \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} (G_{ij} - Q_{ij}) \ln\left(\frac{Q_{ij}(\beta, A_i, \hat{q})}{1 - Q_{ij}(\beta, A_i, \hat{q})} \right) \right| \geqslant \sqrt{\frac{3 \ln n}{n}} \right) \leqslant O\Big(\frac{1}{n^2}\Big) \quad (2.20)$$

and hence $I = o_p(1)$ and (2.19) follows. □

Proof of Theorem 2.2. Let A_0 denote the population vector of heterogeneity terms and $A_1 = \varphi(A_0)$. From (2.12), I have

$$A_{1,i} - A_{0,i} = \ln \sum_{j \neq i} G_{ij} - \ln \sum_{j \neq i} \frac{\exp(\hat{Z}'_{ij}\hat{\beta} + A_{0i})}{1 + \exp(\hat{Z}'_{ij}\hat{\beta} + A_{0i})}.$$

A Taylor expansion of the second term on the right-hand side gives:

$$\ln \sum_{j \neq i} \frac{\exp(\hat{Z}'_{ij}\hat{\beta} + A_{0i})}{1 + \exp(\hat{Z}'_{ij}\hat{\beta} + A_{0i})}$$

$$= \ln \sum_{j \neq i} \frac{\exp(\hat{Z}_{ij}'\beta_0 + A_{0i})}{1 + \exp(\hat{Z}_{ij}'\beta_0 + A_{0i})} + \frac{\sum_{j \neq i} Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})(1 - Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij}))\hat{Z}_{ij}'}{\sum_{i \neq i} Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})}(\hat{\beta} - \beta_0).$$

Using (2.13), the compact support of Z_{ij} , and Theorem 2.1,

$$\begin{aligned}
&\left| \frac{\sum_{j \neq i} Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})(1 - Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij}))\hat{Z}_{ij}'}{\sum_{i \neq i} Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})}(\hat{\beta} - \beta_0) \right| \\
\leqslant &\sum_{j \neq i} \left| \frac{Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})(1 - Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij}))\hat{Z}_{ij}'}{\sum_{i \neq i} Q_{ij}(\bar{\beta}, A_{i0}, \hat{Z}_{ij})} \right| \left| (\hat{\beta} - \beta_0) \right| \\
\leqslant &\frac{\sup_{z \in \mathbb{Z}} |z'|}{4\kappa} \left| (\hat{\beta} - \beta_0) \right| \\
= &O_p(1) \cdot o_p(1) \\
= &o_p(1).
\end{aligned}$$

I can conclude that

$$A_{1,i} - A_{0,i} = \ln \sum_{j \neq i} G_{ij} - \ln \sum_{j \neq i} \frac{\exp(\hat{Z}_{ij}'\beta_0 + A_{0i})}{1 + \exp(\hat{Z}_{ij}'\beta_0 + A_{0i})} + o_p(1).$$

Denote $\hat{Q}_{ij} := \frac{\exp(\hat{Z}_{ij}'\beta_0 + A_{0i})}{1 + \exp(\hat{Z}_{ij}'\beta_0 + A_{0i})}$, a mean value expansion around $\hat{Q}_{ij}$ gives

$$\ln \sum_{j \neq i} G_{ij} = \ln \sum_{j \neq i} \hat{Q}_{ij} + \frac{\sum_{j \neq i} G_{ij} - \hat{Q}_{ij}}{\lambda \sum_{j \neq i} G_{ij} + (1 - \lambda) \sum_{j \neq i} \hat{Q}_{ij}},$$

for some $\lambda \in (0, 1)$. By (2.13), for all i

$$\left| \frac{\sum_{j \neq i}(G_{ij} - \hat{Q}_{ij})}{\lambda \sum_{j \neq i} G_{ij} + (1 - \lambda) \sum_{j \neq i} \hat{Q}_{ij}} \right| \leqslant \frac{\left| \sum_{j \neq i}(G_{ij} - \hat{Q}_{ij}) \right|}{(n - 1)(1 - \lambda)\kappa}.$$

Lemma 2.2 then gives, with probability $1 - O(n^{-2})$, the uniform bound

$$\sup_{1 \leqslant i \leqslant n} \left| A_{1,i} - A_{0,i} \right| < O\left(\sqrt{\frac{\ln n}{n}} \right).$$

Then the conclusion follows by applying Lemma 4 in Graham (2017). $\square$

Proof of Theorem 2.3. *Step 1. Characterizing the probability limit of the Hessian of the concentrated log-likelihood.*

First define the following notations. The Hessian matrix of the joint log-likelihood is given by

$$H_n = \begin{pmatrix} H_{n,\beta\beta} & H_{n,\beta A} \\ H'_{n,\beta A} & H_{n,AA} \end{pmatrix}$$

where

$$H_{n,\beta\beta} = -\sum_i \sum_{j\neq i} Z_{ij} Zij' Q_{ij}(1-Q_{ij}) \tag{2.21}$$

$$H'_{n,\beta A} = -\begin{pmatrix} \sum_{j\neq 1} Q_{1j}(1-Q_{1j})Z'_{1j} \\ \vdots \\ \sum_{j\neq n} Q_{nj}(1-Q_{nj})Z'_{nj} \end{pmatrix} \tag{2.22}$$

$$H_{n,AA} = -\begin{pmatrix} \sum_{j\neq 1} Q_{1j}(1-Q_{1j}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{j\neq n} Q_{nj}(1-Q_{nj}) \end{pmatrix} \tag{2.23}$$

and $\hat{H}_{n,\beta\beta}$, $\hat{H}'_{n,\beta A}$, and $\hat{H}_{n,AA}$ are defined by (2.21), (2.22), (2.23) respectively with Z_{ij} replaced by $\hat{Z}_{ij}$.

Following Amemiya (1985, pp. 125-127), the Hessian of the concentrated likelihood is

$$\frac{\partial^2 \mathcal{L}_n^c(\beta_0, \hat{A}(\beta_0), \hat{q})}{\partial\beta\partial\beta'} = \sum_{i=1}^n \sum_{j\neq i} \frac{\partial}{\partial\beta'} s_{\beta,ij}(\beta_0, \hat{A}_i(\beta_0), \hat{q}_{ij})$$

$$= \hat{H}_{n,\beta\beta} - \hat{H}_{n,\beta A}\hat{H}_{n,AA}^{-1}\hat{H}'_{n,\beta A}$$

$$= -\sum_{i=1}^n \sum_{j\neq i} \hat{Z}_{ij}\hat{Z}'_{ij}\hat{Q}_{ij}(1-\hat{Q}_{ij}) + \sum_{i=1}^n \frac{\left(\sum_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})\hat{Z}_{ij}\right)\left(\sum_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})\hat{Z}'_{ij}\right)}{\sum_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})},$$

which implies

$$
\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\partial}{\partial\beta'}s_{\beta,ij}(\beta_0,\hat{A}_i(\beta_0),\hat{q}_{ij})\\
=&-\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\hat{Z}_{ij}\hat{Z}'_{ij}\hat{Q}_{ij}(1-\hat{Q}_{ij})\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\frac{1}{n-1}\sum\limits_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})\hat{Z}_{ij}\right)\left(\frac{1}{n-1}\sum\limits_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})\hat{Z}'_{ij}\right)}{\frac{1}{n-1}\sum_{j\neq i}\hat{Q}_{ij}(1-\hat{Q}_{ij})}\\
=&-\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}Z_{ij}Z'_{ij}Q_{ij}(1-Q_{ij})\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\frac{1}{n-1}\sum\limits_{j\neq i}Q_{ij}(1-Q_{ij})Z_{ij}\right)\left(\frac{1}{n-1}\sum\limits_{j\neq i}Q_{ij}(1-Q_{ij})Z'_{ij}\right)}{\frac{1}{n-1}\sum_{j\neq i}Q_{ij}(1-Q_{ij})}+o_p(1)\\
=&\mathcal{I}_0+o_p(1),
\end{aligned}
\tag{2.24}
$$

where $\mathcal{I}_0$ is as defined in (2.14). The second equality in (2.24) is given by the same logic as the proof of Lemma 2.2 and more involved calculations.

Step 2. Asymptotic Linear Representation

Consider the first-order condition associated with the concentrated log-likelihood

$$\left.\frac{\partial\mathcal{L}_n^c(\beta,\hat{A}(\beta),\hat{q})}{\partial\beta}\right|_{\beta=\hat{\beta}}=0;$$

a mean value expansion gives

$$
\begin{aligned}
0=\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\hat{\beta},\hat{A}_i(\hat{\beta}),\hat{q}_{ij})=&\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\beta_0,\hat{A}_i(\beta_0),\hat{q}_{ij})\\
&+\sum_{i=1}^{n}\sum_{j\neq i}\frac{\partial}{\partial\beta'}s_{\beta,ij}(\bar{\beta},\hat{A}_i(\bar{\beta}),\hat{q}_{ij})(\hat{\beta}-\beta_0),
\end{aligned}
$$

which implies

$$\sqrt{n}(\hat{\beta}-\beta_0)$$
$$=-\underbrace{\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\frac{\partial}{\partial\beta'}s_{\beta,ij}(\bar{\beta},\hat{A}_i(\bar{\beta}),\hat{q}_{ij})\right]^{-1}}_{I^{-1}}\underbrace{\left[\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\beta_0,\hat{A}_i(\beta_0),\hat{q}_{ij})\right]}_{II}.$$

The first term I converges in probability to $\mathcal{I}_0$ as defined in (2.14). I cannot apply a CLT directly to II because of the strong correlation between summands caused by using the same set of data to get $\hat{q}$, $\hat{A}$ and estimator of β.

A second order Taylor expansion of II gives

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\beta_0,\hat{A}_i(\beta_0),\hat{q}_{ij})\\
=&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\beta_0,A_i(\beta_0),q_{ij})\\
&-\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))Q_{ij}(1-Q_{ij})Z_{ij}\\
&-\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}Q_{ij}(1-Q_{ij})Z_{ij}\beta_0'(\hat{Z}_{ij}-Z_{ij})+Q_{ij}(\hat{Z}_{ij}-Z_{ij})\\
&-\frac{1}{2}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))^2\bar{Q}_{ij}(1-\bar{Q}_{ij})(1-2\bar{Q}_{ij})\bar{Z}_{ij}\\
&-\frac{1}{2}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{Z}_{ij}-Z_{ij})'\nabla_{Z_{ij}Z'_{ij}}s_{\beta,ij}(\beta_0,\bar{A}_i(\beta_0),\bar{q}_{ij})(\hat{Z}_{ij}-Z_{ij})\\
&-\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))\Big[\bar{Q}_{ij}(1-\bar{Q}_{ij})(1-2\bar{Q}_{ij})\bar{Z}_{ij}\beta_0'(\hat{Z}_{ij}-Z_{ij})\\
&\qquad\qquad+\bar{Q}_{ij}(1-\bar{Q}_{ij})(\hat{Z}_{ij}-Z_{ij})\Big],
\end{aligned}\tag{2.25}$$

where $\bar{Q}_{ij}=\frac{\exp(\bar{Z}_{ij}\beta_0+\bar{A}_i)}{1+\exp(\bar{Z}_{ij}\beta_0+\bar{A}_i)}$, with $\bar{A}_i$ between $\hat{A}$ and A, $\bar{Z}_{ij}$ between $\hat{Z}_{ij}$ and Z_{ij}, for all i,j. The main result follows by showing that

(i) A CLT can be applied to the second and third terms of (2.25).

(ii) The first term converges in probability to 0.

(iii) The last three terms (second-order terms) converge in probability to 0.

I start from the last three terms in (2.25). Condition (2.13), compact support and Theorem 2.2 implies that

$$\begin{aligned}
&\left|-\frac{1}{2}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))^2\bar{Q}_{ij}(1-\bar{Q}_{ij})(1-2\bar{Q}_{ij})\bar{Z}_{ij}\right| \\
\leqslant&\frac{1}{2}\sqrt{n}\,|\lambda_n|^2\,O_p\left(\frac{\ln n}{n}\right) \\
=&O_p\left(\frac{\ln n}{\sqrt{n}}\right) \\
=&o_p(1).
\end{aligned}$$

By the same argument, it can be shown that

$$\left|-\frac{1}{2}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{Z}_{ij}-Z_{ij})'\nabla_{Z_{ij}Z_{ij}'}s_{\beta,ij}(\beta_0,\bar{A}_i(\beta_0),\bar{q}_{ij})(\hat{Z}_{ij}-Z_{ij})\right| = o_p(1)$$

$$\begin{aligned}
&\left|-\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))\left[\bar{Q}_{ij}(1-\bar{Q}_{ij})(1-2\bar{Q}_{ij})\bar{Z}_{ij}\beta_0'(\hat{Z}_{ij}-Z_{ij})+\bar{Q}_{ij}(1-\bar{Q}_{ij})(\hat{Z}_{ij}-Z_{ij})\right]\right| \\
&= o_p(1).
\end{aligned}$$

Then I consider the first term in (2.25). By Lemma 2.2,

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}s_{\beta,ij}(\beta_0,A_i(\beta_0),q_{ij}) \\
=&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(G_{ij}-Q_{ij})Z_{ij} \\
\leqslant&\frac{1}{\sqrt{n}}\sup_{Z\in\mathbb{Z}}|Z|\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}(G_{ij}-Q_{ij}) \\
\leqslant&\frac{1}{\sqrt{n}}O_p\left(1\right) \\
=&o_p(1),
\end{aligned}$$

where the second inequality comes from the fact that $\frac{1}{\sqrt{n(n-1)}}\sum_{i=1}^{n}\sum_{j\neq i}(G_{ij}-Q_{ij})=O_p(1)$.

This is true because G_{ij} are independent conditional on A and X. Applying the central limit theorem yields the desired conclusion.

Then look at the second term. Applying Lemma 2.2 and 2.3 yields

$$
\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{A}_i(\beta_0)-A_i(\beta_0))Q_{ij}(1-Q_{ij})Z_{ij}\\
=&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{\sum_{j\neq i}(G_{ij}-Q_{ij})}{\sum_{j\neq i}Q_{ij}(1-Q_{ij})}+\frac{\sum_{j\neq i}\hat{Q}_{ij}-Q_{ij}}{\sum_{j\neq i}Q_{ij}(1-Q_{ij})}+O_P\left(\frac{1}{n}\right)\right)Q_{ij}(1-Q_{ij})Z_{ij}\\
=&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{\sum_{j\neq i}\hat{Q}_{ij}-Q_{ij}}{\sum_{j\neq i}Q_{ij}(1-Q_{ij})}\right)Q_{ij}(1-Q_{ij})Z_{ij}+o_p(1)\\
=&\frac{1}{n^{3/2}}\sum_{i=1}^{n}\left(\frac{\sum_{j\neq i}Q_{ij}(1-Q_{ij})Z_{ij}}{\sum_{j\neq i}Q_{ij}(1-Q_{ij})}\right)\sum_{j\neq i}\left(\frac{\exp(Z_{ij}'\beta_0+A_i(\beta_0))}{1+\exp(Z_{ij}'\beta_0+A_i(\beta_0))}(\hat{Z}_{ij}-Z_{ij})\right)+o_p(1).
\end{aligned}
$$

The sum of the second and third terms can be written as

$$
=\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}M_{ij}(\hat{Z}_{ij}-Z_{ij})+o_p(1), \tag{2.26}
$$

where $M_{ij}=\left(\frac{\sum_{j\neq i}Q_{ij}(1-Q_{ij})Z_{ij}}{\sum_{j\neq i}Q_{ij}(1-Q_{ij})}\right)\frac{\exp(Z_{ij}'\beta_0+A_i(\beta_0))}{1+\exp(Z_{ij}'\beta_0+A_i(\beta_0))}I_d+\left[Q_{ij}(1-Q_{ij})Z_{ij}\beta_0'+Q_{ij}I_d\right]$.

Define $\zeta_{ij}=(W_{ij}',G_{ji},\frac{1}{n-2}\sum_{k\neq i,j}G_{jk})'$. As defined in Section 2.3, $\hat{Z}_{ij}:=(W_{ij}',\hat{q}_{ji},\frac{1}{n-2}\sum_{k\neq i,j}\hat{q}_{jk})'$. I will show that

$$
\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}M_{ij}(\hat{q}_{ij}-G_{ij})=0, \tag{2.27}
$$

so that

$$
\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}M_{ij}(\hat{Z}_{ij}-\zeta_{ij})=0,
$$

and hence I can replace $\hat{Z}_{ij}$ in (2.26) with ζ_{ij}. To see why (2.27) holds, observe that

$$
\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j\neq i}M_{ij}\left(\frac{\sum_{k=1}^{n}\sum_{l\neq k}G_{ij}\mathbf{1}\{W_{k,l}=W_{ij}\}}{\sum_{k=1}^{n}\sum_{l\neq k}\mathbf{1}\{W_{k,l}=W_{ij}\}}-G_{ij}\right)=0,
$$

and the claim follows. Define $V_i = \frac{1}{n}\sum_{j\neq i} M_{ij}(\zeta_{ij} - Z_{ij})$. Then (2.26) can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i + o_p(1).$$

where $\{V_i\}_{i=1}^n$ is independently distributed, conditional on X, σ.

Step 3. Demonstration of Asymptotic Normality of the second and third term in (2.25).

To apply CLT, one needs to check the Lindeberg condition. Take any vector $a \in \mathbb{R}^d$, the conditional mean of $\frac{1}{\sqrt{n}}a'V_i$ is 0

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}a'V_i\middle|X,\sigma\right] = 0.$$

The conditional variance of $\frac{1}{\sqrt{n}}\sum_i a'V_i$ given X, σ is

$$Var\left(\frac{1}{\sqrt{n}}\sum_i a'V_i\middle|X,\sigma\right) = \frac{1}{n}\sum_i \mathbb{E}\left[(a'V_i)^2|X,\sigma\right] := \Omega_n.$$

By compact support, Condition (2.13), and Lemma 2.1,

$$\frac{\max_i |\frac{1}{\sqrt{n}}a'V_i|}{\sqrt{\Omega_n}} \xrightarrow{p} 0.$$

To check the Lindeberg condition, note that for any $\epsilon > 0$

$$\begin{aligned}
&\frac{1}{\Omega_n}\sum_i \mathbb{E}\left[\frac{1}{n}(a'V_i)^2\mathbf{1}\left\{\frac{|\frac{1}{\sqrt{n}}a'V_i|}{\sqrt{\Omega_n}} > \epsilon\right\}\middle|X,\sigma\right] \\
\leqslant &\frac{1}{\Omega_n}\sum_i \mathbb{E}\left[\frac{1}{n}(a'V_i)^2\mathbf{1}\left\{\frac{\max_i |\frac{1}{\sqrt{n}}a'V_i|}{\sqrt{\Omega_n}} > \epsilon\right\}\middle|X,\sigma\right] \\
\leqslant &\frac{1}{\Omega_n}\sum_i \mathbb{E}\left[\frac{1}{n}(a'V_i)^2\middle|X,\sigma\right] = 1.
\end{aligned}$$

By the dominated convergence theorem, the Lindeberg condition follows, i.e. for any $\epsilon > 0$

$$\frac{1}{\Omega_n}\sum_i \mathbb{E}\left[\frac{1}{n}(a'V_i)^2\mathbf{1}\left\{\frac{|\frac{1}{\sqrt{n}}a'V_i|}{\sqrt{\Omega_n}} > \epsilon\right\}\middle|X,\sigma\right] \xrightarrow{p} 0.$$

By Lindeberg-Feller CLT, for any $a \in \mathbb{R}^d$,

$$\Omega_n^{-1/2}\frac{1}{\sqrt{n}}\sum_i a'V_i \xrightarrow{d} N(0,1).$$

Combining with the result in *Step 1*, this yields the desired conclusion that for any $a \in \mathbb{R}^d$,

$$\frac{\sqrt{n}a'(\hat{\beta}-\beta_0)}{\|a\|^{-1/2}(a'\mathcal{I}_0^{-1}\Omega_n\mathcal{I}_0^{-1}a)^{1/2}} \xrightarrow{d} N(0,1).$$

□

Chapter 3

Identification and Estimation in Differentiated Products Markets where Firms Affect Consumers' Attention

3.1 Introduction

Demand estimation in differentiated product markets is an important topic in modern industrial organization studies. The approach proposed by Berry et al. (1995) (hereafter BLP) plays a central role in this area of research. Though allowing heterogeneous tastes, traditional BLP assumes that consumers pay full attention to all the products available on the market. However, there is various empirical evidence that consumers have limited and heterogeneous attention and make choices only from their own "consideration sets" (Goeree (2008), Reutskaja et al. (2011), Draganska and Klapper (2011), Conlon and Mortimer (2013), Honka et al. (2017)), among others). Ignoring this issue may lead to a biased estimation of both demand and supply.

When incorporating limited attention into the model, several features of consideration sets raise concern. First of all, they are usually formed endogenously and are affected by firms' (optimally chosen) marketing inputs such as advertising, packaging, and listing of products on websites (Ursu (2018)). As a result, heterogeneous consideration sets are an equilibrium outcome reflecting both the demand and supply sides. Second, however, different from other equilibrium outcomes, consideration sets are usually unobserved by researchers. Third, consideration sets can be formed through many mechanisms. The modeling of their formation process is often parametric and context-based. Identification and estimation of such a model are usually infeasible once the functional form or modeling assumptions have changed. Due to these concerns, separately recovering preference, attention, and supply side features in the entire demand and supply system is an interesting topic that hasn't been well studied in the previous literature.

Assuming the availability of market-level data, this paper studies the identification and estimation of a nonparametric demand and supply system when consumers have limited attention. On the demand side, I propose a nonparametric model for utility, which allows rich heterogeneity in preferences. To characterize consumers' limited attention, for each consumer-product pair, I propose a latent attention score which is a nonparametric function of the product- and consumer-specific characteristics. The attention score function proposed in this paper is flexible in three aspects. First, it's nonparametric and thus accommodates many reduced-form representations of drivers of inattention used previously in applied work. Second, it allows the unobservable determinants of utility and attention to be arbitrarily correlated with each other. Third, it allows correlated consideration among products through unobservables. These three features distinguish the model in this paper from many limited attention models studied in the previous literature. Another essential feature of the utility and attention models in this paper is that it allows the endogeneity of prices and the firm's marketing inputs. Though the nonparametric identification of models with price endogeneity has been studied previously by Berry and Haile (2014) and Berry and Haile (2018), no previous literature has studied the nonparametric identifiability of a model with consumers' attention influenced by endogenous marketing inputs.

On the supply side, following Berry and Haile (2014), this paper characterizes firms' optimal choices by a set of first-order conditions without specifying the form of the oligopoly model. In addition to the traditional production decision, I add firms' decision for marketing inputs into the model and showed that the optimal marketing decision can be characterized by an additional first-order condition in a similar fashion. I allow for latent cost shocks and unobserved heterogeneity in cost functions.

To identify the model, I combine the demand and supply model in a single system of nonparametric simultaneous equations. In the system, the exogenous demand and cost shifters serve as exclusive regressors, whose variations can be exploited to show identification. The identification of demand and supply functions follows Matzkin (2015). By connecting the elements of the structural system with conditional densities of observed variables, it can be shown that desired features of the structural functions, such as derivatives and ratios of derivatives, can be expressed as easily computed functionals of the conditional densities.

From the identification of the demand and supply system, this paper goes one step forward to identify the latent utility and attention functions on the consumer's side and the marginal cost functions on the firm's side. To separately identify the features of utility and attention functions, I exploit the exclusion restriction that prices only affect utility and marketing inputs only affect attention. In the general model with nonadditive utility and attention functions, I show the identification of ratios of derivatives of the structural functions by exploiting the cross-market variation of the market share of the outside option. Then I consider a more restrictive additive-separable specification for the utility and attention functions and show identification of the derivatives of utility and attention functions with respect to product features. To recover the marginal cost functions, one needs to have information on the specific oligopoly model, since the structural functions identified in the simultaneous system are actually reduced forms combining both the marginal cost functions and the (unknown) oligopoly first-order conditions. This can be achieved by investigating a testable condition proposed by Berry and Haile (2014), once the demand and supply system is identified.

Following the constructive identification results, one can propose a nonparametric estimator for the demand and supply system by replacing the conditional densities of observables with their Kernel estimators. The asymptotic properties of this estimator are studied by Matzkin (2015). I extend the result in Matzkin's paper to the estimators of derivatives of utilities and attention functions. The asymptotic properties of these estimators follow directly by the delta method. The rest of the paper consists of six sections. In Section 2, I briefly discuss the place of this paper in the previous literature. In Section 3, I describe the model studied in this paper. In Section 4, I show the identification of the demand and supply side structural functions. In Section 5, I propose a nonparametric estimator based on the constructive identification results in Section 4. I conclude in Section 6.

3.2 Related Literature

The demand side structure of this paper is related to the large body of literature on discrete choice models when consumers consider subsets (consideration sets) of the entire set of alternatives. This topic has long been studied in the applied literature, where researchers usually assume that consideration sets are observable, or that auxiliary data on consideration sets are available (recent work includes Reutskaja et al. (2011), Draganska and Klapper (2011), Conlon and Mortimer (2013), Honka and Chintagunta (2017))). Different from this strand of research, this paper doesn't require the availability of any auxiliary data on the consideration sets.

Without assuming that consideration sets are observable, many theoretical papers have shown the identifiability of the discrete choice model under different settings. Masatlioglu et al. (2012), Manzini and Mariotti (2014) and Cattaneo et al. (2020) rely on exogenous changes to the set of alternatives (menu hereafter) available on the market. Data satisfying this hypothetical requirement could be difficult to find outside of an experimental setting. Instead of exploiting the change of menus, other researchers put additional assumptions on preferences to obtain point identification. Dardanoni et al. (2020) assumes homogeneous preference and that individuals only have the capacity to consider a certain number of alternatives. In a recent survey by Crawford et al. (2021), the authors mentioned a differencing-out approach dating back to McFadden (1978), where the unobserved consideration sets are differenced out by using a subset of the true choice set in estimation. However, this approach requires multinomial preference and the stability of consideration sets over time. Moreover, it cannot identify the attention parameter and cannot deal with market-level data. A recent paper by Abaluck and Adams-Prassl (2021) showed the identification of the derivatives of demand functions using the asymmetry of demand responses in limited-consideration models. Though there is no explicit assumption on preference, their key identification assumptions - Slutsky symmetry and translation invariance of the demand functions essentially require the utility to be linear in price and the marginal disutility of price is the same for different products. Moreover, their identification result requires a large support assumption of the price vector. They assume that when the price of a product equals ∞, it's conceptually the same as when this product is removed from the menu.

On the demand side, this paper differs from the previous literature on discrete choice with limited attention in the following ways. First, the nonparametric specification of the utility function in this paper imposes no additional restrictions on the preference. Second, the structure of consideration set formation also takes a very general nonparametric form, which allows arbitrary correlation among unobservables driving utility and attention. This framework is more general than the previous work where each product is considered by the consumer independently, and consideration is independent with preference conditional on observables (see, for example, Manzini and Mariotti (2014) and Abaluck and Adams-Prassl (2021)). Barseghyan et al. (2021) also allows correlation between unobservables driving consideration and utility. They remain completely agnostic about the consideration set formation process, which is more general than the model in this paper. However, the cost is that they only get partial identification. The third perspective in which this paper differs from the previous literature is that the data structure I use is market-level. Only the market share and product-specific characteristics are observed by the researcher. I show that the nonparametric structural model can be identified as long as the exogenous variables in the system have enough variation across markets. No exogenous change in the availability of products is required. Fourth, the identification of structural functions in this paper doesn't rely on large support assumptions, which can be unrealistic in applications.

The most important feature of this paper that differs from previous work is that I incorporate price and marketing endogeneity in the model. The formation of consideration sets is not only endogenous but also affected by the firm's endogenous choices. Instead of looking purely at the demand side discrete choice problem, one has to consider the interaction between the demand and supply side to properly identify the model. In a theoretic paper, Eliaz and Spiegler (2011) studies a model in which competing firms use costly marketing devices to influence consideration sets. They assume homogeneous consumers and firms and deterministic consideration while this paper allows rich heterogeneities in preference, consideration, and the supply side. Another difference between the model in this paper from theirs in that they consider firms' joint decisions on product quality and marketing devices, while this paper treats product quality as fixed. In applied literature, the work by Goeree (2008) is under a similar framework as this paper, but their model is parametric. In this paper, I show the nonparametric identifiability of the model which encompasses the model studied in Goeree (2008) and many other potential specifications estimated in the applied literature. This paper thus contributes to the nonparametric identification of demand and supply in differentiated product markets. The methods used in this paper closely follow conclusions in Berry and Haile (2014). Their model incorporates rich preference heterogeneities. I extend their results by adding information heterogeneities into the model, where a consideration stage is added to the demand side and firms' marketing decisions are added to the supply side.

3.3 The Model

3.3.1 The Demand Side

In this part, I describe the demand side of the model consisting of both the preference and the attention of consumers. The preference is characterized by a random utility model. For the attention, I constructed a nonparametric rule which can accommodate many reduced-form representations of drivers of inattention used previously in applied work.

3.3.1.1 The Preference

Suppose there are T markets in the economy. Each of the markets consists of a continuum of consumers with measure one. On each market, there is a menu of products that are available $\mathcal{J} = \{0, 1, ..., J\}$. Consumer i on market t's indirect utility from choosing product $j \in \{1, ..., J\}$ is given by:

$$U_{ijt} = u^j \left(x_{jt}, \xi_{jt}, p_{jt}, \epsilon_{ijt} \right),$$

where u^j is an unknown function; p_{jt} is the price of product j; x_{jt} and ξ_{jt} are, respectively, observed and unobserved product characteristics. In practice, each product may have multiple observed characteristics. Here, for simplicity, I suppress the dependence of utility on other demand shifters and leave only one demand shifter in the model, as the variation of a scalar x_{jt} is enough for identification. Since price p_{jt} is an equilibrium outcome, it's correlated with ξ_{kt} for all $k \in 1, ..., J$. Let the vector $\epsilon_{it} := (\epsilon_{i1t}, ..., \epsilon_{iJt})$ denote individual's heterogeneous tastes towards products. I assume the vector ϵ_{it} is i.i.d. across i, t with an unknown joint distribution. Note that the elements of vector ϵ_{it} i.e. consumer i in market

t's tastes towards different products may be correlated. Let $x_t := (x_{1t}, ..., x_{Jt})$, $\xi_t := (\xi_{1t}, ..., \xi_{Jt})$ and $p_t := (p_{1t}, ..., p_{Jt})$ denote the vector of observed and unobserved product characteristics, and the price vector, respectively. Without loss of generality, I normalize the utilities of good 0 for each consumer on each market: $U_{i0t} = 0$ for all i and all t.

Following Berry and Haile (2014), I restrict that x_{jt} and ξ_{jt} enters the utility through a linear index $\delta_{jt} = x_{jt}\beta + \xi_{jt}$, and normalize the scale by letting $\beta = -1$ for all j, so that

$$\begin{aligned} U_{ijt} &= u^j \left(\delta_{jt}, p_{jt}, \epsilon_{ijt}\right) \\ &= u^j \left(-x_{jt} + \xi_{jt}, p_{jt}, \epsilon_{ijt}\right). \end{aligned} \tag{3.1}$$

As discussed in Berry and Haile (2014), the linear structure of the index is stronger than necessary. The minimum requirement for identification is that x_{jt} and ξ_{jt} enter the utility through an index that is strictly monotone in ξ_{jt}. The normalization of $\beta = -1$ is just for simplicity of notation in the later part of this paper, this is innocuous because theoretically, one could normalize β to any real number.

Moreover, I impose the following assumptions on the function u^j and the error term ϵ_{jt}:

Assumption 3.1. *(i) $\epsilon_{it} := (\epsilon_{i1t}, ..., \epsilon_{iJt})$ is distributed independently with x_t, ξ_t, p_t with a joint CDF that is continuous. (ii) For all $j \in \{1, ..., J\}$, for all p_j and ϵ_{ij} on their support, function u^j is continuously differentiable in δ_j; for all δ_j and ϵ_{ij} on their support, function u^j is continuously differentiable in p_j; for all δ_j and p_j on their support, function u^j is continuously differentiable in ϵ_{ij}. (iii)For all $j \in \{1, ..., J\}$, for all δ_j and p_j on their support, $u^j(\delta_j, p_j, \epsilon_{ij})$ is strictly increasing in ϵ_{ij}.*

3.3.1.2 The Attention

Let Φ_{ijt} denote the latent attention score of consumer i in market t towards product j. It's some unknown function ϕ^i of product characteristics index δ_{jt}, firm's marketing input M_{jt} and a consumer- and product-specific heterogeneous error term η_{ijt}.

$$\begin{aligned}\Phi_{jt} &= \phi^j(\delta_{jt}, M_{jt}, \eta_{jt}) \\ &= \phi^j(-x_{jt} + \xi_{jt}, M_{jt}, \eta_{ijt}),\end{aligned} \tag{3.2}$$

where the vector $\eta_{it} := (\eta_{i1t}, ..., \eta_{iJt})$ is i.i.d. across i, t. I require the following assumptions on function ϕ^j and the error term η_{it}:

Assumption 3.2. *(i) $\eta_{it} := (\eta_{i1t}, ..., \eta_{iJt})$ is distributed independently with x_t, ξ_t, M_t with a joint CDF that is continuous. (ii) For all $j \in \{1, ..., J\}$, for all M_j and η_{ij} on their support, function ϕ^j is continuously differentiable in δ_j; for all δ_j and η_{ij} on their support, function ϕ^j is continuously differentiable in M_j; for all δ_j and M_{ij} on their support, function ϕ^j is continuously differentiable in η_{ij}. (iii)For all $j \in \{1, ..., J\}$, for all δ_j and M_j on their support, $\phi^j(\delta_j, M_j, \eta_{ij})$ is strictly increasing in η_{ij}.*

In the general model in (3.1), where the individual-specific error terms enter the attention function nonadditively, the independence assumption in (i) is much weaker than it would be if the error terms enter additively. Assumption 3.2(ii) and (iii) are required for the identification of utility functions and attention functions in Section 3.4.2.

Consumer i always considers product 0, while he will consider product $j \in \{1, ..., J\}$ if his latent attention score of that product is larger than 0. In other words,

$$\begin{cases} 0 \in \mathcal{C}_t \\ j \in \mathcal{C}_t \text{ iff } \Phi_{jt} > 0, \text{ for } j = 1, ..., J, \end{cases}$$

where $\mathcal{C}_{it}$ denote the consumer's consideration set.

The nonparametric attention model in this paper is very flexible. It can accommodate many reduced-form representations of drivers of inattention used previously in applied work, as shown in the following three examples. The first two examples describe models with independent and correlated considerations among alternatives. The third example describes a model incorporating consumers' limited awareness and costly search.

Example 3.1 (Limited Consumer Awareness)**.** Goeree (2008) studies a model of limited consumer awareness in the U.S. personal computer industry. The model assumes that each product l has an independent (conditional) probability of being considered. Advertising on product l helps inform consumers about that product. To fit the context of aggregate data in this paper, I omit the individual heterogeneities in the original model of Goeree (2008). The conditional probability that a consumer is aware of product j is given by

$$q_j = \frac{\exp\left(\gamma_j\right)}{1 + \exp\left(\gamma_j\right)}$$

where

$$\gamma_j = \varphi M_j + \rho M_j^2 + M_j \Psi_f + \vartheta x_j.$$

M_j stands for the number of advertisements of product j on certain media. Ψ_f is a firm fixed effect and x_j is the PC age measured in quarters. Let η_{ij} follow a logit distribution for all i, j. The information technology proposed by Goeree (2008) can be regarded as a parametric special case of the model in this paper since

$$\phi^j(x_j, M_j, \eta_{ij}) = \varphi M_j + \rho M_j^2 + M_j \Psi_f + \vartheta x_j + \eta_{ij}$$

and

$$q_{ij} = Pr\left(\phi^j(x_j, M_j, \eta_{ij}) > 0 \mid x_j, M_j, \nu_i\right) = \frac{\exp(\varphi M_j + \rho M_j^2 + M_j \Psi_f + \vartheta x_j)}{1 + \exp(\varphi M_j + \rho M_j^2 + M_j \Psi_f + \vartheta x_j)}.$$

□

Example 3.2 (Limited Attention Due to Imperfect Information and Cognitive Limits)**.** In a real choice environment, consumers may not fully know the utility-relevant attributes of products, for example, a consumer buying a car may not be aware of the price; applicants for insurance may not be aware of the premium. Before making the purchase decision, however, they may receive some signal helping them refine their posterior beliefs about the unknown attribute. Moreover, if the consumer is willing to do a further inspection, they can fully discover the true value of this attribute (for example, through quoting). Suppose the consumer's information acquisition process can be characterized by two steps where in the first step he/she pins down the set of alternatives for further inspection by comparing the posterior expected utility of alternatives with the utility of outside options. Let $F_{p_j,\epsilon_{ij}|M_j,\eta_{ij}}$ denote the posterior belief about the joint distribution of (p_j, ϵ_{ij}) conditional on product j's marketing device M_j and a consumer specific signal η_{ij}. The posterior belief may come from certain information structures whose sufficiency is affected by the marketing device put on product j. For example, a car advertisement mentioning the minimum monthly payment for buying/leasing that car reveals more information about the price compared with an advertisement without the minimum payment information. In fact, such an advertisement can be viewed as a signal revealing the lower bound of the price. Upon forming the posterior belief, consumers can calculate their posterior expected utilities denoted as $\phi^j(\delta_j, M_j, \eta_{ij})$ by

$$\phi^j(\delta_j, M_j, \eta_{ij}) = \int u^j(\delta_j, p_j, \epsilon_{ij}) dF_{p_j,\epsilon_{ij}|M_j,\eta_{ij}}.$$

In the second step, the consumer will further investigate the set of products with posterior expected beliefs greater than 0, which is the utility of the outside option. That means the consumer's consideration set will consist of the outside option and all $j \in \{1, ..., J\}$ such that $\phi^j(\delta_j, M_j, \eta_{ij}) > 0$. □

Example 3.3 (Simultaneous Search)**.** Suppose consumer $i's$ indirect utility for product j

takes a linear form

$$U_{ij} = \beta\delta_j + \gamma p_j + \epsilon_{ij}.$$

δ_j is product j's feature. ϵ_{ij} is a consumer- and product-specific heterogeneous taste. δ_j and ϵ_{ij} are observable by the consumer before the search. p_j is the price of product j, which is not known by the consumer until the search. Consumers know the distribution of p_j. Denote their expectation of p_j as μ_j^p. Based on the information they have, consumers can form their expected utilities before the search:

$$\begin{aligned}\mathbb{E}\left[U_{ij}\right] &= \beta\delta_j + \gamma\mathbb{E}\left[p_j\right] + \epsilon_{ij} \\ &= \beta\delta_j + \gamma\mu_j^p + \epsilon_{ij}.\end{aligned}$$

Let c_j denote the cost of the search for product j, which depends on the marketing input of product j and a consumer-specific component η_{ij} and takes a linear form $c_j(M_j, \eta_{ij}) = \beta^c M_j + \eta_{ij}^c$. Consumers simultaneously search all products whose expected utility less search cost is positive. They can only choose from the products that have been searched. That means consumer i will consider product $j \in \{1, ..., J\}$ if and only if

$$\begin{aligned}&\beta\delta_j + \gamma\mu_j^p + \epsilon_{ij} - (\beta^c M_j + \eta_{ij}^c) > 0 \\ \Leftrightarrow &\beta\delta_j + \gamma\mu_j^p - \beta^c M_j + (\epsilon_{ij} - \eta_{ij}^c) > 0 \\ \Leftrightarrow &\phi^j\left(\delta_j, M_j, \eta_{ij}\right) > 0\end{aligned}$$

where $\eta_{ij} := \epsilon_{ij} - \eta_{ij}^c$. In this example, the unobserved consumer-specific component η_{ij} in the utility function and the unobserved consumer-specific component η_{ij} in the attention function are correlated. □

Example 3.4 (Listings by Online Platforms.)**.** Previous research has found the power of listings on consumers' attention and purchasing decisions online (see, for example, Ursu

(2018)). Consumers searching for a laptop online may enter some attribute δ_j into the search box. The underlying listing algorithm of the website will then process the attribute the consumer enters, and generate a list of products with customized ranking, which depends on the marketing inputs of products and the consumer's characteristics. Consumers may search multiple times, but because of cognitive limits, they may only pay attention to the products that have ever appeared on the first page of the search result. Let M_j denote the marketing input of product j. Let Φ_{ij} denote the frequency that product j appears on the first page of consumer i's search result. It's given by $\Phi_{ij} = \phi^j(\delta_j, M_j, \eta_{ij})$. The consumer will then consider product j if and only if $\phi^j(\delta_j, M_j, \eta_{ij}) > 0$. □

3.3.1.3 The Demand System and Invertibility

Given the preference and attention model, the market share of product j in market t is the probability that j is considered by the consumer and its utility is the highest in the consideration set, i.e.

$$s_{jt} = Pr\left(\Phi_{jt} > 0 \text{ and } U_{jt} > 0 \text{ and } U_{jt} > U_{kt} \text{ for } \Phi_{kt} > 0, k = 1, ..., J \mid \delta_t, p_t, M_t\right) \tag{3.3}$$

$$= \sigma_j\left(\delta_t, p_t, M_t\right), \tag{3.4}$$

where σ_j is some unknown function to be identified. The equality in (3.4) comes from the fact that after integrating out the consumer heterogeneities ϵ_{it} and η_{it}, the market share of product j is a function of the product features δ_t, the prices p_t and the marketing inputs M_t of all products.

A key step for identification is the inversion of the demand system as shown in Berry et al. (2013). The invertibility requires the following assumption:

Assumption 3.3 (Connected Substitutes)**.** *Let λ denote either δ, $-p$, or M. The demand for products $(0, 1, ..., J)$ satisfy:*

1. *(Weak substitutes): $\sigma_j(\delta, p, M)$ is nonincreasing in λ_k for all $j \in \{0, 1, ..., J\}$, $k \notin \{0, j\}$*

and for all $(\delta, p, M) \in R^{3J}$.

2. *(Connected Strict Substitution): For any nonempty* $\mathcal{K} \subseteq \{1, ..., J\}$, *there exist* $k \in \mathcal{K}$ *and* $j \notin \mathcal{K}$ *such that* $\sigma_j(\delta, p, M)$ *is strictly decreasing in* λ_k *for all* $(\delta, p, M) \in$ *supp*$((\delta, p, M))$.

Lemma 3.1 (Lemma 1 in Berry and Haile (2014))**.** *For any price vector p, marketing input vector M, and any market share vector* $s = (s_1, ..., s_J)$ *such that* $s_j > 0$ *for all j and* $\sum_{j=1}^{J} s_j < 1$. *Under Assumption 3.1-3.3, there is at most one vector* δ *such that* $\sigma_j(\delta, p, M) = s_j, \forall j$.

With this result, for any (s_t, p_t) in their support, one can write

$$\underbrace{-x_{jt} + \xi_{jt}}_{=:\delta_{jt}} = \sigma_j^{-1}(s_t, p_t, M_t), \ j = 1, ..., J. \tag{3.5}$$

3.3.2 The Supply Side

On the supply side, instead of specifying a particular supply model, following Berry and Haile (2014), I require the less restrictive condition that a set of first-order conditions characterize firms' optimal choices for prices or quantities and marketing inputs. The existence of first-order conditions requires the market share functions $\sigma_j(\delta, p, M)$ to be continuously differentiable with respect to prices and marketing inputs. Formally, the following assumption analogous to Assumption 6 in Berry and Haile (2014) is required. Part (ii) of the assumption slightly strengthens the connected strict substitutes assumption in Assumption 3.3 by ruling out a zero derivative when σ_j is strictly increasing in λ_k.

Assumption 3.4. *(i)* $\sigma_j(\delta, p, M)$ *is continuously differentiable with respect to* p_k *and* M_k, $\forall j, k \in \{1, ..., J\}$; *(ii) For any nonempty* $\mathcal{K} \subseteq \{1, ..., J\}$, *there exist* $k \in \mathcal{K}$ *and* $j \notin \mathcal{K}$ *such that* $\frac{\partial \sigma_j(\delta,p,M)}{\partial p_k} > 0$ *and* $\frac{\partial \sigma_j(\delta,p,M)}{\partial M_k} < 0$ *for all* $(\delta, p, M) \in$ *supp*$((\delta, p, M))$.

Under Assumption 3.3 and 3.4, the first order conditions defining firms' behavior can be expressed as the marginal cost of production mc_{jt} and the marginal cost of marketing input

mc^a_{jt} as functions of equilibrium prices, equilibrium quantities, and equilibrium marketing inputs. Formally, the following high-level assumption is imposed:

Assumption 3.5. *For each $j = 1, ..., J$, there exists a (possibly unknown) function ψ_j and γ_j such that, for all δ_t, s_t in their support,*

$$mc_{jt} = \psi_j(\delta_t, s_t, p_t, M_t);$$
$$mc^a_{jt} = \gamma_j(\delta_t, s_t, p_t, M_t).$$

Building upon similar arguments as Berry and Haile (2014), it can be shown that given Assumption 3.3 (connected substitutes) and 3.4, after adding the optimal marketing inputs decision into the model, Assumption 3.5 follows from the first order conditions of a variety of supply models. To see how the optimal decision on production and marketing inputs interact with each other, I show the supply model under price setting and quantity setting in the following two examples.

Example 3.5 (Price Setting)**.** Consider a complete information simultaneous price-setting game with Nash equilibrium as the solution concept. Let $\mathcal{J}_j$ denote the set of products produced by the firm that produces goods j. The first order conditions for the price and marketing inputs of good j are:

$$s_{jt} + \sum_{k \in \mathcal{J}_j} (p_{kt} - mc_{kt}) \frac{\partial s_{kt}}{\partial p_{jt}} = 0$$
$$\sum_{k \in \mathcal{J}_j} (p_{kt} - mc_{kt}) \frac{\partial s_{jt}}{\partial M_{jt}} - mc^a_{jt} = 0$$

The first-order condition of all firms can be written in matrix form as

$$s_t + \Delta_t(p_t - mc_t) = 0$$
$$\Gamma_t(p_t - mc_t) - mc^a_t = 0,$$

where the (j,k) elements of matrix Δ_t and Γ_t are $\partial s_{kt}/\partial p_{jt}$ and $\partial s_{kt}/\partial M_{jt}$ respectively if j and k are produced by the same firm, and zero otherwise. By simultaneous permutation of rows and columns one can obtain from matrix Δ_t a block diagonal matrix, with each blocking being a principal submatrix of D_t^p, the Jacobian matrix of s_t with respect to p_t. Under Assumption 3.3 and 3.4, one can show that D_t^p is a P-matrix using Theorem 2 in Berry et al. (2013). As a result, Δ_t is invertible and

$$mc_t = p_t + \Delta_t^{-1} s_t$$
$$mc_t^a = -\Gamma_t \Delta_t^{-1} s_t.$$

□

Example 3.6 (Quantity Setting)**.** Consider a complete information simultaneous quantity-setting game with Nash equilibrium as the solution concept. Under Assumption 3.3, by Theorem 1 of Berry et al. (2013), there exists an inverse demand function $p_t = P(\delta_t, s_t, M_t)$. Under Assumption 3.3 and 3.4, by Theorem 2 in Berry et al. (2013), D_t^p is invertible. Then by inverse function theorem, the derivatives of $p_t = P(\delta_t, s_t, M_t)$ exist and are equal to the inverse of D_t^p, that is

$$\frac{\partial p_{kt}}{\partial s_{jt}} = \left[D_t^p\right]_{kj}^{-1}.$$

If the inverse demand function is also differentiable in M, the first order conditions of firms can be written as

$$\sum_{k \in J_j} \frac{\partial p_{kt}}{\partial s_{jt}} s_{kt} + p_{jt} = mc_{jt}$$
$$\sum_{k \in J_j} \frac{\partial p_{kt}}{\partial M_{jt}} s_{kt} = mc_{jt}^a.$$

□

Following Berry and Haile (2014), I define the following cost indices as linear functions of production cost shifter w_{jt} and marketing cost shifter z_{jt} respectively:

$$\kappa_{jt} = w_{jt}\gamma_j + \omega_{jt}$$
$$\lambda_{jt} = z_{jt}\theta_j + \zeta_{jt}.$$

I normalize γ_j and θ_j for each $j \in \{1, ..., J\}$ to be -1 without loss of generality. Moreover, I assume that the cost shifters affect the marginal cost of production and the marginal cost of marketing inputs only through the cost indices κ_{jt} and λ_{jt}.

Assumption 3.6. *For all $j = 1, ..., J$,*

$$mc_{jt} = c_j(s_{jt}, \kappa_{jt})$$
$$mca_{jt} = ca_j(s_{jt}, \lambda_{jt})$$

where c_j is strictly increasing in κ_{jt} and ca_j is strictly increasing in λ_{jt}.

Under Assumption 3.5 and 3.6, the supply-side equilibrium conditions can be written as

$$c_j(s_{jt}, \kappa_{jt}) = \psi_j(\delta_t, s_t, p_t, M_t) \tag{3.6}$$
$$ca_j(s_{jt}, \lambda_{jt}) = \gamma_j(\delta_t, s_t, p_t, M_t). \tag{3.7}$$

The following lemma shows the invertibility of these conditions, a conclusion that can be derived directly from Lemma 2 of Berry and Haile (2014).

Lemma 3.2. *Under Assumption 3.3, 3.5, and 3.6, the cost indices can be written as (unknown) functions of s_t, p_t, M_t:*

$$\underbrace{-w_{jt} + \omega_{jt}}_{=:\kappa_{jt}} = \pi_j^{-1}(s_t, p_t, M_t), \quad j = 1, ..., J. \tag{3.8}$$

$$\underbrace{-z_{jt} + \zeta_{jt}}_{=:\lambda_{jt}} = \rho_j^{-1}(s_t, p_t, M_t), \quad j = 1, ..., J. \tag{3.9}$$

Proof. The index structure of the utility model satisfies Assumption 1 in Berry and Haile (2014). Assumption 3.3, 3.5, and 3.6 are Assumption 2, 7b, and 10 of Berry and Haile (2014) respectively. The conclusion follows directly from Lemma 2 of Berry and Haile (2014). □

3.3.3 The Demand and Supply System

Combining the inverse of demand and supply side equations (3.5), (3.8), and (3.9) yields a system of $3J$ simultaneous equations:

$$\begin{aligned} \underbrace{-x_{jt} + \xi_{jt}}_{=:\delta_{jt}} &= \sigma_j^{-1}(s_t, p_t, M_t), \; j = 1, ..., J. \\ \underbrace{-w_{jt} + \omega_{jt}}_{=:\kappa_{jt}} &= \pi_j^{-1}(s_t, p_t, M_t), \quad j = 1, ..., J. \\ \underbrace{-z_{jt} + \zeta_{jt}}_{=:\lambda_{jt}} &= \rho_j^{-1}(s_t, p_t, M_t), \quad j = 1, ..., J. \end{aligned} \tag{3.10}$$

The researchers observe the exogenous demand and cost shifters $(x_{1t}, ..., x_{Jt}, w_{1t}, ..., w_{Jt}, z_{1t}, ..., z_{Jt})$ and the endogeneous equilibrium outcomes $(s_{1t}, ...s_{Jt}, p_{1t}, ..., p_{Jt}, M_{1t}, ..., M_{Jt})$ for each market t. The joint distribution of ξ_t, ω_t, and ζ_t, as well as the functional forms of σ_j^{-1}, π_j^{-1}, and ρ_j^{-1} for all $j \in \{1, ..., J\}$ are unknown to the researcher and to be identified.

The system of equations in (3.10) takes a form of equations (2.2) in Matzkin (2015), with the demand shifter x_{jt} appearing exclusively in the j-th demand equation, the production cost shifter w_{jt} appearing exclusively in the j-th production equation, and the marketing cost shifter z_{jt} appearing exclusively in the j-th equation derived from optimal choice of marketing inputs.

For the simplicity of discussion, I give the functions, endogeneous and exogeneous variables in (3.10) a unified notation. Since all variables are i.i.d. across markets, I suppress the subscript t from here on. Let the $3J$-dimension vector y denote the vector stacking all the endogenous variables, with the first J elements being $(s_1, ...s_J)$, the $J+1$-th to $2J$-th elements being $(p_1, ..., p_J)$, and the last J elements being $(M_1, ..., M_J)$. Similarly, I denote x as the $3J$-dimension vector stacking all the exogenous variables together, and ξ as the $3J$-dimension vector stacking all the unobservables together. Let function r^j denote σ_j^{-1} for $j = 1, ..., J$, π_{j-J}^{-1} for $j = J+1, ..., 2J$ and ρ_{j-2J}^{-1} for $j = 2J+1, ..., 3J$. That is

$$r^j := \begin{cases} \sigma_j^{-1}, & \text{for } j = 1, ..., J; \\ \pi_{j-J}^{-1}, & \text{for } j = J+1, ..., 2J; \\ \rho_{j-2J}^{-1}, & \text{for } j = 2J+1, ..., 3J; \end{cases}$$

$$y := (s_1, ...s_J, p_1, ..., p_J, M_1, ..., M_J)_{3J\times 1}$$

$$x := (x_1, ..., x_J, w_1, ..., w_J, z_1, ..., z_J)_{3J\times 1}.$$

With the new notations, the system of equations (3.10) can be written as:

$$\begin{aligned} -x_1 + \xi_1 &= r^1(y_1, ..., y_{3J}) \\ &\vdots \\ -x_{3J} + \xi_{3J} &= r^{3J}(y_1, ..., y_{3J}), \end{aligned}$$

where the researcher observes $(x_1, ...x_{3J}, y_1, ...y_{3J})$ in each market and wants to identify the joint distribution of $(\xi_1, ...\xi_{3J})$ and functions $r^1, ..., r^J$.

3.4 Identification

3.4.1 Identifying the Demand and Supply System

3.4.1.1 Identifying the Inverse Function r

The identification follows the argument in Section 2 of Matzkin (2015). Let y denote the vector of all endogenous variables $(y_1, ..., y_{3J})$ and let x denote the vector of all exogeneous variables $(x_1, ..., x_{3J})$. Let r denote the function $r = (r^1, ..., r^{3J})$. Let r_y denote the Jacobian of r. The (j, k)-th element of r_y is $\frac{\partial r^j}{\partial y_k}$, for $j, k \in \{1, ..., 3J\}$. For each value of y in its support, I will show the identification of r_y by exploiting the variation of x. Consider a set where the value of y is held fixed: $\bar{M} \subset \{(y, t_1, ..., t_{3J}) | (t_1, ..., t_{3J}) \in R^{3J}\}$. The constructive identification of the demand and supply system requires the following restrictions on function r and the densities of ξ and x:

Assumption 3.7. *The function $r = (r^1, ..., r^{3J})$ is twice continuously differentiable. The function $r : \mathbf{R}^{3J} \to \mathbf{R}^{3J}$ is injective.*

Assumption 3.8 (Assumption 2.2 in Matzkin (2015))**.** $(\xi_1, ..., \xi_{3J})$ *is distributed independently of* $(x_1, ..., x_{3J})$ *with an everywhere positive and twice continuously differentiable density f_ξ.*

Assumption 3.9 (Assumption 2.3 in Matzkin (2015))**.** $(x_1, ..., x_{3J})$ *has a differentiable density.*

Assumption 3.10 (Assumption 2.5 and 2.6 in Matzkin (2015))**.** *(i) There exist $3J + 1$, not necessarily known values $\xi^{(1)}, ..., \xi^{(3J+1)}$ of ξ, such that the following matrix is invertible:*

$$A(\xi^{(1)}, ..., \xi^{(3J+1)})$$

$$= \begin{pmatrix} \frac{\partial \log f_\xi(\xi^{(1)})}{\partial \xi_1} & \frac{\partial \log f_\xi(\xi^{(1)})}{\partial \xi_2} & \dots & \frac{\partial \log f_\xi(\xi^{(1)})}{\partial \xi_{3J}} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \log f_\xi(\xi^{(3J)})}{\partial \xi_1} & \frac{\partial \log f_\xi(\xi^{(3J)})}{\partial \xi_2} & \dots & \frac{\partial \log f_\xi(\xi^{(3J)})}{\partial \xi_{3J}} & 1 \\ \frac{\partial \log f_\xi(\xi^{(3J+1)})}{\partial \xi_1} & \frac{\partial \log f_\xi(\xi^{(3J+1)})}{\partial \xi_2} & \dots & \frac{\partial \log f_\xi(\xi^{(3J+1)})}{\partial \xi_{3J}} & 1 \end{pmatrix}.$$

(ii)There exist $3J+1$, not necessarily known values $(x^{(1)}, y), ..., (x^{(3J+1)}, y)$ in the set $\bar{M}$, such that for each $k = 1, ..., 3J+1$, $\xi^{(k)} = r(y) + x^{(k)}$, where $\xi^{(k)}$ is as in (i).

The following theorem states that r_y is identified on M under the assumptions above.

Theorem 3.1. *Under Assumption 3.7-3.10, r_y is identified on $\bar{M}$.*

Proof. By Assumption 3.7 and 3.8, r and f_ξ are continuously differentiable, and the function r is bijective. For all (x,y) in the support of (X, Y), the following transformation of variables equation holds

$$f_{Y|X=x}(y) = f_\xi(r(y) + x) \cdot |r_y|.$$

Taking logs and taking derivatives with respect to y_j and x_k on both sides yields:

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial y_j} = \sum_{k=1}^{3J} \frac{\partial \log f_\xi(r(y)+x)}{\partial \xi_k} \frac{\partial r^k(y)}{\partial y_j} + \frac{\partial \log |r_y|}{\partial y_j} \tag{3.11}$$

$$\frac{\partial \log f_{Y|X=x}(y)}{\partial x_k} = \frac{\partial \log f_\xi(r(y)+x)}{\partial \xi_k}. \tag{3.12}$$

Plugging (3.12) into (3.11) and eliminating $\partial \log f_\xi / \partial \xi_k$ yields

$$\underbrace{\frac{\partial \log f_{Y|X=x}(y)}{\partial y_j}}_{=:g_{y_j}(x,y)} = \sum_{k=1}^{3J} \underbrace{\frac{\partial \log f_{Y|X=x}(y)}{\partial x_k}}_{=:g_{x_k}(x,y)} \frac{\partial r^k(y)}{\partial y_j} + \frac{\partial \log |r_y|}{\partial y_j}. \tag{3.13}$$

Denote the derivative of log conditional density over y_j by $g_{y_j}(x,y)$, and the derivative of log conditional density over x_k by $g_{x_k}(x,y)$. Suppose there are $3J+1$ different points on $\bar{M}$:

$(x^{(1)}, y), ..., (x^{(3J+1)}, y)$. Denote the value of $g_{x_k}(x, y)$ evaluated at the point $(x^{(i)}, y)$ as $g_{x_k}^{(i)}$.

By Proposition 2.4 in Matzkin (2015), Assumption 3.10 is equivalent to the condition that there exist $(x^{(1)}, y), ..., (x^{(3J+1)}, y)$ in the set $\bar{M}$ such that the matrix

$$B\left((x^{(1)}, y), ..., (x^{(3J+1)}, y)\right) = \begin{pmatrix} g_{x_1}^{(1)} & g_{x_2}^{(1)} & \dots & g_{x_{3J}}^{(1)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{x_1}^{(3J)} & g_{x_2}^{(3J)} & \dots & g_{x_{3J}}^{(3J)} & 1 \\ g_{x_1}^{(3J+1)} & g_{x_2}^{(3J+1)} & \dots & g_{x_{3J}}^{(3J+1)} & 1 \end{pmatrix}$$

is invertible.

Evaluating (3.13) at these $3J + 1$ points yields $3J + 1$ equations:

$$\begin{aligned} g_{y_j}(x^{(1)}, y) &= g_{x_1}(x^{(1)}, y)\frac{\partial r^1(y)}{\partial y_j} + ... + g_{x_{3J}}(x^{(1)}, y)\frac{\partial r^{3J}(y)}{\partial y_j} + \frac{\partial \log |r_y|}{\partial y_j} \\ g_{y_j}(x^{(2)}, y) &= g_{x_1}(x^{(2)}, y)\frac{\partial r^1(y)}{\partial y_j} + ... + g_{x_{3J}}(x^{(2)}, y)\frac{\partial r^{3J}(y)}{\partial y_j} + \frac{\partial \log |r_y|}{\partial y_j} \\ &\vdots \\ g_{y_j}(x^{(3J+1)}, y) &= g_{x_1}(x^{(3J+1)}, y)\frac{\partial r^1(y)}{\partial y_j} + ... + g_{x_{3J}}(x^{(3J+1)}, y)\frac{\partial r^{3J}(y)}{\partial y_j} + \frac{\partial \log |r_y|}{\partial y_j}. \end{aligned} \tag{3.14}$$

They can also be written in matrix form as

$$\begin{pmatrix} g_{y_j}^{(1)} \\ \vdots \\ g_{y_j}^{(3J)} \\ g_{y_j}^{(3J+1)} \end{pmatrix} = B\left((x^{(1)}, y), ..., (x^{(3J+1)}, y)\right) \begin{pmatrix} \frac{\partial r^1(y)}{\partial y_j} \\ \vdots \\ \frac{\partial r^{3J}(y)}{\partial y_j} \\ \frac{\partial \log |r_y|}{\partial y_j} \end{pmatrix}.$$

The left-hand side vector $\left(g_{y_j}^{(1)}, ..., g_{y_j}^{(3J+1)}\right)'$ and the matrix $B\left((x^{(1)}, y), ..., (x^{(3J+1)}, y)\right)$ are both observable from data. Since matrix B is invertible. The derivatives of r with respect to y_j, $\left(\frac{\partial r^1(y)}{\partial y_j}, ..., \frac{\partial r^{3J}(y)}{\partial y_j}\right)$, are identified. The same argument holds for all $j \in \{1, ..., 3J\}$. $\square$

Define the class of functions r that satisfy Assumption 3.7 as Γ. Given the invertibility of matrix B, Theorem 2.1 in Matzkin (2015) implies that two functions r and $\tilde{r}$ in Γ are observationally equivalent on $\bar{M}$ if and only if their Jacobians are equal. Having identified the Jacobian r_y, one can impose location normalizations to get the exact value of r.

3.4.1.2 Identifying the Demand Function σ

Now I build a relationship between the derivative of the function r and the derivative of σ. Plugging the inverse of functions back into the demand system, I have that for $j \in \{1, 2, ..., J\}$,

$$\begin{aligned} y_j &= \sigma_j \left(-x_1 + \xi_1, -x_2 + \xi_2, ..., -x_J + \xi_J, y_{J+1}, ..., y_{3J}\right) \\ &\equiv \sigma_j \left(r^1(y), r^2(y), ..., r^J(y), y_{J+1}, ..., y_{3J}\right). \end{aligned} \tag{3.15}$$

The right-hand side should be y_j (i.e.s_j). Taking derivatives on both sides of (3.15) with respect to x and y, and by the chain rule, I get the following $3J$ equations:

$$\begin{aligned} 1 &= \frac{\partial \sigma_j}{\partial \delta_1}\frac{\partial r^1(y)}{\partial y_j} + \frac{\partial \sigma_j}{\partial \delta_2}\frac{\partial r^2(y)}{\partial y_j} + ... + \frac{\partial \sigma_j}{\partial \delta_J}\frac{\partial r^J(y)}{\partial y_j} \\ 0 &= \frac{\partial \sigma_j}{\partial \delta_1}\frac{\partial r^1(y)}{\partial y_k} + \frac{\partial \sigma_j}{\partial \delta_2}\frac{\partial r^2(y)}{\partial y_k} + ... + \frac{\partial \sigma_j}{\partial \delta_J}\frac{\partial r^J(y)}{\partial y_k} \text{ for } k \neq j \text{ and } k \in \{1, ..., J\} \\ 0 &= \frac{\partial \sigma_j}{\partial \delta_1}\frac{\partial r^1(y)}{\partial y_l} + \frac{\partial \sigma_j}{\partial \delta_2}\frac{\partial r^2(y)}{\partial y_l} + ... + \frac{\partial \sigma_j}{\partial \delta_J}\frac{\partial r^J(y)}{\partial y_l} + \frac{\partial \sigma_j}{\partial y_l} \text{ for } l \in \{J+1, ..., 3J\}. \end{aligned} \tag{3.16}$$

Writing (3.16) in matrix form:

$$\underbrace{\begin{pmatrix} \frac{\partial r^1}{\partial y_1} & \cdots & \frac{\partial r^J}{\partial y_1} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial r^1}{\partial y_J} & \cdots & \frac{\partial r^J}{\partial y_J} & 0 & \cdots & 0 \\ \frac{\partial r^1}{\partial y_{J+1}} & \cdots & \frac{\partial r^J}{\partial y_{J+1}} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r^1}{\partial y_{3J}} & \cdots & \frac{\partial r^J}{\partial y_{3J}} & 0 & \cdots & 1 \end{pmatrix}}_{RT(r)'} \underbrace{\begin{pmatrix} \frac{\partial \sigma_j}{\partial \delta_1} \\ \vdots \\ \frac{\partial \sigma_j}{\partial \delta_J} \\ \frac{\partial \sigma_j}{\partial y_{J+1}} \\ \vdots \\ \frac{\partial \sigma_j}{\partial y_{3J}} \end{pmatrix}}_{\partial \sigma_j} = l_j \tag{3.17}$$

where l_j is the $3J$-dimensional unit vector with the j-th element equals to 1. As long as the matrix $RT(r)$ is invertible, one can recover the vector of derivatives $\left(\frac{\partial \sigma_j}{\partial \delta_1}, ..., \frac{\partial \sigma_j}{\partial \delta_J}, \frac{\partial \sigma_j}{\partial y_{J+1}}, ..., \frac{\partial \sigma_j}{\partial y_{3J}}\right)$ evaluated at the point $\left(r^1(y), ..., r^J(y), y_{J+1}, ..., y_{3J}\right)$. Formally,

Assumption 3.11. *On the set $\bar{M}$, the matrix $RT(r)$ in (3.17) is nonsingular.*

Corollary 3. *Under Assumption 3.4, 3.7-3.11, the following derivatives are identified on $\bar{M}$ for all $j, k \in \{1, ..., J\}$ and $l \in \{J+1, ..., 3J\}$:*

$$\frac{\partial \sigma_j \left(r^1(y), ..., r^J(y), y_{J+1}, ..., y_{3J}\right)}{\partial \delta_k}$$
$$\frac{\partial \sigma_j \left(r^1(y), ..., r^J(y), y_{J+1}, ..., y_{3J}\right)}{\partial y_l}.$$

Proof. The conclusion follows from Theorem 3.1 and the previous argument in this section.

□

3.4.2 Identifying the Preference and Attention

The previous section established the identification result of the derivatives of $\sigma_j, j = 1, ..., J$ with respect to p and M. Given this information, features of the utility and attention functions can be separately identified by exploiting the exclusion restriction that p only affects utility and M only affects attention. In this section, I first show that the ratios of derivatives $\frac{\partial u^j/\partial \delta_j}{\partial u^j/\partial p_j}$ and $\frac{\partial \phi^j/\partial \delta_j}{\partial \phi^j/\partial M_j}$ are identified when p_j and ϵ_{ij} are not separable in function u^j, and M_j and η_{ij} are not separable in ϕ^j, that is, u^j and ϕ^j take the general form in (3.1) and (3.2) respectively. Then I let p_j and ϵ_{ij}, M_j and η_{ij} enter u^j and ϕ^j in an additive separable way, and show identification of the derivatives $\partial u^j/\partial \delta_j$ and $\partial \phi^j/\partial \delta_j$.

3.4.2.1 Identification of a Nonseparable Model

The identification of the ratios of derivatives $\frac{\partial u^j/\partial \delta_j}{\partial u^j/\partial p_j}$ and $\frac{\partial \phi^j/\partial \delta_j}{\partial \phi^j/\partial M_j}$ exploits the variation of the market share of the outside option and the fact that functions $u^j(\delta_j, p_j, \epsilon_{ij})$ and $\phi^j(\delta_j, M_j, \eta_{ij})$ are invertible with respect to their last arguments. The following assumption states the rank condition necessary for the identification:

Assumption 3.12. *For each j = 1,...,J, and for any (δ_j, p_j, M_j) belonging to its support, there exist 2 not necessarily known values $(\delta_j, \delta_{-j}^{(1)}, p_j, p_{-j}^{(1)}, M_j, M_{-j}^{(1)})$ and $(\delta_j, \delta_{-j}^{(2)}, p_j, p_{-j}^{(2)}, M_j, M_{-j}^{(2)})$ such that $\left(\delta_j, \delta_{-j}^{(i)}\right) = \sigma^{-1}\left(s, \left(p_j, p_{-j}^{(i)}\right), \left(M_j, M_{-j}^{(i)}\right)\right)$, and the matrix*

$$\Pi_j = \begin{pmatrix} \frac{\partial \sigma_0^{(1)}}{\partial p_j} & \frac{\partial \sigma_0^{(1)}}{\partial M_j} \\ \frac{\partial \sigma_0^{(2)}}{\partial p_j} & \frac{\partial \sigma_0^{(2)}}{\partial M_j} \end{pmatrix}$$

has full rank, where $\frac{\partial \sigma_0^{(i)}}{\partial p_j} := \frac{\partial \sigma_0\left(\delta_j, \delta_{-j}^{(i)}, p_j, p_{-j}^{(i)}, M_j, M_{-j}^{(i)}\right)}{\partial p_j}$ and $\frac{\partial \sigma_0^{(i)}}{\partial M_j} := \frac{\partial \sigma_0\left(\delta_j, \delta_{-j}^{(i)}, p_j, p_{-j}^{(i)}, M_j, M_{-j}^{(i)}\right)}{\partial M_j}$ for $i = 1, 2$.

Denote by $\Pi_{\delta_j}^k$ the matrix formed by replacing the k-th column of Π_j with the vector $(\frac{\partial \sigma_0^{(1)}}{\partial \delta_j}, \frac{\partial \sigma_0^{(2)}}{\partial \delta_j})'$. The following Proposition (proved in the Appendix) states that the ratios of derivatives $\frac{\partial u^j/\partial \delta_j}{\partial u^j/\partial p_j}$ and $\frac{\partial \phi^j/\partial \delta_j}{\partial \phi^j/\partial M_j}$ are constructively identified.

Proposition 3.1. *Suppose for each $j = 1, ...J$, the derivatives of σ_j are identified. Under*

Assumption 3.1, 3.2, and 3.12, for all $j = 1, ...J$,

$$\frac{u^j_{\delta_j}(\delta_j, p_j, \epsilon_{ij})}{u^j_{p_j}(\delta_j, p_j, \epsilon_{ij})} = \frac{\left|\Pi^1_{\delta_j}\right|}{\left|\Pi_j\right|},$$
$$\frac{\phi^j_{\delta_j}(\delta_j, M_j, \eta_{ij})}{\phi^j_{M_j}(\delta_j, M_j, \eta_{ij})} = \frac{\left|\Pi^2_{\delta_j}\right|}{\left|\Pi_j\right|}.$$

where ϵ_{ij} is the value at which $u^j(\delta_j, p_j, \epsilon_{ij}) = 0$, and $\delta_j = \sigma_j^{-1}(s_1, ...s_J, p_1, ..., p_J, M_1, ..., M_J)$.

Proof. See Appendix. □

3.4.2.2 Identification of an Additively Separable Model

Till now, additivity between p and ϵ is not required, neither is the separability of them with other parts of the utility. For simplicity, in the following parts of this paper, I will let p_j and ϵ_j enter the utility function as an additive index, and let M_j and η_j enter the attention function as an additive index. Moreover, I will let the two additive indices be separable from other parts of the utility and attention, respectively. With a slight abuse of notations, the following assumption is maintained for the rest of this paper:

Assumption 3.13. *The indirect utility and attention functions take the following additive and separable forms:*

$$U_{ij} = u^j\left(\delta_j\right) - p_j + \epsilon_{ij}$$
$$\Phi_{ij} = \phi^j\left(\delta_j\right) + M_j + \eta_{ij}.$$

Under the separable and additive setting, instead of using only the outside option, I can utilize the demand functions of all alternatives $j = 0, ..., J$ for identification:

$$s_j = \sigma_j\left(\delta_1, ..., \delta_J, p_1, ..., p_J, M_1, ..., M_J\right)$$
$$= \sum_{c\in 2^{\mathcal{J}}} Pr\left(\mathcal{C} = c \mid \delta, M\right) Pr\left(j = \arg\max_{k\in c} U_k \mid \delta, p, \mathcal{C} = c\right)$$

$$= \sum_{c \in 2^{\mathcal{J}}} \lambda_c \left(\phi^1(\delta_1) + M_1, ..., \phi^J(\delta_J) + M_{3J}\right) g_{j,c} \left(u^k(\delta_k) - p_k, k \in c\right)$$

$$=: \Lambda^j \left(u^1(\delta_1) - p_1, ..., u^J(\delta_J) - p_J, \phi^1(\delta_1) + M_1, ..., \phi^J(\delta_J) + M_{3J}\right). \tag{3.18}$$

The second equality in (3.18) comes from a mixture representation of the demand function when the consideration set could be a subset of $\mathcal{J}$. The third equality in (3.18) comes from the fact that the conditional probability $Pr\left(\mathcal{C} = c \mid \delta, M\right)$ can be represented as a c-specific function of the systematic components of attention functions and that the choice probability conditional on the consideration set c, $Pr\Big(j = \arg\max_{k \in c} U_k | \delta, p, \mathcal{C} = c\Big)$ can be written as a (j,c)-specific function of the systematic utilities of the products in the consideration set c. The last equality in (3.18) is just a simplification of notation since one doesn't need to investigate the choice probability under each possible consideration set for identification.

To see how derivatives of utility and attention functions are identified, let $\Lambda^j_{(l)}$ denote the derivative of Λ^j with respect to its l-th argument. By the chain rule, taking derivative with respect to δ_1, p_1, and M_1 on both sides of (3.18) gives:

$$\frac{\partial \sigma_j(\delta, p, M)}{\partial \delta_1} = \Lambda^j_{(1)} \frac{\partial u^1(\delta_1)}{\partial \delta_1} + \Lambda^j_{(J+1)} \frac{\partial \phi^1(\delta_1)}{\partial \delta_1}$$
$$\frac{\partial \sigma_j(\delta, p, M)}{\partial p_1} = -\Lambda^j_{(1)}$$
$$\frac{\partial \sigma_j(\delta, p, M)}{\partial M_1} = \Lambda^j_{(J+1)}.$$

Eliminating $\Lambda^j_{(1)}$ and $\Lambda^j_{(J+1)}$ yields

$$\frac{\partial \sigma_j(\delta, p, M)}{\partial \delta_1} = -\frac{\partial \sigma_j(\delta, p, M)}{\partial p_1} \frac{\partial u^1(\delta_1)}{\partial \delta_1} + \frac{\partial \sigma_j(\delta, p, M)}{\partial M_1} \frac{\partial \phi^1(\delta_1)}{\partial \delta_1}. \tag{3.19}$$

Evaluating (3.19) at $\xi_j = \sigma_j^{-1}(s, p, M) + x_j$, for $j \in \{1, ..., J\}$ gives:

$$\frac{\partial \sigma_j(\sigma^{-1}(s, p, M), p, M)}{\partial \delta_1} = -\frac{\partial \sigma_j\left(\sigma^{-1}(s, p, M), p, M\right)}{\partial p_1} \frac{\partial u^1(\sigma_1^{-1}(s, p, M))}{\partial \delta_1}$$

$$+\frac{\partial \sigma_j\left(\sigma^{-1}(s,p,M),p,M\right)}{\partial M_1}\frac{\partial \phi^1(\sigma_1^{-1}(s,p,M))}{\partial \delta_1} \tag{3.20}$$

which holds for all $j \in \{1, ..., J\}$. Out of the J equations in (3.20), as long as the following rank condition for at least 2 of them are satisfied. one can identify the derivatives $\frac{\partial u^1(\sigma_1^{-1}(s,p,M))}{\partial \delta_1}$ and $\frac{\partial \phi^1(\sigma_1^{-1}(s,p,M))}{\partial \delta_1}$. The same is true for all $\frac{\partial u^j(\sigma_j^{-1}(s,p,M))}{\partial \delta_j}$ and $\frac{\partial \phi^j(\sigma_j^{-1}(s,p,M))}{\partial \delta_j}$, where $j = 2, ..., J$. Formally,

Assumption 3.14. *For each $j \in \{1, ...J\}$, there exist $k_j, l_j \in \{1, ..., J\}$ such that on the set $\bar{M}$, the matrix*

$$\begin{pmatrix} -\frac{\partial \sigma_{k_j}\left(\sigma^{-1}(s,p,M),p,M\right)}{\partial p_j} & \frac{\partial \sigma_{k_j}\left(\sigma^{-1}(s,p,M),p,M\right)}{\partial M_j} \\ -\frac{\partial \sigma_{l_j}\left(\sigma^{-1}(s,p,M),p,M\right)}{\partial p_j} & \frac{\partial \sigma_{l_j}\left(\sigma^{-1}(s,p,M),p,M\right)}{\partial M_j} \end{pmatrix}$$

has full rank.

Theorem 3.2. *Suppose for each $j = 1, ...J$, the derivatives of σ_j are identified. Under Assumption 3.1, 3.2, and 3.14, for all $j = 1, ...J$, $\frac{\partial u^j(\sigma_j^{-1}(s,p,M))}{\partial \delta_j}$ and $\frac{\partial \phi^j(\sigma_j^{-1}(s,p,M))}{\partial \delta_j}$ are identified.*

Proof. The conclusion follows from the above argument in this section. □

After identifying the derivatives, one can impose location normalizations to get the exact values of the functions.

3.4.3 Identifying the Joint Distribution of Unobservables

In this part, I analyze how to recover the distribution of ξ, and the joint distribution of ϵ and η.

The identification of the distribution of ξ is directly given by the transformation of variables equation:

$$f_{Y|X=x}(y) = f_\xi\left(r(y)+x\right)\left|\frac{\partial r(y)}{\partial y}\right|.$$

Next I analyze the joint distribution of ϵ and η. Consider the market share of the outside option $j = 0$:

$$\begin{aligned} s_0 &= \sigma_0(\delta, p, M) \\ &= Pr\left(\cap_{j=1}^{J}\left\{\left\{u^j(\delta_j) - p_j + \epsilon_{ij} \leqslant 0\right\} \cup \left\{\phi^j(\delta_j) + M_j + \eta_{ij} \leqslant 0\right\}\right\} \mid x, p, M, \xi\right) \\ &= \Lambda^0(-u^1(\delta_1) + p_1, ..., -u^J(\delta_J) + p_J, -\phi^1(\delta_1) - M_1, ..., -\phi^J(\delta_J) - M_J). \end{aligned}$$

Proposition 3.2. *Suppose for each j=1,...,J, σ_j, u^j and ϕ^j are identified. Then F, the joint distribution of $(\epsilon_{i1}, ..., \epsilon_{iJ}, \eta_{i1}, ..., \eta_{iJ})$ is identified.*

Proof. I prove the identification of F for $J = 2$. The argument carries over for all finite integers $J > 2$. Since the demand functions $\sigma_j(\delta, p, M)$ is identified for $j = 1, ..., J$, then the demand function of the outside option $j = 0$ can be recovered from $\sigma_0(\delta, p, M) = 1 - \sum_{j=1}^{J} \sigma_j(\delta, p, M)$. Once u^j, ϕ^j are identified, $\Lambda^0()$ is also identified.

Let $\bar{\epsilon}_j = -u^j(\delta_j) + p_j, \quad j = 1, 2$ and $\bar{\eta}_j = -\phi^j(\delta_j) - M_j, \quad j = 1, 2$. Then

$$\begin{aligned} &\Lambda^0\left(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2\right) \\ &= P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) + P(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) \\ &+ P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) + P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) \\ &+ P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 > \bar{\eta}_2) + P(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) \\ &+ P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) + P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1, \eta_2 > \bar{\eta}_2) \\ &+ P(\epsilon_1 > \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 > \bar{\eta}_2) \\ &= P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) + P(\epsilon_1 > \bar{\epsilon}_1, \eta_1 \leqslant \bar{\eta}_1, \eta_2 \leqslant \bar{\eta}_2) \\ &+ P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 > \bar{\epsilon}_2, \eta_2 \leqslant \bar{\eta}_2) + P(\epsilon_1 \leqslant \bar{\epsilon}_1, \epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 > \bar{\eta}_1) + P(\epsilon_2 \leqslant \bar{\epsilon}_2, \eta_1 \leqslant \bar{\eta}_1, \eta_2 > \bar{\eta}_2) \\ &= F(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2) + F_{\eta_1,\eta_2}(\bar{\eta}_1, \bar{\eta}_2) - F_{\epsilon_1\eta_1\eta_2}(\bar{\epsilon}_1, \bar{\eta}_1, \bar{\eta}_2) + F_{\epsilon_1\eta_2}(\bar{\epsilon}_1, \bar{\eta}_2) - F_{\epsilon_1\epsilon_2\eta_2}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_2) \\ &+ F_{\epsilon_1\epsilon_2}(\bar{\epsilon}_1, \bar{\epsilon}_2) - F_{\epsilon_1\epsilon_2\eta_1}(\bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\eta}_1) + F_{\epsilon_2\eta_1}(\bar{\epsilon}_2, \bar{\eta}_1) - F_{\epsilon_2,\eta_1,\eta_2}(\bar{\epsilon}_2, \bar{\eta}_1, \bar{\eta}_2) \end{aligned}$$

Assuming $-u^j(\delta_j)+p_j$ and $-\phi^j(\delta_j)-M_j$, $j=1,2$ have full support, then

$$
\begin{aligned}
F_{\epsilon_1\epsilon_2}(\bar{\epsilon}_1,\bar{\epsilon}_2) &= \Lambda^0(\bar{\epsilon}_1,\bar{\epsilon}_2,-\infty,-\infty)\\
F_{\eta_1\eta_2}(\bar{\eta}_1,\bar{\eta}_2) &= \Lambda^0(-\infty,-\infty,\bar{\eta}_1,\bar{\eta}_2)\\
F_{\epsilon_1\eta_2}(\bar{\epsilon}_1,\bar{\eta}_2) &= \Lambda^0(\bar{\epsilon}_1,-\infty,-\infty,\bar{\eta}_2)\\
F_{\epsilon_2\eta_1}(\bar{\epsilon}_2,\bar{\eta}_1) &= \Lambda^0(-\infty,\bar{\epsilon}_2,\bar{\eta}_1,-\infty)\\
F_{\epsilon_1\epsilon_2\eta_1}(\bar{\epsilon}_1,\bar{\epsilon}_2,\bar{\eta}_1) &= \Lambda^0(\bar{\epsilon}_1,\bar{\epsilon}_2,-\infty,-\infty)+\Lambda^0(-\infty,\bar{\epsilon}_2,\bar{\eta}_1,-\infty)-\Lambda^0(\bar{\epsilon}_1,\bar{\epsilon}_2,\bar{\eta}_1,-\infty)\\
F_{\epsilon_1\epsilon_2\eta_2}(\bar{\epsilon}_1,\bar{\epsilon}_2,\bar{\eta}_2) &= \Lambda^0(\bar{\epsilon}_1,-\infty,-\infty,\bar{\eta}_2)+\Lambda^0(\bar{\epsilon}_1,\bar{\epsilon}_2,-\infty,-\infty)-\Lambda^0(\bar{\epsilon}_1,\bar{\epsilon}_2,-\infty,\bar{\eta}_2)\\
F_{\epsilon_1\eta_1\eta_2}(\bar{\epsilon}_1,\bar{\eta}_1,\bar{\eta}_2) &= \Lambda^0(-\infty,-\infty,\bar{\eta}_1,\bar{\eta}_2)+\Lambda^0(\bar{\epsilon}_1,-\infty,-\infty,\bar{\eta}_2)-\Lambda^0(\bar{\epsilon}_1,-\infty,\bar{\eta}_1,\bar{\eta}_2)\\
F_{\epsilon_2\eta_1\eta_2}(\bar{\epsilon}_2,\bar{\eta}_1,\bar{\eta}_2) &= \Lambda^0(-\infty,-\infty,\bar{\eta}_1,\bar{\eta}_2)+\Lambda^0(-\infty,\bar{\epsilon}_2,\bar{\eta}_1,-\infty)-\Lambda^0(-\infty,\bar{\epsilon}_2,\bar{\eta}_1,\bar{\eta}_2).
\end{aligned}
$$

Therefore, $F(\bar{\epsilon}_1,\bar{\epsilon}_2,\bar{\eta}_1,\bar{\eta}_2)$ is also identified. □

3.4.4 Identifying the Supply-side Features

Section 3.4.1.1 shows the identification of the inverse function r. Note that the $J+1$-th to $3J$-th components of r are actually $\pi_1^{-1},...,\pi_J^{-1}$ and $\rho_1^{-1},...,\rho_J^{-1}$. Given the identification of π^{-1} and ρ^{-1}, in this section, I discuss how to identify the marginal cost functions c_j and ca_j for $j\in\{1,...,J\}$.

By Assumption 3.6, the marginal price function c_j is invertible in the production cost index κ_j, and ca_j is invertible in the marketing cost index λ_j, so that the following invertibility results hold

$$
\begin{aligned}
\kappa_{jt} &= c_j^{-1}\left(s_{jt},\varphi_j(\delta_t,s_t,p_t,M_t)\right)=\pi_j^{-1}(\delta_t,s_t,p_t,M_t)\\
\lambda_{jt} &= ca_j^{-1}\left(s_{jt},\gamma_j(\delta_t,s_t,p_t,M_t)\right)=\rho_j^{-1}(\delta_t,s_t,p_t,M_t).
\end{aligned}
$$

Since π_j^{-1} and ρ_j^{-1} are identified, to identify c_j^{-1} and ca_j^{-1}, one needs to specify a model of oligopoly competition so that the functional forms of φ_j and γ_j are known. Berry and

Haile (2014) discussed a testable condition based on (3.6) and (3.7) to discriminate between alternative models of oligopoly. After specifying φ_j and γ_j, the inverse cost functions c_j^{-1} and ca_j^{-1} are identified on the support of their arguments. Then the marginal cost functions c_j and ca_j are also identified on the support of their arguments.

3.5 Nonparametric Estimation

In this section, I propose nonparametric estimators for the derivatives of the demand function σ_j, the utility function u^j, and the attention function ϕ^j, $j = 1, ..., J$ based on their constructive identification conditions. Then I show the asymptotic properties of the estimators.

3.5.1 The Estimators

First, I consider the estimation of the derivatives of the inverse function r. They can be estimated by the average derivatives estimator proposed by Matzkin (2015). The estimator employs the fact that the identification equation (3.13) holds for all values of $\left(g_{x_1}, ..., g_{x_{3J}}\right)$ over the set $\bar{M}$. For convenience, rewrite the $3J$ equations in (3.13) as follows:

$$\begin{aligned} g_{y_1} &= g_{x_1} r^1_{y_1} + g_{x_2} r^2_{y_1} + ... + g_{x_{3J}} r^{3J}_{y_1} + d_{y_1} \\ g_{y_2} &= g_{x_1} r^1_{y_2} + g_{x_2} r^2_{y_2} + ... + g_{x_{3J}} r^{3J}_{y_2} + d_{y_2} \\ &\vdots \\ g_{y_{3J}} &= g_{x_1} r^1_{y_{3J}} + g_{x_2} r^2_{y_{3J}} + ... + g_{x_{3J}} r^{3J}_{y_{3J}} + d_{y_{3J}}, \end{aligned} \tag{3.21}$$

where $d_{y_j} := \frac{\partial \log |r_y|}{\partial y_j}$. Then (r_y, d_y) for $j = \{1, ..., 3J\}$ can be characterized as the unique solution to the minimization of an integrated square distance between the left-hand side and the right-hand side of (3.21). Before formally introducing the estimator, I introduce some additional notations. Let $\mu(y, t_1, ..., t_{3J})$ be a strictly positive and continuous weighting function such that $\int_{\bar{M}} \mu(y,t)dt = 1$. Denote the average of g_{y_j} and g_{x_j} over $\bar{M}$ by

$$\int_{\bar{M}} g_{y_j} := \int_{\bar{M}} g_{y_j}(y,t)\mu(y,t)dt$$
$$\int_{\bar{M}} g_{x_j} := \int_{\bar{M}} g_{x_j}(y,t)\mu(y,t)dt.$$

Denote the average centered cross products between g_{y_j} and g_{x_k}, and between g_{x_j} and g_{x_k}, by

$$T_{y_j,x_k} := \int_{\bar{M}} \left(g_{y_j}(y,t) - \int_{\bar{M}} g_{y_j}\right)\left(g_{x_k}(y,t) - \int_{\bar{M}} g_{x_k}\right) \times \mu(y,t)dt$$
$$T_{x_j,x_k} := \int_{\bar{M}} \left(g_{x_j}(y,t) - \int_{\bar{M}} g_{x_j}\right)\left(g_{x_k}(y,t) - \int_{\bar{M}} g_{x_k}\right) \times \mu(y,t)dt.$$

Then define the matrices of average centered cross products by:

$$T_{YX} := \begin{pmatrix} T_{y_1,x_1} & \cdots & T_{y_{3J},x_1} \\ \vdots & & \vdots \\ T_{y_1,x_{3J}} & \cdots & T_{y_{3J},x_{3J}} \end{pmatrix} \quad \text{and} \quad T_{XX} := \begin{pmatrix} T_{x_1,x_1} & \cdots & T_{x_{3J},x_1} \\ \vdots & & \vdots \\ T_{x_1,x_{3J}} & \cdots & T_{x_{3J},x_{3J}} \end{pmatrix}.$$

The following theorem states the conditions under which (r_y, d_y) can be expressed in terms of T_{YX} and T_{XX}.

Theorem 3.3 (Matzkin (2015) Theorem 2.3)**.** *Under Assumptions 3.7-3.10, if the nonnegative and continuous function $\mu(y, t_1, ..., t_{3J})$ is strictly positive at least at one set of points $(x^{(1)}, y), ..., (x^{(3J+1)}, y)$ satisfying Assumption 3.10(ii), then (r_y, d_y) is the unique minimizer of*

$$S\left(\widetilde{r_y}, \widetilde{d_y}\right) = \int_{\bar{M}} \left[\sum_{j=1}^{3J} \left(g_{y_j} - \widetilde{r_{y_j}^1} g_{x_1} - \widetilde{r_{y_j}^2} g_{x_2} - ... - \widetilde{r_{y_j}^{3J}} g_{x_{3J}} - \widetilde{d_{y_j}}\right)^2\right] \times \mu(y, t_1, ..., t_{3J}) d(t_1, ..., t_{3J}),$$

and r_y is given by

$$r_y = T_{XX}^{-1} T_{YX}.$$

Proof. The functional forms of the inverse demand and supply system satisfy Assumption 2.4' in Matzkin (2015). The rest of the proof is the same as the proof of Theorem 2.3 in Matzkin (2015). □

To obtain the estimator of r, in T_{YX} and T_{XX}, I replace the conditional density $f_{Y|X=x}(y)$ with its Kernel estimator $\hat{f}_{Y|X=x}(y)$. Denote the estimators of the matrices by $\widehat{T_{XX}}$ and $\widehat{T_{YX}}$ respectively. The estimator of r_y can be obtained as $\widehat{r_y} = \widehat{T_{XX}}^{-1}\widehat{T_{YX}}$.

Next I look at the estimator of $\frac{\partial u^j}{\partial \delta_j}$ and $\frac{\partial \phi^j}{\partial \delta_j}$. From (3.17), on $\bar{M}$ where y is fixed, by solving the linear system of equations, I can write the vector of derivatives of σ_j (denoted as $\partial\sigma_j$) as a known function H_j of vector rr:

$$\partial\sigma_j := \left(\frac{\partial\sigma_j}{\partial\delta_1}, ..., \frac{\partial\sigma_j}{\partial\delta_J}, \frac{\partial\sigma_j}{\partial y_{J+1}}, ..., \frac{\partial\sigma_j}{\partial y_{3J}}\right)' = H_j(rr(y)).$$

Moreover, picking the equations corresponding to the product k_j and l_j, $k_j, l_j \in \{1, ..., J\}$ in (3.20), on $\bar{M}$ where y is fixed, I can write the vector of derivatives of u^j and ϕ^j as a known function L_j of vector $\partial\sigma_{k_j}$ and $\partial\sigma_{l_j}$:

$$\left(\frac{\partial u^j}{\partial\delta_j}, \frac{\partial\phi^j}{\partial\delta_j}\right) = L_j(\partial\sigma_j, \partial\sigma_k) = L_j(H_{k_j}(rr(y)), H_{l_j}(rr(y))).$$

Denote the known composite function $L_j(H_{k_j}(rr(y)), H_{l_j}(rr(y)))$ as $LH_j(rr(y))$, so that

$$\left(\frac{\partial u^j}{\partial\delta_j}, \frac{\partial\phi^j}{\partial\delta_j}\right) = LH_j(rr(y)). \tag{3.22}$$

The exact forms of functions H_j, L_j, and LH_j are shown in the appendix.

Then for all $j \in \{1, ..., J\}$, the estimator for $\frac{\partial u^j}{\partial \delta_j}$ and $\frac{\partial \phi^j}{\partial \delta_j}$ can be obtained by replacing rr in (3.22) with its estimator, that is:

$$\left(\widehat{\frac{\partial u^j}{\partial\delta_j}}, \widehat{\frac{\partial\phi^j}{\partial\delta_j}}\right) = LH_j(\widehat{rr}(y)).$$

3.5.2 Asymptotic Properties of the Estimators

In this part, I discuss the asymptotic properties of the estimator. Let $\{Y^i, X^i\}_{i=1}^N$ denote N i.i.d. observations from $f_{Y,X}$. The Kernel estimator of conditional density $\hat{f}_{Y|X=x}(y)$ is

$$\hat{f}_{Y|X=x}(y) = \frac{\sum_{i=1}^N K\left(\frac{Y^i - y}{\sigma_N}, \frac{X^i - x}{\sigma_N}\right)}{\sigma_N^{3J} \sum_{i=1}^N K\left(\frac{X^i - x}{\sigma_N}\right)},$$

where K is a kernel function and σ_N is a bandwidth. I assume that, for any $t \in \bar{M}^t$, $t = x \in \mathbf{R}^{3J}$. Thus, $\bar{M} = \{y\} \times \bar{M}^t$. Let $\bar{M}^y$ be a convex compact set that contains y in its interior. Let $\bar{M}^x$ be a convex compact set whose interior strictly includes $\bar{M}^t$. I impose the following assumptions.

Assumption 3.15 (Matzkin (2015) Assumption 2.9)**.** *The kernel function K is of order s. equals zero outside a compact set, integrates to 1, is differentiable of order Δ, and its derivatives of order Δ are Lipschitz, where $\Delta \geqslant 2$.*

Assumption 3.16 (Matzkin (2015) Assumption 2.7)**.** *The density $f_{Y,X}$ generated by f_ϵ, f_X and r is bounded and continuously differentiable of order $d \geqslant s+2$ where s is the order of the kernel function. Moreover, there exists $\delta > 0$ such that for all $(y,x) \in \bar{M}^y \times \bar{M}^x$, $f_X(x) > \delta$ and $f_{Y,X}(y,x) > \delta$.*

Assumption 3.17 (Matzkin (2015) Assumption 2.8)**.** *The set $\bar{M}^t$ is compact. The function $\mu(y, \cdot)$ is bounded and continuously differentiable. It has strictly positive values at all x belonging to the interior of $\bar{M}^t$, values and derivatives equal to zero on the boundary and the complement of $\bar{M}^t$. The set $\bar{M}$ contains at least one set of points $(y, x^{(1)}), ..., (y, x^{(3J+1)})$ satisfying Assumption 3.10(ii) and such that μ is strictly positive at each of the points.*

Assumption 3.18 (Matzkin (2015) Assumption 2.10)**.** *The sequence of bandwidths, h_N, is such that $h_N \to 0$, $N h_N^{3J+2} \to \infty$, $\sqrt{N} h_N^{(3J/2)+1+s} \to 0$, $\frac{N h_N^{6J+2}}{\ln N} \to \infty$, and $\sqrt{N} h_N^{(3J/2)+1} \times \left[\sqrt{\ln N/(N h_N^{6J+2})} + h_N^s\right]^2 \to 0$.*

Under the assumptions above, the asymptotic properties follow from Matzkin (2015)'s Theorem 2.4. Before formally stating the theorem, I define the following additional notations. Let rr be the vector formed by stacking the columns of matrix r_y. Let TT_{YX} denote the vector formed by stacking the columns of matrix T_{YX} and let $TT_{XX} := I_{3J}$.

$$\begin{aligned} rr &:= vec(r_y) = \left(r^1_{y_1}, ..., r^{3J}_{y_1}; r^1_{y_2}, ..., r^{3J}_{y_2}; ...; r^1_{y_{3J}}, ..., r^{3J}_{y_{3J}}\right)' \\ TT_{YX} &:= vec(T_{YX}) = \left(T_{y_1,x_1}, ..., T_{y_1,x_{3J}}; T_{y_2,x_1}, ..., T_{y_2,x_{3J}}; ...; T_{y_{3J},x_1}, ..., T_{y_{3J},x_{3J}}\right)' \\ TT_{XX} &:= I_{3J} \otimes T_{XX}. \end{aligned}$$

For each $s \in \{1, ..., 3J\}$, denote

$$\Delta\partial_{x_s} \log f_{Y|X=x}(y) := \frac{\partial \log f_{Y|X=x}(y)}{\partial x_s} - \int_{\bar{M}} \frac{\partial \log f_{Y|X=x}(y)}{\partial x_s} \mu(y,x)dx,$$

and for each $j, k \in \{1, ..., 3J\}$, denote

$$\widetilde{KK}_{y_j,y_k} := \int \left(\int \frac{\partial K(\tilde{y},\tilde{x})}{\partial y_j} d\tilde{x}\right) \left(\int \frac{\partial K(\tilde{y},\tilde{x})}{\partial y_k} d\tilde{x}\right) d\tilde{y}.$$

Denote by $\hat{V}_{T_{YX}}$ the matrix whose elements are

$$\left[\int_{\bar{M}} \left(\Delta\partial_{x_s} \log \hat{f}_{Y|X=x}(y)\right) \left(\Delta\partial_{x_l} \log \hat{f}_{Y|X=x}(y)\right) \left(\frac{\mu(y,x)^2}{\hat{f}_{Y,X(y,x)}}\right) dx\right] \widetilde{KK}_{y_j,y_k}.$$

The following theorem in Matzkin (2015) states the asymptotic property of the estimator $\widehat{rr}$.

Theorem 3.4 (Matzkin (2015) Theorem 2.4)**.** *Under Assumption 3.7-3.10, 3.15-3.18,*

$$\sqrt{N h_N^{3J+2}} \left(\widehat{rr} - rr\right) \xrightarrow{d} N\left(0, (TT_{XX})^{-1} V_{T_{YX}} (TT_{XX})^{-1}\right)$$

and $(\widehat{TT_{XX}})^{-1}\hat{V}_{T_{YX}}(\widehat{TT_{XX}})^{-1}$ *is a consistent estimator for* $(TT_{XX})^{-1}V_{T_{YX}}(TT_{XX})^{-1}$.

Proof of Proposition 3.4. The functional form of r satisfies Assumption 2.4' in Matzkin (2015). The rest of the proof is the same as the proof of Theorem 2.4 in Matzkin (2015). □

Next, I consider the asymptotic properties of the estimator of the derivatives of utility functions $\widehat{\frac{\partial u^j}{\partial \delta_j}}, j \in \{1, ..., J\}$ and the estimator of derivatives of attention functions $\widehat{\frac{\partial \phi^j}{\partial \delta_j}}, j \in \{1, ..., J\}$. Denote

$$\beta_j(rr(y)) := \frac{\partial u^j(r^j(y))}{\partial \delta_j}$$
$$\gamma_j(rr(y)) := \frac{\partial \phi^j(r^j(y))}{\partial \delta_j}.$$

The following corollary follows directly from Theorem 3.4.

Corollary 4. *Under Assumption 3.7-3.11, 3.13-3.18, on $\bar{M}$ where y is fixed, for each $j =$ 1,...,J*

$$\sqrt{Nh_N^{3J+2}}\left(\left(\beta_j(\hat{rr}(y)), \gamma_j(\hat{rr}(y))\right) - (\left(\beta_j(rr(y)), \gamma_j(rr(y))\right)\right) \xrightarrow{d} N\left(0, \frac{\partial LH_j}{\partial rr}' \Sigma_{rr} \frac{\partial LH_j}{\partial rr}\right)$$

where $\Sigma_{rr} = (TT_{XX})^{-1} V_{T_{YX}} (TT_{XX})^{-1}$.

Proof. It can be shown that function LH_j in (3.22) is continuously differentiable with respect to $rr(y)$. By delta method, the conclusion follows from Theorem 3.4. □

3.6 Conclusion

Assuming the availability of market-level data, this paper studies the nonparametric identification and estimation of a demand and supply system where firms affect consumers' consideration sets using costly marketing inputs. On the demand side, I characterize preferences and considerations nonparametrically, allowing rich heterogeneities and correlations between them. On the supply side, I characterize firms' optimal choices by a set of first-order conditions without specifying the form of the oligopoly model. The demand and supply sides form a simultaneous system of equations in the spirit of Berry and Haile (2014). I then show identification of the system using the method proposed by Matzkin (2015). Moreover, using the variations of exclusive regressors entering preferences and considerations respectively, I separately identify features of the utility functions and the attention functions. Based on the constructive identification results, I propose nonparametric estimators of the demand, utility, and attention functions and show their asymptotic properties.

3.7 Appendix

Proof of Proposition 3.1. By Assumption 3.1(iii), function $u^j(\delta_j, p_j, \epsilon_{ij})$ has an inverse with respect to the last argument. I can write $\tilde{u}^j(\delta_j, p_j, 0)$ as the value of ϵ_{ij} at which the utility of j equals the utility of the outside option, which is equal to 0. Similarly, I can write $\tilde{\phi}^j(\delta_j, M_j, 0)$ as the value of η_{ij} at which the attention of j equals 0, which equals the attention of the outside option,i.e.

$$u^j\left(\delta_j, p_j, \tilde{u}^j(\delta_j, p_j, 0)\right) = 0 \tag{3.23}$$

$$\phi^j\left(\delta_j, M_j, \tilde{\phi}^j(\delta_j, M_j, 0)\right) = 0. \tag{3.24}$$

Consider the market share of the outside option:

$$\begin{aligned}
s_0 &= \sigma_0\left(\delta_1, ..., \delta_J, p_1, ..., p_J, M_1, ..., M_J\right) \\
&= Pr\Big(u^j(\delta_j, p_j, \epsilon_{ij}) \leqslant 0 \text{ or } \phi^j(\delta_j, M_j, \eta_{ij}) \leqslant 0, \text{ for all } j = 1, ..., J \mid \delta, p, M\Big) \\
&= Pr\Big(\epsilon_{ij} \leqslant \tilde{u}^j(\delta_j, p_j, 0) \text{ or } \eta_{ij} \leqslant \tilde{\phi}^j(\delta_j, M_j, 0), \text{ for all } j = 1, ..., J \mid \delta, p, M\Big) \\
&= \Lambda^0\left(\tilde{u}^1(\delta_1, p_1, 0), ..., \tilde{u}^J(\delta_J, p_J, 0), \tilde{\phi}^1(\delta_1, M_1, 0), ..., \tilde{\phi}^J(\delta_J, M_J, 0)\right).
\end{aligned}$$

Let $\Lambda^0_{(l)}$ denote the derivative of Λ^0 with respect to its l-th argument. By the chain rule, for all $j = 1, ..., J$, one can derive the following equations:

$$\begin{aligned}
\frac{\partial\sigma_0(\delta, p, M)}{\partial\delta_j} &= \Lambda^0_{(j)}\frac{\partial\tilde{u}^j(\delta_j, p_j, 0)}{\partial\delta_j} + \Lambda^0_{(J+j)}\frac{\partial\tilde{\phi}^j(\delta_j, M_j, 0)}{\partial\delta_j} \\
\frac{\partial\sigma_0(\delta, p, M)}{\partial p_j} &= \Lambda^0_{(j)}\frac{\partial\tilde{u}^j(\delta_j, p_j, 0)}{\partial p_j} \\
\frac{\partial\sigma_0(\delta, p, M)}{\partial M_j} &= \Lambda^0_{(J+j)}\frac{\partial\tilde{\phi}^j(\delta_j, M_j, 0)}{\partial M_j}.
\end{aligned}$$

Solving for $\Lambda^0_{(j)}$ and $\Lambda^0_{(J+j)}$ from the equations for $\partial\sigma^0/\partial p_j$ and $\partial\sigma^0/\partial M_j$ and substituting them into the first equation, I get for all $j = 1, ..., J$

$$\frac{\partial\sigma_0(\delta,p,M)}{\partial\delta_j} = \frac{\partial\sigma_0(\delta,p,M)}{\partial p_j}\frac{\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial\delta_j}}{\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial p_j}} + \frac{\partial\sigma_0(\delta,p,M)}{\partial M_j}\frac{\frac{\partial\tilde{\phi}^j\delta_j,M_j,0)}{\partial\delta_j}}{\frac{\partial\tilde{\phi}^j\delta_j,M_j,0)}{\partial M_j}}. \tag{3.25}$$

Differentiating (3.23), I get

$$\begin{aligned}
\frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial\delta_j} + \frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial\epsilon_{ij}}\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial\delta_j} &= 0\\
\frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial p_j} + \frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial\epsilon_{ij}}\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial p_j} &= 0,
\end{aligned}$$

which implies

$$\frac{\frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial\delta_j}}{\frac{\partial u^j(\delta_j,p_j,\tilde{u}^j(\delta_j,p_j,0))}{\partial p_j}} = \frac{\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial\delta_j}}{\frac{\partial\tilde{u}^j(\delta_j,p_j,0)}{\partial p_j}}.$$

Differentiating (3.24) and by the same reasoning, I get similar results for ϕ^j

$$\frac{\frac{\partial\phi^j(\delta_j,M_j,\tilde{\phi}^j(\delta_j,M_j,0))}{\partial\delta_j}}{\frac{\partial\phi^j(\delta_j,M_j,\tilde{\phi}^j(\delta_j,M_j,0))}{\partial M_j}} = \frac{\frac{\partial\tilde{\phi}^j(\delta_j,M_j,0)}{\partial\delta_j}}{\frac{\partial\tilde{\phi}^j(\delta_j,M_j,0)}{\partial M_j}}.$$

For notation simplicity, denote the derivatives of u^j as $u^j_{\delta_j}$ and $u^j_{p_j}$, and derivatives of ϕ^j as $\phi^j_{\delta_j}$ and $\phi^j_{M_j}$, respectively. The above equation (3.25) is equivalent to

$$\frac{\partial\sigma_0(\delta,p,M)}{\partial\delta_j} = \frac{\partial\sigma_0(\delta,p,M)}{\partial p_j}\frac{u^j_{\delta_j}}{u^j_{p_j}} + \frac{\partial\sigma_0(\delta,p,M)}{\partial M_j}\frac{\phi^j_{\delta_j}}{\phi^j_{M_j}}.$$

Assumption 3.12 ensures that the coefficient matrix of the following system of equations has full rank:

$$\frac{\partial\sigma_0^{(1)}}{\partial\delta_j} = \frac{\partial\sigma_0^{(1)}}{\partial p_j}\frac{u^j_{\delta_j}}{u^j_{p_j}} + \frac{\partial\sigma_0^{(1)}}{\partial M_j}\frac{\phi^j_{\delta_j}}{\phi^j_{M_j}}$$

$$\frac{\partial \sigma_0^{(2)}}{\partial \delta_j} = \frac{\partial \sigma_0^{(2)}}{\partial p_j}\frac{u^j_{\delta_j}}{u^j_{p_j}} + \frac{\partial \sigma_0^{(2)}}{\partial M_j}\frac{\phi^j_{\delta_j}}{\phi^j_{M_j}}.$$

The conclusion follows from Cramer's rule. □

The Exact Form of Function LH_j in (3.22)

Take $J = 2$ for example. Results for more general values of J are of similar logic.

$$rr = \left(\frac{\partial r^1}{\partial y_1}, \frac{\partial r^2}{\partial y_1}; \frac{\partial r^1}{\partial y_2}, \frac{\partial r^2}{\partial y_2}; \frac{\partial r^1}{\partial y_{J+1}}, \frac{\partial r^2}{\partial y_{J+1}}; \frac{\partial r^1}{\partial y_{J+2}}, \frac{\partial r^2}{\partial y_{J+2}}; \frac{\partial r^1}{\partial y_{2J+1}}, \frac{\partial r^2}{\partial y_{2J+1}}; \frac{\partial r^1}{\partial y_{2J+2}}, \frac{\partial r^2}{\partial y_{2J+2}}\right)$$

Consider $j = 1$ and let $k_1 = 1$ and $l_1 = 2$. Equation (3.17) for $k_1 = 1$ and $l_1 = 2$ are respectively:

$$\begin{pmatrix} \frac{\partial r^1}{\partial y_1} & \frac{\partial r^2}{\partial y_1} & 0 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_2} & \frac{\partial r^2}{\partial y_2} & 0 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{J+1}} & \frac{\partial r^2}{\partial y_{J+1}} & 1 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{J+2}} & \frac{\partial r^2}{\partial y_{J+2}} & 0 & 1 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{2J+1}} & \frac{\partial r^2}{\partial y_{2J+1}} & 0 & 0 & 1 & 0 \\ \frac{\partial r^1}{\partial y_{2J+2}} & \frac{\partial r^2}{\partial y_{2J+2}} & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial \sigma_1}{\partial \delta_1} \\ \frac{\partial \sigma_1}{\partial \delta_2} \\ \frac{\partial \sigma_1}{\partial y_{J+1}} \\ \frac{\partial \sigma_1}{\partial y_{J+2}} \\ \frac{\partial \sigma_1}{\partial y_{2J+1}} \\ \frac{\partial \sigma_1}{\partial y_{2J+2}} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$\begin{pmatrix} \frac{\partial r^1}{\partial y_1} & \frac{\partial r^2}{\partial y_1} & 0 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_2} & \frac{\partial r^2}{\partial y_2} & 0 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{J+1}} & \frac{\partial r^2}{\partial y_{J+1}} & 1 & 0 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{J+2}} & \frac{\partial r^2}{\partial y_{J+2}} & 0 & 1 & 0 & 0 \\ \frac{\partial r^1}{\partial y_{2J+1}} & \frac{\partial r^2}{\partial y_{2J+1}} & 0 & 0 & 1 & 0 \\ \frac{\partial r^1}{\partial y_{2J+2}} & \frac{\partial r^2}{\partial y_{2J+2}} & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial \sigma_2}{\partial \delta_1} \\ \frac{\partial \sigma_2}{\partial \delta_2} \\ \frac{\partial \sigma_2}{\partial y_{J+1}} \\ \frac{\partial \sigma_2}{\partial y_{J+2}} \\ \frac{\partial \sigma_2}{\partial y_{2J+1}} \\ \frac{\partial \sigma_2}{\partial y_{2J+2}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

By Cramer's Rule,

$$\begin{aligned}
\frac{\partial \sigma_1}{\partial \delta_1} &= \frac{\frac{\partial r^2}{\partial y_2}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}} \\
\frac{\partial \sigma_1}{\partial y_{J+1}} &= \frac{\frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_{J+1}} - \frac{\partial r^1}{\partial y_{J+1}}\frac{\partial r^2}{\partial y_2}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}} \\
\frac{\partial \sigma_1}{\partial y_{2J+1}} &= \frac{\frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_{2J+1}} - \frac{\partial r^1}{\partial y_{2J+1}}\frac{\partial r^2}{\partial y_2}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}} \\
\frac{\partial \sigma_2}{\partial \delta_1} &= \frac{-\frac{\partial r^2}{\partial y_1}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}} \\
\frac{\partial \sigma_2}{\partial y_{J+1}} &= \frac{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_{J+1}} - \frac{\partial r^1}{\partial y_{J+1}}\frac{\partial r^2}{\partial y_1}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}} \\
\frac{\partial \sigma_2}{\partial y_{2J+1}} &= \frac{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_{2J+1}} - \frac{\partial r^1}{\partial y_{2J+1}}\frac{\partial r^2}{\partial y_1}}{\frac{\partial r^1}{\partial y_1}\frac{\partial r^2}{\partial y_2} - \frac{\partial r^1}{\partial y_2}\frac{\partial r^2}{\partial y_1}}.
\end{aligned} \tag{3.26}$$

Equation (3.20) for $k_1 = 1$ and $l_1 = 2$ can be written as

$$\begin{pmatrix} \frac{\partial \sigma_1}{\partial \delta_1} \\ \frac{\partial \sigma_2}{\partial \delta_1} \end{pmatrix} = -\begin{pmatrix} \frac{\partial \sigma_1}{\partial y_{J+1}} & \frac{\partial \sigma_1}{\partial y_{2J+1}} \\ \frac{\partial \sigma_2}{\partial y_{J+1}} & \frac{\partial \sigma_2}{\partial y_{2J+1}} \end{pmatrix} \begin{pmatrix} \frac{\partial u^1}{\partial \delta_1} \\ \frac{\partial \phi^1}{\partial \delta_1} \end{pmatrix}.$$

Again by Cramer's Rule

$$\begin{aligned}
\frac{\partial u^1}{\partial \delta_1} &= -\frac{\frac{\partial \sigma_1}{\partial \delta_1}\frac{\partial \sigma_2}{\partial y_{2J+1}} - \frac{\partial \sigma_2}{\partial \delta_1}\frac{\partial \sigma_1}{\partial y_{2J+1}}}{\frac{\partial \sigma_1}{\partial y_{J+1}}\frac{\partial \sigma_2}{\partial y_{2J+1}} - \frac{\partial \sigma_2}{\partial y_{J+1}}\frac{\partial \sigma_1}{\partial y_{2J+1}}} \\
\frac{\partial \phi^1}{\partial \delta_1} &= -\frac{\frac{\partial \sigma_1}{\partial y_{J+1}}\frac{\partial \sigma_2}{\partial \delta_1} - \frac{\partial \sigma_2}{\partial y_{J+1}}\frac{\partial \sigma_1}{\partial \delta_1}}{\frac{\partial \sigma_1}{\partial y_{J+1}}\frac{\partial \sigma_2}{\partial y_{2J+1}} - \frac{\partial \sigma_2}{\partial y_{J+1}}\frac{\partial \sigma_1}{\partial y_{2J+1}}}.
\end{aligned} \tag{3.27}$$

The exact functional form of LH_1 can be obtained by plugging (3.26) into (3.27).

Bibliography

Abaluck, J. and Adams-Prassl, A. (2021). What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics*, 136(3):1611–1663.

Altonji, J. G. and Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053–1102.

Andrews, D. W. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, 11(3):560–586.

Aradillas-Lopez, A. (2022). Inference in models with partially identified control functions.

Arellano and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Barseghyan, L., Coughlin, M., Molinari, F., and Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111.

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

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797.

Berry, S. T. and Haile, P. A. (2018). Identification of nonparametric simultaneous equations models with a residual index structure. *Econometrica*, 86(1):289–315.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

Brancaccio, G., Kalouptsidi, M., and Papageorgiou, T. (2020a). Geography, transportation, and endogenous trade costs. *Econometrica*, 88(2):657–691.

Brancaccio, G., Kalouptsidi, M., and Papageorgiou, T. (2020b). A guide to estimating matching functions in spatial models. *International Journal of Industrial Organization*, 70:102533.

Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, 128(7):2796–2836.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5):1405–1441.

Conlon, C. T. and Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Crawford, G. S., Griffith, R., and Iaria, A. (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, 222(1):4–43.

Dardanoni, V., Manzini, P., Mariotti, M., and Tyson, C. J. (2020). Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88(3):1269–1296.

Draganska, M. and Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, 48(4):653–669.

Eliaz, K. and Spiegler, R. (2011). Consideration sets and competitive marketing. *The Review of Economic Studies*, 78(1):235–262.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Honka, E. and Chintagunta, P. (2017). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science*, 36(1):21–42.

Honka, E., Hortaçsu, A., and Vitorino, M. A. (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics*, 48(3):611–646.

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

Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, 188(1):182–195.

Li, T. (2002). Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics*, 110(1):1–26.

Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5):1339–1375.

Matzkin, R. L. (2015). Estimation of nonparametric models with simultaneity. *Econometrica*, 83(1):1–66.

Matzkin, R. L. (2016). On independence conditions in nonseparable models: Observable and unobservable instruments. *Journal of Econometrics*, 191(2):302–311.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. *Journal of Econometrics*, 195(2):224–235.

Politis, D. N. and Romano, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis*, 68(1):1–25.

Reutskaja, E., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926.

Ridder, G. and Sheng, S. (2022). Two-step estimation of a strategic network formation model with clustering. *arXiv preprint arXiv:2001.03838*.

Schennach, S., White, H., and Chalak, K. (2012). Local indirect least squares and average marginal effects in nonseparable structural systems. *Journal of Econometrics*, 166(2):282–302.

Schennach, S. M. (2004a). Estimation of nonlinear models with measurement error. *Econometrica*, 72(1):33–75.

Schennach, S. M. (2004b). Nonparametric regression in the presence of measurement error. *Econometric Theory*, 20(6):1046–1093.

Shaikh, A. M. and Vytlacil, E. J. (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica*, 79(3):949–955.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Song, S., Schennach, S. M., and White, H. (2015). Estimating nonseparable models with mismeasured endogenous variables. *Quantitative Economics*, 6(3):749–794.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.