Title
Endogeneity and measurement error in nonparametric and semiparametric models

Author
Song, Suyong

Publication Date
2010

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Endogeneity and Measurement Error in Nonparametric and Semiparametric Models

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

in

Economics

by

Suyong Song

Committee in charge:

Professor Halbert L. White, Chair
Professor Gordon Dahl
Professor Anthony Gamst
Professor Dimitris Politis
Professor Susanne M. Schennach
Professor Yixiao Sun

2010

The dissertation of Suyong Song is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

DEDICATION

To my wife Jin Young.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

From the bottom of my heart, I would like to thank my advisor Halbert White, a great researcher and teacher, for his guidance, encouragement, and patience throughout my studies at University of California, San Diego. His enthusiasm and insights for academic life is both inspiring and contagious. I have been amazingly fortunate to have him as an advisor.

I am also grateful to other advisors. I learned from Yixiao Sun a great deal of economic intuition and ideas in nonparametric and semiparametric econometrics. Susanne Schennach at University of Chicago was very generous and insightful to discuss and to give constructive suggestions about nonlinear measurement error. I think it was impossible to finish my dissertation without their encouragement through the years. Gordon Dahl has offered many good comments on applied works to labor economics. I also would like to thank other dissertation committee members, Dimitris Politis and Anthony Gamst.

I have been very fortunate to have had many other great professors at UCSD. Among them, I would like to thank Graham Elliott, James Hamilton, Ivana Komunjer and Andres Santos.

My advisors during master program in Seoul, Chang-Jin Kim and Yoon-Jae Whang, also deserve my warm thanks. My interest in endogeneity and nonparametrics came from both econometricians, and they have offered much advice and support since my undergraduate studies. I am also indebt to my lifetime mentors, Songwon Cho, Junghee Park, and Sangsoo Yang for their endless support.

Chapter one and part of chapter three are coauthored by Halbert White and Susanne Schennach. I would like to thank them for their generous permission to use the paper for this dissertation. Specially, Susanne Schennach provided me her programming codes. Based on those codes, we could complete codes for all chapters. I am indebted to Richard Blundell at University of College, London and Gordon Dahl because they generously provided their data set, British FES and NLSY, respectively. I am also grateful to Xiaohong Chen at Yale University. It was impossible to start chapter two without her encouragement.

I gratefully acknowledge helpful discussion from an econometrics reading group, of which members are Tom Daula, Ben Fissel, Hiroaki Kaido, Min Seong Kim and Xun Lu. I also acknowledge deep friendships from my colleagues at UCSD. My thanks go to Ayelen Banegas, Michael Callen, Jess Diamond, Youjin Hahn, Chulyoung Kim, Robert MacKay, Ian McConnaha, Nidhi Mirani, Danielken Molina, Philip Neary, Augusto Nieto, Young Joon Park, Sarada, Lucas Siga, Hee-Seung Yang, and other graduate students.

My warmest thanks go to my family. I am grateful to my parents, brother and sisters for supporting and encouraging me during many years of my education. I also gratefully acknowledge the support from my parents-in-law and brother-in-law. I also thank my daughter, Serene, and second daughter-to-be, Serah, for their love. My final but the deepest thanks go to my wife, Jin Young, for her patience, support, encouragement, and love since 2004. I dedicate this dissertation to her.

VITA

2002	B. A. in Economics, Korea University, Seoul, Korea
2004	M. A. in Economics, Korea University, Seoul, Korea
2005	Economist, Korea Development Bank, Seoul, Korea
2010	Ph. D. in Economics, University of California, San Diego

FIELD OF STUDY

Major Field: Econometric Theory, Applied Econometrics

ABSTRACT OF THE DISSERTATION

Endogeneity and Measurement Error in Nonparametric and Semiparametric Models

by

Suyong Song

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor Halbert L. White, Chair

It has long been an area of interest to consider a consistent estimation of nonlinear models with measurement error or endogeneity in the explanatory variables. Contrast to linear parametric models, both topics in nonlinear models are difficult to correct for. As a result, many of studies have addressed only one of them in nonlinear models, although controlling for only one mostly fails to identify economically meaningful structural parameters. Thus, this dissertation presents solutions to simultaneously control for both endogeneity and measurement error in general nonlinear regression models.

Chapter one of this dissertation studies the identification and estimation of covariate-conditioned average marginal effects of endogenous regressors in nonsep-

arable models when the regressors are mismeasured. Endogeneity is controlled for by making use of covariates as conditioning instruments; this ensures independence between the endogenous causes and other unobservable drivers of the dependent variable. Moreover, distributions of the underlying true causes from their error-laden measurements are recovered. Specifically, it is shown that two error-laden measurements of the unobserved true causes are sufficient to identify objects of interest and to deliver consistent estimators.

Chapter two develops semiparametric estimation of models defined by conditional moment restrictions, where the unknown functions depend on endogenous variables which are contaminated by nonclassical measurement errors. A two-stage estimation procedure is proposed to recover the true conditional density of endogenous variables given conditioning variables masked by measurement errors, and to rectify the difficulty associated with endogeneity of the unknown functions.

Chapter three investigates empirical importance of endogeneity and measurement error in economic examples. The proposed methods in chapter one and two are applied to topics of interest, the impact of family income on children's achievement and the estimation of Engel curves, respectively. The first application finds that the effects of family income on both math and reading scores from the proposed estimator are positive and that the magnitudes of the income effects are substantially larger than previously recognized. From the second application, findings indicate that correcting for both endogeneity and measurement error obtains significantly different shapes of Engel curves, compared to the method which ignores measurement error on total expenditure.

Chapter 1

Identification and Estimation of Nonseparable Models with Measurement Errors

1.1 Introduction

In this paper, we examine the identification and estimation of covariate-conditioned average marginal effects of endogenous regressors in nonseparable structural systems when the regressors are mismeasured. We control for the endogeneity by making use of covariates as conditioning instruments; this ensures independence between the endogenous causes of interest and other unobservable drivers of the dependent variable. Moreover, we recover distributions of the underlying true causes from their error-laden measurements. Our approach relies on a useful property of the Fourier transform, namely, its ability to convert complicated integral equations that relate unobservables to observables into simple algebraic equations. Even though our structural relations are nonparametric and nonseparable, we show that we can identify and estimate objects of interest, specifically, covariate-conditioned average marginal effects and weighted averages of covariate-conditioned average marginal effects.

Researchers have previously imposed linearity or separability on systems of structural equations because of the resulting ease of interpretation and implementation. But realistic models of economic behavior need not exhibit these convenient features. When these simplifying assumptions fail, serious errors of inference may result. To overcome such difficulties, researchers have devoted increasing attention to relaxing some or all of these assumptions. For example, additively separable nonparametric models for endogenous regressors with observable instruments, possibly with a limited or qualitative dependent variable, have been intensively studied under various sets of assumptions. Examples are Newey, Powell and Vella (1999), Darolles, Florens and Renault (2003), Newey and Powell (2003), Blundell and Powell (2004), Hall and Horowitz (2005), Das (2005), Severini and Tripathi (2005), and Blundell and Powell (2007) and the references therein.

Other recent work has studied identification and estimation of models with nonseparable structural equations, e.g., Matzkin (2003), Chesher (2003, 2005), Altonji and Matzkin (2005), Chernozhukov and Hansen (2005), Imbens and Newey (2006), Imbens (2006), White and Chalak (2006), Chernozhukov, Imbens and

Newey (2007), Hoderlein (2007), Hoderlein and Mammen (2007), Chalak and White (2007a, b), Schennach, White and Chalak (2007) (SWC), and Hahn and Ridder (2007).

Here we use a conditional independence assumption to achieve structural identification, as considered, for example, by Altonji and Matzkin (2005), White and Chalak (2006), Chalak and White (2007a, b), Hoderlein (2007), and Hoderlein and Mammen (2007). Altonji and Matzkin (2005) propose methods for estimating nonseparable models with observable endogenous regressors and unobservable errors in cross-section and panel data. One of their objects of interest is a local average response. A similar structure is considered here for cross-section data. Nevertheless, our framework differs from that of Altonji and Matzkin (2005) in that in our setting, the endogenous cause of interest is unobservable. Instead, we suppose we have available two error-laden measurements of the true underlying variable.

SWC also study identification and estimation of average marginal effects in nonseparable structural systems. They consider estimating causal effects from a nonseparable data generating process using either an observed standard exogenous instrument or an unobserved exogenous instrument for which two error-laden measurements are available. We extend the approach of SWC to the case in which the instrument is no longer exogenous, but is instead a conditioning instrument. This ensures that the cause of interest is independent of other unobservable drivers of the dependent variable, conditional on the instrument. Here, this instrument is observable. Nevertheless, the endogenous cause of interest is unobservable; to handle this, we employ nonlinear errors-in-variables methods, employing a Fourier transform approach.

We first nonparametrically estimate quantities of a general form and construct objects of interest from these. This covers such objects as the average counterfactual response function, the covariate-conditioned average marginal effect, Altonji and Matzkin's (2005) "local average response", corresponding to the effect of treatment on the treated for continuous treatments (Florens, Heckman, Meghir, and Vytlacil, 2008), and the average treatment effect. We establish uniform convergence rates and asymptotic normality for estimators of covariate-conditioned average marginal effects, faster convergence rates for estimators of their weighted averages over instruments, and $\sqrt{n}$ consistency and asymptotic normality for estimators of their weighted averages over instruments and regressors.

In Section 1.2, we describe the data generating process for the triangular structural system studied here. We also study the identification of a specific object of interest, the covariate-conditioned average marginal effect. A nonparametric estimator for quantities of a general form used to construct this object is presented in Section 1.3, and asymptotic properties of the estimator are analyzed in Section 1.4. The practical usefulness of the proposed estimator is illustrated by Monte Carlo experiments in Section 1.5. Section 1.6 concludes. All technical proofs are included in the Mathematical Appendix.

1.2 Data Generation and Identification

1.2.1 The Data Generating Process

We first specify the data generating process (DGP) for the recursive structural system studied here. There is an inherent ordering of the variables in such systems: in the language of White and Chalak (2008), "predecessor" variables may determine "successor" variables, but not vice versa. For instance, when X causes Y, then Y cannot cause X. In such cases, we say that Y *succeeds* X, and we write $Y \Leftarrow X$ as a shorthand notation. (See also Chalak and White (2007a, b), and SWC.) Throughout, random variables are defined on a complete probability space $(\Omega, \mathcal{F}, P)$.

Assumption 2.1 *(i) Let (U, W, X, Y) be random variables such that $E(|Y|) < \infty$; (ii) (U, W, X, Y) is generated by a recursive structural system such that $Y \Leftarrow (U, X)$ and $X \Leftarrow (U, W)$ with Y generated by the structural equation*

$$Y = r(X, U_y),$$

where r is an unknown measurable scalar-valued function and $U_y \equiv \upsilon_y(U)$ is a *random vector of countable dimension l_y, with v_y a measurable function; and (iii) the realizations of Y and W are observed, whereas those of U, X, and U_y are not.*

For now, U, X, and W can be viewed as random vectors; we let Y be scalar. Although X and W have finite dimension, the dimensions of U and U_y may be countably infinite. The specified structural relations are directional causal links; thus, variations in X and U_y structurally determine Y, as in Strotz and Wold (1960) (see also White and Chalak, 2008, and Chalak and White, 2007a, b). We do not assume that r is linear or monotone in its arguments or separable between X and U_y.

A primary object of interest is the marginal effect of X on Y. As there is no restriction to the contrary, X and U_y are generally correlated, so that X is endogenous. In classical treatments, the effects of endogenous variables are identified with the aid of instrumental variables. These are "standard" or "proper" when they are (i) correlated with X and (ii) exogenous (i.e., uncorrelated with or independent of unobservables, corresponding to U_y here). Nevertheless, standard instrumental variables are absent here, as the covariates W are also generally endogenous. However, identification of certain average marginal effects is possible when X satisfies a particular conditional form of exogeneity. To state this, we follow Dawid (1979), and write $X \perp U_y \mid W$ to denote that X is independent of U_y given W.[1]

Assumption 2.2 $X \perp U_y \mid W$.

Assumption 2.2 is analogous to structure imposed by Altonji and Matzkin (2005), White and Chalak (2006), Chalak and White (2007a, b), Hoderlein (2007), and Hoderlein and Mammen (2007). Given its instrumental role in ensuring conditional exogeneity, we call W conditioning instruments, following White and Chalak (2006) and Chalak and White (2007a, b).

[1]Conditional independence implies a similar 'common support assumption' in Imbens and Newey (2006). We can see this from the following argument. $\text{supp}(U_y \mid X = x, W = w) \equiv \bigcap\{S \in \mathcal{F} : P[U_y \in S \mid X = x, W = w] = 1\} = \bigcap\{S \in \mathcal{F} : P[U_y \in S \mid X = x] = 1\} \equiv \text{supp}(U_y \mid W = w)$, where the second equality follows by $X \perp U_y \mid W$.

Figure 1.1 provides a convenient graphical depiction of a structure consistent with Assumptions 2.1 and 2.2. Here, arrows denote direct causal relationships. Dashed circles denote unobservables and complete circles denote observables. Here, because of the indicated causal relations, U_w, U_x, and U_y are dependent, which generally leads to dependence between X, W, and U_y.

In contrast to Altonji and Matzkin (2005) and the other references just given, we do not assume that X is observable. Instead, we suppose that we observe two error-contaminated measurements of X, permitting us to employ methods of Schennach (2004a, b). The following assumption expresses this formally.

Assumption 2.3 *Observables X_1 and X_2 are determined by the structural equations $X_1 = X + U_1$ and $X_2 = X + U_2$, where $U_1 \equiv \upsilon_1(U)$ and $U_2 \equiv \upsilon_2(U)$ for measurable functions υ_1 and υ_2.*

Figure 1.2 depicts structural relations consistent with Assumptions 2.1 - 2.3. A line without an arrow denotes dependence arising from a causal relation in either direction or the existence of an underlying common cause. Later, we will rule out correlation (more precisely, conditional mean dependence) between U_1 and U_2 but permit dependence otherwise. We will also impose further restrictions on the relations between the measurement errors and the other variables of the system.

1.2.2 Structural Identification

Before going further, it is important to understand how conditional exogeneity ensures the identification of effects of interest for the structures of Assumption 2.1, regardless of the observability of X. Given this, we can consider how to proceed when X is unobservable.

To study identification of the effects of interest, we start with a representation of the conditional expectation of the response given X and W,

$$\mu(X, W) \equiv E(Y \mid X, W). \tag{1.1}$$

The function μ exists whenever $E(|Y|) < \infty$, as ensured by Assumption 2.1(i), regardless of underlying structural relations. When the structural relations of

Assumption 2.1(ii) hold, we have the representation

$$\mu(x,w) = \int r(x,u_y)\ dF(u_y \mid x,w),$$

where $dF(u_y \mid x,w)$ denotes the conditional density of U_y given $X = x$ and $W = w$. This represents $\mu(X,W)$ as the average response given $(X,W) = (x,w)$. With no further restrictions, this is a purely stochastic object. It provides no information about the causal effect of X on Y.

When $X \Leftarrow (U,W)$, as assumed here, we can define a particular conditional expectation that has a clear counterfactual meaning, supporting causal interpretations. Specifically, the average counterfactual response at x given $W = w$ is

$$\rho(x \mid w) \equiv E(r(x,U_y) \mid W = w) = \int r(x,u_y)\ dF(u_y \mid w),$$

where $dF(u_y \mid w)$ denotes the conditional density of U_y given $W = w$. The function $\rho(x \mid w)$ is a conditional analog of the average structural function of Blundell and Powell (2004), and a stepping stone to the analysis of various causally informative quantities of interest. Let $D_x \equiv (\partial/\partial x)$. The covariate-conditioned average marginal effect of X on Y at x given $W = w$ is

$$\beta^*(x \mid w) \equiv D_x\rho(x \mid w) = D_x \int r(x,u_y)\ dF(u_y \mid w) = \int D_x r(x,u_y)\ dF(u_y \mid w),$$

provided the derivative and integral can be interchanged. This function is a weighted average of the unobservable marginal effect $D_x r(x,u_y)$ over unobserved causes, given observed covariates. As described in the next section, it can be used to construct various effect measures of interest; for instance, the average treatment effect, the effect of treatment on the treated (Florens, Heckman, Meghir, and Vytlacil, 2008), and the weighted average of the local average response (Altonji and Matzkin, 2005). When Assumption 2.2 holds, we have

$$\int r(x,u_y)\ dF(u_y \mid x,w) = \int r(x,u_y)\ dF(u_y \mid w),$$

as $X \perp U_y \mid W$ implies $dF(u_y \mid x,w) = dF(u_y \mid w)$. That is, $\mu(x,w) = \rho(x \mid w)$, so μ acquires causal meaning from ρ. We call this a "structural identification" result because it identifies an aspect of the causal structure, ρ, with μ, a standard stochastic object. When $\mu(x, w)$ is differentiable, let $\beta(x, w) \equiv D_x\mu(x, w)$. With μ structurally identified ($\mu = \rho$), we also have $\beta(x, w) = \beta^*(x \mid w)$, so that $\beta(x, w)$ is also structurally identified. (See White and Chalak (2008) for additional formal conditions ensuring these identifications.)

If X were observable, we could estimate the covariate-conditioned average marginal effect $\beta^*(x \mid w)$ by first estimating $\mu(x, w)$ using standard techniques. Differentiating this with respect to x then yields $\beta(x, w) = \beta^*(x \mid w)$. Here, however, X is not observable, so such a direct approach is not available. Instead, we estimate $\mu(x, w)$ and its derivatives using the Fourier transform approach exploited in simpler settings by Schennach (2004a, b).

1.2.3 Weighted Averages of Effects

In addition to $\beta^*(x \mid w)$, we are interested in weighted averages of $\beta^*(x \mid w)$, such as

$$\beta^*_m(x) \equiv \int \beta^*(x \mid w)m(w)dw, \tag{1.2}$$

$$\beta^*_{mf_W}(x) \equiv \int \beta^*(x \mid w)m(w)f_W(w)dw, \tag{1.3}$$

$$\beta^*_{mf_{W|X}}(x) \equiv \int \beta^*(x \mid w)m(w)f_{W|X}(w \mid x)dw, \tag{1.4}$$

$$\beta^*_{\tilde{m}} \equiv \int\int \beta^*(x \mid w)\tilde{m}(x, w)dwdx, \tag{1.5}$$

$$\beta^*_{\tilde{m}f_{W|X}} \equiv \int\int \beta^*(x \mid w)\tilde{m}(x, w)f_{W|X}(w \mid x)dwdx, \tag{1.6}$$

$$\beta^*_{\tilde{m}f_{W,X}} \equiv \int\int \beta^*(x \mid w)\tilde{m}(x, w)f_{W,X}(w, x)dwdx, \tag{1.7}$$

where $m(\cdot)$ and $\tilde{m}(\cdot, \cdot)$ are user-supplied weight functions, and where f_W, $f_{W|X}$, and $f_{W,X}$ are the marginal density of W, conditional density of W given X, and joint density of W and X, respectively. When $m(w) = 1$, for instance, $\beta^*_{mf_W}(x)$ is analogous to the derivative of the average structural function of Blundell and Powell (2004) and the average treatment effect of Florens, Heckman, Meghir, and Vytlacil (2008). When $m(w) = 1$, $\beta^*_{mf_{W|X}}(x)$ corresponds to the local average response of Altonji and Matzkin (2005) and the effect of treatment on the treated

(Florens, Heckman, Meghir, and Vytlacil, 2008). When $\tilde{m}(x,w) = m(w)$, $\beta^*_{\tilde{m}f_{W|X}}$ corresponds to the weighted average of the local average response (Altonji and Matzkin, 2005).

Under structural identification, we have $\beta^*_m(x) = \beta_m(x)$, $\beta^*_{mf_W}(x) = \beta_{mf_W}(x)$, $\beta^*_{mf_{W|X}}(x) = \beta_{mf_{W|X}}(x)$, $\beta^*_{\tilde{m}} = \beta_{\tilde{m}}$, $\beta^*_{\tilde{m}f_{W|X}} = \beta_{\tilde{m}f_{W|X}}$, and $\beta^*_{\tilde{m}f_{W,X}} = \beta_{\tilde{m}f_{W,X}}$, where all quantities on the right-hand side are analogs of those on the left, obtained by replacing β^* with β in the defining integrals above. We thus are interested in estimating structurally identified $\beta(x,w)$, $\beta_m(x)$, $\beta_{mf_W}(x)$, $\beta_{mf_{W|X}}(x)$, $\beta_{\tilde{m}}$, $\beta_{\tilde{m}f_{W|X}}$, and $\beta_{\tilde{m}f_{WX}}$, relying only on observations of W, X_1, X_2, and Y.

1.2.4 Stochastic Identification

In what follows we take X and W to be scalars for simplicity. Analogous to the approach taken in SWC, we first focus on estimating quantities of the general form

$$g_{V,\lambda}(x,w) \equiv D^\lambda_x(E[V \mid X = x, W = w] f_{X|W}(x \mid w)),$$

where $D^\lambda_x \equiv (\partial^\lambda / \partial x^\lambda)$ denotes the derivative operator of degree λ, V is a generic random variable that will stand either for Y or for the constant ($V \equiv 1$), and $f_{X|W}$ is the conditional density of X given W. For example, special cases of the general form above are $f_{X|W}(x \mid w) = g_{1,0}(x,w)$, $E[Y \mid X = x, W = w] f_{X|W}(x \mid w) = g_{Y,0}(x,w)$, and $\mu(x,w) = g_{Y,0}(x,w)/g_{1,0}(x,w)$. Thus, with structural identification, the covariate-conditioned average marginal effect of X on Y at x given $W = w$ is

$$\beta(x,w) = \frac{g_{Y,1}(x,w)}{g_{1,0}(x,w)} - \frac{g_{Y,0}(x,w)}{g_{1,0}(x,w)} \frac{g_{1,1}(x,w)}{g_{1,0}(x,w)}.$$

We first analyze the asymptotic properties of estimators of $g_{V,\lambda}$ with generic V when we observe two error-contaminated measurements of X, as in Assumption 2.3. We can then straightforwardly obtain the asymptotic properties of estimators of $\beta(x,w)$ and weighted averages of $\beta(x,w)$. We denote the support of a random variable by supp$(\cdot)$. By convention, we take the value of any referenced function to be zero except when the indicated random variable lies in supp$(\cdot)$. We impose the following conditions on Y, X, W, U_1, and U_2.

Assumption 3.1 *$E[|X|] < \infty$ and $E[|U_1|] < \infty$.*

Assumption 3.2 *(i) $E[U_1 \mid X, U_2] = 0$; (ii) $U_2 \perp (X, W)$; (iii) $E[Y \mid X, U_2, W] = E[Y \mid X, W]$.*

Assumption 3.3 *(i) $\inf_{w \in \text{supp}(W)} f_W(w) > 0$; (ii) $\sup_{(x,w) \in \text{supp}(X,W)} f_{X|W}(x \mid w) < \infty$.*

Assumption 3.4 *For any finite $\zeta \in \mathbb{R}$, $|E[\exp(i\zeta X_2)]| > 0$.*

Assumption 3.1 imposes mild conditions regarding the existence of the first moments of the cause of interest and the measurement error of the first measurement error-laden observation. Assumption 3.4 is commonly imposed in the deconvolution literature (e.g., Fan, 1991; Fan and Truong, 1993; Li and Vuong, 1998; Li, 2002;, Schennach, 2004a,b), which requires a nonvanishing characteristic function for X_2. Assumptions 3.1, 3.3, and 3.4 jointly ensure that $g_{V,0}(x, w)$ is well defined.

Assumption 3.2 has been imposed in a similar fashion in the repeated measurements literature (e.g., Hausman, Ichimura, Newey, and Powell, 1991; and Schennach, 2004a, b); however, the presence of W is new here. Assumption 3.2(i) imposes a mild conditional moment restriction, while Assumption 3.2(ii) is crucial but plausible. The conditional mean restriction in Assumption 3.2(i) is imposed instead of independence to ensure the weakest possible assumptions. The independence in Assumption 3.2(ii) is necessary because of the nonlinearity of the model. Note that $E[U_1 \mid U_2] = E[E[U_1 \mid X, U_2] \mid U_2] = 0$, so that U_1 is mean independent of U_2. On the other hand, the mean of U_2 does not have to be zero. These relatively mild requirements on the measurement errors are plausible for many practical applications, but are asymmetric between U_1 and U_2. If symmetry is plausible, one can obtain analogous estimators, interchanging the roles of X_1 and X_2.

Let $\mathbb{N} \equiv \{0, 1, ...\}$ and $\overline{\mathbb{N}} \equiv \mathbb{N} \cup \{\infty\}$.

Assumption 3.5 *For $V = 1, Y$, $g_{V,0}(\cdot, w)$ is continuously differentiable of order $\Lambda \in \overline{\mathbb{N}}$ on $\mathbb{R}$ for each $w \in \text{supp}(W)$.*

This assumption imposes smoothness on $g_{V,0}$. If $g_{V,\lambda}$ can be defined solely in terms of the joint distribution of observable variables V, X_1, and X_2, we say it is "stochastically identified." This is shown in the next lemma.[2]

Lemma 3.1 *Suppose Assumptions 2.1(i), 2.3, and 3.1 - 3.5 hold. Then for $V = 1, Y$ and for each $\lambda \in \{0, ..., \Lambda\}$* and $(x, w) \in \text{supp}(X, W)$,

$$g_{V,\lambda}(x,w) = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta,$$

where for each real ζ,

$$\phi_V(\zeta,w) \equiv E[Ve^{i\zeta X} \mid W=w] = \frac{E[Ve^{i\zeta X_2} \mid W=w]}{E[e^{i\zeta X_2}]}\exp\left(\int_0^{\zeta}\frac{iE[X_1e^{i\xi X_2}]}{E[e^{i\xi X_2}]}d\xi\right).$$

Note that $g_{V,\lambda}$ is empirically accessible when it involves only observable variables. Thus, knowledge of $E[Ve^{i\zeta X_2} \mid W = w]$, $E[e^{i\zeta X_2}]$, and $E[X_1e^{i\xi X_2}]$ is sufficient to obtain stochastic identification of $g_{V,\lambda}$.

1.3 Estimation

Our nonparametric estimators of $g_{V,\lambda}$ make use of the following class of flat-top kernels of infinite order proposed by Politis and Romano (1999).

Assumption 3.6 *The real-valued kernel $x \to k(x)$ is measurable and symmetric, $\int k(x)dx = 1$, and its Fourier transform $\xi \to \kappa(\xi)$ is such that: (i) κ is compactly supported (without loss of generality, we take the support to be [-1,1]); and (ii) there exists $\bar{\xi} > 0$ such that $\kappa(\xi) = 1$ for $|\xi| < \bar{\xi}$.*

The above assumption is similar to that used in SWC. The fact that the kernel is continously differentiable to any order is ensured by the requirement of Assumption 3.6(i) that the Fourier transform of the kernel is compactly supported.

[2] Derivation of a part of the expression for ϕ_V is similar to that of an identity due to Kotlarski (see Rao, 1992, p. 21), which enables one to recover the densities of X, U_1, and U_2 from the joint density of X_1 and X_2 under the assumption that X, U_1, and U_2 are independent. Our identification strategy for the density of X relies on weaker assumptions than independence. In fact, we only require $E[U_1 \mid X, U_2] = 0$ and $U_2 \perp X$ for the result, instead of mutual independence of X, U_1, and U_2. As a result, our setup allows dependence between X and U_1, and between U_1 and U_2.

The assumption of compact support of κ is commonly used in the kernel deconvolution estimator (e.g., Fan and Truong, 1993; Schennach, 2004a). Because the kernel deconvolution estimator involves a division by an asymptotically vanishing characteristic function as frequency increases toward infinity, it suffers from the well-known ill-posed inverse problem that occurs when one tries to invert a convolution operation. This problem can be rectified by estimating an associated numerator using a kernel whose Fourier transform is compactly supported, which guarantees that the numerator will decay well before the denominator causes the ratio to diverge, ensuring that the divergence is kept under control.

Compact support of the Fourier transform of the kernel is a weak requirement because one can transform any given kernel $\tilde{k}$ into a modified kernel k with compact Fourier support, having most of the properties of the original kernel, as mentioned in Schennach (2004a). To construct the modified Fourier transform κ from the original Fourier transform $\tilde{\kappa}$ of $\tilde{k}$ put

$$\kappa(\xi) = \mathcal{W}(\xi)\tilde{\kappa}(\xi),$$
$$\mathcal{W}(\xi) = \begin{cases} 1 & \text{if} \quad |\xi| \leq \bar{\xi} \\ (1 + \exp((1-\bar{\xi})((1-|\xi|)^{-1} - (|\xi| - \bar{\xi})^{-1})))^{-1} & \text{if} \quad \bar{\xi} < |\xi| \leq 1. \\ 0 & \text{if} \quad 1 < |\xi| \end{cases} \tag{1.8}$$

Here $\mathcal{W}(\cdot)$ is a window function that is constant in the neighborhood of the origin and vanishes beyond a given frequency, determined by $\bar{\xi} \in (0, 1)$.

Flat-top kernels of infinite order have the property that their Fourier transforms are "flat" over an open neighborhood of the origin, as described in Politis and Romano (1999). When a flat-top kernel of infinite order is used, the smoothness of the function to be estimated is the only factor controlling the rate of decrease of the bias, whereas when a finite-order kernel is used, both the smoothness of the function and the order of the kernel affect the rate of decrease of the bias. When the function to be estimated is infinitely many times differentiable, a flat-top kernel of infinite order guarantees that the bias of the kernel estimator goes to zero faster than any power of the bandwidth. For instance, the bias from a flat-top kernel of infinite order could be an exponentially shrinking function of the inverse bandwidth, even though the bias from a traditional finite-order kernel is a decaying function of the inverse bandwidth to a negative power.

The estimator for $g_{V,\lambda}(x,w)$ is motivated by a smoothed version of $g_{V,\lambda}(x,w)$. The next lemma incorporates the kernel into the expression for $g_{V,\lambda}(x,w)$.

Lemma 3.2 *Suppose Assumptions 2.1(i), 2.3, 3.1, and 3.3 - 3.5 hold, and let k satisfy Assumption 3.6. For $V = 1, Y$ and for each $\lambda \in \{0, ..., \Lambda\}$, $(x,w) \in \operatorname{supp}(X,W)$, and $h_1 > 0$, let*

$$g_{V,\lambda}(x,w,h_1) \equiv \int \frac{1}{h} k\left(\frac{\tilde{x}-x}{h_1}\right) g_{V,\lambda}(\tilde{x},w) d\tilde{x}.$$

Then

$$g_{V,\lambda}(x,w,h_1) = \frac{1}{2\pi}\int (-i\zeta)^{\lambda} \kappa(h_1\zeta)\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta.$$

We let $h \equiv (h_1, h_2)$ denote the kernel bandwidth or smoothing parameter. Because $\lim_{h_1\to 0} g_{V,\lambda}(x,w,h_1) = g_{V,\lambda}(x,w)$ by lemma 1 of the appendix of Pagan and Ullah (1999, p.362), we also define $g_{V,\lambda}(x,w,0) \equiv g_{V,\lambda}(x,w)$. Motivated by Lemma 3.2, we define our estimator for $g_{V,\lambda}(x,w)$ as follows.

Definition 3.3 *Let $h_n \equiv (h_{1n}, h_{2n})$. The estimator for $g_{V,\lambda}(x,w)$ is defined as*

$$\hat{g}_{V,\lambda}(x,w,h_n) \equiv \frac{1}{2\pi}\int (-i\zeta)^{\lambda} \kappa(h_{1n}\zeta)\hat{\phi}_V(\zeta,w,h_{2n})\exp(-i\zeta x)d\zeta,$$

for $h_n \to 0$ as $n \to \infty$, where

$$\hat{\phi}_V(\zeta,w,h_{2n}) \equiv \frac{\hat{E}[Ve^{i\zeta X_2} \mid W = w]}{\hat{E}[e^{i\zeta X_2}]}\exp\left(\int_0^{\zeta} \frac{i\hat{E}[X_1 e^{i\xi X_2}]}{\hat{E}[e^{i\xi X_2}]}d\xi\right),$$

$$\hat{E}[Ve^{i\zeta X_2} \mid W = w] \equiv \frac{(nh_{2n})^{-1}\sum_{j=1}^n V_j e^{i\zeta X_{2j}} k\left(\frac{W_j - w}{h_{2n}}\right)}{(nh_{2n})^{-1}\sum_{j=1}^n k\left(\frac{W_j - w}{h_{2n}}\right)} = \frac{\hat{E}[Ve^{i\zeta X_2} k_{h_{2n}}(W-w)]}{\hat{E}[k_{h_{2n}}(W-w)]},$$

and where $k_{h_{2n}}(\cdot) = h_{2n}^{-1} k(\cdot/h_{2n})$ and $\hat{E}[\cdot]$ denotes a sample average.[3]

[3] There are two kernels in the expression of $\hat{g}_{V,\lambda}(x,w,h_n)$: one is associated with the regressor X and the other is needed for the conditioning instrument W. Even though we do not explicitly use different notations for the purpose of notational convenience, they could be different (note that nevertheless, we use different bandwidths for different kernels). So $\kappa(\cdot)$ is the Fourier transform of a flat-top kernel associated with X and $k(\cdot)$ is another flat-top kernel for W. Indeed, different flat-top kernels are incorporated in the empirical parts.

With $\hat{E}[\cdot]$ denoting a sample average, for any random variable X, $\hat{E}[X] \equiv n^{-1}\sum_{i=1}^{n} X_i$, where $X_1, ..., X_n$ is a sample of random variables, distributed identically as X. We replace $\phi_V(\zeta, w)$ by its sample analog, $\hat{\phi}_V(\zeta, w, h_{2n})$. $\hat{E}[Ve^{i\zeta X_2} \mid W = w]$ is a kernel estimator of $E[Ve^{i\zeta X_2} \mid W = w]$.

1.4 Asymptotics

1.4.1 Asymptotics for the General Form

SWC extensively generalize Schennach (2004a, b) to encompass (i) the $\lambda \neq 0$ case; (ii) uniform convergence results; and (iii) general semiparametric functionals of $g_{V,\lambda}$. Here, we use the approach of Schennach (2004a, b) to achieve counterparts of these three results in the context of models where endogeneity is handled with conditional independence, as in the treatment effect literature, and where the cause of interest is contaminated by measurement error. The analysis of estimator properties is complicated by the presence of the kernel estimator of the conditional expectation. We begin by deriving the asymptotic behavior of the estimator for the quantities of the general form $\hat{g}_{V,\lambda}(x, w, h_n)$. The first result decomposes the estimation error into a "bias term," a "variance term," and a "remainder term."

Lemma 4.1 *Suppose that $\{U_i, W_i, X_i, Y_i\}$ is an independent and identically distributed (IID) sequence satisfying Assumptions 2.1(i), 2.3, 3.1 - 3.5, and that Assumption 3.6 holds. Then for $V = 1, Y$ and for each $\lambda \in \{0, \ldots, \Lambda\}$, $(x, w) \in \text{supp}(X, W)$, and $h \equiv (h_1, h_2) > 0$,*

$$\hat{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w) = B_{V,\lambda}(x, w, h_1) + L_{V,\lambda}(x, w, h) + R_{V,\lambda}(x, w, h),$$

where $B_{V,\lambda}(x, w, h_1)$ is a nonrandom "bias term" defined as

$$B_{V,\lambda}(x, w, h_1) \equiv g_{V,\lambda}(x, w, h_1) - g_{V,\lambda}(x, w);$$

$L_{V,\lambda}(x, w, h)$ is a "variance term" admitting the linear representation

$$L_{V,\lambda}(x, w, h) \equiv \bar{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h_1) = \hat{E}\left[\ell_{V,\lambda}(x, w, h; V, X_1, X_2, W)\right],$$

where $\bar{g}_{V,\lambda}(x,w,h)$ *is the linearization of* $\hat{g}_{V,\lambda}(x,w,h)$ *in terms of* $(\hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}])$,
$(\hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}])$, $(\hat{E}[V e^{i\zeta X_2} k_{h_2}(W-w)] - E[V e^{i\zeta X_2} k_{h_2}(W-w)])$, *and* $(\hat{E}[k_{h_2}(W-w)] - E[k_{h_2}(W-w)])$, *where*

$$\begin{aligned}
&\ell_{V,\lambda}(x,w,h;v,x_1,x_2,\tilde{w}) \\
\equiv &\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)\left(e^{i\zeta x_2} - E\left[e^{i\zeta X_2}\right]\right) d\zeta \\
&+ \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)\left(x_1 e^{i\zeta x_2} - E\left[X_1 e^{i\zeta X_2}\right]\right) d\zeta \\
&+ \int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)\left(v e^{i\zeta x_2} k_{h_2}(\tilde{w}-w) - E\left[V e^{i\zeta X_2} k_{h_2}(W-w)\right]\right) d\zeta \\
&+ \int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)\left(k_{h_2}(\tilde{w}-w) - E\left[k_{h_2}(W-w)\right]\right) d\zeta,
\end{aligned}$$

and where, letting $\theta_A(\zeta) \equiv E\left[A e^{i\zeta X_2}\right]$ *for* $A = 1, X_1$ *and* $\chi_V(\zeta,w) \equiv \int\int v e^{i\zeta x_2}$ $f_{V,X_2,W}(v,x_2,w)dvdx_2$, *we define*

$$\begin{aligned}
\Psi_{V,\lambda,1}(\zeta,x,w,h_1) \equiv &-\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}(-i\xi)^\lambda \kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi \\
&-\frac{1}{2\pi}(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\theta_1(\zeta)} \\
\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1) \equiv &\frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_\zeta^{\pm\infty}(-i\xi)^\lambda \kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi \\
\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1) \equiv &\frac{1}{2\pi}(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)} \\
\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1) \equiv &-\frac{1}{2\pi}(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{f_W(w)},
\end{aligned}$$

where for a given function $\xi \to f(\xi)$, *we write* $\int_\zeta^{\pm\infty} f(\xi)d\xi \equiv \lim_{c\to+\infty}\int_\zeta^{c\zeta} f(\xi)d\xi$; *and* $R_{V,\lambda}(x,w,h)$ *is a "remainder term,"*

$$R_{V,\lambda}(x,w,h) \equiv \hat{g}_{V,\lambda}(x,w,h) - \bar{g}_{V,\lambda}(x,w,h).$$

Because $\hat{g}_{V,\lambda}(x,w,h)$ takes the form of a nonlinear functional of the data generating process, the above linearization facilitates the analysis of the asymptotic behavior of the estimator. In fact, the limiting distribution of $\hat{g}_{V,\lambda}(x,w,h) -$

$g_{V,\lambda}(x, w)$ is equivalent to that of $L_{V,\lambda}(x, w, h)$, as long as $B_{V,\lambda}(x, w, h_1)$ and $R_{V,\lambda}$ (x, w, h) are asymptotically negligible. Thus we first establish bounds on the bias, the variance, and the remainder terms; we then establish the asymptotic normality of the variance term.

To obtain rate of convergence results for our kernel estimators, we impose bounds on the tail behavior of the Fourier transforms. These conditions describe the smoothness of the corresponding densities. The deconvolution literature (e.g., Fan, 1991; Fan and Truong, 1993; Li and Vuong, 1998; Li, 2002; Schennach, 2004a; and Caroll, Ruppert, Stefanski, and Crainiceanu, 2006) commonly distinguishes between "ordinarily smooth" and "supersmooth" functions. Specifically, ordinarily smooth functions admit a finite number of continuous derivatives and have a Fourier transform whose tail decays to zero at a geometric rate, $|\zeta|^{\gamma}, \gamma < 0$, as the frequency, $|\zeta|$, goes to infinite (e.g., uniform, gamma, and double exponential); whereas supersmooth functions admit an infinite number of continuous derivatives and have a Fourier transform whose tail decays to zero at an exponential rate as $\exp(\alpha|\zeta|^{\beta}), \alpha < 0, \beta > 0$ as the frequency goes to infinite (e.g., Cauchy and normal). For conciseness, our smoothness restrictions encompass both the ordinarily smooth and supersmooth cases; for this, our regularity conditions are expressed in terms of $(1 + |\zeta|)^{\gamma} \exp(\alpha|\zeta|^{\beta})$.

Assumption 4.1 *Let* $\phi_1(\zeta) \equiv E[e^{i\zeta X}]$.

(i) There exist constants $C_1 > 0$ *and* $\gamma_1 \geq 0$ *such that*

$$|D_\zeta \ln \phi_1(\zeta)| = \left| \frac{D_\zeta \phi_1(\zeta)}{\phi_1(\zeta)} \right| \leq C_1(1 + |\zeta|)^{\gamma_1};$$

(ii) 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* $V = 1, Y$

$$\sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \leq C_\phi(1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_\phi|\zeta|^{\beta_\phi}),$$

and if $\alpha_\phi = 0$, *then* $\gamma_\phi < -\lambda - 1$ *for given* $\lambda \in \{0, \ldots, \Lambda\}$;

(iii) There exist constants $C_\theta > 0$,$\alpha_\theta \leq 0$, $\beta_\theta \geq \beta_\phi \geq 0$, *and* $\gamma_\theta \in \mathbb{R}$ *such that* $\beta_\theta \gamma_\theta \geq 0$ *and for* $V = 1, Y$

$$\min\{\inf_{w \in \text{supp}(W)} |\chi_V(\zeta, w)|, |\theta_1(\zeta)|\} \geq C_\theta(1 + |\zeta|)^{\gamma_\theta} \exp(\alpha_\theta|\zeta|^{\beta_\theta}).$$

We omit a term $\exp(\alpha_1|\zeta|^{\beta_1})$ in Assumption 4.1(i) with negligible loss of generality because $\ln\phi_1$ is typically a power of ζ for large ζ, even when the density of $\phi_1(\zeta)$ is supersmooth, as pointed out in Schennach (2004a) and SWC. Note that the rate of decay of $\phi_V(\zeta,w)$ is governed by the smoothness of $g_{V,0}(x,w) = E[V \mid X = x, W = w]f_{X|W}(x \mid w)$, as $\phi_V(\zeta,w) = \int g_{V,0}(x,w)e^{i\zeta x}dx$. Note that a lower bound, instead of an upper bound, is imposed on $\chi_V(\zeta,w)$ and $\theta_1(\zeta)$, because these appear in the denominator of the expression for $\hat{g}_{V,\lambda}(x,w,h)$. Individual lower bounds on the modulus of the characteristic functions of X and U_2 imply the lower bound on $\theta_1(\zeta)$, as $\theta_1(\zeta) = E[e^{i\zeta X_2}] = E[e^{i\zeta X}]E[e^{i\zeta U_2}]$ by Assumption 3.2(ii). We group together $\chi_V(\zeta,w)$ and $\theta_1(\zeta)$ (in fact, $E[e^{i\zeta X}]$ and $E[e^{i\zeta U_2}]$) in a single assumption for the lower bound for notational convenience. We explicitly impose $\beta_\theta \geq \beta_\phi$ because

$$
\begin{aligned}
&C_\phi(1+|\zeta|)^{\gamma_\phi}\exp(\alpha_\phi|\zeta|^{\beta_\phi}) \geq \sup_{w\in\mathrm{supp}(W)}|\phi_1(\zeta,w)| = \sup_{w\in\mathrm{supp}(W)}|E[e^{i\zeta X} \mid W = w]| \\
&\geq \left|\int E[e^{i\zeta X} \mid W = w]f_W(w)dw\right| = |E[e^{i\zeta X}]| \geq |E[e^{i\zeta X}]||E[e^{i\zeta U_2}]| \geq |E[e^{i\zeta X_2}]| \\
&= |\theta_1(\zeta)| \geq C_\theta(1+|\zeta|)^{\gamma_\theta}\exp(\alpha_\theta|\zeta|^{\beta_\theta}).
\end{aligned}
$$

The next theorem describes the asymptotic properties of the bias term defined in Lemma 4.1.

Theorem 4.2 *Let the conditions of Lemma 4.1 hold, and suppose in addition that Assumption 4.1 (ii) holds. Then for $V = 1, Y$, and each $\lambda \in \{0,\dots,\Lambda\}$ and $h_1 > 0$,*

$$
\sup_{(x,w)\in\mathrm{supp}(X,W)}|B_{V,\lambda}(x,w,h_1)| = O\left(\left(h_1^{-1}\right)^{\gamma_{\lambda,B}}\exp\left(\alpha_B\left(h_1^{-1}\right)^{\beta_B}\right)\right),
$$

where $\alpha_B \equiv \alpha_\phi\bar{\xi}^{\beta_\phi}$, $\beta_B \equiv \beta_\phi$, and $\gamma_{\lambda,B} \equiv \gamma_\phi + \lambda + 1$.

Note that the bias term behaves identically to that of a conventional kernel estimator employed when X is measurement error-free, because $B_{V,\lambda}(x,w,h_1)$ only involves the kernel and error-free variables.[4]

[4]When X is perfectly observed, one can propose an estimator of $g_{V,\lambda}$ using a similar Fourier

To establish a divergence rate and asymptotic normality for the variance term, $L_{V,\lambda}(x, w, h)$, we impose some regularity conditions. We first impose conditions ensuring finite variance of $L_{V,\lambda}(x, w, h)$.

Assumption 4.2 *$E[|X_1|^2] < \infty$ and $E[|Y|^2] < \infty$.*

We next impose bounds on some moments that are useful for establishing asymptotic normality of $L_{V,\lambda}(x, w, h)$.

Assumption 4.3 *For some $\delta > 0$, $E[|X_1|^{2+\delta}] < \infty$, $\sup_{x_2 \in \text{supp}(X_2)} E[|X_1|^{2+\delta} \mid X_2 = x_2] < \infty$, $E[|Y|^{2+\delta}] < \infty$, and $\sup_{w \in \text{supp}(W)} E[|Y|^{2+\delta} \mid W = w] < \infty$.*

We also suitably control the bandwidth to establish asymptotic normality.

Assumption 4.4 *$h_n \to 0$ as $n \to \infty$, such that: if $\beta_\theta \neq 0$ in Assumption 4.1(iii), then $h_{2n}^{-1} = O\left(\exp(\frac{1}{2}\ln n^{3/2-\eta} - \frac{1}{2}(\alpha_\phi 1_{\beta_\theta = \beta_\phi} - \alpha_\theta)(\ln n)^{1-\eta\beta_\theta})\right)$ and $h_{1n}^{-1} = O\left((\ln n)^{1/\beta_\theta - \eta}\right)$ for some $\eta > 0$; otherwise, for each $\lambda \in \{0, \ldots, \Lambda\}$, $h_{2n}^{-1} = O\left(n^{(3-2\eta)/4} n^{3\eta(\gamma_\phi + \lambda - \gamma_\theta + 1)/(2(\gamma_\phi + \lambda + \gamma_1 - \gamma_\theta + 3))}\right)$ and $h_{1n}^{-1} = O\left(n^{-\eta} n^{(3/2)/(\gamma_\phi + \lambda + \gamma_1 - \gamma_\theta + 3)}\right)$ for some $\eta > 0$.*

The bandwidth sequences given above can be selected by ensuring that a regularity condition in Lemma A.2 holds (see Lemma A.2 and the proof of Theorem 4.3 in the Appendix). The bandwidth sequences imply that if densities appearing

---

transform as

$$\hat{g}_{V,\lambda}(x, w, h_n) \equiv \frac{1}{2\pi} \int (-i\zeta)^\lambda \kappa(h_{1n}\zeta) \hat{\phi}_V(\zeta, w, h_{2n}) \exp(-i\zeta x) d\zeta,$$

for $h_n \to 0$ as $n \to \infty$, where

$$\hat{\phi}_V(\zeta, w, h_{2n}) \equiv \hat{E}[V e^{i\zeta X} \mid W = w] = \frac{\hat{E}[V e^{i\zeta X} k_{h_{2n}}(W - w)]}{\hat{E}[k_{h_{2n}}(W - w)]}.$$

Then one can easily derive the order of the bias, which is the same as that in Theorem 4.2. Note that this estimator for $g_{V,\lambda}$ has the same asymptotic properties as a traditional kernel estimator of $g_{V,\lambda}$ with the flat-top kernel of infinite order when X is perfectly observed; but this estimator using the Fourier transform approach makes possible easy comparisons with our estimator in Definition 3.3.

in quantities in the denominator ($\chi_V(\zeta, w)$ and θ_1) are supersmooth, one must choose a larger bandwidth than in the case of ordinary smoothness. The achievable convergence rates will thus be slower than for ordinary smoothness. Similar but simpler results have also been observed in the kernel deconvolution literature (see Fan (1991), Fan and Truong (1993), Li and Vuong (1998), Li (2002), and Schennach (2004a)).

We are ready to state a uniform rate and asymptotic normality for the variance term.

Theorem 4.3 *Let the conditions of Lemma 4.1 hold. (i) Then for* $V = 1, Y$ *and for each* $\lambda \in \{0, ..., \Lambda\}$, $(x, w) \in \operatorname{supp}(X, W)$, *and* $h > 0$, $E[L_{V,\lambda}(x, w, h)] = 0$, *and if Assumption 4.2 also holds, then*

$$E\left[(L_{V,\lambda}(x, w, h_n))^2\right] = n^{-1}\Omega_{V,\lambda}(x, w, h_n),$$

where

$$\Omega_{V,\lambda}(x, w, h_n) \equiv E\left[(\ell_{V,\lambda}(x, w, h_n; V, X_1, X_2, W))^2\right]$$

is finite. Further, if Assumption 4.1 holds, then

$$\begin{aligned}&\sqrt{\sup_{(x,w)\in\operatorname{supp}(X,W)} \Omega_{V,\lambda}(x, w, h_n)}\\ &= O\left(\max\{\left(h_{1n}^{-1}\right)^{\delta_L}, h_{2n}^{-1}\} \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right)\right),\end{aligned}$$

with $\alpha_L \equiv \alpha_\phi 1_{\{\beta_\phi = \beta_\theta\}} - \alpha_\theta$, $\beta_L \equiv \beta_\theta$, $\gamma_{\lambda,L} \equiv 1 + \gamma_\phi - \gamma_\theta + \lambda$, *and* $\delta_L \equiv 1 + \gamma_1$. *We also have*

$$\begin{aligned}&\sup_{(x,w)\in\operatorname{supp}(X,W)} |L_{V,\lambda}(x, w, h_n)|\\ &= O_p\left(n^{-1/2}\left(\max\{\left(h_{1n}^{-1}\right)^{\delta_L}, h_{2n}^{-1}\}\right) \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right)\right);\end{aligned}$$

(ii) If Assumptions 4.3 and 4.4 also hold, and if for $V = 1, Y$ *and for each* $\lambda \in \{0, ..., \Lambda\}$, $(x, w) \in \operatorname{supp}(X, W)$, $\Omega_{V,\lambda}(x, w, h_n) > 0$ *for all* n *sufficiently large, then*

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

A few remarks are in order. The rate of divergence of the variance term is controlled by the smoothness of the density of the measurement error U_2 and

$E[\varphi(x_2, w) \mid X_2 = x_2]$ (through $\gamma_\theta, \alpha_\theta, \beta_\theta$) as well as by the smoothness of the density of X and $E[V \mid X = x, W = w]$ (through $\gamma_\phi, \alpha_\phi, \beta_\phi$, and γ_1), where $\varphi(x_2, w) = \int v f_{V,X_2,W}(v, x_2, w) dv$. As expected, the order of the variance term is larger than that of a traditional kernel estimator with error-free variables.[5] As a result, the rate of convergence of the estimator $\hat{g}_{V,\lambda}$ will be slower than that of a standard kernel estimator, because the bias term is identical to that of a standard kernel estimator with measurement error-free X.

We now establish a uniform convergence rate and asymptotic normality of the estimator $\hat{g}_{V,\lambda}(x, w, h_n)$. We first provide bounds on the remainder term that are used to obtain a convergence rate. The next assumption puts restrictions on the moments of X_2 that are useful for establishing a bound on the remainder term, $R_{V,\lambda}(x, w, h_n)$.

Assumption 4.5 $E[|X_2|] < \infty$, $E[|X_1 X_2|] < \infty$, *and* $E[|Y X_2|] < \infty$.

The following assumption provides a uniform convergence rate for the kernel density estimator, $\hat{f}_W(w)$, in the denominator of $\hat{g}_{V,\lambda}(x, w, h)$. This assumption is also used to get the bound on the remainder term and is satisfied by density estimation with conventional choice of kernel. Even though flat-top kernels of infinite order attain a faster convergence rate than that below (e.g., Politis and Romano, 1999), the faster rate is not necessary for our result.

Assumption 4.6 $\sup_{w \in \text{supp}(W)} |\hat{f}_W(w) - f_W(w)| = O_p\left(\sqrt{\frac{\ln n}{n h_2}} + h_2^2\right)$.

The following assumption gives a lower bandwidth bound that slightly differs from that of Assumption 4.4. Note that neither Assumption 4.4 nor 4.7 is necessarily stronger than the other.

Assumption 4.7 *If* $\beta_\theta \neq 0$ *in Assumption 4.1,* $h_{1n}^{-1} = O\left((\ln n)^{1/\beta_\theta - \eta}\right)$ *and* $h_{2n}^{-1} = O\left(\exp(\frac{\alpha_\theta}{4}(\ln n)^{1-\eta\beta_\theta})\right)$ *for some* $\eta > 0$; *otherwise*

[5] With perfectly observed X, the order of the variance term of the estimator in footnote 2 can be derived as $n^{-1/2} h_{2n}^{-1} \left(h_{1n}^{-1}\right)^{1+\gamma_\phi+\lambda} \exp\left(\alpha_\phi \left(h_{1n}^{-1}\right)^{\beta_\phi}\right)$. Thus if $\beta_\phi > 0, \beta_L \equiv \beta_\theta \geq \beta_\phi$ by construction, and if $\beta_L \equiv \beta_\theta = \beta_\phi = 0, \gamma_{\lambda,L} \equiv 1 + \gamma_\phi - \gamma_\theta + \lambda > 1 + \gamma_\phi + \lambda$ since $(-\gamma_\theta) > 0$, and $\max\{\left(h_{1n}^{-1}\right)^{\delta_L}, h_{2n}^{-1}\} \geq h_{2n}^{-1}$. Then the order of the variance term in Theorem 4.3 is greater than that of the kernel estimator with perfectly observed variables.

$h_{1n}^{-1} = O\left(n^{-\eta} n^{1/(2\gamma_1 - 2\gamma_\theta)}\right)$ *and* $h_{2n}^{-1} = O\left(n^{\eta(\gamma_1 - \gamma_\theta - 1)/4}\right)$ *for some* $\eta > 0$.

The bandwidth sequences above can be selected to ensure that the nonlinear remainder term, $R_{V,\lambda}(x, w, h_n)$, is indeed asymptotically negligible so that the decomposition of the estimation error into bias, variance, and remainder terms is justified, thus implying that the linear approximation of $\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)$ using the variance term, $L_{V,\lambda}(x, w, h_n)$, is appropriate. The basic intuition behind the selection of the bandwidth is similar to that for Assumption 4.4. We now state uniform bounds on the nonlinear remainder.

Theorem 4.4 *(i) Suppose the conditions of Theorem 4.3 hold, together with Assumptions 4.5, 4.6. Then for* $V = 1, Y$*, each* $\lambda \in \{0, ..., \Lambda\}$*, and some* $\epsilon > 0$,

$$\begin{aligned}
&\sup_{(x,w)\in \operatorname{supp}(X,W)} |R_{V,\lambda}(x, w, h_n)| \\
&= O_p\left(n^{-1/2+\epsilon} (h_{2n}^{-1})^3 (h_{1n}^{-1})^{\gamma_1 - \gamma_\theta} \exp\left(-\alpha_\theta (h_{1n}^{-1})^{\beta_\theta}\right)\right) \\
&\times O_p\left(n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}\right)(h_{1n}^{-1})^{\gamma_{\lambda,L}} \exp\left(\alpha_L (h_{1n}^{-1})^{\beta_L}\right)\right);
\end{aligned}$$

(ii) If Assumption 4.7 holds in place of Assumption 4.4, then for $V = 1, Y$ *and each* $\lambda \in \{0, ..., \Lambda\}$,

$$\begin{aligned}
&\sup_{(x,w)\in \operatorname{supp}(X,W)} |R_{V,\lambda}(x, w, h_n)| \\
&= o_p\left(n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}\right)(h_{1n}^{-1})^{\gamma_{\lambda,L}} \exp\left(\alpha_L (h_{1n}^{-1})^{\beta_L}\right)\right).
\end{aligned}$$

Theorem 4.4 (i) is used to establish the asymptotic normality of $\hat{g}_{V,\lambda}$, and (ii) is relevant to obtaining a convergence rate. The next corollary establishes a uniform convergence rate by combining Theorems 4.2, 4.3, and 4.4(ii).

Corollary 4.5 *If the conditions of Theorem 4.4(ii) hold, then for* $V = 1, Y$ *and each* $\lambda \in \{0, ..., \Lambda\}$,

$$\begin{aligned}
&\sup_{(x,w)\in \operatorname{supp}(X,W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w, 0)| \\
=& O\left(\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h_{1n}^{-1}\right)^{\beta_B}\right)\right) \\
&+ O_p\left(n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right)\right).
\end{aligned}$$

In the next assumption, we ensure that the bias term and remainder term do not dominate the variance term admitting the linear representation.

Assumption 4.8 *$h_n \to 0$ at a rate such that for $V = 1, Y$ and for each $\lambda \in \{0, ..., \Lambda\}$ and $(x, w) \in \operatorname{supp}(X, W)$ we have: (i) $\Omega_{V,\lambda}(x, w, h_n) > 0$ for all n sufficiently large; (ii) $n^{1/2} (\Omega_{V,\lambda}(x, w, h_n))^{-1/2} \; |B_{V,\lambda}(x, w, h_{1n})| \to 0$; and (iii) $n^{1/2} (\Omega_{V,\lambda}(x, w, h_n))^{-1/2} |R_{V,\lambda}(x, w, h_n)| \xrightarrow{p} 0$.*

This assumption provides a lower bound on $\Omega_{V,\lambda}(x, w, h_n)$ such that $B_{V,\lambda}$ (x, w, h_{1n}) and $R_{V,\lambda}(x, w, h_n)$ are small relative to this lower bound. Note that the bound on $\Omega_{V,\lambda}(x, w, h_n)$ given in Theorem 4.3(i) is an upper bound on the convergence rate, so is not sufficient to obtain our next result, Corollary 4.6. As a result, the bias term and nonlinear remainder term must be asymptotically negligible relative to $n^{-1/2}(\Omega_{V,\lambda}(x, w, h_n))^{1/2}$, the standard deviation of $L_{V,\lambda}(x, w, h_n)$, in order to ensure that they have no effect on the limiting distribution of the estimator.

The following corollary establishes asymptotic normality by collecting together Assumption 4.8, Theorem 4.3, and Theorem 4.4(i).

Corollary 4.6 *If the conditions of Theorem 4.4 (i) and Assumption 4.8 hold, then for $V = 1, Y$ and each $\lambda \in \{0, ..., \Lambda\}$ and $(x, w) \in \operatorname{supp}(X, W)$, we have*

$$n^{1/2} (\Omega_{V,\lambda}(x, w, h_n))^{-1/2} (\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w, 0)) \xrightarrow{d} N(0, 1).$$

1.4.2 Asymptotics for Functionals of the General Form

We now consider functionals b of J-vectors $g_x \equiv (g_{V_1,\lambda_1}(x, \cdot), ..., g_{V_J,\lambda_J}(x, \cdot))$ and $g \equiv (g_{V_1,\lambda_1}, ..., g_{V_J,\lambda_J})$ with finite J, and establish the asymptotic properties of $b(\hat{g}_x(h)) - b(g_x) \equiv b((\hat{g}_{V_1,\lambda_1}(x, \cdot, h), ..., \hat{g}_{V_J,\lambda_J}(x, \cdot, h)) - b((g_{V_1,\lambda_1}(x, \cdot), ..., g_{V_J,\lambda_J}(x, \cdot))$ and $b(\hat{g}(h)) - b(g) \equiv b((\hat{g}_{V_1,\lambda_1}(\cdot, h), ..., \hat{g}_{V_J,\lambda_J}(\cdot, h)) - b((g_{V_1,\lambda_1}, ..., g_{V_J,\lambda_J}))$. The first of the following theorems is relevant to estimating $\beta_m(x)$, $\beta_{mf_W}(x)$, and $\beta_{mf_{W|X}}(x)$. Because the weighted average of coordinates of g_x is taken only over w, functionals of g_x obtain a rate between $\sqrt{n}-$ and that obtained in Corollary 4.5. It is not easy to use a functional delta method to obtain asymptotic normality of the functional because we need to show tightness of integrands by introducing trimming of the tails of characteristic functions in the theorem. We therefore leave formal treatment of asymptotic normality results to future research. The second theorem is useful for estimating $\beta_{\tilde{m}}$, $\beta_{\tilde{m}f_{W,X}}$, and $\beta_{\tilde{m}f_{W|X}}$ and delivers $\sqrt{n}-$ consistency and asymptotic normality results for the weighted averages of interest. Because it involves a weighted average over both x and w, it achieves the standard parametric rate of convergence. Each theorem relies on the validity of an asymptotically linear representation, useful for analyzing a scalar estimator constructed as a functional of a vector of estimators. To obtain a faster rate for functionals of g_x than that for $g_{V,\lambda}(x,w)$, we first impose a bound on the tail behavior of the Fourier transforms involved, as in Assumption 4.1.

Assumption 4.9 *Suppose that for each* $x \in \operatorname{supp}(X)$, $\sup_{x\in\operatorname{supp}(X)} \int |s(x,w)|dw < \infty$. *Then for* $V = 1, Y$, *there exist constants* $C_{\phi s} > 0$, $\alpha_{\phi s} \leq 0$, $\beta_{\phi s} \geq \beta_\phi \geq 0$, *and* $\gamma_{\phi s} \in \mathbb{R}$ *such that* $\beta_{\phi s}\gamma_{\phi s} \geq 0$ *and if* $\beta_{\phi s} = \beta_\phi = 0$, $\gamma_\phi \geq \gamma_{\phi s}$, *and*

$$\sup_{x\in\operatorname{supp}(X)} \left| \int \phi_V(\zeta,w)s(x,w)dw \right| \leq C_{\phi s}(1+|\zeta|)^{\gamma_{\phi s}} \exp(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}}),$$

and in addition if $\alpha_{\phi s} = 0$, *then* $\gamma_{\phi s} < -\lambda - 1$ *for given* $\lambda \in \{0, \ldots, \Lambda\}$.

The assumption above relies on the intuition that averaging a quantity generates a faster convergence rate. It is natural to assume $\beta_{\phi s} \geq \beta_\phi$ and if $\beta_{\phi s} = \beta_\phi = 0$, $\gamma_\phi \geq \gamma_{\phi s}$, because

$$\begin{aligned} \sup_{w\in\operatorname{supp}(W)} |\phi_V(\zeta,w)| &\succeq \left(\sup_{w\in\operatorname{supp}(W)} |\phi_V(\zeta,w)| \right) \left(\sup_{x\in\operatorname{supp}(X)} \left| \int s(x,w)dw \right| \right) \\ &\geq \sup_{x\in\operatorname{supp}(X)} \left| \int \phi_V(\zeta,w)s(x,w)dw \right|. \end{aligned}$$

Observe, however, that the inequality above can hold even when $\beta_{\phi s} < \beta_\phi$ or $\gamma_\phi < \gamma_{\phi s}$, because both bounds on $\sup_{w\in\operatorname{supp}(W)} |\phi_V(\zeta,w)|$ and on $\sup_{x\in\operatorname{supp}(X)} | \int \phi_V(\zeta,w) s(x,w)dw \mid$ given in Assumption 4.1(*ii*) and 4.9, respectively, are upper bounds. Thus, a faster convergence rate due to averaging over W is not a necessary result.

We next impose minimum convergence rates in a high-level form for conciseness.

Assumption 4.10 *$h_n \to 0$ as $n \to \infty$ such that for all $\lambda \in \{0, ..., \Lambda\}$, we have: (i) if $\beta_{\phi s} = \beta_\phi > 0$ or $\gamma_\phi = \gamma_{\phi s}$ for $\beta_{\phi s} = \beta_\phi = 0$, $\sup_{(x,w)\in \text{supp}(X,W)} |B_{V,\lambda}(x, w, h_{1n})| = o(\alpha_{1n})$, $\sup_{(x,w)\in \text{supp}(X,W)} |L_{V,\lambda}(x, w, h_n)| = o_p\left(\alpha_{1n}^{1/2}\right)$, and $\sup_{(x,w)\in \text{supp}(X,W)} |R_{V,\lambda}(x, w, h_n)| = o_p(\alpha_{1n})$ where $\alpha_{1n} \equiv \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h_{1n}^{-1}\right)^{\beta_B}\right) + n^{-1/2} \left(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}\right) \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right)$ and where α_B, β_B, $\gamma_{\lambda,B}$, α_L, β_L, $\gamma_{\lambda,L}$, and δ_L are as defined in Theorem 4.2 and 4.3.*

(ii) if $\beta_{\phi s} > \beta_\phi > 0$ or $\gamma_\phi > \gamma_{\phi s}$ for $\beta_{\phi s} = \beta_\phi = 0$, $\sup_{(x,w)\in \text{supp}(X,W)} |B_{V,\lambda}(x, w, h_{1n})| = o(\alpha_{2n})$, $\sup_{(x,w)\in \text{supp}(X,W)} |L_{V,\lambda}(x, w, h_n)| = o_p\left(\alpha_{2n}^{1/2}\right)$, and $\sup_{(x,w)\in \text{supp}(X,W)} |R_{V,\lambda}(x, w, h_n)| = o_p(\alpha_{2n})$ where $\alpha_{2n} \equiv \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,B,s}} \exp\left(\alpha_{B,s} \left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right) + n^{-1/2} \left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right) \left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L,s}} \exp\left(\alpha_{L,s} \left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)$, and where $\alpha_{B,s} \equiv \alpha_{\phi s}\bar{\xi}^{\beta_{\phi s}}$, $\beta_{B,s} \equiv \beta_{\phi s}$, $\gamma_{\lambda,B,s} \equiv \gamma_{\phi s} + \lambda + 1$, $\alpha_{L,s} \equiv \alpha_{\phi s} 1_{\{\beta_{\phi s} \geq \beta_\theta\}} - \alpha_\theta 1_{\{\beta_{\phi s} \leq \beta_\theta\}}$, $\beta_{L,s} \equiv \max\{\beta_\theta, \beta_{\phi s}\}$, $\gamma_{\lambda,L,s} \equiv 1 + \gamma_{\phi s} - \gamma_\theta + \lambda$, and $\delta_{L,s} \equiv 1 + \gamma_1$.

We now establish a faster convergence rate for functionals of g_x than that for $g_{V,\lambda}(x, w)$, which is useful for analyzing $\beta_m(x)$, $\beta_{mf_W}(x)$, and $\beta_{mf_{W|X}}(x)$.

Theorem 4.7 *For given Λ, $J \in \mathbb{N}$, let $\lambda_1, ..., \lambda_J$ belong to $\{0, ..., \Lambda\}$, let $V_1, ..., V_J$ belong to $\{1, Y\}$, and suppose that the conditions of Corollary 4.5 and Assumption 4.9 hold. For each $x \in \text{supp}(X)$, let the real-valued functional b satisfy, for any $\tilde{g}_x \equiv (\tilde{g}_{V_1,\lambda_1}(x, \cdot), ..., \tilde{g}_{V_J,\lambda_J}(x, \cdot))$ in an L_∞ neighborhood of the J-vector $g_x \equiv (g_{V_1,\lambda_1}(x, \cdot), ..., g_{V_J,\lambda_J}(x, \cdot))$,*

$$b(\tilde{g}_x) - b(g_x) = \sum_{j=1}^{J} \int \left(\tilde{g}_{V_j,\lambda_j}(x, w) - g_{V_j,\lambda_j}(x, w)\right) s_j(x, w) dw + \sum_{j=1}^{J} O\left(\| \tilde{g}_{V_j,\lambda_j}(x, \cdot) - g_{V_j,\lambda_j}(x, \cdot) \|_\infty^2\right) \tag{1.9}$$

for some real-valued functions s_j, $j = 1, ..., J$. In addition, suppose that s_j is such that $\sup_{x\in \text{supp}(X)} \int |s_j(x, w)| dw < \infty$, and let $\hat{g}_x(h_n) \equiv (\hat{g}_{V_1,\lambda_1}(x, \cdot, h_n), ..., \hat{g}_{V_J,\lambda_J}(x, \cdot, h_n))$. (i) If Assumption 4.10(i) holds, then

$$\sup_{x\in \text{supp}(X)} |b(\hat{g}_x(h_n)) - b(g_x)| = O\left(\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h_{1n}^{-1}\right)^{\beta_B}\right)\right)$$

$$+ O_p\left(n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right)\right);$$

(ii) If Assumption 4.10(ii) holds, then

$$\begin{aligned}
&\sup_{x \in \text{supp}(X)} |b(\hat{g}_x(h_n)) - b(g_x)| \\
&= O\left(\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,B,s}} \exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad + O_p\left(n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{\lambda,L,s}} \exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\right).
\end{aligned}$$

Note that Eqn. (1.9) of this result is Fréchet differentiability of $b(\tilde{g}_x)$ with respect to $\tilde{g}_x$ in the norm $\| \tilde{g}_{V_j,\lambda_j}(x,\cdot) \|_\infty^2$, where the derivative is $s_j(x, w)$.

We impose minimum convergence rates for the next theorem in a high-level form.

Assumption 4.11 *$h_n \to 0$ as $n \to \infty$ such that for all $\lambda \in \{0, ..., \Lambda\}$, we have* $\sup_{(x,w)\in\text{supp}(X,W)} |B_{V,\lambda}(x, w, h_{1n})| = o\left(n^{-1/2}\right)$, $\sup_{(x,w)\in\text{supp}(X,W)} |L_{V,\lambda}(x, w, h_n)| = o_p\left(n^{-1/4}\right)$, $\sup_{(x,w)\in\text{supp}(X,W)} |R_{V,\lambda}(x, w, h_n)| = o_p\left(n^{-1/2}\right)$, *and* $\sup_{w\in\text{supp}(W)} |\hat{f}_W(w) - f_W(w)| = o_p\left(n^{-1/4}\right)$.

The following theorem gives a convenient asymptotic normality and $\sqrt{n}-$consistency result useful for analyzing $\beta_{\tilde{m}}$, $\beta_{\tilde{m}f_{W,X}}$, and $\beta_{\tilde{m}f_{W|X}}$.

Theorem 4.8 *For given Λ, $J \in \mathbb{N}$, let $\lambda_1, ..., \lambda_J$ belong to $\{0, ..., \Lambda\}$, let $V_1, ..., V_J$ belong to $\{1, Y\}$, and suppose that the conditions of Corollary 4.6 and Assumption 4.8 hold. Let the real-valued functional b satisfy, for any $\tilde{g} \equiv (\tilde{g}_{V_1,\lambda_1}, ..., \tilde{g}_{V_J,\lambda_J})$ in an L_∞ neighborhood of the J-vector $g \equiv (g_{V_1,\lambda_1}, ..., g_{V_J,\lambda_J})$ and for any $\tilde{f} \equiv \tilde{f}_W$ in a neighborhood of $f \equiv f_W$,*

$$\begin{aligned}
b(\tilde{g}, \tilde{f}) - b(g, f) = &\sum_{j=1}^{J} \int\int \left(\tilde{g}_{V_j,\lambda_j}(x, w) - g_{V_j,\lambda_j}(x, w)\right) s_j(x, w) dw dx \\
&+ \int\int \left(\tilde{f}_W(w) - f_W(w)\right) s_{J+1}(x, w) dw dx \qquad (1.10) \\
&+ \sum_{j=1}^{J} O\left(\| \tilde{g}_{V_j,\lambda_j} - g_{V_j,\lambda_j} \|_\infty^2\right) + O\left(\| \tilde{f}_W - f_W \|_\infty^2\right)
\end{aligned}$$

for some real-valued functions s_j, $j = 1, ..., J+1$. If s_j is such that $\int\int |s_j(x,w)| dw\, dx < \infty$ and $\bar{\Psi}_{V,\lambda,s} \equiv \sum_{j=1}^{J} \int \Psi_{V_j,\lambda_j,s_j}(\zeta) d\zeta + |\sigma_{f_W,s}| < \infty$, where

$$\Psi_{V,\lambda,s}(\zeta) \equiv \frac{1}{|\theta_1(\zeta)|}\left(1+\frac{|\theta_{X_1}(\zeta)|}{|\theta_1(\zeta)|}\right)\int_{|\zeta|}^{\infty}|\sigma_{V,1,s}(\xi)||\xi|^{\lambda}d\xi + |\zeta|^{\lambda}\left(\frac{|\sigma_{V,1,s}(\zeta)|}{|\theta_1(\zeta)|}+|\sigma_{V,\chi_V,s}(\zeta)|+|\sigma_{V,f_W,s}(\zeta)|\right)$$

$$\sigma_{V,1,s}(\zeta) \equiv \int \exp(i\zeta x)\int s(x,w)\phi_V(\zeta,w)dwdx$$

$$\sigma_{V,\chi_V,s}(\zeta) \equiv \int \exp(i\zeta x)\lim_{h_2\to 0}\int \frac{1}{\chi_V(\zeta,w)}s(x,w)\phi_V(\zeta,w)ve^{i\zeta x_2}k_{h_2}(\tilde{w}-w)dwdx$$

$$\sigma_{V,f_W,s}(\zeta) \equiv \int \exp(i\zeta x)\lim_{h_2\to 0}\int \frac{1}{f_W(w)}s(x,w)\phi_V(\zeta,w)k_{h_2}(\tilde{w}-w)dwdx$$

$$\sigma_{f_W,s} \equiv \int \lim_{h_2\to 0}\int s_{J+1}(x,w)k_{h_2}(\tilde{w}-w)dwdx,$$

then, letting $\hat{g}(h_n) \equiv (\hat{g}_{V_1,\lambda_1}(\cdot,h_n), ..., \hat{g}_{V_J,\lambda_J}(\cdot,h_n))$ and $\hat{f}(h_{2n}) \equiv \hat{E}[k_{h_{2n}}(\cdot)]$,

$$b(\hat{g}(h_n),\hat{f}(h_{2n})) - b(g,f) = \hat{E}\left[\psi_s(V,X_1,X_2,W)\right] + o_p\left(n^{-1/2}\right),$$

where

$$\psi_s(v,x_1,x_2,\tilde{w}) \equiv \sum_{j=1}^{J}\psi_{V_j,\lambda_j}(s_j;v_j,x_1,x_2,\tilde{w}) + \psi_f(s_{J+1};\tilde{w})$$

and where

$$\begin{aligned}\psi_{V,\lambda}(s;v,x_1,x_2,\tilde{w}) \equiv \int \Big\{ & \Psi_{V,\lambda,1,s}(\zeta)\left(e^{i\zeta x_2}-E[e^{i\zeta X_2}]\right) \\ & + \Psi_{V,\lambda,X_1,s}(\zeta)\left(x_1 e^{i\zeta x_2}-E[X_1 e^{i\zeta X_2}]\right) \\ & + \left(\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta;v,x_2,\tilde{w}) - E[\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta;V,X_2,W)]\right) \\ & + \left(\mathcal{Z}_{V,\lambda,f_W}(s,\zeta;\tilde{w}) - E[\mathcal{Z}_{V,\lambda,f_W}(s,\zeta;W)]\right)\Big\} d\zeta\end{aligned}$$

$$\psi_f(s_{J+1};\tilde{w}) \equiv \int \lim_{h_2\to 0}\int s_{J+1}(x,w)\left(k_{h_2}(\tilde{w}-w)-E[k_{h_2}(W-w)]\right)dwdx,$$

with

$$\Psi_{V,\lambda,1,s}(\zeta) \equiv -\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_{\zeta}^{\pm\infty}\left(\int \exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^{\lambda} - \frac{1}{2\pi}\frac{(-i\zeta)^{\lambda}}{\theta_1(\zeta)}\left(\int \exp(-i\zeta x)\int s(x,w)\phi_V(\zeta,w)dwdx\right)$$

$$\Psi_{V,\lambda,X_1,s}(\zeta) \equiv \frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_{\zeta}^{\pm\infty}\left(\int \exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^{\lambda}d\xi$$

$$\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta;v,x_2,\tilde{w}) \equiv \frac{1}{2\pi}(-i\zeta)^{\lambda}\int \exp(-i\zeta x)\lim_{h_2\to 0}\int \frac{1}{\chi_V(\zeta,w)}s(x,w)\phi_V(\zeta,w)$$
$$\times\, ve^{i\zeta x_2}k_{h_2}(\tilde{w}-w)dwdx$$

$$\mathcal{Z}_{V,\lambda,f_W}(s,\zeta;\tilde{w}) \equiv -\frac{1}{2\pi}(-i\zeta)^{\lambda}\int \exp(-i\zeta x)\lim_{h_2\to 0}\int \frac{1}{f_W(w)}s(x,w)\phi_V(\zeta,w)$$
$$\times\, k_{h_2}(\tilde{w}-w)dwdx.$$

Moreover,

$$n^{1/2}(b(\hat{g}(h_n),\hat{f}(h_{2n}))-b(g,f)) \xrightarrow{d} N(0,\Omega_b),$$

where

$$\Omega_b \equiv E\left[(\psi_s(V,X_1,X_2,W))^2\right] < \infty.$$

1.4.3 Asymptotics for Average Marginal Effects

We now apply our previous general results to obtain the asymptotic properties of estimators of the objects of interest here. First, consider the plug-in estimator for the covariate-conditioned average marginal effect,

$$\hat{\beta}(x,w,h) \equiv \frac{\hat{g}_{Y,1}(x,w,h)}{\hat{g}_{1,0}(x,w,h)} - \frac{\hat{g}_{Y,0}(x,w,h)}{\hat{g}_{1,0}(x,w,h)}\frac{\hat{g}_{1,1}(x,w,h)}{\hat{g}_{1,0}(x,w,h)}$$

for each $(x,w)\in \text{supp}(X,W)$, where the nonparametric estimators $\hat{g}$ are as given above.

The results above and a straightforward Taylor expansion yield the following result.

Theorem 4.9 *Suppose the conditions of Theorem 4.4(ii) hold for* $\Lambda = 1$ *and that* $\max_{V=1,Y}\max_{\lambda=0,1}$
$\sup_{(x,w)\in\text{supp}(X,W)}|g_{V,\lambda}(x,w)| < \infty$. *Further, for* $\tau=\tau_n>0$, *define*

$$\Gamma_\tau \equiv \{(x,w)\in\mathbb{R}^2 : f_{X|W}(x\mid w)\geq \tau_n\}.$$

Then we have

$$\begin{aligned}
&\sup_{(x,w)\in\Gamma_\tau}\left|\hat{\beta}(x,w,h_n)-\beta(x,w)\right| \\
&= O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B}}\exp\left(\alpha_B\left(h_{1n}^{-1}\right)^{\beta_B}\right)\right) \\
&\quad + O_p\left(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_L},h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L}}\exp\left(\alpha_L\left(h_{1n}^{-1}\right)^{\beta_L}\right)\right),
\end{aligned}$$

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

$$\sup_{(x,w)\in\Gamma_\tau}\left|\hat{\beta}(x,w,h_n)-\beta(x,w)\right| = o_p(1).$$

The delta method gives us the next result.

Theorem 4.10 *Suppose the conditions of Corollary 4.6 hold for* $\Lambda = 1$ *and that* $\max_{V=1,Y}\max_{\lambda=0,1}\sup_{(x,w)\in\mathrm{supp}(X,W)}|g_{V,\lambda}(x,w)| < \infty$. *Then for all* $(x,w) \in \mathrm{supp}(X,W)$,

$$n^{1/2}\left(\Omega_\beta(x,w,h_n)\right)^{-1/2}\left(\hat{\beta}(x,w,h_n)-\beta(x,w)\right) \xrightarrow{d} N(0,1),$$

provided that

$$\Omega_\beta(x,w,h_n) \equiv E\left[\left(\ell_\beta(x,w,h_n;V,X_1,X_2,W)\right)^2\right]$$

is finite and positive for all n sufficiently large, where

$$\begin{aligned}
&\ell_\beta(x,w,h;v,x_1,x_2,\tilde{w}) \\
&= s_{Y,1}(x,w)\ell_{Y,1}(x,w,h;y,x_1,x_2,\tilde{w}) + s_{Y,0}(x,w)\ell_{Y,0}(x,w,h;y,x_1,x_2,\tilde{w}) \\
&\quad + s_{1,1}(x,w)\ell_{1,1}(x,w,h;1,x_1,x_2,\tilde{w}) + s_{1,0}(x,w)\ell_{1,0}(x,w,h;1,x_1,x_2,\tilde{w}),
\end{aligned}$$

where $\ell_{V,\lambda}$ *is as defined in Lemma 4.1,and*

$$\begin{aligned}
s_{Y,1}(x,w) &= \frac{1}{g_{1,0}(x,w)}, \\
s_{Y,0}(x,w) &= -\frac{g_{1,1}(x,w)}{g_{1,0}(x,w)}\frac{1}{g_{1,0}(x,w)}, \\
s_{1,1}(x,w) &= -\frac{g_{Y,0}(x,w)}{g_{1,0}(x,w)}\frac{1}{g_{1,0}(x,w)}, \\
s_{1,0}(x,w) &= \left(2\frac{g_{Y,0}(x,w)}{g_{1,0}(x,w)}\frac{g_{1,1}(x,w)}{g_{1,0}(x,w)} - \frac{g_{Y,1}(x,w)}{g_{1,0}(x,w)}\right)\frac{1}{g_{1,0}(x,w)}.
\end{aligned}$$

Because we are interested in weighted averages of $\beta(x,w)$ as well as $\beta(x,w)$ itself, we now consider the asymptotic properties of the following estimators of the weighted averages in eqns. (1.2)$\sim$(1.7):

$$\hat{\beta}_m(x,h_n) = \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)m(w)dw, \tag{1.11}$$

$$\hat{\beta}_{mf_W}(x,h_n) = \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)m(w)\hat{f}_W(w)dw, \tag{1.12}$$

$$\begin{aligned}
\hat{\beta}_{mf_{W|X}}(x,h_n) &= \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)m(w)\hat{f}_{W|X}(w \mid x)dw \qquad (1.13)\\
&= \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)m(w)\frac{\hat{g}_{1,0}(x,w,h_n)\hat{f}_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n)dw}dw,
\end{aligned}$$

$$\hat{\beta}_{\tilde{m}}(h_n) = \int_{S^{x,w}_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)\tilde{m}(x,w)dwdx, \tag{1.14}$$

$$\begin{aligned}
\hat{\beta}_{\tilde{m}f_{W|X}}(h_n) &= \int_{S^{x,w}_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)\tilde{m}(x,w)\hat{f}_{W|X}(w \mid x)dwdx \qquad (1.15)\\
&= \int_{S^{x,w}_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)\tilde{m}(x,w)\frac{\hat{g}_{1,0}(x,w,h_n)\hat{f}_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n)dw}dwdx,
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_{\tilde{m}f_{W,X}}(h_n) &= \int_{S^{x,w}_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)\tilde{m}(x,w)\hat{f}_{W,X}(w,x)dwdx \qquad (1.16)\\
&= \int_{S^{x,w}_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(x,w,h_n)\tilde{m}(x,w)\hat{g}_{1,0}(x,w,h_n)\hat{f}_W(w)dwdx,
\end{aligned}$$

where $S^w_{\hat{\beta}(\cdot,h_n)} \equiv \{w \in \mathbb{R} : \hat{g}_{1,0}(x,w,h_n) > 0\}$, $S^{x,w}_{\hat{\beta}(\cdot,h_n)} \equiv \{(x,w) \in \mathbb{R}^2 : \hat{g}_{1,0}(x,w,h_n) > 0\}$, and where $\hat{f}_W(w)$ is a nonparametric estimator of the density of W. The next assumption restricts the weight functions, m and $\tilde{m}$.

Assumption 4.12 *Let $\mathbb{M}$ and $\tilde{\mathbb{M}}$ be bounded measurable subsets of $\mathbb{R}$ and $\mathbb{R}^2$*, respectively. *(i) The weight functions $m : \mathbb{R} \to \mathbb{R}$ and $\tilde{m} : \mathbb{R}^2 \to \mathbb{R}$ are measurable and supported on $\mathbb{M}$ and $\tilde{\mathbb{M}}$, respectively*; *(ii)* $\inf_{(x,w)\in\tilde{\mathbb{M}}} f_{X|W}(x \mid w) > 0$; *(iii)* $\max_{V=1,Y} \max_{\lambda=0,1} \sup_{(x,w)\in\tilde{\mathbb{M}}} |g_{V,\lambda}(x,w)| < \infty$.

The next two theorems establish asymptotic properties for these estimators by applying Theorem 4.7 and 4.8. We first establish asymptotic results for the semiparametric functionals taking the forms of eqns. (1.11)∼(1.13) by applying Theorem 4.7.

Theorem 4.11 *Suppose the conditions of Theorem 4.7 hold for* $\Lambda = 1$ *and that Assumption 4.12 holds. Then* (i)

$$\begin{aligned}
&\sup_{x\in\mathbb{M}} |\hat{\beta}_m(x,h_n) - \beta_m(x)| \\
&= O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}} \exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad + O_p\left(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}} \exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\right),
\end{aligned}$$

(ii)

$$\begin{aligned}
&\sup_{x\in\mathbb{M}} |\hat{\beta}_{mf_W}(x,h_n) - \beta_{mf_W}(x)| \\
&= O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}} \exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad + O_p\left(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}} \exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\right),
\end{aligned}$$

and (iii)

$$\begin{aligned}
&\sup_{x\in\mathbb{M}} |\hat{\beta}_{mf_{W|X}}(x,h_n) - \beta_{mf_{W|X}}(x)| \\
&= O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}} \exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad + O_p\left(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}} \exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\right),
\end{aligned}$$

where $\alpha_{B,s}$, $\beta_{B,s}$, $\gamma_{\lambda,B,s}$, $\alpha_{L,s}$, $\beta_{L,s}$, $\gamma_{\lambda,L,s}$, *and* $\delta_{L,s}$ *are as defined in Theorem 4.7.*

The following theorem establishes asymptotic results for the semiparametric functionals taking the forms of eqns. (1.14)∼(1.16) by straightforward application of Theorem 4.8.

Theorem 4.12 *Suppose the conditions of Theorem 4.8 hold for* $\Lambda = 1$ *and that Assumption 4.12 holds. Then* (i)

$$n^{1/2}\left(\Omega_{\tilde{m}}\right)^{-1/2}\left(\hat{\beta}_{\tilde{m}}(h_n) - \beta_{\tilde{m}}\right) \xrightarrow{d} N(0,1),$$

provided that

$$\Omega_{\tilde{m}} \equiv E\left[(\psi_{\beta_{\tilde{m}}}(V, X_1, X_2, W))^2\right]$$

is finite and positive for all n *sufficiently large, where*

$$\psi_{\beta_{\tilde{m}}}(v, x_1, x_2, \tilde{w}) \equiv \sum_{V=1,Y} \sum_{\lambda=0,1} \psi_{V,\lambda}(\tilde{m}s_{V,\lambda}; v, x_1, x_2, \tilde{w}),$$

where $\tilde{m}s_{V,\lambda}$ *denotes the function mapping* (x, w) *to* $\tilde{m}(x, w)s_{V,\lambda}(x, w)$ *and where* $\psi_{V,\lambda}$ *is defined in Theorem 4.7;* (*ii*)

$$n^{1/2}\left(\Omega_{\tilde{m}f_{W|X}}\right)^{-1/2}\left(\hat{\beta}_{\tilde{m}f_{W|X}}(h_n) - \beta_{\tilde{m}f_{W|X}}\right) \xrightarrow{d} N(0, 1),$$

provided that

$$\Omega_{\tilde{m}f_{W|X}} \equiv E\left[(\psi_{\beta_{\tilde{m}f_{W|X}}}(V, X_1, X_2, W))^2\right]$$

is finite and positive for all n *sufficiently large, where*

$$\begin{aligned}\psi_{\beta_{\tilde{m}f_{W|X}}}(v, x_1, x_2, \tilde{w}) \equiv & \sum_{V=1,Y} \sum_{\lambda=0,1} \psi_{V,\lambda}(\tilde{m}f_{W|X}s_{V,\lambda}; v, x_1, x_2, \tilde{w}) \\ & + \psi_{1,0}(P_1; 1, x_1, x_2, \tilde{w}) - \psi_{1,0}(P_2; 1, x_1, x_2, \tilde{w}) + \psi_f(P_3; \tilde{w}),\end{aligned}$$

and where $\tilde{m}f_{W|X}s_{V,\lambda}$, P_1, P_2, *and* P_3 *denote the functions mapping* (x, w) *to* $\tilde{m}(x, w)f_{W|X}(w \mid x)s_{V,\lambda}(x, w)$, $\beta(x, w)\tilde{m}(x, w)f_W(w)/f_X(x)$, $\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x, w)$ $\tilde{m}(x, w)f_{W|X}(w \mid x)dw/f_X(x)$, *and* $\beta(x, w)\tilde{m}(x, w)f_{X|W}(x \mid w)/f_X(x)$, *respectively;* (*iii*)

$$n^{1/2}\left(\Omega_{\tilde{m}f_{W,X}}\right)^{-1/2}\left(\hat{\beta}_{\tilde{m}f_{W,X}}(h_n) - \beta_{\tilde{m}f_{W,X}}\right) \xrightarrow{d} N(0, 1),$$

provided that

$$\Omega_{\tilde{m}f_{W,X}} \equiv E\left[(\psi_{\beta_{\tilde{m}f_{W,X}}}(V, X_1, X_2, W))^2\right]$$

is finite and positive for all n *sufficiently large, where*

$$\begin{aligned}\psi_{\beta_{\tilde{m}f_{W,X}}}(v, x_1, x_2, \tilde{w}) \equiv & \sum_{V=1,Y} \sum_{\lambda=0,1} \psi_{V,\lambda}(\tilde{m}f_{W,X}s_{V,\lambda}; v, x_1, x_2, \tilde{w}) \\ & + \psi_{1,0}(\beta\tilde{m}f_W; 1, x_1, x_2, \tilde{w}) + \psi_f(\beta\tilde{m}f_{X|W}; \tilde{w}),\end{aligned}$$

where $\tilde{m}f_{W,X}s_{V,\lambda}$, $\beta\tilde{m}f_W$, *and* $\beta\tilde{m}f_{X|W}$ *denote the functions mapping* (x, w) *to* $\tilde{m}(x, w)f_{W,X}(w, x)s_{V,\lambda}(x, w)$, $\beta(x, w)\tilde{m}(x, w)f_W(w)$, *and* $\beta(x, w)\tilde{m}(x, w)f_{X|W}(x \mid w)$, *respectively.*

1.5 Monte Carlo Simulations

This section investigates the finite-sample properties of the proposed estimator through various Monte Carlo experiments. We consider the following nonseparable data generating process:

$$Y = f_1(X)U_y, \quad X = 0.5W + U_x, \quad U_y = f_2(W) + U_u,$$
$$X_1 = X + U_1, \quad X_2 = X + U_2,$$

where the distributions of each random variable and the explicit forms of f_1, f_2 are specified below and where Y, W, X_1, and X_2 are standardized to have mean zero and standard deviation one. We assume $U_x \perp U_u \mid W$ which implies $X \perp U_y \mid W$.[6] The variables (Y, X_1, X_2, W) are used as an input for our estimator, and the variables (Y, X_1, W) are used for the local linear estimator that neglects the measurement error. We also use the variables (Y, X, W) to construct an infeasible local linear estimator, and (Y, X, X_2, W) and (Y, X_1, X, W) to construct infeasible versions of our estimator for purposes of comparison. For those estimators, we consider flat-top kernels of infinite order. In our estimators[7], the expression of Fourier transform, $\kappa(\cdot)$, associated with X is given in eqn.(1.8) with $\bar{\xi} = .5$. We use a different flat-top kernel for W, which is introduced in Politis and Romano (1999): $k_h(x) \equiv \frac{h}{2\pi}\frac{\sin^2(2\pi x/h) - \sin^2(\pi x/h)}{\pi^2 x^2}$. All estimates are constructed at values $x = 0$ and $w = 1$. For our estimators, we scan a set of bandwidths[8] ranging from 7 to 12.5 for X and from 3.5 to 6 for W in increments of 0.05 in order to find the optimal bandwidth minimizing the root mean square error (RMSE). For both local linear estimators, we scan a set of bandwidths ranging from 2.5 to 6 for X and from 1.5 to 3.5 for W, with the same increments. All simulations draw 500 samples of $1,000$, $2,000$, or $8,000$ observations.

We examine a total 16 combinations of ordinary and supersmooth distributions for random variables and functions f_1 and f_2, as given in Table 1.1. As

[6] In the simulations, we assume $U_x \perp (U_u, W)$ which implies $U_x \perp U_u \mid W$ by Lemma 4.3 of Dawid (1979). Lemma 4.1 of Dawid then ensures that $U_x \perp U_u \mid W$ implies $X \perp U_y \mid W$.

[7] For the local linear estimator the same flat-top kernel is used for X and W since estimation results are not sensitive to the choice of the kernel.

[8] Note that the flat-top kernel has a very narrow central peak, so that even moderately large bandwidths result in highly local smoothing.

in Schennach (2004a), we also consider the Laplace distribution as an example of an ordinarily smooth distribution. The Laplace distribution density, denoted by $L(t; \mu, \sigma^2)$, is defined by

$$\frac{1}{\sigma\sqrt{2}} \exp\left(-\sigma|t-\mu|\sqrt{2}\right)$$

for any $t \in \mathbb{R}$ with mean μ and variance σ^2. Its characteristic function has a tail of the form $|\zeta|^{-2}$. The normal distribution with variance σ^2 is used as an example of a supersmooth distribution. The tail of the characteristic function of the normal distribution is of the form $\exp(-(\sigma^2/2)|\zeta|^2)$. Our example of an ordinarily smooth function for $f_2(W)$ is a piecewise linear continuous function with a discontinuous first derivative

$$S(W) \equiv \begin{cases} -1 & \text{if} \quad W < -1 \\ W & \text{if} \quad W \in [-1, 1] \\ 1 & \text{if} \quad W > 1, \end{cases}$$

whose Fourier transform decays at the rate $|\zeta|^{-2}$ as $|\zeta| \to \infty$. As an example of a supersmooth function for $f_1(X)$ or $f_2(W)$, we consider the error function

$$\text{erf}(V) \equiv \frac{2}{\sqrt{\pi}} \int_0^V e^{-t^2} dt$$

having a Fourier transform decaying at the rate $|\zeta|^{-1} \exp(-\frac{1}{4}|\zeta|^2)$ as $|\zeta| \to \infty$ for $V = X$ or W.

Table 1.2 $\sim$ 1.6 reports the bias squared, variance, and RMSE of the five estimators, which are functions of bandwidth for a sample size of $1,000$, for example 1.[9] Fourier 1, 2 and 3 refer to our estimators which are based on variables (Y, X_1, X_2, W), (Y, X, X_2, W), and (Y, X_1, X, W), respectively. Local linear without correction and local linear without errors refer to local linear estimators which use variables (Y, X_1, W) and (Y, X, W), respectively. We show results from only a subset of the bandwidths for conciseness. For each choice of bandwidths, the bias squared, variance, and RMSE are reported in the first, second, and third row, respectively. The results from the optimal bandwidth are reported at the bottom of each estimator.

[9] We only report this example due to space limitations, but results from all examples give similar messages on the performance of the estimators.

A few remarks are in order. It is shown that our estimator is as effective in reducing bias as the infeasible local linear estimator using the true covariate X is. However, the bias from the local linear estimator ignoring the measurement error does not shrink toward zero as bandwidth decreases. Our estimator also gives smaller variance than the local linear estimator based on error-contaminated covariates. As a result, our estimator outperforms the local linear estimator in terms of RMSE. All Fourier estimators perform better than the infeasible local linear estimator. So it would be interesting to investigate under what conditions and why Fourier-based estimators outperform local linear estimators. By comparing among Fourier estimators, we can see the role of clean data as well as the asymmetry between two measurement errors in Assumption 3.2. Interestingly, Fourier 1 and Fourier 2 obtain quite similar estimation results, but Fourier 3 outperforms these estimators. So one would want to use more clean one for X_2 among two error-laden observations in order to get better estimation results.

Table 1.7 reports Monte Carlo simulation results for the convergence rate as a function of sample size for each example. RMSE's in all examples decrease as sample size increases, corroborating our theoretical results.

1.6 Summary and Concluding Remarks

We examine the identification and estimation of covariate-conditioned average marginal effects in a nonseparable data generating process with an endogenous and mismeasured cause of interest. We use conditioning instruments to ensure the conditional independence between the cause of interest and other unobservable drivers, permitting identification of causal effects of interest. Although the endogenous cause of interest is unobserved, two error-laden measurements are available. We extend methods of the deconvolution literature for nonlinear measurement errors to obtain estimates of the distribution functions of the underlying cause of interest from its error-laden measurements and to recover parameters of interest. These parameters include covariate-conditioned average marginal effects and weighted averages of these. We obtain uniform convergence rates and asymptotic normality for estimators of covariate-conditioned average marginal effects, faster convergence rates for estimators of their weighted averages over conditioning instruments, and $\sqrt{n}$ consistency and asymptotic normality for estimators of their weighted averages over conditioning instruments and causes.

1.7 Acknowledgements

Chapter 1 is coauthored with Susanne M. Schennach and Halbert L. White.

1.8 Mathematical Appendix

Proof of Lemma 3.1 By Assumption 3.1, all expectations below exist and are finite. We first observe that $U_2 \perp (X, W)$ implies $U_2 \perp X$ and $U_2 \perp W$. Given Assumptions 2.3, 3.2 and 3.4, we get

$$\begin{aligned}
\frac{iE[X_1 e^{i\xi X_2}]}{E[e^{i\xi X_2}]} &= \frac{iE[Xe^{i\xi(X+U_2)}] + iE[U_1 e^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]} \\
&= \frac{iE[Xe^{i\xi(X+U_2)}] + iE[E(U_1 e^{i\xi(X+U_2)} \mid X, U_2)]}{E[e^{i\xi(X+U_2)}]} \\
&= \frac{iE[Xe^{i\xi(X+U_2)}] + iE[E(U_1 \mid X, U_2) e^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]} \\
&= \frac{iE[Xe^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]} \\
&= \frac{iE[Xe^{i\xi X}]E[e^{i\xi U_2}]}{E[e^{i\xi X}]E[e^{i\xi U_2}]} \\
&= \frac{iE[Xe^{i\xi X}]}{E[e^{i\xi X}]} \\
&= D_\xi \ln(E[e^{i\xi X}]),
\end{aligned} \tag{1.17}$$

as considered by SWC. We use $E[U_1 \mid X, U_2] = 0$ in the step from the third to the fourth equality and use $U_2 \perp X$ in the step from the fourth to the fifth equality.

We note that $U_2 \perp (X, W)$ if and only if $U_2 \perp W$ and $U_2 \perp X \mid W$ because $f(U_2, X, W) = f(U_2, X \mid W)f(W) = f(U_2 \mid W)f(X \mid W)f(W) = f(U_2)f(X \mid W)f(W) = f(U_2)f(X, W)$. And we note that $U_2 \perp (X, W) \mid W$ if and only if $U_2 \perp X \mid W$. The 'only if' part of the assertion follows immediately because $U_2 \perp (X, W) \mid W$ implies $U_2 \perp X \mid W$ and $U_2 \perp W \mid W$. The 'if' part can be proven by the fact that $U_2 \perp X \mid W$ if and only if $(U_2, W) \perp (X, W) \mid W$ from Lemma 4.1 in Dawid (1979) and by the fact that if $(U_2, W) \perp (X, W) \mid W$, then $U_2 \perp (X, W) \mid W$ from Lemma 4.2(ii)

in Dawid (1979). Then for each real ζ, we have

$$\begin{aligned}
\phi_V(\zeta, W) &\equiv E[Ve^{i\zeta X} \mid W] \\
&= \frac{E[Ve^{i\zeta X} \mid W]E[e^{i\zeta U_2}]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[Ve^{i\zeta X} \mid W]E[e^{i\zeta U_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[E[Ve^{i\zeta X} \mid X, W] \mid W]E[e^{i\zeta U_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}]
\end{aligned}$$

$$\begin{aligned}
&= \frac{E[E[V \mid X, W]e^{i\zeta X} \mid W]E[e^{i\zeta U_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[E[V \mid X, W]e^{i\zeta X}e^{i\zeta U_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[E[V \mid X, U_2, W]e^{i\zeta X}e^{i\zeta U_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[E[Ve^{i\zeta X}e^{i\zeta U_2} \mid X, U_2, W] \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[E[Ve^{i\zeta X_2} \mid X, U_2, W] \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[Ve^{i\zeta X_2} \mid W]}{E[e^{i\zeta X}]E[e^{i\zeta U_2}]}E[e^{i\zeta X}] \\
&= \frac{E[Ve^{i\zeta X_2} \mid W]}{E[e^{i\zeta X_2}]} \exp\left(\ln(E[e^{i\zeta X}]) - \ln 1\right) \\
&= \frac{E[Ve^{i\zeta X_2} \mid W]}{E[e^{i\zeta X_2}]} \exp\left(\int_0^\zeta D_\xi \ln(E[e^{i\zeta X}])d\xi\right) \\
&= \frac{E[Ve^{i\zeta X_2} \mid W]}{E[e^{i\zeta X_2}]} \exp\left(\int_0^\zeta \frac{iE[X_1e^{i\xi X_2}]}{E[e^{i\xi X_2}]}d\xi\right),
\end{aligned}$$

where $U_2 \perp W$, $U_2 \perp (X, W) \mid W$ and $E[V \mid X, U_2, W] = E[V \mid X, W]$ are used in the steps from the second to the third line, from the fifth to the sixth line, and from the sixth to the seventh line, respectively.

Given Assumptions 3.3 - 3.5, integral by parts gives

$$\begin{aligned}
(-i\zeta)^\lambda E[Ve^{i\zeta X} \mid W = w] &= (-i\zeta)^\lambda \int E[V \mid W = w, X = x]f_{X|W}(x \mid w)e^{i\zeta x}dx \\
&= (-1)^\lambda \int E[V \mid W = w, X = x]f_{X|W}(x \mid w)D_x^\lambda e^{i\zeta x}dx \\
&= \int D_x^\lambda(E[V \mid W = w, X = x]f_{X|W}(x \mid w))e^{i\zeta x}dx \\
&= \int g_{V,\lambda}(x, w)e^{i\zeta x}dx.
\end{aligned}$$

The last expression is the Fourier transform of $g_{V,\lambda}(x,w)$. For each $\lambda \in \{0,...,\Lambda\}$ and $(x,w) \in \mathrm{supp}(X,W)$, we have

$$\frac{1}{2\pi}\int(-i\zeta)^{\lambda}\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}E[Ve^{i\zeta X} \mid W=w]\exp(-i\zeta x)d\zeta.$$

Since the right hand side is the inverse Fourier transform of $(-i\zeta)^{\lambda}E[Ve^{i\zeta X} \mid W=w]$, the result follows. $\square$

Proof of Lemma 3.2 Assumptions 3.1, 3.3 - 3.5, and 3.6 ensure the existence of

$$\begin{aligned} g_{V,\lambda}(x,w,h_1) &\equiv \int \frac{1}{h_1}k\left(\frac{\tilde{x}-x}{h_1}\right)g_{V,\lambda}(\tilde{x},w)d\tilde{x} \\ &= \int \frac{1}{h_1}k\left(\frac{\tilde{x}-x}{h_1}\right)D_{\tilde{x}}^{\lambda}(E[V \mid X=\tilde{x},W=w]f_{X|W}(\tilde{x} \mid w))d\tilde{x}. \end{aligned}$$

By the convolution theorem, the inverse Fourier Transform of the product of $\kappa(h_1\zeta)$ and $(-i\zeta)^{\lambda} \times$
$E[Ve^{i\zeta X} \mid W=w]$ is the convolution between the inverse Fourier Transform of $\kappa(h\zeta)$ and the inverse Fourier Transform of $(-i\zeta)^{\lambda}E[Ve^{i\zeta X} \mid W=w]$. The inverse Fourier Transform of $\kappa(h_1\zeta)$ is $h_1^{-1}k(x/h_1)$, and the inverse Fourier Transform of $(-i\zeta)^{\lambda}E[Ve^{i\zeta X} \mid W=w]$ is $D_x^{\lambda}(E[V \mid X=x,W=w]f_{X|W}(x \mid w))$. It follows that

$$\begin{aligned} g_{V,\lambda}(x,w,h_1) &= \frac{1}{2\pi}\int \kappa(h_1\zeta)\left((-i\zeta)^{\lambda}E[Ve^{i\zeta X} \mid W=w]\right)\exp(-i\zeta x)d\zeta \\ &= \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta. \ \square \end{aligned}$$

Proof of Lemma 4.1 For $A = 1, X_1$, we let $\theta_A(\zeta) \equiv E\left[Ae^{i\zeta X_2}\right]$ and for $V = 1, Y$,

$$\begin{aligned} \theta_V(\zeta,w) &\equiv E\left[Ve^{i\zeta X_2} \mid W=w\right] \\ &= \int\int ve^{i\zeta x_2}f_{V,X_2|W}(v,x_2 \mid w)dvdx_2 \\ &= \frac{\chi_V(\zeta,w)}{f_W(w)}, \end{aligned}$$

where $\chi_V(\zeta,w) \equiv \int\int ve^{i\zeta x_2}f_{V,X_2,W}(v,x_2,w)dvdx_2$, $f_{V,X_2|W}(v,x_2 \mid w)$ is the conditional density of (V,X_2) given $W=w$, and $f_{V,X_2,W}(v,x_2,w)$ is the joint density of (V,X_2,W).

Also we let $\hat{\theta}_A(\zeta) \equiv \hat{E}\left[Ae^{i\zeta X_2}\right]$ and $\delta\hat{\theta}_A(\zeta) \equiv \hat{\theta}_A(\zeta) - \theta_A(\zeta)$. Similarly $\hat{\theta}_V(\zeta, w) \equiv \hat{E}\left[Ve^{i\zeta X_2} \mid W = w\right] \equiv \hat{\chi}_V(\zeta, w)/\hat{f}_W(w)$ where

$$\hat{\chi}_V(\zeta, w) = \frac{1}{n}\sum_{j=1}^{n} k_{h_2}(W_j - w)V_j e^{i\zeta X_{2j}} = \hat{E}\left[Ve^{i\zeta X_2}k_{h_2}(W - w)\right]$$
$$\hat{f}_W(w) = \frac{1}{n}\sum_{j=1}^{n} k_{h_2}(W_j - w) = \hat{E}\left[k_{h_2}(W - w)\right]$$

so that $\delta\hat{\chi}_V(\zeta, w) \equiv \hat{\chi}_V(\zeta, w) - \chi_V(\zeta, w)$ and $\delta\hat{f}_W(w) \equiv \hat{f}_W(w) - f_W(w)$. As used in Schennach (2004a, b) and SWC, we state a useful representation for $\hat{\theta}_{X_1}(\zeta)/\hat{\theta}_1(\zeta)$:

$$\frac{\hat{\theta}_{X_1}(\zeta)}{\hat{\theta}_1(\zeta)} = \frac{\theta_{X_1}(\zeta) + \delta\hat{\theta}_{X_1}(\zeta)}{\theta_1(\zeta) + \delta\hat{\theta}_1(\zeta)} = q_{X_1}(\zeta) + \delta\hat{q}_{X_1}(\zeta) \tag{1.18}$$

where $q_{X_1}(\zeta) = \theta_{X_1}(\zeta)/\theta_1(\zeta)$ and where $\delta\hat{q}_{X_1}(\zeta)$ can be written as either

$$\delta\hat{q}_{X_1}(\zeta) = \left(\frac{\delta\hat{\theta}_{X_1}(\zeta)}{\theta_1(\zeta)} - \frac{\theta_{X_1}(\zeta)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1}$$

or $\delta\hat{q}_{X_1}(\zeta) = \delta_1\hat{q}_{X_1}(\zeta) + \delta_2\hat{q}_{X_1}(\zeta)$ with

$$\delta_1\hat{q}_{X_1}(\zeta) \equiv \frac{\delta\hat{\theta}_{X_1}(\zeta)}{\theta_1(\zeta)} - \frac{\theta_{X_1}(\zeta)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}$$
$$\delta_2\hat{q}_{X_1}(\zeta) \equiv \frac{\theta_{X_1}(\zeta)}{\theta_1(\zeta)}\left(\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^2\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1} - \frac{\delta\hat{\theta}_{X_1}(\zeta)}{\theta_1(\zeta)}\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1}.$$

For $\hat{\chi}_V(\zeta, w)/\hat{\theta}_1(\zeta)$,

$$\frac{\hat{\chi}_V(\zeta, w)}{\hat{\theta}_1(\zeta)} = \frac{\chi_V(\zeta, w) + \delta\hat{\chi}_V(\zeta, w)}{\theta_1(\zeta) + \delta\hat{\theta}_1(\zeta)} = q_V(\zeta, w) + \delta\hat{q}_V(\zeta, w) \tag{1.19}$$

where $q_V(\zeta, w) \equiv \chi_V(\zeta, w)/\theta_1(\zeta)$ and where $\delta\hat{q}_V(\zeta, w)$ can be written as either

$$\delta\hat{q}_V(\zeta, w) = \left(\frac{\delta\hat{\chi}_V(\zeta, w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta, w)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1}$$

or $\delta\hat{q}_V(\zeta, w) = \delta_1\hat{q}_V(\zeta, w) + \delta_2\hat{q}_V(\zeta, w)$ with

$$\delta_1\hat{q}_V(\zeta, w) \equiv \frac{\delta\hat{\chi}_V(\zeta, w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta, w)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}$$
$$\delta_2\hat{q}_V(\zeta, w) \equiv \frac{\chi_V(\zeta, w)}{\theta_1(\zeta)}\left(\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^2\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1} - \frac{\delta\hat{\chi}_V(\zeta, w)}{\theta_1(\zeta)}\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1}.$$

Similarly for $1/\hat{f}_W(w)$,

$$\frac{1}{\hat{f}_W(w)} = \frac{1}{f_W(w) + \delta\hat{f}_W(w)} = q_1(w) + \delta\hat{q}_1(w) \tag{1.20}$$

where $q_1(w) \equiv 1/f_W(w)$ and where $\delta\hat{q}_1(w)$ can be written as either

$$\delta\hat{q}_1(w) = \left(-\frac{\delta\hat{f}_W(w)}{(f_W(w))^2}\right)\left(1 + \frac{\delta\hat{f}_W(w)}{f_W(w)}\right)^{-1}$$

or $\delta\hat{q}_1(w) = \delta_1\hat{q}_1(w) + \delta_2\hat{q}_1(w)$ with

$$\begin{aligned}\delta_1\hat{q}_1(w) &\equiv -\frac{\delta\hat{f}_W(w)}{(f_W(w))^2}\\ \delta_2\hat{q}_1(w) &\equiv \frac{1}{f_W(w)}\left(\frac{\delta\hat{f}_W(w)}{f_W(w)}\right)^2\left(1 + \frac{\delta\hat{f}_W(w)}{f_W(w)}\right)^{-1}.\end{aligned}$$

For $Q_{X_1}(\zeta) \equiv \int_0^\zeta (i\theta_{X_1}(\xi)/\theta_1(\xi))d\xi$, $\delta\hat{Q}_{X_1}(\zeta) \equiv \int_0^\zeta (i\hat{\theta}_{X_1}(\xi)/\hat{\theta}_1(\xi))d\xi - Q_{X_1}(\zeta)$ and some random function $\delta\bar{Q}_{X_1}(\zeta)$ such that $|\delta\bar{Q}_{X_1}(\zeta)| \leq |\delta\hat{Q}_{X_1}(\zeta)|$ for all ζ,

$$\begin{aligned}&\exp\left(Q_{X_1}(\zeta) + \delta\hat{Q}_{X_1}(\zeta)\right) \qquad (1.21)\\ &= \exp(Q_{X_1}(\zeta))\left(1 + \delta\hat{Q}_{X_1}(\zeta) + \frac{1}{2}\left[\exp(\delta\bar{Q}_{X_1}(\zeta))\right]\left(\delta\hat{Q}_{X_1}(\zeta)\right)^2\right).\end{aligned}$$

By substituting eqn.(1.18)$\sim$(1.21) into

$$\begin{aligned}&\hat{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1)\\ &= \frac{1}{2\pi}\int(-i\zeta)^\lambda\kappa(h_1\zeta)\exp(-i\zeta x)\\ &\times\left[\frac{\hat{\theta}_V(\zeta,w)}{\hat{\theta}_1(\zeta)}\exp\left(\int_0^\zeta\frac{i\hat{\theta}_{X_1}(\xi)}{\hat{\theta}_1(\zeta)}d\xi\right) - \frac{\theta_V(\zeta,w)}{\theta_1(\zeta)}\exp\left(\int_0^\zeta\frac{i\theta_{X_1}(\xi)}{\theta_1(\zeta)}d\xi\right)\right]d\zeta,\end{aligned}$$

we have

$$\begin{aligned}&\hat{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1)\\ =&\frac{1}{2\pi}\int(-i\zeta)^\lambda\kappa(h_1\zeta)\exp(-i\zeta x)\Bigg[-\frac{\theta_V(\zeta,w)}{\theta_1(\zeta)}\exp\left(\int_0^\zeta\frac{i\theta_{X_1}(\xi)}{\theta_1(\xi)}d\xi\right)\\ &+\left\{\frac{\chi_V(\zeta,w)}{\theta_1(\zeta)} + \frac{\delta\hat{\chi}_V(\zeta,w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta,w)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2} + \delta_2\hat{q}_V(\zeta,w)\right\}\\ &\times\left\{\frac{1}{f_W(w)} - \frac{\delta\hat{f}_W(w)}{(f_W(w))^2} + \delta_2\hat{q}_1(w)\right\}\exp(Q_{X_1}(\zeta))\\ &\times\left\{1 + \int_0^\zeta i\delta_1\hat{q}_{X_1}(\xi)d\xi + \int_0^\zeta i\delta_2\hat{q}_{X_1}(\xi)d\xi + \frac{1}{2}\exp(\delta\bar{Q}_{X_1}(\zeta))\left(\int_0^\zeta i\delta\hat{q}_{X_1}(\xi)d\xi\right)^2\right\}\Bigg]d\zeta.\end{aligned}$$

Keeping the terms linear in $\delta\hat{\theta}_1(\zeta)$, $\delta\hat{\theta}_{X_1}(\zeta)$, $\delta\hat{\chi}_V(\zeta,w)$, and $\delta\hat{f}_W(w)$ gives the linearization of $\hat{g}_{V,\lambda}(x,w,h)$, denoted $\bar{g}_{V,\lambda}(x,w,h)$:

$$
\begin{aligned}
&\bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1)\\
=&\frac{1}{2\pi}\int(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\left[\frac{\theta_V(\zeta,w)}{\theta_1(\zeta)}\exp(Q_{X_1})\int_0^\zeta\left(\frac{i\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)} - \frac{i\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right)d\xi\right.\\
&-\exp(Q_{X_1}(\zeta))\frac{\chi_V(\zeta,w)}{\theta_1(\zeta)}\frac{\delta\hat{f}_W(w)}{(f_W(w))^2}\\
&\left.+\exp(Q_{X_1}(\zeta))\frac{1}{f_W(w)}\left(\frac{\delta\hat{\chi}_V(\zeta,w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta,w)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)\right]d\zeta
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{2\pi}\int(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\phi_V(\zeta,w)\int_0^\zeta\left(\frac{i\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)} - \frac{i\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right)d\xi\\
&+\frac{1}{2\pi}\int(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\phi_V(\zeta,w)\left(-\frac{\delta\hat{f}_W(w)}{f_W(w)} + \frac{\delta\hat{\chi}_V(\zeta,w)}{\chi_V(\zeta,w)} - \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)d\zeta.
\end{aligned}
$$

Using the identity

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_0^\zeta f(\zeta,\xi)d\xi d\zeta &= \int_0^\infty\int_\xi^\infty f(\zeta,\xi)d\zeta d\xi + \int_{-\infty}^0\int_\xi^{-\infty} f(\zeta,\xi)d\zeta d\xi\\
&\equiv \int\int_\xi^{\pm\infty} f(\zeta,\xi)d\zeta d\xi
\end{aligned}
$$

for any absolutely integrable function f, we get

$$
\begin{aligned}
& L_{V,\lambda}(x,w,h) \\
\equiv & \bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1) \\
= & \frac{1}{2\pi}\int\int_{\xi}^{\pm\infty}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\phi_V(\zeta,w)d\zeta\left(\frac{i\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)} - \frac{i\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right)d\xi \\
& + \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\phi_V(\zeta,w)\left(\frac{\delta\hat{\chi}_V(\zeta,w)}{\chi_V(\zeta,w)} - \frac{\delta\hat{f}_W(w)}{f_W(w)} - \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)d\zeta \\
= & \int\Bigg[\Bigg\{-\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_{\zeta}^{\pm\infty}(-i\xi)^{\lambda}\kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi \\
& - \frac{1}{2\pi}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\theta_1(\zeta)}\Bigg\}\delta\hat{\theta}_1(\zeta) \\
& + \Bigg\{\frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_{\zeta}^{\pm\infty}(-i\xi)^{\lambda}\kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi\Bigg\}\delta\hat{\theta}_{X_1}(\zeta) \\
& + \Bigg\{\frac{1}{2\pi}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)}\Bigg\}\delta\hat{\chi}_V(\zeta,w) \\
& + \Bigg\{-\frac{1}{2\pi}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{f_W(w)}\Bigg\}\delta\hat{f}_W(w)\Bigg]d\zeta \\
= & \int\Bigg[\Psi_{V,\lambda,1}(\zeta,x,w,h_1)\left(\hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}]\right) \\
& + \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)\left(\hat{E}[X_1e^{i\zeta X_2}] - E[X_1e^{i\zeta X_2}]\right) \\
& + \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)\left(\hat{E}[Ve^{i\zeta X_2}k_{h_2}(W-w)] - E[Ve^{i\zeta X_2}k_{h_2}(W-w)]\right) \\
& + \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)\left(\hat{E}[k_{h_2}(W-w)] - E[k_{h_2}(W-w)]\right)\Bigg]d\zeta \\
= & \hat{E}\Bigg[\int\Psi_{V,\lambda,1}(\zeta,x,w,h_1)\left(e^{i\zeta X_2} - E[e^{i\zeta X_2}]\right)d\zeta \\
& + \int\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)\left(X_1e^{i\zeta X_2} - E[X_1e^{i\zeta X_2}]\right)d\zeta \\
& + \int\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)\left(Ve^{i\zeta X_2}k_{h_2}(W-w) - E[Ve^{i\zeta X_2}k_{h_2}(W-w)]\right)d\zeta \\
& + \int\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)\left(k_{h_2}(W-w) - E[k_{h_2}(W-w)]\right)d\zeta\Bigg] \\
= & \hat{E}\left[\ell_{V,\lambda}(x,w,h;V,X_1,X_2,W)\right]
\end{aligned}
$$

where $\Psi_{V,\lambda,A}(\zeta,x,w,h_1)$ and $\ell_{V,\lambda}(x,w,h;V,X_1,X_2,W)$ are defined in the statement of the Lemma 4.1. □

We define the following convenient notation as employed in SWC.

Definition A.1 *We write $f(\zeta) \preceq g(\zeta)$ for $f, g : \mathbb{R} \mapsto \mathbb{R}$ when there exists a constant $C > 0$, independent of ζ, such that $f(\zeta) \leq Cg(\zeta)$ for all $\zeta \in \mathbb{R}$ (and similarly for $\succeq$). Analogously, we write $a_n \preceq b_n$ for two sequences a_n, b_n when there exists a constant C independent of n such that $a_n \leq Cb_n$ for all $n \in \mathbb{N}$.*

Proof of Theorem 4.2 Using Parseval's identity, we have

$$
\begin{aligned}
&|B_{V,\lambda}(x,w,h_1)| \\
&= |g_{V,\lambda}(x,w,h_1) - g_{V,\lambda}(x,w)| \\
&= |g_{V,\lambda}(x,w,h_1) - g_{V,\lambda}(x,w,0)| \\
&= \left| \frac{1}{2\pi} \int \kappa(h_1\zeta)(-i\zeta)^\lambda \phi_V(\zeta,w) \exp(-i\zeta x) d\zeta - \frac{1}{2\pi} \int (-i\zeta)^\lambda \phi_V(\zeta,w) \exp(-i\zeta x) d\zeta \right| \\
&= \left| \frac{1}{2\pi} \int (\kappa(h_1\zeta) - 1)(-i\zeta)^\lambda \phi_V(\zeta,w) \exp(-i\zeta x) d\zeta \right| \\
&\leq \frac{1}{2\pi} \int |(\kappa(h_1\zeta) - 1)| \, |\zeta|^\lambda \, |\phi_V(\zeta,w)| \, d\zeta \\
&= \frac{1}{\pi} \int_{\bar{\xi}/h_1}^{\infty} |(\kappa(h_1\zeta) - 1)| \, |\zeta|^\lambda \, |\phi_V(\zeta,w)| \, d\zeta \\
&\preceq \int_{\bar{\xi}/h_1}^{\infty} |\zeta|^\lambda \, |\phi_V(\zeta,w)| \, d\zeta,
\end{aligned}
$$

since Assumption 3.6 ensures $\kappa(\zeta) = 1$ for $|\zeta| \leq \bar{\xi}$ and $\sup_\zeta |\kappa(h_1\zeta)| < \infty$. Thus, by Assumption 4.1(ii), we have

$$
\begin{aligned}
\sup_{(x,w)\in\mathrm{supp}(X,W)} |B_{V,\lambda}(x,w,h_1)| &\preceq \int_{\bar{\xi}/h_1}^{\infty} |\zeta|^\lambda C_\phi (1+|\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) d\zeta \\
&\preceq \int_{\bar{\xi}/h_1}^{\infty} |\zeta|^\lambda (1+|\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) d\zeta \\
&= O\left(\left(\bar{\xi}/h_1\right)^{\gamma_\phi+\lambda+1} \exp\left(\alpha_\phi \left(\bar{\xi}/h_1\right)^{\beta_\phi} \right) \right) \\
&= O\left(\left(h_1^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h_1^{-1}\right)^{\beta_B} \right) \right). \ \square
\end{aligned}
$$

Lemma A.1 *Suppose the conditions of Lemma 4.1 hold. For each ζ and $h \equiv (h_1, h_2)$, and for $A = 1, X_1, \chi_V, f_W$, let $\Psi^+_{V,\lambda,A}(\zeta, h_1) \equiv \sup_{(x,w)\in\mathrm{supp}(X,W)} |\Psi_{V,\lambda,A}(\zeta, x, w, h_1)|$, and define*

$$
\Psi^+_{V,\lambda}(h) \equiv \sum_{A=1,X_1} \int \Psi^+_{V,\lambda,A}(\zeta,h_1) d\zeta + h_2^{-1} \sum_{B=\chi_V, f_W} \int \Psi^+_{V,\lambda,B}(\zeta,h_1) d\zeta.
$$

If Assumption 4.1 also holds, then for $h > 0$

$$\begin{aligned}
&\Psi^+_{V,\lambda}(h)\\
&= O\left(\max\{(1+h_1^{-1})^{\gamma_1+1}, h_2^{-1}\}\left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda-\gamma_\theta+1}\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)(h_1^{-1})^{\beta_\theta}\right)\right).
\end{aligned}$$

Proof We obtain rates for each term of $\Psi^+_{V,\lambda}(h)$. First,

$$\begin{aligned}
\Psi^+_{V,\lambda,1}(\zeta,h_1) \equiv& \sup_{(x,w)\in\mathrm{supp}(X,W)} |\Psi_{V,\lambda,1}(\zeta,x,w,h_1)|\\
=& \sup_{(x,w)\in\mathrm{supp}(X,W)} \Bigg| -\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi\\
&-\frac{1}{2\pi}(-i\zeta)^\lambda\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\theta_1(\zeta)}\Bigg|\\
\le& \sup_{(x,w)\in\mathrm{supp}(X,W)} \frac{1}{2\pi}\frac{|\theta_{X_1}(\zeta)|}{|\theta_1(\zeta)|^2}\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)||\exp(-i\xi x)||\phi_V(\xi,w)|d\xi\\
&+\sup_{(x,w)\in\mathrm{supp}(X,W)} \frac{1}{2\pi}|\zeta|^\lambda|\kappa(h_1\zeta)||\exp(-i\zeta x)|\frac{|\phi_V(\zeta,w)|}{|\theta_1(\zeta)|}\\
\preceq& \frac{|\theta_{X_1}(\zeta)|}{|\theta_1(\zeta)|^2}\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi\\
&+|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\frac{1}{|\theta_1(\zeta)|}\\
=& \frac{1}{|\theta_1(\zeta)|}\left[\frac{|\theta_{X_1}(\zeta)|}{|\theta_1(\zeta)|}\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi\right.\\
&\left.+|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\right]\\
=& \frac{1}{|\theta_1(\zeta)|}\left[|D_\zeta\ln\phi_1(\zeta)|\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi\right.\\
&\left.+|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\right]
\end{aligned}$$

because we have $\theta_{X_1}(\zeta)/\theta_1(\zeta) = -iD_\zeta\ln\phi_1(\zeta)$ by eqn.(1.20) in the proof of Lemma 3.1.

Then

$$
\begin{aligned}
\Psi^{+}_{V,\lambda,1}(\zeta,h_1) \preceq{}& \frac{1}{|\theta_1(\zeta)|}\Bigg[|D_\zeta \ln\phi_1(\zeta)| \int_\zeta^{\pm\infty} |\xi|^\lambda 1_{\{|\xi|\le\bar{\xi}h_1^{-1}\}}\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi \\
&+ |\zeta|^\lambda 1_{\{|\zeta|\le\bar{\zeta}h_1^{-1}\}}\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\Bigg] \\
\preceq{}& \frac{1}{|\theta_1(\zeta)|}1_{\{|\zeta|\le h_1^{-1}\}}\Bigg[|D_\zeta \ln\phi_1(\zeta)| \int_\zeta^{h_1^{-1}} |\xi|^\lambda \left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi \\
&+ |\zeta|^\lambda \left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\Bigg].
\end{aligned}
$$

By using Assumption 4.1 and integrating $\Psi^{+}_{V,\lambda,1}(\zeta,h_1)$ with respect to ζ, we obtain

$$
\begin{aligned}
&\int \Psi^{+}_{V,\lambda,1}(\zeta,h_1)d\zeta \\
\preceq{}& \int \frac{1}{|\theta_1(\zeta)|}1_{\{|\zeta|\le h_1^{-1}\}}\Bigg[|D_\zeta \ln\phi_1(\zeta)| \int_\zeta^{h_1^{-1}} |\xi|^\lambda \left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\xi,w)|\right)d\xi \\
&+ |\zeta|^\lambda \left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\Bigg]d\zeta \\
\preceq{}& \int (1+|\zeta|)^{-\gamma_\theta}\exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right)1_{\{|\zeta|\le h_1^{-1}\}} \\
&\times\left[(1+|\zeta|)^{\gamma_1}\int_0^{h_1^{-1}}|\xi|^\lambda(1+|\xi|)^{\gamma_\phi}\exp\left(\alpha_\phi|\xi|^{\beta_\phi}\right)d\xi + |\zeta|^\lambda(1+|\zeta|)^{\gamma_\phi}\exp\left(\alpha_\phi|\zeta|^{\beta_\phi}\right)\right]d\zeta \\
\preceq{}& \int_0^{h_1^{-1}} (1+|\zeta|)^{-\gamma_\theta}\exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right)\Bigg[(1+|\zeta|)^{\gamma_1}\int_0^{h_1^{-1}}|\xi|^\lambda(1+|\xi|)^{\gamma_\phi}\exp\left(\alpha_\phi|\xi|^{\beta_\phi}\right)d\xi \\
&+ |\zeta|^\lambda(1+|\zeta|)^{\gamma_\phi}\exp\left(\alpha_\phi|\zeta|^{\beta_\phi}\right)\Bigg]d\zeta \\
\preceq{}&(1+h_1^{-1})^{1-\gamma_\theta}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)\Bigg[(1+h_1^{-1})^{\gamma_1}(1+h_1^{-1})^{\lambda+\gamma_\phi+1}\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right) \\
&+ (1+h_1^{-1})^{\gamma_\phi+\lambda}\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right)\Bigg] \\
\preceq{}&(1+h_1^{-1})^{1-\gamma_\theta}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)(1+h_1^{-1})^{\lambda+\gamma_\phi}\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right)\left((1+h_1^{-1})^{\gamma_1+1}+1\right) \\
\preceq{}&(1+h_1^{-1})^{\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+2}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right).
\end{aligned}
$$

Second,

$$
\begin{aligned}
\Psi^{+}_{V,\lambda,X_1}(\zeta,h_1) \equiv{}& \sup_{(x,w)\in\mathrm{supp}(X,W)}|\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)| \\
={}& \sup_{(x,w)\in\mathrm{supp}(X,W)}\left|\frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi\right|
\end{aligned}
$$

$$
\begin{aligned}
&\preceq \sup_{(x,w)\in\text{supp}(X,W)} \frac{1}{|\theta_1(\zeta)|}\int_{\zeta}^{\pm\infty} |\xi|^{\lambda}|\kappa(h_1\xi)||\exp(-i\xi x)||\phi_V(\xi,w)|d\xi \\
&= \frac{1}{|\theta_1(\zeta)|}\int_{\zeta}^{\pm\infty} |\xi|^{\lambda}|\kappa(h_1\xi)|\left(\sup_{w\in\text{supp}(W)}|\phi_V(\xi,w)|\right)d\xi \\
&= \frac{1}{|\theta_1(\zeta)|}\int_{\zeta}^{\pm\infty} |\xi|^{\lambda}1_{\{|\xi|\le\bar{\xi}h_1^{-1}\}}\left(\sup_{w\in\text{supp}(W)}|\phi_V(\xi,w)|\right)d\xi \\
&\preceq \frac{1}{|\theta_1(\zeta)|}1_{\{|\zeta|\le h_1^{-1}\}}\int_{\zeta}^{h_1^{-1}} |\xi|^{\lambda}\left(\sup_{w\in\text{supp}(W)}|\phi_V(\xi,w)|\right)d\xi
\end{aligned}
$$

so that

$$
\begin{aligned}
\int \Psi^{+}_{V,\lambda,X_1}(\zeta,h_1)d\zeta \preceq& \int_0^{h_1^{-1}} (1+|\zeta|)^{-\gamma_\theta}\exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right) \\
&\times\left(\int_0^{h_1^{-1}} |\xi|^{\lambda}(1+|\xi|)^{\gamma_\phi}\exp\left(\alpha_\phi|\xi|^{\beta_\phi}\right)d\xi\right)d\zeta \\
\preceq& (1+h_1^{-1})^{1-\gamma_\theta}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)(1+h_1^{-1})^{\lambda+\gamma_\phi+1}\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right) \\
\preceq& (1+h_1^{-1})^{\gamma_\phi+\lambda-\gamma_\theta+2}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right).
\end{aligned}
$$

Third,

$$
\begin{aligned}
\Psi^{+}_{V,\lambda,\chi_V}(\zeta,h_1) \equiv& \sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)| \\
=& \sup_{(x,w)\in\text{supp}(X,W)} \left|\frac{1}{2\pi}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)}\right| \\
\preceq& |\zeta|^{\lambda}1_{\{|\zeta|\le h_1^{-1}\}}\left(\sup_{w\in\text{supp}(W)}\left|\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)}\right|\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
h_2^{-1}\int \Psi^{+}_{V,\lambda,\chi_V}(\zeta,h_1)d\zeta \preceq& h_2^{-1}\int_0^{h_1^{-1}} |\zeta|^{\lambda}(1+|\zeta|)^{-\gamma_\theta}\exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right)(1+|\zeta|)^{\gamma_\phi} \\
&\times\exp\left(\alpha_\phi|\zeta|^{\beta_\phi}\right)d\zeta \\
\preceq& h_2^{-1}(1+h_1^{-1})^{\gamma_\phi+\lambda-\gamma_\theta+1}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right).
\end{aligned}
$$

Because $\inf_{w\in\text{supp}(W)} f_W(w) > 0$ by Assumption 3.3 (i), finally we have

$$
\begin{aligned}
\Psi^{+}_{V,\lambda,f_W}(\zeta,h_1) \equiv& \sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)| \\
=& \sup_{(x,w)\in\text{supp}(X,W)} \left|-\frac{1}{2\pi}(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{f_W(w)}\right| \\
\preceq& |\zeta|^{\lambda}1_{\{|\zeta|\le h_1^{-1}\}}\left(\sup_{w\in\text{supp}(W)}|\phi_V(\zeta,w)|\right)
\end{aligned}
$$

so that

$$h_2^{-1}\int \Psi^{+}_{V,\lambda,f_W}(\zeta,h)d\zeta \preceq h_2^{-1}\int_0^{h_1^{-1}} |\zeta|^{\lambda}(1+|\zeta|)^{\gamma_\phi}\exp\left(\alpha_\phi|\zeta|^{\beta_\phi}\right)d\zeta$$
$$\preceq h_2^{-1}(1+h_1^{-1})^{\gamma_\phi+\lambda+1}\exp\left(\alpha_\phi(h_1^{-1})^{\beta_\phi}\right).$$

Putting together these rates for each term of $\Psi^{+}_{V,\lambda}(h)$ gives the desired result. □

Lemma A.2 *For a finite integer J and K, let $P_{n,j}(x_2)$ define a sequence of nonrandom real-valued continuously differentiable functions of a real variable x_2, $j=1,...,J$, and $Q_{n,k}(w)$ define a sequence of nonrandom real-valued continuously differentiable functions of a real variable w, $k=1,...,K$. For some C_1, C_2 and $\delta>0$, let A_j and X_2 be random variables satisfying $E\left[A_j^{2+\delta}\mid X_2=x_2\right]\leq C_1$ for all $x_2\in\operatorname{supp}(X_2)$, $j=1,...,J$, and let B_k and W be random variables satisfying $E\left[B_k^{2+\delta}\mid W=w\right]\leq C_2$ for all $w\in\operatorname{supp}(W)$, $k=1,...,K$, such that $\sup_{n\geq N}\sigma_n<\infty$ and $\inf_{n\geq N}\sigma_n>0$ for some $N\in\mathbb{N}^+$, where*

$$\sigma_n \equiv \left(var\left[\sum_{j=1}^{J} A_j P_{n,j}(X_2)+\sum_{k=1}^{K} B_k Q_{n,k}(W)\right]\right)^{1/2}.$$

If there exists some $\eta>0$ such that $\max\{\sup_{x_2\in\operatorname{supp}(X_2)}|D_{x_2}P_{n,j}(x_2)|,\sup_{w\in\operatorname{supp}(W)}$ $|D_w Q_{n,k}(w)|\}=O(n^{(3/2)-\eta})$ for $j=1,...J$, and $k=1,...,K$, then

$$\sigma_n^{-1}n^{1/2}\Bigg(\hat{E}\left[\sum_{j=1}^{J} A_j P_{n,j}(X_2)+\sum_{k=1}^{K} B_k Q_{n,k}(W)\right]$$
$$-E\left[\sum_{j=1}^{J} A_j P_{n,j}(X_2)+\sum_{k=1}^{K} B_k Q_{n,k}(W)\right]\Bigg)\xrightarrow{d} N(0,1).$$

Proof Apply the argument of Lemma 9 in Schennach (2004a) and the Lindeberg-Feller central limit theorem. □

Proof of Theorem 4.3 (i) It follows that $E[L_{V,\lambda}(x,w,h)]=0$ by the definition of $L_{V,\lambda}(x,w,h)$. Assumption 4.2 guarantees that $L_{V,\lambda}(x,w,h)$ has a finite variance so that

$$E\left[(L_{V,\lambda}(x,w,h))^2\right]=E\left[(\hat{E}[\ell_{V,\lambda}(x,w,h;V,X_1,X_2,W)])^2\right]$$
$$=n^{-1}E\left[(\ell_{V,\lambda}(x,w,h;V,X_1,X_2,W))^2\right]$$
$$=n^{-1}\Omega_{V,\lambda}(x,w,h).$$

Because $L_{V,\lambda}(x,w,h) \equiv \bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1)$, we have by Minkowski inequality that

$$
\begin{aligned}
&\Omega_{V,\lambda}(x,w,h) \\
&= nE\left[(\bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1))^2\right] \\
&= E\Bigg[\bigg(\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_1(\zeta)d\zeta + \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_{X_1}(\zeta)d\zeta \\
&\quad + \int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\chi}_V(\zeta,w)d\zeta + \int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)n^{1/2}\delta\hat{f}_W(w)d\zeta\bigg)^2\Bigg] \\
&\le E\Bigg[\bigg(\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_1(\zeta)d\zeta + \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_{X_1}(\zeta)d\zeta \\
&\quad + h_2^{-1}\int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)n^{1/2}\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{\chi}_V(\zeta,w)\right)d\zeta \\
&\quad + h_2^{-1}\int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)n^{1/2}\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{f}_W(w)\right)d\zeta\bigg)^2\Bigg] \\
&\le \Bigg[\bigg\{E\bigg[\bigg(\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_1(\zeta)d\zeta\bigg)^2\bigg]\bigg\}^{1/2} \\
&\quad + \bigg\{E\bigg[\bigg(\int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)n^{1/2}\delta\hat{\theta}_{X_1}(\zeta)d\zeta\bigg)^2\bigg]\bigg\}^{1/2} \\
&\quad + \bigg\{E\bigg[\bigg(h_2^{-1}\int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)n^{1/2}\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{\chi}_V(\zeta,w)\right)d\zeta\bigg)^2\bigg]\bigg\}^{1/2} \\
&\quad + \bigg\{E\bigg[\bigg(h_2^{-1}\int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)n^{1/2}\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{f}_W(w)\right)d\zeta\bigg)^2\bigg]\bigg\}^{1/2}\Bigg]^2 \\
&= \Bigg[\bigg\{\int\int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)E\left[n\delta\hat{\theta}_{X_1}(\zeta)\delta\hat{\theta}^{\dagger}_{X_1}(\xi)\right](\Psi_{V,\lambda,X_1}(\xi,x,w,h_1))^{\dagger}d\zeta d\xi\bigg\}^{1/2} \\
&\quad + \bigg\{\int\int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)E\left[n\delta\hat{\theta}_{X_1}(\zeta)\delta\hat{\theta}^{\dagger}_{X_1}(\xi)\right](\Psi_{V,\lambda,X_1}(\xi,x,w,h_1))^{\dagger}d\zeta d\xi\bigg\}^{1/2} \\
&\quad + \bigg\{h_2^{-2}\int\int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)E\bigg[n\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{\chi}_V(\zeta,w)\right) \\
&\qquad \times\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{\chi}^{\dagger}_V(\xi,w)\right)\bigg](\Psi_{V,\lambda,\chi_V}(\xi,x,w,h_1))^{\dagger}d\zeta d\xi\bigg\}^{1/2} \\
&\quad + \bigg\{h_2^{-2}\int\int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)E\bigg[n\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{f}_W(w)\right) \\
&\qquad \times\left(\sup_{w\in\mathrm{supp}(W)} h_2\delta\hat{f}_W(w)\right)\bigg](\Psi_{V,\lambda,f_W}(\xi,x,w,h_1))^{\dagger}d\zeta d\xi\bigg\}^{1/2}\Bigg]^2.
\end{aligned}
$$

Note that by Assumption 4.2

$$\begin{aligned}
E\left[n\delta\hat{\theta}_1(\zeta)\delta\hat{\theta}_1^\dagger(\xi)\right] &= E\left[n\left(\hat{\theta}_1(\zeta)-\theta_1(\zeta)\right)\left(\hat{\theta}_1^\dagger(\xi)-\theta_1^\dagger(\xi)\right)\right] \\
&= E\left[\left(e^{i\zeta X_2}-\theta_1(\zeta)\right)\left(e^{-i\xi X_2}-\theta_1^\dagger(\zeta)\right)\right] \\
&= E\left[e^{i\zeta X_2}e^{-i\xi X_2}\right]-\theta_1(\zeta)E\left[e^{-i\xi X_2}\right]-E\left[e^{i\zeta X_2}\right]\theta_1^\dagger(\zeta)-\theta_1(\zeta)\theta_1^\dagger(\xi) \\
&= E\left[e^{i(\zeta-\xi)X_2}\right]-\theta_1(\zeta)\theta_1^\dagger(\xi)-\theta_1(\zeta)\theta_1^\dagger(\xi)+\theta_1(\zeta)\theta_1^\dagger(\xi) \\
&= \theta_1(\zeta-\xi)-\theta_1(\zeta)\theta_1(-\xi)
\end{aligned}$$

so that

$$\begin{aligned}
\left|E\left[n\delta\hat{\theta}_1(\zeta)\delta\hat{\theta}_1^\dagger(\xi)\right]\right| &= |\theta_1(\zeta-\xi)-\theta_1(\zeta)\theta_1(-\xi)| \\
&\le E\left[|e^{i(\zeta-\xi)X_2}|\right]+E\left[|e^{i\zeta X_2}|\right]E\left[|e^{-i\xi X_2}|\right] \\
&\preceq 1;
\end{aligned}$$

$$\begin{aligned}
E\left[n\delta\hat{\theta}_{X_1}(\zeta)\delta\hat{\theta}_{X_1}^\dagger(\xi)\right] &= E\left[n\left(\hat{\theta}_{X_1}(\zeta)-\theta_{X_1}(\zeta)\right)\left(\hat{\theta}_{X_1}^\dagger(\xi)-\theta_{X_1}^\dagger(\xi)\right)\right] \\
&= E\left[\left(X_1e^{i\zeta X_2}-\theta_{X_1}(\zeta)\right)\left(X_1e^{-i\xi X_2}-\theta_{X_1}^\dagger(\xi)\right)\right] \\
&= E\left[X_1e^{i\zeta X_2}X_1e^{-i\xi X_2}\right]-\theta_{X_1}(\zeta)E\left[X_1e^{-i\xi X_2}\right] \\
&\quad -E\left[X_1e^{i\zeta X_2}\right]\theta_{X_1}^\dagger(\xi)+\theta_{X_1}(\zeta)\theta_{X_1}^\dagger(\xi) \\
&= E\left[X_1X_1e^{i(\zeta-\xi)X_2}\right]-\theta_{X_1}(\zeta)\theta_{X_1}^\dagger(\xi)
\end{aligned}$$

so that

$$\begin{aligned}
\left|E\left[n\delta\hat{\theta}_{X_1}(\zeta)\delta\hat{\theta}_{X_1}^\dagger(\xi)\right]\right| &= \left|E\left[X_1X_1e^{i(\zeta-\xi)X_2}\right]-\theta_{X_1}(\zeta)\theta_{X_1}^\dagger(\xi)\right| \\
&\le E\left[|X_1X_1|\left|e^{i(\zeta-\xi)X_2}\right|\right]+E\left[|X_1|\left|e^{i\zeta X_2}\right|\right]E\left[|X_1|\left|e^{-i\xi X_2}\right|\right] \\
&\le E\left[|X_1X_1|\right]+E\left[|X_1|\right]E\left[|X_1|\right] \\
&\preceq 1;
\end{aligned}$$

$$\begin{aligned}
&E\left[n\left(\sup_{w\in\text{supp}(W)}h_2\delta\hat{\chi}_V(\zeta,w)\right)\left(\sup_{w\in\text{supp}(W)}h_2\delta\hat{\chi}_V^\dagger(\zeta,w)\right)\right] \\
&= E\left[n\left(\sup_{w\in\text{supp}(W)}h_2(\hat{\chi}_V(\zeta,w)-\chi_V(\zeta,w))\right)\left(\sup_{w\in\text{supp}(W)}h_2(\delta\hat{\chi}_V^\dagger(\zeta,w)-\chi_V^\dagger(\zeta,w))\right)\right] \\
&= E\Bigg[\sup_{w\in\text{supp}(W)}h_2\left(Ve^{i\zeta X_2}k_{h_2}(W-w)-\chi_V(\zeta,w)\right) \\
&\qquad \times\sup_{w\in\text{supp}(W)}h_2\left(Ve^{-i\xi X_2}k_{h_2}(W-w)-\chi_V^\dagger(\xi,w)\right)\Bigg]
\end{aligned}$$

so that

$$
\begin{aligned}
&\left| E\left[n \left(\sup_{w\in\text{supp}(W)} h_2\delta\hat{\chi}_V(\zeta,w)\right)\left(\sup_{w\in\text{supp}(W)} h_2\delta\hat{\chi}_V^{\dagger}(\zeta,w)\right)\right]\right| \\
&\le E\Bigg[\sup_{w\in\text{supp}(W)} |Ve^{i\zeta X_2}h_2k_{h_2}(W-w) - h_2\chi_V(\zeta,w)| \\
&\qquad \sup_{w\in\text{supp}(W)} |Ve^{-i\xi X_2}h_2k_{h_2}(W-w) - h_2\chi_V^{\dagger}(\xi,w)|\Bigg] \\
&\le E\Bigg[\sup_{w\in\text{supp}(W)} (|Ve^{i\zeta X_2}h_2k_{h_2}(W-w)| + |h_2\chi_V(\zeta,w)|) \\
&\qquad \sup_{w\in\text{supp}(W)} (|Ve^{-i\xi X_2}h_2k_{h_2}(W-w)| + |h_2\chi_V^{\dagger}(\xi,w)|\Bigg] \\
&\le E\Bigg[\Bigg(\left| Ve^{i\zeta X_2}\left(\sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right)\right| \\
&\qquad\qquad + E\left[\left| Ve^{i\zeta X_2}\left(\sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right)\right|\right]\Bigg) \\
&\qquad \times \Bigg(\left| Ve^{-i\xi X_2}\left(\sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right)\right| \\
&\qquad\qquad + E\left[\left| Ve^{-i\zeta X_2}\left(\sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right)\right|\right]\Bigg)\Bigg] \\
&= E\left[|V|^2\left| e^{i(\zeta-\xi)X_2}\right| \left| \sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right|^2\right] \\
&\quad + 3E\left[|V|\left| e^{i\zeta X_2}\right| \left| \sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right|\right] \\
&\qquad \times E\left[|V|\left| e^{-i\xi X_2}\right| \left| \sup_{w\in\text{supp}(W)} h_2k_{h_2}(W-w)\right|\right] \\
&\preceq 1,
\end{aligned}
$$

where the last line is obtained by Assumption 4.2 and the following note:

$$
\begin{aligned}
\sup_{w\in\text{supp}(W)} |h_2k_{h_2}(w)| &= \sup_{w\in\text{supp}(W)} \left| \frac{h_2}{2\pi}\int \kappa(h_2\zeta)e^{-i\zeta w}d\zeta\right| \\
&\le \frac{h_2}{2\pi}\sup_{w\in\text{supp}(W)} \int |\kappa(h_2\zeta)||e^{-i\zeta w}|d\zeta \\
&= \frac{h_2}{2\pi}\int |\kappa(h_2\zeta)|d\zeta = \frac{1}{2\pi}\int |\kappa(\bar{\zeta})|d\bar{\zeta} = \frac{1}{2\pi}\int_{-1}^{1} |\kappa(\bar{\zeta})|d\bar{\zeta} \\
&\preceq 1;
\end{aligned}
$$

Finally,

$$\begin{aligned}
&E\left[n\left(\sup_{w\in\operatorname{supp}(W)} h_2\delta\hat{f}_W(w)\right)\left(\sup_{w\in\operatorname{supp}(W)} h_2\delta\hat{f}_W(w)\right)\right]\\
&= E\left[n\left(\sup_{w\in\operatorname{supp}(W)} h_2(\hat{f}_W(w)-f_W(w))\right)\left(\sup_{w\in\operatorname{supp}(W)} h_2(\hat{f}_W(w)-f_W(w))\right)\right]\\
&= E\Bigg[\left(\sup_{w\in\operatorname{supp}(W)} h_2\left(k_{h_2}(W-w)-E[k_{h_2}(W-w)]\right)\right)\\
&\qquad\times\left(\sup_{w\in\operatorname{supp}(W)} h_2\left(k_{h_2}(W-w)-E[k_{h_2}(W-w)]\right)\right)\Bigg]
\end{aligned}$$

so that

$$\begin{aligned}
&\left|E\left[n\left(\sup_{w\in\operatorname{supp}(W)} h_2\delta\hat{f}_W(w)\right)\left(\sup_{w\in\operatorname{supp}(W)} h_2\delta\hat{f}_W(w)\right)\right]\right|\\
&\le E\left[\left|\sup_{w\in\operatorname{supp}(W)} h_2k_{h_2}(W-w)\right|^2\right]\\
&\quad+E\left[\left|\sup_{w\in\operatorname{supp}(W)} h_2k_{h_2}(W-w)\right|\right]E\left[\left|\sup_{w\in\operatorname{supp}(W)} h_2k_{h_2}(W-w)\right|\right]\\
&\preceq 1.
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\Omega_{V,\lambda}(x,w,h)\\
&\preceq\Bigg(\left\{\int\int|\Psi_{V,\lambda,1}(\zeta,x,w,h_1)|\,|\Psi_{V,\lambda,1}(\xi,x,w,h_1)|\,d\zeta d\xi\right\}^{1/2}\\
&\quad+\left\{\int\int|\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)|\,|\Psi_{V,\lambda,X_1}(\xi,x,w,h_1)|\,d\zeta d\xi\right\}^{1/2}\\
&\quad+\left\{h_2^{-2}\int\int|\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)|\,|\Psi_{V,\lambda,\chi_V}(\xi,x,w,h_1)|\,d\zeta d\xi\right\}^{1/2}\\
&\quad+\left\{h_2^{-2}\int\int|\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)|\,|\Psi_{V,\lambda,f_W}(\xi,x,w,h_1)|\,d\zeta d\xi\right\}^{1/2}\Bigg)^2\\
&=\left(\sum_{A=1,X_1}\int|\Psi_{V,\lambda,A}(\zeta,x,w,h_1)|\,d\zeta+h_2^{-1}\sum_{B=\chi_V,f_W}\int|\Psi_{V,\lambda,B}(\zeta,x,w,h_1)|\,d\zeta\right)^2\\
&\le\left(\sum_{A=1,X_1}\int\Psi^+_{V,\lambda,A}(\zeta,h_1)d\zeta+h_2^{-1}\sum_{B=\chi_V,f_W}\int\Psi^+_{V,\lambda,B}(\zeta,h_1)d\zeta\right)^2\\
&=\left(\Psi^+_{V,\lambda}(h)\right)^2,
\end{aligned}$$

where for $A = 1, X_1, \chi_V, f_W$

$$\begin{aligned}
\Psi^+_{V,\lambda,A}(\zeta, h_1) &\equiv \sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,A}(\zeta, x, w, h_1)| \\
\Psi^+_{V,\lambda}(h) &\equiv \sum_{A=1,X_1} \int \Psi^+_{V,\lambda,A}(\zeta, h_1)d\zeta + h_2^{-1} \sum_{B=\chi_V,f_W} \int \Psi^+_{V,\lambda,B}(\zeta, h_1)d\zeta \\
&= O\Bigg(\max\{\left(1+h_1^{-1}\right)^{\gamma_1+1}, h_2^{-1}\} \left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda-\gamma_\theta+1} \\
&\qquad \times \exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}} - \alpha_\theta)(h_1^{-1})^{\beta_\theta} \right) \Bigg).
\end{aligned}$$

Thus it follows that

$$\sqrt{\sup_{(x,w)\in\text{supp}(X,W)} \Omega_{V,\lambda}(x,w,h)} = O\left(\max\{\left(h_1^{-1}\right)^{\delta_L}, h_2^{-1}\} \left(h_1^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h_1^{-1}\right)^{\beta_L}\right)\right),$$

with $\alpha_L \equiv \alpha_\phi 1_{\{\beta_\phi=\beta_\theta\}} - \alpha_\theta$, $\beta_L \equiv \beta_\theta$, $\gamma_{\lambda,L} \equiv 1 + \gamma_\phi - \gamma_\theta + \lambda$, and $\delta_L \equiv \gamma_1 + 1$.

To show uniform convergence,

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |\bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1)| \\
&= \sup_{(x,w)\in\text{supp}(X,W)} \Bigg| \int \Bigg[\Psi_{V,\lambda,1}(\zeta,x,w,h_1) \left(\hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}]\right) \\
&\quad + \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1) \left(\hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}]\right) \\
&\quad + \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1) \left(\hat{E}[Ve^{i\zeta X_2} k_{h_2}(W-w)] - E[Ve^{i\zeta X_2} k_{h_2}(W-w)]\right) \\
&\quad + \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1) \left(\hat{E}[k_{h_2}(W-w)] - E[k_{h_2}(W-w)]\right) \Bigg] d\zeta \Bigg| \\
&\le \int \Bigg[\left(\sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,1}(\zeta,x,w,h_1)| \right) \left|\hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}]\right| \\
&\quad + \left(\sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)| \right) \left|\hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}]\right| \\
&\quad + \left(\sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)| \right) \\
&\qquad \times \left(\sup_{w\in\text{supp}(W)} \left|\hat{E}[Ve^{i\zeta X_2} k_{h_2}(W-w)] - E[Ve^{i\zeta X_2} k_{h_2}(W-w)]\right| \right) \\
&\quad + \left(\sup_{(x,w)\in\text{supp}(X,W)} |\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)| \right) \\
&\qquad \times \left(\sup_{w\in\text{supp}(W)} \left|\hat{E}[k_{h_2}(W-w)] - E[k_{h_2}(W-w)]\right| \right) \Bigg] d\zeta
\end{aligned}$$

$$= \int \Bigg[\Psi^{+}_{V,\lambda,1}(\zeta, h_1) \left| \hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right| + \Psi^{+}_{V,\lambda,X_1}(\zeta, h_1) \left| \hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}] \right|$$
$$+ h_2^{-1} \Psi^{+}_{V,\lambda,\chi_V}(\zeta, h_1)$$
$$\times \left(\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)] - E[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)] \right| \right)$$
$$+ h_2^{-1} \Psi^{+}_{V,\lambda,f_W}(\zeta, h_1) \left(\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[h_2 k_{h_2}(W-w)] - E[h_2 k_{h_2}(W-w)] \right| \right) \Bigg] d\zeta$$

where the integrals are finite since $\left| \hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right| \preceq 1$, $\left| \hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}] \right| \preceq 1$, $\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)] - E[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)] \right| \preceq 1$, and $\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[h_2 k_{h_2}(W-w)] - E[h_2 k_{h_2}(W-w)] \right| \preceq 1$, and since Lemma A.1 implies that $\Psi^{+}_{V,\lambda}(h) < \infty$. Then we have

$$E\left[\sup_{(x,w) \in \mathrm{supp}(X,W)} \left| \bar{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1) \right| \right]$$
$$\leq \int \Bigg[\Psi^{+}_{V,\lambda,1}(\zeta, h_1) E\left\{ \left(\left| \hat{E}[e^{i\zeta X_2} - E[e^{i\zeta X_2}]] \right|^2 \right)^{1/2} \right\}$$
$$+ \Psi^{+}_{V,\lambda,X_1}(\zeta, h_1) E\left\{ \left(\left| \hat{E}[X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}]] \right|^2 \right)^{1/2} \right\}$$
$$+ h_2^{-1} \Psi^{+}_{V,\lambda,\chi_V}(\zeta, h_1)$$
$$\times E\left\{ \left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[V e^{i\zeta X_2} h_2 k_{h_2}(W-w) - E[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)]] \right) \right|^2 \right)^{1/2} \right\}$$
$$+ h_2^{-1} \Psi^{+}_{V,\lambda,f_W}(\zeta, h_1)$$
$$\times E\left\{ \left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[h_2 k_{h_2}(W-w) - E[h_2 k_{h_2}(W-w)]] \right) \right|^2 \right)^{1/2} \right\} \Bigg] d\zeta$$
$$\leq \int \Bigg[\Psi^{+}_{V,\lambda,1}(\zeta, h_1) \left\{ E\left(\left| \hat{E}[e^{i\zeta X_2} - E[e^{i\zeta X_2}]] \right|^2 \right) \right\}^{1/2}$$
$$+ \Psi^{+}_{V,\lambda,X_1}(\zeta, h_1) \left\{ E\left(\left| \hat{E}[X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}]] \right|^2 \right) \right\}^{1/2}$$
$$+ h_2^{-1} \Psi^{+}_{V,\lambda,\chi_V}(\zeta, h_1)$$
$$\times \left\{ E\left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[V e^{i\zeta X_2} h_2 k_{h_2}(W-w) - E[V e^{i\zeta X_2} h_2 k_{h_2}(W-w)]] \right) \right|^2 \right) \right\}^{1/2}$$

$$
\begin{aligned}
&+ h_2^{-1}\Psi^+_{V,\lambda,f_W}(\zeta,h_1)\\
&\times\left\{E\left(\left|\sup_{w\in\mathrm{supp}(W)}\left(\hat{E}[h_2k_{h_2}(W-w)-E[h_2k_{h_2}(W-w)]]\right)\right|^2\right)\right\}^{1/2}\Bigg]d\zeta\\
&=\int\Bigg[\Psi^+_{V,\lambda,1}(\zeta,h_1)\left\{n^{-1}E\left(\left|e^{i\zeta X_2}-E[e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2}\\
&+\Psi^+_{V,\lambda,X_1}(\zeta,h_1)\left\{n^{-1}E\left(\left|X_1e^{i\zeta X_2}-E[X_1e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2}\\
&+h_2^{-1}\Psi^+_{V,\lambda,\chi_V}(\zeta,h_1)\\
&\times\left\{n^{-1}E\left(\left|\sup_{w\in\mathrm{supp}(W)}\left(Ve^{i\zeta X_2}h_2k_{h_2}(W-w)-E[Ve^{i\zeta X_2}h_2k_{h_2}(W-w)]\right)\right|^2\right)\right\}^{1/2}\\
&+h_2^{-1}\Psi^+_{V,\lambda,f_W}(\zeta,h_1)\\
&\times\left\{n^{-1}E\left(\left|\sup_{w\in\mathrm{supp}(W)}(h_2k_{h_2}(W-w)-E[h_2k_{h_2}(W-w)])\right|^2\right)\right\}^{1/2}\Bigg]d\zeta\\
&=n^{-1/2}\int\Bigg[\Psi^+_{V,\lambda,1}(\zeta,h_1)\left\{E\left(\left|e^{i\zeta X_2}-E[e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2}\\
&+\Psi^+_{V,\lambda,X_1}(\zeta,h_1)\left\{E\left(\left|X_1e^{i\zeta X_2}-E[X_1e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2}\\
&+h_2^{-1}\Psi^+_{V,\lambda,\chi_V}(\zeta,h_1)\\
&\times\left\{E\left(\left|\sup_{w\in\mathrm{supp}(W)}\left(Ve^{i\zeta X_2}h_2k_{h_2}(W-w)-E[Ve^{i\zeta X_2}h_2k_{h_2}(W-w)]\right)\right|^2\right)\right\}^{1/2}\\
&+h_2^{-1}\Psi^+_{V,\lambda,f_W}(\zeta,h_1)\\
&\times\left\{E\left(\left|\sup_{w\in\mathrm{supp}(W)}(h_2k_{h_2}(W-w)-E[h_2k_{h_2}(W-w)])\right|^2\right)\right\}^{1/2}\Bigg]d\zeta\\
&\preceq n^{-1/2}\left(\sum_{A=1,X_1}\int\Psi^+_{V,\lambda,A}(\zeta,h_1)d\zeta+h_2^{-1}\sum_{B=\chi_V,f_W}\int\Psi^+_{V,\lambda,B}(\zeta,h_1)d\zeta\right)\\
&=n^{-1/2}\Psi^+_{V,\lambda}(h),
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi^+_{V,\lambda}(h)=O\Bigg(&\max\{\left(1+h_1^{-1}\right)^{\gamma_1+1},h_2^{-1}\}\left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda-\gamma_\theta+1}\\
&\times\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)(h_1^{-1})^{\beta_\theta}\right)\Bigg).
\end{aligned}
$$

It follows that by Markov's inequality

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |L_{V,\lambda}(x,w,h)| \\
&= O_p\Bigg(n^{-1/2}\left(\max\{\left(1+h_1^{-1}\right)^{\gamma_1+1}, h_2^{-1}\}\right) \\
&\quad \times \left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda-\gamma_\theta+1} \exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}} - \alpha_\theta)(h_1^{-1})^{\beta_\theta}\right)\Bigg).
\end{aligned}$$

(ii) To show asymptotic normality, for fixed x and w, we apply Lemma A.2 to

$$\begin{aligned}
&\sum_{j=1}^{2} A_j P_{n,j}(X_2) + \sum_{k=1}^{2} B_k Q_{n,k}(W) \\
\equiv &\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)\left(e^{i\zeta X_2}\right)d\zeta + \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)\left(X_1 e^{i\zeta X_2}\right)d\zeta \\
&+ \int \Psi_{V,\lambda,\chi V}(\zeta,x,w,h_1)\left(Ve^{i\zeta X_2}k_{h_2}(W-w)\right)d\zeta \\
&+ \int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)\left(k_{h_2}(W-w)\right)d\zeta,
\end{aligned}$$

with

$$\begin{aligned}
P_{n,1}(x_2) &= \int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)e^{i\zeta x_2}d\zeta, \\
P_{n,2}(x_2) &= \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)e^{i\zeta x_2}d\zeta, \\
Q_{n,1}(\tilde{w}) &= \int \Psi_{V,\lambda,\chi V}(\zeta,x,w,h_1)e^{i\zeta X_2}k_{h_2}(\tilde{w}-w)d\zeta, \\
Q_{n,2}(\tilde{w}) &= \int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)k_{h_2}(\tilde{w}-w)d\zeta,
\end{aligned}$$

corresponding to $A_1 = 1$, $A_2 = X_1$, $B_1 = V$, and $B_2 = 1$, respectively. We assume that $\inf_{n>N}\Omega_{V,\lambda}(x,w,h) > 0$, and previous conditions ensure that for some finite N, $\sup_{n>N}\Omega_{V,\lambda}(x,w,h) = \sup_{n>N}\text{var}[\ell_{V,\lambda}(x,w,h_n;V,X_1,X_2)] < \infty$. We need to verify that $\max\{\sup_{x_2\in\text{supp}(X_2)}|D_{x_2}P_{n,j}(x_2)|, \sup_{\tilde{w}\in\text{supp}(W)}|D_{\tilde{w}}Q_{n,k}(\tilde{w})|\} = O(n^{(3/2)-\eta})$ for $j = 1,2$ and $k = 1,2$. To do this, we use Lemma A.1. For $j = 1,2$,

$$\begin{aligned}
\sup_{x_2\in\text{supp}(X_2)} |D_{x_2}P_{n,j}(x_2)| &= \sup_{x_2\in\text{supp}(X_2)} \left|\int i\zeta \Psi_{V,\lambda,j}(\zeta,x,w,h_1)e^{i\zeta x_2}d\zeta\right| \\
&= \int |\zeta||\Psi_{V,\lambda,j}(\zeta,x,w,h_1)|d\zeta \\
&\preceq \sup_{(x,w)\in\text{supp}(X,W)} \int_0^{h_{1n}^{-1}} |\zeta||\Psi_{V,\lambda,j}(\zeta,x,w,h_1)|d\zeta \\
&\le h_{1n}^{-1}\int_0^{h_{1n}^{-1}} \Psi^{+}_{V,\lambda,j}(\zeta,h_1)d\zeta
\end{aligned}$$

$$\preceq h_{1n}^{-1}\left(1+h_{1n}^{-1}\right)^{\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+2}\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)(h_{1n}^{-1})^{\beta_\theta}\right)$$
$$=\left(1+h_{1n}^{-1}\right)^{\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+3}\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)(h_{1n}^{-1})^{\beta_\theta}\right).$$

By Assumption 4.4, if $\beta_\theta \neq 0$, we have $h_{1n}^{-1}=O\left((\ln n)^{1/\beta_\theta-\eta}\right)$ for some $\eta>0$. Thus we have for $j=1,2$

$$\sup_{x_2\in\mathrm{supp}(X_2)}|D_{x_2}P_{n,j}(x_2)|$$
$$\preceq\left(1+(\ln n)^{1/\beta_\theta-\eta}\right)^{\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+3}\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)((\ln n)^{1/\beta_\theta-\eta})^{\beta_\theta}\right).$$

Because the right-hand side grows more slowly than any power of n, we certainly have $\sup_{x_2\in\mathrm{supp}(X_2)}|D_{x_2}P_{n,j}(x_2)|=O(n^{(3/2)-\eta})$ for $j=1,2$. If $\beta_\theta=0$, we have $h_{1n}^{-1}=O\left(n^{-\eta}n^{(3/2)/(\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+3)}\right)$ for some $\eta>0$. Thus we have

$$\sup_{x_2\in\mathrm{supp}(X_2)}|D_{x_2}P_{n,j}(x_2)|\preceq\left(1+n^{-\eta}n^{(3/2)/(\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+3)}\right)^{\gamma_\phi+\lambda+\gamma_1-\gamma_\theta+3}$$
$$\preceq\left(1+n^{-\eta}n^{(3/2)}\right)=O(n^{(3/2)-\eta}).$$

Because the Fourier transform of $D_x^\lambda k_{h_1}(x)$ is $(-i\zeta)^\lambda\kappa(h_1\zeta)$, we have

$$\left|h_1^{\lambda+1}D_x^\lambda k_{h_1}(x)\right|=\left|\frac{h_1^{\lambda+1}}{2\pi}\int(-i\zeta)^\lambda\kappa(h_1\zeta)e^{-i\zeta x}d\zeta\right|\leq\frac{h_1^{\lambda+1}}{2\pi}\int|\zeta|^\lambda|\kappa(h_1\zeta)|d\zeta$$
$$=\frac{1}{2\pi}\int|\bar\zeta|^\lambda|\kappa(\bar\zeta)|d\bar\zeta=\frac{1}{2\pi}\int_{-1}^{1}|\bar\zeta|^\lambda|\kappa(\bar\zeta)|d\bar\zeta<\infty.$$

Therefore we get

$$\sup_{\tilde w\in\mathrm{supp}(W)}|D_{\tilde w}Q_{n,1}(\tilde w)|$$
$$=\sup_{\tilde w\in\mathrm{supp}(W)}\left|D_{\tilde w}\int\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)e^{i\zeta x_2}k_{h_2}(\tilde w-w)d\zeta\right|$$
$$=h_2^{-2}\int|\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)|\,|e^{i\zeta x_2}|\left(\sup_{\tilde w\in\mathrm{supp}(W)}\left|h_2^2D_{\tilde w}k_{h_2}(\tilde w-w)\right|\right)d\zeta$$
$$\preceq h_2^{-2}\int\sup_{(x,w)\in\mathrm{supp}(X,W)}|\Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)|\,d\zeta$$
$$=h_2^{-2}\int\Psi^+_{V,\lambda,\chi_V}(\zeta,h_1)d\zeta$$
$$=O\left(\left(1+h_2^{-1}\right)^2\left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda-\gamma_\theta+1}\exp\left((\alpha_\phi 1_{\{\beta_\theta=\beta_\phi\}}-\alpha_\theta)(h_1^{-1})^{\beta_\theta}\right)\right).$$

Bandwidth sequences in Assumption 4.4 guarantee that $\sup_{\tilde w\in\mathrm{supp}(W)}|D_{\tilde w}Q_{n,1}(\tilde w)|=O(n^{(3/2)-\eta})$. Similarly,

$$\begin{aligned}
\sup_{\tilde{w}\in\operatorname{supp}(W)} |D_{\tilde{w}} Q_{n,2}(\tilde{w})| &= \sup_{\tilde{w}\in\operatorname{supp}(W)} \left| D_{\tilde{w}} \int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1) k_{h_2}(\tilde{w}-w) d\zeta \right| \\
&= h_2^{-2} \int |\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)| \left(\sup_{\tilde{w}\in\operatorname{supp}(W)} \left| h_2^2 D_{\tilde{w}} k_{h_2}(\tilde{w}-w) \right| \right) d\zeta \\
&\preceq h_2^{-2} \int \sup_{(x,w)\in\operatorname{supp}(X,W)} |\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)| \, d\zeta \\
&= h_2^{-2} \int \Psi^{+}_{V,\lambda,f_W}(\zeta,h_1) d\zeta \\
&= O\left(\left(1+h_2^{-1}\right)^2 \left(1+h_1^{-1}\right)^{\gamma_\phi+\lambda+1} \exp\left((\alpha_\phi (h^{-1})^{\beta_\phi} \right) \right).
\end{aligned}$$

Because $\sup_{\tilde{w}\in\operatorname{supp}(W)} |D_{\tilde{w}} Q_{n,2}(\tilde{w})| \preceq \sup_{\tilde{w}\in\operatorname{supp}(W)} |D_{\tilde{w}} Q_{n,1}(\tilde{w})|$, the result follows. □

Lemma A.3 *Let A and X_2 be random variables satisfying $E[|A|^2] < \infty$ and $E[|A||X_2|] < \infty$, and let $\{A_i, X_{2,i}\}_{i=1,\ldots,n}$ be a corresponding IID sample. Then for any u, $U \geq 0$, and $\epsilon > 0$,*

$$\sup_{\zeta\in[-Un^u,Un^u]} \left| \hat{E}[A\exp(i\zeta X_2)] - E[A\exp(i\zeta X_2)] \right| = O_p(n^{-1/2+\epsilon}).$$

Proof See Lemma 6 in Schennach (2004b). □

Proof of Theorem 4.4 By substituting eqn.(1.18)∼(1.21) into

$$\begin{aligned}
&\hat{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h_1) \\
=&\frac{1}{2\pi} \int (-i\zeta)^\lambda \kappa(h_1\zeta) \exp(-i\zeta x) \\
&\times \left[\frac{\hat{\theta}_V(\zeta,w)}{\hat{\theta}_1(\zeta)} \exp\left(\int_0^\zeta \frac{i\hat{\theta}_{X_1}(\xi)}{\hat{\theta}_1(\zeta)} d\xi \right) - \frac{\theta_V(\zeta,w)}{\theta_1(\zeta)} \exp\left(\int_0^\zeta \frac{i\theta_{X_1}(\xi)}{\theta_1(\zeta)} d\xi \right) \right] d\zeta,
\end{aligned}$$

and removing the terms linear in $\delta\hat{\theta}_1(\zeta)$, $\delta\hat{\theta}_{X_1}(\zeta)$, $\delta\hat{\chi}_V(\zeta,w)$, and $\delta\hat{f}_W(w)$, we obtain the nonlinear remainder term such that $R_{V,\lambda}(x,w,h) \equiv \hat{g}_{V,\lambda}(x,w,h) - \bar{g}_{V,\lambda}(x,w,h) = \sum_{i=1}^{22} R_i$ where

$$\begin{aligned}
R_1 = \frac{1}{4\pi} &\int (-i\zeta)^\lambda \kappa(h_1\zeta) \exp(-i\zeta x) q_V(\zeta,w) q_1(w) \exp(Q_{X_1}(\zeta)) \exp\left(\delta\bar{Q}_{X_1}(\zeta)\right) \\
&\times \left(\int_0^\zeta i\delta\hat{q}_{X_1}(\xi) d\xi \right)^2 d\zeta \\
R_2 = \frac{1}{4\pi} &\int (-i\zeta)^\lambda \kappa(h_1\zeta) \exp(-i\zeta x) \delta\hat{q}_V(\zeta,w) q_1(w) \exp(Q_{X_1}(\zeta)) \exp\left(\delta\bar{Q}_{X_1}(\zeta)\right) \\
&\times \left(\int_0^\zeta i\delta\hat{q}_{X_1}(\xi) d\xi \right)^2 d\zeta
\end{aligned}$$

$$R_3 = \frac{1}{4\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)\delta\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\exp\left(\delta\bar{Q}_{X_1}(\zeta)\right)$$
$$\times\left(\int_0^{\zeta} i\delta\hat{q}_{X_1}(\xi)d\xi\right)^2 d\zeta$$
$$R_4 = \frac{1}{4\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta\hat{q}_V(\zeta,w)\delta\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\exp\left(\delta\bar{Q}_{X_1}(\zeta)\right)$$
$$\times\left(\int_0^{\zeta} i\delta\hat{q}_{X_1}(\xi)d\xi\right)^2 d\zeta$$
$$R_5 = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)q_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_2\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_6 = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta\hat{q}_V(\zeta,w)q_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_2\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_7 = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)\delta\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_2\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_8 = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta\hat{q}_V(\zeta,w)\delta\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_2\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_9 = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_1\hat{q}_V(\zeta,w)q_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{10} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)q_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{11} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)\delta_1\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{12} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_1\hat{q}_V(\zeta,w)\delta_1\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{13} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)\delta_1\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{14} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{15} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_1\hat{q}_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{16} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))\int_0^{\zeta} i\delta_1\hat{q}_{X_1}(\xi)d\xi d\zeta$$
$$R_{17} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)q_1(w)\exp(Q_{X_1}(\zeta))d\zeta$$
$$R_{18} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_1\hat{q}_V(\zeta,w)\delta_1\hat{q}_1(w)\exp(Q_{X_1}(\zeta))d\zeta$$
$$R_{19} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)\delta_1\hat{q}_1(w)\exp(Q_{X_1}(\zeta))d\zeta$$
$$R_{20} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)q_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))d\zeta$$
$$R_{21} = \frac{1}{2\pi}\int(-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_1\hat{q}_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))d\zeta$$

$$R_{22} = \frac{1}{2\pi}\int (-i\zeta)^{\lambda}\kappa(h_1\zeta)\exp(-i\zeta x)\delta_2\hat{q}_V(\zeta,w)\delta_2\hat{q}_1(w)\exp(Q_{X_1}(\zeta))d\zeta.$$

Because $E[Y^2] < \infty$ by assumption 4.2 and $E[|YX_2|] < \infty$ by assumption 4.5, Lemma A.3 gives that for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{w\in\mathrm{supp}(W)}\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}|\hat{\chi}_V(\zeta,w)-\chi_V(\zeta,w)| \\
&= \sup_{w\in\mathrm{supp}(W)}\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{E}[Vk_{h_{2n}}(W-w)\exp(i\zeta X_2)] - E[Vk_{h_{2n}}(W-w)\exp(i\zeta X_2)]\right| \\
&= h_{2n}^{-1}\sup_{w\in\mathrm{supp}(W)}|h_{2n}k_{h_{2n}}(W-w)|\sup_{\zeta\in[-Un^u,Un^u]}\left|\hat{E}[V\exp(i\zeta X_2)] - E[V\exp(i\zeta X_2)]\right| \\
&= O_p(h_{2n}^{-1}n^{-1/2+\epsilon}).
\end{aligned}$$

We define $\Upsilon(h_n)$ and $\hat{\Phi}_n$ as follows:

$$\begin{aligned}
\Upsilon(h_n) &\equiv \left(1+h_{2n}^{-1}\right)\left(\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}|D_\zeta\ln\phi_1(\zeta)|\right) \\
&\quad\times\left(\max\left\{\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}\sup_{w\in\mathrm{supp}(W)}|\chi_V(\zeta,w)|^{-1},\ \sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}|\theta_1(\zeta)|^{-1}\right\}\right) \\
&= O\left(\left(1+h_{2n}^{-1}\right)\left(1+h_{1n}^{-1}\right)^{\gamma_1-\gamma_\theta}\exp\left(-\alpha_\theta\left(h_{1n}^{-1}\right)^{\beta_\theta}\right)\right),
\end{aligned}$$

$$\begin{aligned}
\hat{\Phi}_n &\equiv \max\Bigg\{\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{\theta}_1(\zeta)-\theta_1(\zeta)\right|,\ \sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}\left|\hat{\theta}_{X_1}(\zeta)-\theta_{X_1}(\zeta)\right|, \\
&\qquad \sup_{w\in\mathrm{supp}(W)}\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}|\hat{\chi}_V(\zeta,w)-\chi_V(\zeta,w)|,\ \sup_{w\in\mathrm{supp}(W)}\left|\hat{f}_W(w)-f_W(w)\right|\Bigg\} \\
&= O_p\left(h_{2n}^{-1}n^{-1/2+\epsilon}\right)
\end{aligned}$$

for any $\epsilon > 0$. Note that the supremums associated with ζ can be taken over $[-h_{1n}^{-1}, h_{1n}^{-1}]$ since $\kappa(h_{1n}\zeta)$ vanishes outside the interval by Assumption 3.6 (ii). The second order of magnitude follows from Lemma A.3 and Assumption 4.6 since $h_{2n}^{-1}n^{-1/2+\epsilon} = h_{2n}^{-1/2}n^{-1/2}(n^{\epsilon}h_{2n}^{-1/2}) > h_{2n}^{-1/2}n^{-1/2}(\ln n)^{1/2} + h_{2n}^2$ for any choices of h_{2n} from Assumption 4.4 and 4.7. Then those terms in the nonlinear remainder can be bounded in terms of $\Psi_{V,\lambda}^{+}(h_n)$, $\Upsilon(h_n)$, and $\hat{\Phi}_n$. We note that

$$\begin{aligned}
&\hat{\Phi}_n\times\left(\max\left\{\sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}\sup_{w\in\mathrm{supp}(W)}|\chi_V(\zeta,w)|^{-1},\ \sup_{\zeta\in[-h_{1n}^{-1},h_{1n}^{-1}]}|\theta_1(\zeta)|^{-1}\right\}\right) \\
&\preceq \hat{\Phi}_n\Upsilon(h_n) \\
&= O_p\left(h_{2n}^{-1}n^{-1/2+\varepsilon}\right)O\left(\left(1+h_{2n}^{-1}\right)\left(1+h_{1n}^{-1}\right)^{\gamma_1-\gamma_\theta}\exp\left(-\alpha_\theta\left(h_{1n}^{-1}\right)^{\beta_\theta}\right)\right) \\
&= o_p(1).
\end{aligned}$$

We find upper bounds for each term, R_i, $i = 1, ..., 22$.

$$
\begin{aligned}
&\sup_{(x,w)\in \operatorname{supp}(X,W)} |R_1| \\
&\preceq \sup_{w\in \operatorname{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||q_1(w)| \exp\left(Q_{X_1}(\zeta)\right) \exp\left(\left|\delta\bar{Q}_{X_1}(\zeta)\right|\right) \\
&\qquad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|\, d\xi\right)^2 d\zeta \\
&\le \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|\, d\xi\right) \\
&\qquad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|\, d\xi\right)^2 d\zeta \\
&\preceq \exp(o_p(1)) \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|\, d\xi\right) \\
&\qquad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|\, d\xi\right) d\zeta \\
&\preceq \exp(o_p(1)) \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) \\
&\qquad \times \int_0^\zeta \left| \left(\frac{\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)} - \frac{\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right) \left(1 + \frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\right)^{-1} \right| d\xi \\
&\qquad \times \int_0^\zeta \left| \left(\frac{\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)} - \frac{\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right) \left(1 + \frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\right)^{-1} \right| d\xi d\zeta \\
&\preceq \exp(o_p(1)) \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-2} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) \\
&\qquad \times \int_0^\zeta \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right) d\xi d\zeta \\
&= \exp(o_p(1)) \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-2} \int_0^\infty \left(\int_\xi^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) d\zeta\right) \\
&\qquad \times \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right) d\xi \\
&= \exp(o_p(1)) \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-2} \int_0^\infty \left[\int_\xi^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) d\zeta\right. \\
&\qquad \left. \frac{1}{|\theta_1(\xi)|} + \int_\xi^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in \operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) d\zeta \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right] d\xi \\
&= O_p(1) \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h) \\
&\preceq \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h).
\end{aligned}
$$

When the conditions of Theorem 4.3 hold, we have

$$\begin{aligned}\sup_{(x,w)\in\mathrm{supp}(X,W)}|R_1| &= O_p\Bigg((h_2^{-1})(h_1^{-1})^{\gamma_1-\gamma_\theta}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)(h_2^{-1})^2 n^{-1+2\epsilon}\\ &\qquad\times(\max\{(h_1^{-1})^{\delta_L},h_2^{-1}\})(h_1^{-1})^{\gamma_{\lambda,L}}\exp\left(\alpha_L(h_1^{-1})^{\beta_L}\right)\Bigg)\end{aligned}$$

which is needed for part (i). Because all other terms are also bounded by the upper bound for R_1 as shown below, we focus on the bound for R_1.

In order to get the bound for $R_{V,\lambda}(x,w,h_n)$ when Assumption 4.7 holds in place of Assumption 4.4 in the conditions of Theorem 4.3, we note that

$$\begin{aligned}\Upsilon(h)\hat{\Phi}_n^2\Psi_{V,\lambda}^+(h) &= \left(\Upsilon(h)\hat{\Phi}_n^2 n^{1/2}\right)n^{-1/2}\Psi_{V,\lambda}^+(h),\\ n^{-1/2}\Psi_{V,\lambda}^+(h) &= O_p\left(n^{-1/2}(\max\{(h_1^{-1})^{\delta_L},h_2^{-1}\})(h_1^{-1})^{\gamma_{\lambda,L}}\exp\left(\alpha_L(h_1^{-1})^{\beta_L}\right)\right)\end{aligned}$$

where the second equality is obtained by Lemma A.1. Now we show that $\Upsilon(h_n)\hat{\Phi}_n^2 n^{1/2} = o_p(1)$. When $\beta_\theta\neq 0$, we have $h_{1n}^{-1}\preceq(\ln n)^{1/\beta_\theta-\eta}$ and $h_{2n}^{-1}=O\left(\exp(\frac{\alpha_\theta}{4}(\ln n)^{1-\eta\beta_\theta})\right)$ by the Assumption 4.7 so that

$$\begin{aligned}&\Upsilon(h_n)\hat{\Phi}_n^2 n^{1/2}\\ &=\Upsilon(h_n)O_p\left(h_{2n}^{-2}n^{-1+2\epsilon}\right)n^{1/2}\\ &=O_p\left(\left(1+h_{2n}^{-1}\right)^3\left(1+h_{1n}^{-1}\right)^{\gamma_1-\gamma_\theta}\exp\left(-\alpha_\theta\left(h_{1n}^{-1}\right)^{\beta_\theta}\right)n^{-1/2+2\epsilon}\right)\\ &=O_p\Bigg(\left(1+\exp(\frac{\alpha_\theta}{4}(\ln n)^{1-\eta\beta_\theta})\right)^3\left(1+(\ln n)^{1/\beta-\eta}\right)^{\gamma_1-\gamma_\theta}\\ &\qquad\times\exp\left(-\alpha_\theta(\ln n)^{1-\eta\beta_\theta}\right)n^{-1/2+2\epsilon}\Bigg)\\ &=O_p\left(\exp(\frac{3\alpha_\theta}{4}(\ln n)^{1-\eta\beta_\theta})(\ln n)^{(1/\beta_\theta-\eta)(\gamma_1-\gamma_\theta)}\exp\left(-\alpha_\theta(\ln n)^{1-\eta\beta_\theta}\right)n^{-1/2+2\epsilon}\right)\\ &=O_p\Bigg(\exp\Bigg[\frac{3\alpha_\theta}{4}(\ln n)^{1-\eta\beta_\theta}+(1/\beta_\theta-\eta)(\gamma_1-\gamma_\theta)\ln(\ln n)-\alpha_\theta(\ln n)^{1-\eta\beta_\theta}\\ &\qquad+(-1/2+2\epsilon)\ln n\Bigg]\Bigg)\\ &=o_p(1),\end{aligned}$$

where the last equality follows by the fact that $\ln n$ dominates $(\ln n)^{1-\eta\beta_\theta}$ and $\ln(\ln n)$, and by $-1/2+2\epsilon<0$. When $\beta_\theta=0$, we have $h_{1n}^{-1}\preceq n^{-\eta}n^{1/(2\gamma_1-2\gamma_\theta+6)}$ and $h_{2n}^{-1}=O\left(n^{\eta(\gamma_1-\gamma_\theta-1)/4}\right)$ so that

$$\begin{aligned}\Upsilon(h_n)\hat{\Phi}_n^2 n^{1/2} &= \Upsilon(h_n)O_p\left(h_{2n}^{-2}n^{-1+2\epsilon}\right)n^{1/2}\\ &=O_p\left(\left(1+h_{2n}^{-1}\right)^3\left(1+h_{1n}^{-1}\right)^{\gamma_1-\gamma_\theta}n^{-1/2+2\epsilon}\right)\end{aligned}$$

$$
\begin{aligned}
&= O_p\left(\left(n^{\eta(\gamma_1-\gamma_\theta-1)/4}\right)^3\left(n^{-\eta}n^{1/(2\gamma_1-2\gamma_\theta)}\right)^{\gamma_1-\gamma_\theta}n^{-1/2+2\epsilon}\right)\\
&= O_p\left(n^{-\eta(\gamma_1-\gamma_\theta-3(\gamma_1-\gamma_\theta-1)/4)+2\epsilon}\right)\\
&\preceq O_p\left(n^{-\eta+2\epsilon}\right)\\
&= o_p(1),
\end{aligned}
$$

by selecting $\eta > 2\epsilon$. Now we get the bounds for the remaining terms. Because they all contain the same leading term, $\Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h)$, they can be similarly bounded:

$$
\begin{aligned}
&\sup_{(x,w)\in\mathrm{supp}(X,W)}|R_2|\\
&\preceq \sup_{w\in\mathrm{supp}(W)}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)||\delta\hat{q}_V(\zeta,w)||q_1(w)|\exp(Q_{X_1}(\zeta))\exp\left(|\delta\bar{Q}_{X_1}(\zeta)|\right)\\
&\quad\times\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2d\zeta\\
&\preceq \sup_{w\in\mathrm{supp}(W)}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right)\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}|q_1(w)|\exp(Q_{X_1}(\zeta))\\
&\quad\times\exp\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2d\zeta\\
&= \sup_{w\in\mathrm{supp}(W)}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right)|q_1(w)|\exp(Q_{X_1}(\zeta))\\
&\quad\times\exp\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2d\zeta\\
&\preceq \sup_{w\in\mathrm{supp}(W)}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)||q_1(w)|\exp(Q_{X_1}(\zeta))\\
&\quad\times\exp\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2d\zeta\\
&\quad+\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\mathrm{supp}(W)}|\phi_V(\zeta,w)|\right)\\
&\quad\times\exp\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta|\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2d\zeta\\
&\preceq \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\left(\sup_{(x,w)\in\mathrm{supp}(X,W)}|R_1|\right)\\
&= o_p(1)\left(\sup_{(x,w)\in\mathrm{supp}(X,W)}|R_1|\right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_3| \\
&\preceq \sup_{w\in\operatorname{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||\delta\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \exp\left(|\delta\bar{Q}_{X_1}(\zeta)|\right) \\
&\quad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&\preceq \sup_{w\in\operatorname{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)| \frac{1}{|f_W(w)|} |1+o_p(1)|^{-1} \Upsilon(h)\hat{\Phi}_n \exp(Q_{X_1}(\zeta)) \\
&\quad \times \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&= \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\operatorname{supp}(W)} |\phi_V(\zeta,w)|\right) \frac{1}{2} \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right) \\
&\quad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n (1+o_p(1)) \left(\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_1|\right) \\
&= o_p(1) \left(\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_1|\right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_4| \\
&\preceq \sup_{w\in\operatorname{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta\hat{q}_V(\zeta,w)||\delta\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \exp\left(|\delta\bar{Q}_{X_1}(\zeta)|\right) \\
&\quad \times \left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&\preceq \sup_{w\in\operatorname{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right) \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \frac{1}{|f_W(w)|} \Upsilon(h)\hat{\Phi}_n \\
&\quad \times |1+o_p(1)|^{-1} \exp(Q_{X_1}(\zeta)) \frac{1}{2} \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&\quad \times \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&= \sup_{w\in\operatorname{supp}(W)} \Upsilon^2(h)\hat{\Phi}_n^2 |1+o_p(1)|^{-2} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right) |q_1(w)| \exp(Q_{X_1}(\zeta)) \\
&\quad \times \exp\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)d\xi|\right)\left(\int_0^\zeta |\delta\hat{q}_{X_1}(\xi)|d\xi\right)^2 d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \left(\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_2|\right) = o_p(1) \left(\sup_{(x,w)\in\operatorname{supp}(X,W)} |R_2|\right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_5| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)| \right) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)| \right) \\
&\quad \times \int_0^\zeta \left(\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|} \frac{1}{|\theta_1(\xi)|^2} \hat{\Phi}_n^2 |1+o_p(1)|^{-1} + \frac{1}{|\theta_1(\xi)|^2} \hat{\Phi}_n^2 |1+o_p(1)|^{-1} \right) d\xi d\zeta \\
&\preceq \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)| \right) \Upsilon(h) \hat{\Phi}_n^2 |1+o_p(1)|^{-1} \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|} \left(\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|} + 1 \right) d\xi d\zeta \\
&= \Upsilon(h) \hat{\Phi}_n^2 |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)| \right) \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|} \left(\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|} + 1 \right) d\xi d\zeta \\
&= \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h)(1+o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_6| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta\hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\frac{1}{|\theta_1(\zeta)|} + \frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|^2} \right) \hat{\Phi}_n |1+o_p(1)|^{-1} |q_1(w)| \\
&\quad \times \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \Upsilon(h) \hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(1 + \frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|} \right) |q_1(w)| \\
&\quad \times \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \Upsilon(h) \hat{\Phi}_n |1+o_p(1)|^{-1} \Bigg[\sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_1(w)| \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta + \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)| \right) \int_0^\zeta |\delta_2 \hat{q}_{X_1}(\xi)| d\xi d\zeta \Bigg]
\end{aligned}
$$

$$\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1))\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right)$$
$$= o_p(1)\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right);$$

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)}|R_7| \\
&\preceq \sup_{w\in\text{supp}(W)}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)||q_V(\zeta,w)||\delta\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)||q_V(\zeta,w)|\frac{1}{|f_W(w)|}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\exp(Q_{X_1}(\zeta)) \\
&\quad\times\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\text{supp}(W)}|\phi_V(\zeta,w)|\right)\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1))\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right) \\
&= o_p(1)\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right);
\end{aligned}$$

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)}|R_8| \\
&\preceq \sup_{w\in\text{supp}(W)}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)||\delta\hat{q}_V(\zeta,w)||\delta\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right)\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\frac{1}{|f_W(w)|}\Upsilon(h)\hat{\Phi}_n \\
&\quad\times|1+o_p(1)|^{-1}\exp(Q_{X_1}(\zeta))\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \Upsilon^2(h)\hat{\Phi}_n^2|1+o_p(1)|^{-2}\left[\sup_{w\in\text{supp}(W)}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|f_W(w)|}\exp(Q_{X_1}(\zeta))\right. \\
&\quad\left.\times\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta+\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{w\in\text{supp}(W)}|\phi_V(\zeta,w)|\right)\int_0^\zeta|\delta_2\hat{q}_{X_1}(\xi)|d\xi d\zeta\right] \\
&\preceq \Upsilon^2(h)\hat{\Phi}_n^2|1+o_p(1)|^{-2}\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right) \\
&= o_p(1)\left(\sup_{(x,w)\in\text{supp}(X,W)}|R_5|\right);
\end{aligned}$$

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_9| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||q_1(w)|\exp(Q_{X_1}(\zeta))\int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right)\Upsilon(h)\hat{\Phi}_n|q_1(w)|\exp(Q_{X_1}(\zeta)) \\
&\quad\times \int_0^\zeta \left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)\frac{1}{|\theta_1(\xi)|}\hat{\Phi}_n d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2 \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\left(|q_1(w)|\exp(Q_{X_1}(\zeta))+|\phi_V(\zeta,w)|\right) \\
&\quad\times \int_0^\zeta \left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)\frac{1}{|\theta_1(\xi)|}d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2 \int_0^\infty \left[\int_0^\zeta |\zeta|^\lambda|\kappa(h_1\zeta)|\left(|q_1(w)|\exp(Q_{X_1}(\zeta))+|\phi_V(\zeta,w)|\right)d\zeta\right] \\
&\quad\times \left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)\frac{1}{|\theta_1(\xi)|}d\xi \\
&\preceq \Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h);
\end{aligned}$$

$$\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{10}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||q_1(w)|\exp(Q_{X_1}(\zeta))\int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\Bigg(\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\frac{1}{|\theta_1(\zeta)|^2}\hat{\Phi}_n^2|1+o_p(1)|^{-1} \\
&\quad+\frac{1}{|\theta_1(\zeta)|^2}\hat{\Phi}_n^2|1+o_p(1)|^{-1}\Bigg)|q_1(w)|\exp(Q_{X_1}(\zeta))\int_0^\zeta\left(\frac{1}{|\theta_1(\xi)|}+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right)\hat{\Phi}_n d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1}\frac{1}{|\theta_1(\zeta)|}\left(\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}+1\right)|q_1(w)| \\
&\quad\times\exp(Q_{X_1}(\zeta))\int_0^\zeta\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)\Upsilon(h)\hat{\Phi}_n d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n\Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1} \\
&\quad\times\Bigg[\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)|}|\phi_V(\zeta,w)|\int_0^\zeta\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)d\xi d\zeta \\
&\qquad+\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\frac{|q_1(w)|}{|\theta_1(\zeta)|}\exp(Q_{X_1}(\zeta))\int_0^\zeta\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)d\xi d\zeta\Bigg] \\
&= o_p(1)\Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h);
\end{aligned}$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{11}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)| \frac{1}{|f_W(w)|} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right)\hat{\Phi}_n d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2 |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)| \frac{1}{|f_W(w)|} \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta \\
&= \Upsilon(h)\hat{\Phi}_n^2 |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)|\right) \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta \\
&= \Upsilon(h)\hat{\Phi}_n^2 \Psi^+_{V,\lambda}(h)(1+o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{12}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(\frac{1}{|\theta_1(\zeta)|}+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|^2}\right)\hat{\Phi}_n \frac{1}{|f_W(w)|^2}\hat{\Phi}_n \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta \left(\frac{1}{|\theta_1(\xi)|}+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right)\hat{\Phi}_n d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \Upsilon^2(h)\hat{\Phi}_n^3 \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \left(1+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\right) \frac{1}{|f_W(w)|} \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta \\
&= \Upsilon^2(h)\hat{\Phi}_n^3 \\
&\quad \times \Bigg[\sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \frac{1}{|f_W(w)|} \exp(Q_{X_1}(\zeta)) \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta \\
&\qquad + \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta) \left(\sup_{w\in\text{supp}(W)} |\phi_V(\zeta,w)|\right) \int_0^\zeta \frac{1}{|\theta_1(\xi)|}\left(1+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta\Bigg] \\
&= o_p(1)\Upsilon(h)\hat{\Phi}_n^2 \Psi^+_{V,\lambda}(h);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{13}| \\
&\preceq \sup_{w\in\mathrm{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \sup_{w\in\mathrm{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \bigg(\frac{|\chi_V(\zeta,w)||\delta\hat{\theta}_1(\zeta)|}{|\theta_1(\zeta)|^2} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \\
&\quad + \frac{|\delta\hat{\chi}_V(\zeta,w)|}{|\theta_1(\zeta)|} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \bigg) |\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \sup_{w\in\mathrm{supp}(W)} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \bigg(\frac{|\chi_V(\zeta,w)||\delta\hat{\theta}_1(\zeta)|}{|\theta_1(\zeta)|^2} + \frac{|\delta\hat{\chi}_V(\zeta,w)|}{|\theta_1(\zeta)|} \bigg) \\
&\quad \times |\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \sup_{w\in\mathrm{supp}(W)} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n (1+o_p(1)) \left(\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{12}| \right) \\
&= o_p(1) \left(\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{12}| \right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{14}| \\
&\preceq \sup_{w\in\mathrm{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||\delta_2\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \sup_{w\in\mathrm{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)| \frac{|\delta\hat{f}_W(w)|}{|f_W(w)|^2} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&= \sup_{w\in\mathrm{supp}(W)} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)| d\xi d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n (1+o_p(1)) \left(\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{11}| \right) \\
&= o_p(1) \left(\sup_{(x,w)\in\mathrm{supp}(X,W)} |R_{11}| \right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{15}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)|\frac{|\delta\hat{f}_W(w)|}{|f_W(w)|^2}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1))\left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{12}|\right) \\
&= o_p(1)\left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{12}|\right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{16}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\left(\frac{|\chi_V(\zeta,w)||\delta\hat{\theta}_1(\zeta)|}{|\theta_1(\zeta)|^2}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\right. \\
&\quad \left.+ \frac{|\delta\hat{\chi}_V(\zeta,w)|}{|\theta_1(\zeta)|}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\right)|\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)|\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta)) \\
&\quad \times \int_0^\zeta |\delta_1\hat{q}_{X_1}(\xi)|d\xi d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1))\left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{15}|\right) \\
&= o_p(1)\left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{15}|\right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{17}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||q_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\bigg(\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}\frac{1}{|\theta_1(\zeta)|^2}\hat{\Phi}_n^2|1+o_p(1)|^{-1} \\
&\quad + \frac{1}{|\theta_1(\zeta)|^2}\hat{\Phi}_n^2|1+o_p(1)|^{-1}\bigg)|q_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)|}\bigg(\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|}+1\bigg) \\
&\quad \times |q_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1}\bigg(\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)|}|\phi_V(\zeta,w)|d\zeta \\
&\quad + \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{|q_1(w)|}{|\theta_1(\zeta)|}\exp(Q_{X_1}(\zeta))d\zeta\bigg) \\
&= \Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h)(1+o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{18}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\left(\frac{1}{|\theta_1(\zeta)|}+\frac{|\chi_V(\zeta,w)|}{|\theta_1(\zeta)|^2}\right)\hat{\Phi}_n\frac{1}{|f_W(w)|}\Upsilon(h)\hat{\Phi}_n\exp(Q_{X_1}(\zeta))d\zeta \\
&= \Upsilon(h)\hat{\Phi}_n^2\bigg(\sup_{w\in\text{supp}(W)}\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)||f_W(w)|}\exp(Q_{X_1}(\zeta))d\zeta \\
&\quad + \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)|}\left(\sup_{w\in\text{supp}(W)}|\phi_V(\zeta,w)|\right)d\zeta\bigg) \\
&\preceq \Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{19}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\Bigg(\frac{|\chi_V(\zeta,w)||\delta\hat{\theta}_1(\zeta)|}{|\theta_1(\zeta)|^2} \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1} \\
&\quad + \frac{|\delta\hat{\chi}_V(\zeta,w)|}{|\theta_1(\zeta)|}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\Bigg)|\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1)) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{18}| \right) \\
&= o_p(1) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{18}| \right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{20}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)||\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||q_V(\zeta,w)|\frac{1}{|f_W(w)|^2}\Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1}\exp(Q_{X_1}(\zeta))d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)|\frac{1}{|f_W(w)|}|\phi_V(\zeta,w)|d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda}(h)(1+o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{21}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_2\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)|\frac{|\delta\hat{f}_W(w)|}{|f_W(w)|^2}\Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1}\exp(Q_{X_1}(\zeta))d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n|1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_1\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)|\exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1)) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{18}| \right) \\
&= o_p(1) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{18}| \right);
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{(x,w)\in\text{supp}(X,W)} |R_{22}| \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||\delta_2\hat{q}_1(w)| \exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \sup_{w\in\text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)| \frac{|\delta\hat{f}_W(w)|}{|f_W(w)|^2} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \exp(Q_{X_1}(\zeta))d\zeta \\
&= \sup_{w\in\text{supp}(W)} \Upsilon(h)\hat{\Phi}_n |1+o_p(1)|^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h_1\zeta)||\delta_2\hat{q}_V(\zeta,w)||\delta_1\hat{q}_1(w)| \exp(Q_{X_1}(\zeta))d\zeta \\
&\preceq \Upsilon(h)\hat{\Phi}_n(1+o_p(1)) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{19}| \right) \\
&= o_p(1) \left(\sup_{(x,w)\in\text{supp}(X,W)} |R_{19}| \right). \quad \square
\end{aligned}
$$

Proof of Corollary 4.5 Combining Theorem 4.2, Theorem 4.3 and Theorem 4.4(ii) immediately yields the result. $\square$

Proof of Corollary 4.6 Because the bias and the remainder term will never dominate the variance term by Assumption 4.11, the result immediately follows from Theorem 4.3, Theorem 4.4(i) and the fact that $\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w) = B_{V,\lambda}(x,w,h_{1n}) + L_{V,\lambda}(x,w,h_n) + R_{V,\lambda}(x,w,h_n)$. $\square$

Lemma A.4 *Suppose the conditions of Lemma 4.1 hold. For each ζ and $h \equiv (h_1,h_2)$, and for $A = 1, X_1, \chi_V, f_W$, let*

$$
\Psi^+_{V,\lambda,A,s}(\zeta,h_1) \equiv \sup_{x\in\text{supp}(X)} \left| \int \Psi_{V,\lambda,A}(\zeta,x,w,h_1)s(x,w)dw \right|,
$$

and define

$$
\Psi^+_{V,\lambda,s}(h) \equiv \sum_{A=1,X_1} \int \Psi^+_{V,\lambda,A,s}(\zeta,h_1)d\zeta + h_2^{-1} \sum_{B=\chi_V,f_W} \int \Psi^+_{V,\lambda,B,s}(\zeta,h_1)d\zeta.
$$

If Assumption 4.9 also holds, then for $h > 0$

$$
\begin{aligned}
\Psi^+_{V,\lambda,s}(h) = O\Bigg(&\max\{(1+h_1^{-1})^{\gamma_1+1}, h_2^{-1}\} \left(1+h_1^{-1}\right)^{\gamma_{\phi s}+\lambda-\gamma_\theta+1} \\
&\times \exp\left((\alpha_{\phi s} 1_{\{\beta_{\phi s}\geq\beta_\theta\}} - \alpha_\theta 1_{\{\beta_{\phi s}\leq\beta_\theta\}})(h_1^{-1})^{\max\{\beta_\theta,\beta_{\phi s}\}} \right) \Bigg).
\end{aligned}
$$

Proof We obtain rates for each term of $\Psi^+_{V,\lambda,s}(h)$. First,

$$
\begin{aligned}
&\Psi^+_{V,\lambda,1,s}(\zeta,h_1)\\
&\equiv \sup_{x\in\mathrm{supp}(X)}\left|\int \Psi_{V,\lambda,1}(\zeta,x,w,h_1)s(x,w)dw\right|\\
&= \sup_{x\in\mathrm{supp}(X)}\Bigg|\int\Bigg(-\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(h_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi\\
&\quad-\frac{1}{2\pi}(-i\zeta)^\lambda\kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\theta_1(\zeta)}\Bigg)s(x,w)dw\Bigg|\\
&= \sup_{x\in\mathrm{supp}(X)}\Bigg|-\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(h_1\xi)\exp(-i\xi x)\left(\int\phi_V(\xi,w)s(x,w)dw\right)d\xi\\
&\quad-\frac{1}{2\pi}(-i\zeta)^\lambda\kappa(h_1\zeta)\exp(-i\zeta x)\frac{1}{\theta_1(\zeta)}\left(\int\phi_V(\zeta,w)s(x,w)dw\right)\Bigg|\\
&\preceq \frac{|\theta_{X_1}(\zeta)|}{|\theta_1(\zeta)|^2}\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)|\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\xi,w)s(x,w)dw\right|\right)d\xi\\
&\quad+|\zeta|^\lambda|\kappa(h_1\zeta)|\frac{1}{|\theta_1(\zeta)|}\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)\\
&= \frac{1}{|\theta_1(\zeta)|}\Bigg[|D_\zeta\ln\phi_1(\zeta)|\int_\zeta^{\pm\infty}|\xi|^\lambda|\kappa(h_1\xi)|\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\xi,w)s(x,w)dw\right|\right)d\xi\\
&\quad+|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)\Bigg]\\
&\preceq \frac{1}{|\theta_1(\zeta)|}\Bigg[|D_\zeta\ln\phi_1(\zeta)|\int_\zeta^{\pm\infty}|\xi|^\lambda 1_{\{|\xi|\le\bar{\xi}h_1^{-1}\}}\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\xi,w)s(x,w)dw\right|\right)d\xi\\
&\quad+|\zeta|^\lambda 1_{\{|\zeta|\le\bar{\zeta}h_1^{-1}\}}\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)\Bigg]\\
&\preceq \frac{1}{|\theta_1(\zeta)|}1_{\{|\zeta|\le h_1^{-1}\}}\Bigg[|D_\zeta\ln\phi_1(\zeta)|\int_\zeta^{h_1^{-1}}|\xi|^\lambda\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\xi,w)s(x,w)dw\right|\right)d\xi\\
&\quad+|\zeta|^\lambda\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)\Bigg].
\end{aligned}
$$

By Assumption 4.1 and 4.9, we obtain

$$
\begin{aligned}
&\int\Psi^+_{V,\lambda,1,s}(\zeta,h_1)d\zeta\\
&\preceq\int\frac{1}{|\theta_1(\zeta)|}1_{\{|\zeta|\le h_1^{-1}\}}\Bigg[|D_\zeta\ln\phi_1(\zeta)|\int_\zeta^{h_1^{-1}}|\xi|^\lambda\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\phi_V(\xi,w)s(x,w)dw\right|\right)d\xi
\end{aligned}
$$

$$+ |\zeta|^\lambda \left(\sup_{x\in\mathrm{supp}(X)} \left| \int \phi_V(\zeta,w)s(x,w)dw \right| \right) \Big] d\zeta$$

$$\preceq \int (1+|\zeta|)^{-\gamma_\theta} \exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right) 1_{\{|\zeta|\le h_1^{-1}\}} \Big[(1+|\zeta|)^{\gamma_1} \int_0^{h_1^{-1}} |\xi|^\lambda (1+|\xi|)^{\gamma_{\phi s}}$$

$$\times \exp\left(\alpha_{\phi s}|\xi|^{\beta_{\phi s}}\right) d\xi + |\zeta|^\lambda (1+|\zeta|)^{\gamma_{\phi s}} \exp\left(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}}\right) \Big] d\zeta$$

$$\preceq \int_0^{h_1^{-1}} (1+|\zeta|)^{-\gamma_\theta} \exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right) \Big[(1+|\zeta|)^{\gamma_1} \int_0^{h_1^{-1}} |\xi|^\lambda (1+|\xi|)^{\gamma_{\phi s}} \exp\left(\alpha_{\phi s}|\xi|^{\beta_{\phi s}}\right) d\xi$$

$$+ |\zeta|^\lambda (1+|\zeta|)^{\gamma_{\phi s}} \exp\left(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}}\right) \Big] d\zeta$$

$$\preceq (1+h_1^{-1})^{1-\gamma_\theta} \exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right) \Big[(1+h_1^{-1})^{\gamma_1} (1+h_1^{-1})^{\lambda+\gamma_{\phi s}+1} \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right)$$

$$+ (1+h_1^{-1})^{\lambda+\gamma_{\phi s}} \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right) \Big]$$

$$\preceq (1+h_1^{-1})^{1-\gamma_\theta} \exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right) (1+h_1^{-1})^{\lambda+\gamma_{\phi s}} \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right) \left((1+h_1^{-1})^{\gamma_1+1}+1\right)$$

$$\preceq (1+h_1^{-1})^{\gamma_{\phi s}+\lambda+\gamma_1-\gamma_\theta+2} \exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right) \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right).$$

Second,

$$\begin{aligned}
&\Psi^+_{V,\lambda,X_1,s}(\zeta,h_1)\\
&\equiv \sup_{x\in\mathrm{supp}(X)} \left| \int \Psi_{V,\lambda,X_1}(\zeta,x,w,h_1) s(x,w) dw \right|\\
&= \sup_{x\in\mathrm{supp}(X)} \left| \int \left(\frac{1}{2\pi} \frac{i}{\theta_1(\zeta)} \int_\zeta^{\pm\infty} (-i\xi)^\lambda \kappa(h_1\xi) \exp(-i\xi x) \phi_V(\xi,w) d\xi \right) s(x,w) dw \right|\\
&= \sup_{x\in\mathrm{supp}(X)} \left| \frac{1}{2\pi} \frac{i}{\theta_1(\zeta)} \int_\zeta^{\pm\infty} (-i\xi)^\lambda \kappa(h_1\xi) \exp(-i\xi x) \left(\int \phi_V(\xi,w)s(x,w)dw \right) d\xi \right|\\
&\preceq \frac{1}{|\theta_1(\zeta)|} \int_\zeta^{\pm\infty} |\xi|^\lambda |\kappa(h_1\xi)| \left(\sup_{x\in\mathrm{supp}(X)} \left| \int \phi_V(\xi,w)s(x,w)dw \right| \right) d\xi\\
&= \frac{1}{|\theta_1(\zeta)|} \int_\zeta^{\pm\infty} |\xi|^\lambda 1_{\{|\xi|\le\bar\xi h_1^{-1}\}} \left(\sup_{x\in\mathrm{supp}(X)} \left| \int \phi_V(\xi,w)s(x,w)dw \right| \right) d\xi\\
&\preceq \frac{1}{|\theta_1(\zeta)|} 1_{\{|\zeta|\le h_1^{-1}\}} \int_\zeta^{h_1^{-1}} |\xi|^\lambda \left(\sup_{x\in\mathrm{supp}(X)} \left| \int \phi_V(\xi,w)s(x,w)dw \right| \right) d\xi
\end{aligned}$$

so that

$$\begin{aligned}
&\int \Psi^+_{V,\lambda,X_1,s}(\zeta,h_1) d\zeta\\
&\preceq \int_0^{h_1^{-1}} (1+|\zeta|)^{-\gamma_\theta} \exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right) \left(\int_0^{h_1^{-1}} |\xi|^\lambda (1+|\xi|)^{\gamma_{\phi s}} \exp\left(\alpha_{\phi s}|\xi|^{\beta_{\phi s}}\right) d\xi \right) d\zeta
\end{aligned}$$

$$\preceq (1+h_1^{-1})^{1-\gamma_\theta} \exp\left(-\alpha_\theta (h_1^{-1})^{\beta_\theta}\right) (1+h_1^{-1})^{\lambda+\gamma_{\phi s}+1} \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right)$$
$$\preceq (1+h_1^{-1})^{\gamma_{\phi s}+\lambda-\gamma_\theta+2} \exp\left(-\alpha_\theta (h_1^{-1})^{\beta_\theta}\right) \exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right).$$

Third,

$$\begin{aligned}
&\Psi^+_{V,\lambda,\chi_V,s}(\zeta,h_1)\\
&\equiv \sup_{x\in\text{supp}(X)} \left|\int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,h_1)s(x,w)dw\right|\\
&= \sup_{x\in\text{supp}(X)} \left|\int \left(\frac{1}{2\pi}(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)}\right) s(x,w)dw\right|\\
&\preceq |\zeta|^\lambda 1_{\{|\zeta|\le h_1^{-1}\}} \left(\sup_{x\in\text{supp}(X)} \left|\int \frac{\phi_V(\zeta,w)s(x,w)}{\chi_V(\zeta,w)}dw\right|\right)\\
&\le |\zeta|^\lambda 1_{\{|\zeta|\le h_1^{-1}\}} \left(\sup_{x\in\text{supp}(X)} \left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right)\left(\frac{1}{\inf_{w\in\text{supp}(W)}|\chi_V(\zeta,w)|}\right)
\end{aligned}$$

so that

$$\begin{aligned}
&h_2^{-1}\int \Psi^+_{V,\lambda,\chi_V,s}(\zeta,h_1)d\zeta\\
&\preceq h_2^{-1}\int_0^{h_1^{-1}} |\zeta|^\lambda (1+|\zeta|)^{-\gamma_\theta}\exp\left(-\alpha_\theta|\zeta|^{\beta_\theta}\right)(1+|\zeta|)^{\gamma_{\phi s}}\exp\left(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}}\right)d\zeta\\
&\preceq h_2^{-1}(1+h_1^{-1})^{\gamma_{\phi s}-\gamma_\theta+\lambda+1}\exp\left(-\alpha_\theta(h_1^{-1})^{\beta_\theta}\right)\exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right).
\end{aligned}$$

Finally,

$$\begin{aligned}
\Psi^+_{V,\lambda,f_W,s}(\zeta,h_1) &\equiv \sup_{x\in\text{supp}(X)} \left|\int \Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)s(x,w)dw\right|\\
&= \sup_{x\in\text{supp}(X)} \left|\int \left(-\frac{1}{2\pi}(-i\zeta)^\lambda \kappa(h_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{f_W(w)}\right) s(x,w)dw\right|\\
&\preceq |\zeta|^\lambda 1_{\{|\zeta|\le h_1^{-1}\}} \left(\sup_{x\in\text{supp}(X)} \left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right)
\end{aligned}$$

so that

$$\begin{aligned}
h_2^{-1}\int \Psi^+_{V,\lambda,f_W,s}(\zeta,h_1)d\zeta &\preceq h_2^{-1}\int_0^{h_1^{-1}} |\zeta|^\lambda(1+|\zeta|)^{\gamma_{\phi s}}\exp\left(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}}\right)d\zeta\\
&\preceq h_2^{-1}(1+h_1^{-1})^{\gamma_{\phi s}+\lambda+1}\exp\left(\alpha_{\phi s}(h_1^{-1})^{\beta_{\phi s}}\right).
\end{aligned}$$

Putting four terms together gives the desired result. $\square$

Proof of Theorem 4.7 (i) By the assumption 4.10(i), we have

$$
\begin{aligned}
&\max_{j=1,\ldots,J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w,h_{1n})\right| \\
&= \max_{j=1,\ldots,J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|B_{V_j,\lambda_j}(x,w,h_{1n}) + L_{V_j,\lambda_j}(x,w,h_n) + R_{V_j,\lambda_j}(x,w,h_n)\right| \\
&= o(\alpha_{1n}) + o_p(\alpha_{1n}^{1/2}) + o_p(\alpha_{1n}) \\
&= o_p(\alpha_{1n}^{1/2}).
\end{aligned}
$$

Thus the remainder term in eqn.(1.9) is $o_p\left(\left(\alpha_{1n}^{1/2}\right)^2\right) = o_p(\alpha_{1n})$ by letting $\tilde{g}_{V_j,\lambda_j}(x,w) = \hat{g}_{V_j,\lambda_j}(x,w,h_n)$. We also have

$$
\begin{aligned}
&\left|\sum_{j=1}^{J} \int \left(\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right) s_j(x,w)dw\right| \\
&\leq \sum_{j=1}^{J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right| \int |s_j(x,w)|\,dw \\
&\leq \sum_{j=1}^{J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right| \sup_{x\in \mathrm{supp}(X)} \int |s_j(x,w)|\,dw \\
&= \sum_{j=1}^{J} \left\|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right\|_\infty \sup_{x\in \mathrm{supp}(X)} \int |s_j(x,w)|\,dw \\
&\preceq \left\|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right\|_\infty,
\end{aligned}
$$

since $\sup_{x\in \mathrm{supp}(X)} \int |s_j(x,w)|\,dw < \infty$. Then the result immediately follows.

(ii) By the assumption 4.10(ii), we have

$$
\begin{aligned}
&\max_{j=1,\ldots,J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w,h_{1n})\right| \\
&= \max_{j=1,\ldots,J} \sup_{(x,w)\in \mathrm{supp}(X,W)} \left|B_{V_j,\lambda_j}(x,w,h_{1n}) + L_{V_j,\lambda_j}(x,w,h_n) + R_{V_j,\lambda_j}(x,w,h_n)\right| \\
&= o(\alpha_{2n}) + o_p(\alpha_{2n}^{1/2}) + o_p(\alpha_{2n}) \\
&= o_p(\alpha_{2n}^{1/2}).
\end{aligned}
$$

Thus the remainder term in eqn.(1.9) is $o_p\left(\left(\alpha_{2n}^{1/2}\right)^2\right) = o_p(\alpha_{2n})$ by letting $\tilde{g}_{V_j,\lambda_j}(x,w) = \hat{g}_{V_j,\lambda_j}(x,w,h_n)$. We also have

$$
\begin{aligned}
&\sum_{j=1}^{J}\int\left(\hat{g}_{V_j,\lambda_j}(x,w,h_n)-g_{V_j,\lambda_j}(x,w)\right)s_j(x,w)dw\\
&=\sum_{j=1}^{J}\int B_{V_j,\lambda_j}(x,w,h_{1n})s_j(x,w)dw+\sum_{j=1}^{J}\int L_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\\
&\quad+\sum_{j=1}^{J}\int R_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw.
\end{aligned}
$$

For the first term,

$$
\begin{aligned}
&\sup_{x\in\operatorname{supp}(X)}\left|\sum_{j=1}^{J}\int B_{V_j,\lambda_j}(x,w,h_{1n})s_j(x,w)dw\right|\\
&\leq\sup_{x\in\operatorname{supp}(X)}\sum_{j=1}^{J}\left|\int B_{V_j,\lambda_j}(x,w,h_{1n})s_j(x,w)dw\right|.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\sup_{x\in\operatorname{supp}(X)}\left|\int B_{V,\lambda}(x,w,h_{1n})s(x,w)dw\right|\\
&=\sup_{x\in\operatorname{supp}(X)}\left|\int(g_{V,\lambda}(x,w,h_1)-g_{V,\lambda}(x,w,0))s(x,w)dw\right|\\
&=\sup_{x\in\operatorname{supp}(X)}\left|\int\left(\frac{1}{2\pi}\int\kappa(h_1\zeta)(-i\zeta)^{\lambda}\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta\right.\right.\\
&\qquad\left.\left.-\frac{1}{2\pi}\int(-i\zeta)^{\lambda}\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta\right)s(x,w)dw\right|\\
&=\sup_{x\in\operatorname{supp}(X)}\left|\int\left(\frac{1}{2\pi}\int(\kappa(h_1\zeta)-1)(-i\zeta)^{\lambda}\phi_V(\zeta,w)\exp(-i\zeta x)d\zeta\right)s(x,w)dw\right|\\
&=\sup_{x\in\operatorname{supp}(X)}\left|\frac{1}{2\pi}\int(\kappa(h_1\zeta)-1)(-i\zeta)^{\lambda}\exp(-i\zeta x)\left(\int\phi_V(\zeta,w)s(x,w)dw\right)d\zeta\right|\\
&\leq\frac{1}{\pi}\int_{\bar{\xi}/h_1}^{\infty}|(\kappa(h_1\zeta)-1)|\,|\zeta|^{\lambda}\left(\sup_{x\in\operatorname{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)d\zeta\\
&\preceq\int_{\bar{\xi}/h_1}^{\infty}|\zeta|^{\lambda}\left(\sup_{x\in\operatorname{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right)d\zeta\\
&\preceq\int_{\bar{\xi}/h_1}^{\infty}|\zeta|^{\lambda}(1+|\zeta|)^{\gamma_{\phi s}}\exp(\alpha_{\phi s}|\zeta|^{\beta_{\phi s}})d\zeta\\
&=O\left((\bar{\xi}/h_1)^{\gamma_{\phi s}+\lambda+1}\exp\left(\alpha_{\phi s}\left(\bar{\xi}/h_1\right)^{\beta_{\phi s}}\right)\right)\\
&=O\left(\left(h_1^{-1}\right)^{\gamma_{\lambda,B,s}}\exp\left(\alpha_{B,s}\left(h_1^{-1}\right)^{\beta_{B,s}}\right)\right).
\end{aligned}
$$

Thus we have

$$\sup_{x\in\mathrm{supp}(X)}\left|\sum_{j=1}^{J}\int B_{V_j,\lambda_j}(x,w,h_1)s_j(x,w)dw\right| = O\left(\left(h_1^{-1}\right)^{\gamma_{\lambda,B,s}}\exp\left(\alpha_{B,s}\left(h_1^{-1}\right)^{\beta_{B,s}}\right)\right).$$

For the second term,

$$\begin{aligned}&\sup_{x\in\mathrm{supp}(X)}\left|\sum_{j=1}^{J}\int L_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right|\\ &\le \sup_{x\in\mathrm{supp}(X)}\sum_{j=1}^{J}\left|\int L_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right|.\end{aligned}$$

Note that

$$\begin{aligned}&\sup_{x\in\mathrm{supp}(X)}\left|\int L_{V,\lambda}(x,w,h_n)s(x,w)dw\right|\\
&= \sup_{x\in\mathrm{supp}(X)}\left|\int\int\left[\Psi_{V,\lambda,1}(\zeta,x,w,h_1)\left(\hat{E}[e^{i\zeta X_2}]-E[e^{i\zeta X_2}]\right)\right.\right.\\
&\quad+\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)\left(\hat{E}[X_1e^{i\zeta X_2}]-E[X_1e^{i\zeta X_2}]\right)\\
&\quad+\Psi_{V,\lambda,\chi V}(\zeta,x,w,h_1)\left(\hat{E}[Ve^{i\zeta X_2}k_{h_2}(W-w)]-E[Ve^{i\zeta X_2}k_{h_2}(W-w)]\right)\\
&\quad\left.\left.+\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)\left(\hat{E}[k_{h_2}(W-w)]-E[k_{h_2}(W-w)]\right)\right]d\zeta s(x,w)dw\right|\\
&\le \int\left[\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\Psi_{V,\lambda,1}(\zeta,x,w,h_1)s(x,w)dw\right|\right)\left|\hat{E}[e^{i\zeta X_2}]-E[e^{i\zeta X_2}]\right|\right.\\
&\quad+\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\Psi_{V,\lambda,X_1}(\zeta,x,w,h_1)s(x,w)dw\right|\right)\left|\hat{E}[X_1e^{i\zeta X_2}]-E[X_1e^{i\zeta X_2}]\right|\\
&\quad+\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\Psi_{V,\lambda,\chi V}(\zeta,x,w,h_1)s(x,w)dw\right|\right)\\
&\qquad\times\left(\sup_{w\in\mathrm{supp}(W)}\left|\hat{E}[Ve^{i\zeta X_2}k_{h_2}(W-w)]-E[Ve^{i\zeta X_2}k_{h_2}(W-w)]\right|\right)\\
&\quad+\left(\sup_{x\in\mathrm{supp}(X)}\left|\int\Psi_{V,\lambda,f_W}(\zeta,x,w,h_1)s(x,w)dw\right|\right)\\
&\qquad\left.\times\left(\sup_{w\in\mathrm{supp}(W)}\left|\hat{E}[k_{h_2}(W-w)]-E[k_{h_2}(W-w)]\right|\right)\right]d\zeta\\
&= \int\left[\Psi^{+}_{V,\lambda,1,s}(\zeta,h_1)\left|\hat{E}[e^{i\zeta X_2}]-E[e^{i\zeta X_2}]\right|+\Psi^{+}_{V,\lambda,X_1,s}(\zeta,h_1)\left|\hat{E}[X_1e^{i\zeta X_2}]-E[X_1e^{i\zeta X_2}]\right|\right.\\
&\quad+h_2^{-1}\Psi^{+}_{V,\lambda,\chi V,s}(\zeta,h_1)\end{aligned}$$

$$\times \left(\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w)] - E[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w)] \right| \right)$$

$$+ h_2^{-1} \Psi^{+}_{V,\lambda,f_W,s}(\zeta, h_1) \left(\sup_{w \in \mathrm{supp}(W)} \left| \hat{E}[h_2 k_{h_2}(W-w)] - E[h_2 k_{h_2}(W-w)] \right| \right) \Bigg] d\zeta.$$

Then we have

$$E\left[\sup_{x \in \mathrm{supp}(X)} \left| \int L_{V,\lambda}(x, w, h_n) s(x, w) dw \right| \right]$$

$$\leq \int \Bigg[\Psi^{+}_{V,\lambda,1,s}(\zeta, h_1) E\left\{ \left(\left| \hat{E}[e^{i\zeta X_2} - E[e^{i\zeta X_2}]] \right|^2 \right)^{1/2} \right\}$$

$$+ \Psi^{+}_{V,\lambda,X_1,s}(\zeta, h_1) E\left\{ \left(\left| \hat{E}[X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}]] \right|^2 \right)^{1/2} \right\}$$

$$+ h_2^{-1} \Psi^{+}_{V,\lambda,\chi_V,s}(\zeta, h_1)$$

$$\times E\left\{ \left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w) - E[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w)]] \right) \right|^2 \right)^{1/2} \right\}$$

$$+ h_2^{-1} \Psi^{+}_{V,\lambda,f_W,s}(\zeta, h_1)$$

$$\times E\left\{ \left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[h_2 k_{h_2}(W-w) - E[h_2 k_{h_2}(W-w)]] \right) \right|^2 \right)^{1/2} \right\} \Bigg] d\zeta$$

$$\leq \int \Bigg[\Psi^{+}_{V,\lambda,1,s}(\zeta, h_1) \left\{ E\left(\left| \hat{E}[e^{i\zeta X_2} - E[e^{i\zeta X_2}]] \right|^2 \right) \right\}^{1/2}$$

$$+ \Psi^{+}_{V,\lambda,X_1,s}(\zeta, h_1) \left\{ E\left(\left| \hat{E}[X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}]] \right|^2 \right) \right\}^{1/2}$$

$$+ h_2^{-1} \Psi^{+}_{V,\lambda,\chi_V,s}(\zeta, h_1)$$

$$\times \left\{ E\left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w) - E[Ve^{i\zeta X_2} h_2 k_{h_2}(W-w)]] \right) \right|^2 \right) \right\}^{1/2}$$

$$+ h_2^{-1} \Psi^{+}_{V,\lambda,f_W,s}(\zeta, h_2)$$

$$\times \left\{ E\left(\left| \sup_{w \in \mathrm{supp}(W)} \left(\hat{E}[h_2 k_{h_2}(W-w) - E[h_2 k_{h_2}(W-w)]] \right) \right|^2 \right) \right\}^{1/2} \Bigg] d\zeta$$

$$= \int \Bigg[\Psi^{+}_{V,\lambda,1,s}(\zeta, h_1) \left\{ n^{-1} E\left(\left| e^{i\zeta X_2} - E[e^{i\zeta X_2}] \right|^2 \right) \right\}^{1/2}$$

$$
\begin{aligned}
&+ \Psi^+_{V,\lambda,X_1,s}(\zeta,h_1)\left\{n^{-1}E\left(\left|X_1e^{i\zeta X_2} - E[X_1e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2} \\
&+ h_2^{-1}\Psi^+_{V,\lambda,\chi_V,s}(\zeta,h_1) \\
&\quad\times\left\{n^{-1}E\left(\left|\sup_{w\in\text{supp}(W)}\left(Ve^{i\zeta X_2}h_2k_{h_2}(W-w) - E[Ve^{i\zeta X_2}h_2k_{h_2}(W-w)]\right)\right|^2\right)\right\}^{1/2} \\
&+ h_2^{-1}\Psi^+_{V,\lambda,f_W,s}(\zeta,h_1) \\
&\quad\times\left\{n^{-1}E\left(\left|\sup_{w\in\text{supp}(W)}(h_2k_{h_2}(W-w) - E[h_2k_{h_2}(W-w)])\right|^2\right)\right\}^{1/2}\Bigg]d\zeta \\
&= n^{-1/2}\int\Bigg[\Psi^+_{V,\lambda,1,s}(\zeta,h_1)\left\{E\left(\left|e^{i\zeta X_2} - E[e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2} \\
&+ \Psi^+_{V,\lambda,X_1,s}(\zeta,h_1)\left\{E\left(\left|X_1e^{i\zeta X_2} - E[X_1e^{i\zeta X_2}]\right|^2\right)\right\}^{1/2} \\
&+ h_2^{-1}\Psi^+_{V,\lambda,\chi_V,s}(\zeta,h_1) \\
&\quad\times\left\{E\left(\left|\sup_{w\in\text{supp}(W)}\left(Ve^{i\zeta X_2}h_2k_{h_2}(W-w) - E[Ve^{i\zeta X_2}h_2k_{h_2}(W-w)]\right)\right|^2\right)\right\}^{1/2} \\
&+ h_2^{-1}\Psi^+_{V,\lambda,f_W,s}(\zeta,h_1) \\
&\quad\times\left\{E\left(\left|\sup_{w\in\text{supp}(W)}(h_2k_{h_2}(W-w) - E[h_2k_{h_2}(W-w)])\right|^2\right)\right\}^{1/2}\Bigg]d\zeta \\
&\preceq n^{-1/2}\left(\sum_{A=1,X_1}\int\Psi^+_{V,\lambda,A,s}(\zeta,h_1)d\zeta + h_2^{-1}\sum_{B=\chi_V,f_W}\int\Psi^+_{V,\lambda,B,s}(\zeta,h_1)d\zeta\right) \\
&= n^{-1/2}\Psi^+_{V,\lambda,s}(h),
\end{aligned}
$$

where $\Psi^+_{V,\lambda,s}(h) = O(\max\{(1+h_1^{-1})^{\gamma_1+1}, h_2^{-1}\}(1+h_1^{-1})^{\gamma_{\phi s}+\lambda-\gamma_\theta+1}\exp((\alpha_{\phi s}1_{\{\beta_{\phi s}\geq\beta_\theta\}} - \alpha_\theta 1_{\{\beta_{\phi s}\leq\beta_\theta\}})(h_1^{-1})^{\max\{\beta_\theta,\beta_{\phi s}\}}))$. It follows by Markov's inequality that

$$
\begin{aligned}
&\sup_{x\in\text{supp}(X)}\left|\sum_{j=1}^J\int L_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right| \\
=&O_p\Bigg(n^{-1/2}(\max\{(1+h_1^{-1})^{\gamma_1+1}, h_2^{-1}\})(1+h_1^{-1})^{\gamma_{\phi s}+\lambda-\gamma_\theta+1} \\
&\quad\times\exp((\alpha_{\phi s}1_{\{\beta_{\phi s}\geq\beta_\theta\}} - \alpha_\theta 1_{\{\beta_{\phi s}\leq\beta_\theta\}})(h_1^{-1})^{\max\{\beta_\theta,\beta_{\phi s}\}})\Bigg).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\sup_{x\in\text{supp}(X)}\left|\sum_{j=1}^{J}\int R_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right| \\
&\leq \sup_{x\in\text{supp}(X)}\sum_{j=1}^{J}\left|\int R_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right| \\
&= \sup_{x\in\text{supp}(X)}\sum_{j=1}^{J}\left|\int \sum_{i=1}^{22} R_{ij}s_j(x,w)dw\right|
\end{aligned}
$$

We exploit upper bounds for each term, $\sup_{x\in\text{supp}(X)}\left|\int R_1 s(x,w)dw\right|, i=1,...,22.$

$$
\begin{aligned}
&\sup_{x\in\text{supp}(X)}\left|\int R_1 s(x,w)dw\right| \\
&\preceq \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)\left(\sup_{x\in\text{supp}(X)}\left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right)\exp\left(\left|\delta\bar{Q}_{X_1}(\zeta)\right|\right) \\
&\qquad\times\left(\int_0^\zeta\left|\delta\hat{q}_{X_1}(\xi)\right|d\xi\right)^2 d\zeta \\
&\leq \int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{x\in\text{supp}(X)}\left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right) \\
&\qquad\times\exp\left(\int_0^\zeta\left|\delta\hat{q}_{X_1}(\xi)\right|d\xi\right)\left(\int_0^\zeta\left|\delta\hat{q}_{X_1}(\xi)\right|d\xi\right)^2 d\zeta \\
&\preceq \exp(o_p(1))\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{x\in\text{supp}(X)}\left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right) \\
&\qquad\times\left(\int_0^\zeta\left|\delta\hat{q}_{X_1}(\xi)\right|d\xi\right)\left(\int_0^\zeta\left|\delta\hat{q}_{X_1}(\xi)\right|d\xi\right)d\zeta \\
&\preceq \exp(o_p(1))\int_0^\infty |\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{x\in\text{supp}(X)}\left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right) \\
&\quad\times\int_0^\zeta\left|\left(\frac{\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)}-\frac{\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right)\left(1+\frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\right)^{-1}\right|d\xi \\
&\quad\times\int_0^\zeta\left|\left(\frac{\delta\hat{\theta}_{X_1}(\xi)}{\theta_1(\xi)}-\frac{\theta_{X_1}(\xi)\delta\hat{\theta}_1(\xi)}{(\theta_1(\xi))^2}\right)\left(1+\frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\right)^{-1}\right|d\xi\zeta \\
&\preceq \exp(o_p(1))\Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-2}\int_0^\infty|\zeta|^\lambda|\kappa(h_1\zeta)|\left(\sup_{x\in\text{supp}(X)}\left|\int\phi_V(\zeta,w)s(x,w)dw\right|\right) \\
&\quad\times\int_0^\zeta\left(\frac{1}{|\theta_1(\xi)|}+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right)d\xi d\zeta
\end{aligned}
$$

$$= \exp(o_p(1))\Upsilon(h)\hat{\Phi}_n^2|1+o_p(1)|^{-2}\int_0^\infty \left(\int_\xi^\infty |\zeta|^\lambda |\kappa(h_1\zeta)| \right.$$
$$\left. \times \left(\sup_{x\in\text{supp}(X)}\left|\int \phi_V(\zeta,w)s(x,w)dw\right|\right) d\zeta\right)\left(\frac{1}{|\theta_1(\xi)|}+\frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right) d\xi$$
$$\preceq \Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda,s}(h).$$

we note that

$$\Upsilon(h)\hat{\Phi}_n^2\Psi^+_{V,\lambda,s}(h)$$
$$= \left(\Upsilon(h)\hat{\Phi}_n^2 n^{1/2}\right) n^{-1/2}\Psi^+_{V,\lambda,s}(h)$$
$$= o_p(1)O_p\left(n^{-1/2}(\max\{(1+h_1^{-1})^{\delta_{L,s}}, h_2^{-1}\})(h_1^{-1})^{\gamma_{\lambda,L,s}}\exp\left(\alpha_{L,s}(h_1^{-1})^{\beta_{L,s}}\right)\right)$$
$$= o_p\left(n^{-1/2}(\max\{(1+h_1^{-1})^{\delta_{L,s}}, h_2^{-1}\})(h_1^{-1})^{\gamma_{\lambda,L,s}}\exp\left(\alpha_{L,s}(h_1^{-1})^{\beta_{L,s}}\right)\right).$$

Because all other terms are also bounded by the upper bound for $\sup_{x\in\text{supp}(X)}$ $\left|\int R_1 s(x,w)dw\right|$ as shown in the proof of Theorem 4.4, we have

$$\sup_{x\in\text{supp}(X)}\left|\sum_{j=1}^J \int R_{V_j,\lambda_j}(x,w,h_n)s_j(x,w)dw\right|$$
$$= o_p\left(n^{-1/2}(\max\{(1+h_1^{-1})^{\delta_{L,s}}, h_2^{-1}\})(h_1^{-1})^{\gamma_{\lambda,L,s}}\exp\left(\alpha_{L,s}(h_1^{-1})^{\beta_{L,s}}\right)\right).$$

Thus putting all together gives the desired result. $\square$

Proof of Theorem 4.8 By the assumption 4.11, we have

$$\max_{j=1,\ldots,J}\sup_{(x,w)\in\text{supp}(X,W)}\left|\hat{g}_{V_j,\lambda_j}(x,w,h_n)-g_{V_j,\lambda_j}(x,w,h_{1n})\right|$$
$$= \max_{j=1,\ldots,J}\sup_{(x,w)\in\text{supp}(X,W)}\left|B_{V_j,\lambda_j}(x,w,h_{1n})+L_{V_j,\lambda_j}(x,w,h_n)+R_{V_j,\lambda_j}(x,w,h_n)\right|$$
$$= o(n^{-1/2})+o_p(n^{-1/4})+o_p(n^{-1/2})$$
$$= o_p(n^{-1/4}).$$

Thus the remainder term in eqn.(1.10) is $o_p\left(\left(n^{-1/4}\right)^2\right)+o_p\left(\left(n^{-1/4}\right)^2\right) = o_p(n^{-1/2})$

when we let $\tilde{g}_{V_j,\lambda_j}(x,w) = \hat{g}_{V_j,\lambda_j}(x,w,h_n)$ and $\tilde{f}_W(w) = \hat{f}_W(w)$. We also have

$$\begin{aligned}
&\sum_{j=1}^{J} \int\int \left(\hat{g}_{V_j,\lambda_j}(x,w,h_n) - g_{V_j,\lambda_j}(x,w)\right) s_j(x,w)dwdx \\
&= \sum_{j=1}^{J} \int\int L_{V_j,\lambda_j}(x,w,h_n) s_j(x,w)dwdx \\
&\quad + \sum_{j=1}^{J} \int\int \left(B_{V_j,\lambda_j}(x,w,h_{1n}) + R_{V_j,\lambda_j}(x,w,h_n)\right) s_j(x,w)dwdx.
\end{aligned}$$

Note that

$$\begin{aligned}
&\left|\sum_{j=1}^{J} \int\int \left(B_{V_j,\lambda_j}(x,w,h_{1n}) + R_{V_j,\lambda_j}(x,w,h_n)\right) s_j(x,w)dwdx\right| \\
&\leq \left(\max_{j=1,\ldots,J} \sup_{(x,w)\in(X,W)} \left|B_{V_j,\lambda_j}(x,w,h_{1n}) + R_{V_j,\lambda_j}(x,w,h_n)\right|\right) \sum_{j=1}^{J} \int\int |s_j(x,w)|dwdx \\
&= o_p(n^{-1/2}),
\end{aligned}$$

since $\max_{j=1,\ldots,J} \sup_{(x,w)\in(X,W)} \max\{|B_{V_j,\lambda_j}(x,w,h_{1n})|, |R_{V_j,\lambda_j}(x,w,h_n)|\} = o_p(n^{-1/2})$ and $\int\int |s_j(x,w)|dwdx < \infty$. Therefore we have

$$\begin{aligned}
b(\hat{g}(h_n), \hat{f}(h_n)) - b(g,f) &= \sum_{j=1}^{J} \int\int L_{V_j,\lambda_j}(x,w,h_n) s_j(x,w)dwdx \\
&\quad + \int\int \left(\hat{f}_W(w) - f_W(w)\right) s_{J+1}(x,w)dwdx + o_p(n^{-1/2}).
\end{aligned}$$

We also note that

$$\begin{aligned}
&\sum_{j=1}^{J} \int\int L_{V_j,\lambda_j}(x,w,h_n) s_j(x,w)dwdx + \int\int \left(\hat{f}_W(w) - f_W(w)\right) s_{J+1}(x,w)dwdx \\
&= \left\{ \lim_{\tilde{h}\to 0} \sum_{j=1}^{J} \int\int L_{V_j,\lambda_j}(x,w,\tilde{h}) s_j(x,w)dwdx \right. \\
&\qquad \left. + \lim_{\tilde{h}_2\to 0} \int\int \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)]\right) s_{J+1}(x,w)dwdx \right\} \\
&\quad + \left\{ \lim_{\tilde{h}_2\to 0} \int\int \left(L_{V_j,\lambda_j}(x,w,h_n) - L_{V_j,\lambda_j}(x,w,\tilde{h})\right) s_j(x,w)dwdx \right. \\
&\qquad + \lim_{\tilde{h}_2\to 0} \int\int \left\{ \left(\hat{E}[k_{h_{2n}}(W-w)] - E[k_{h_{2n}}(W-w)]\right) \right. \\
&\qquad\quad \left.\left. - \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)]\right) \right\} s_{J+1}(x,w)dwdx \right\}. \qquad (1.22)
\end{aligned}$$

We will show that the first term in the right-hand side is a standard sample average while the second is asymptotically negligible. By the definition of $L_{V_j,\lambda_j}(x,w,\tilde{h})$ in Lemma 4.1, we have

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\sum_{j=1}^{J}\int\int L_{V_j,\lambda_j}(x,w,\tilde{h})s_j(x,w)dwdx \\
&+\lim_{\tilde{h}_2\to 0}\int\int\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)s_{J+1}(x,w)dwdx \\
&=\lim_{\tilde{h}\to 0}\sum_{j=1}^{J}\int\int\left\{\int\left[\Psi_{V_j,\lambda_j,1}(\zeta,x,w,\tilde{h}_1)\left(\hat{E}[e^{i\zeta X_2}]-E[e^{i\zeta X_2}]\right)\right.\right. \\
&\qquad+\Psi_{V_j,\lambda_j,X_1}(\zeta,x,w,\tilde{h}_1)\left(\hat{E}[X_1e^{i\zeta X_2}]-E[X_1e^{i\zeta X_2}]\right) \\
&\qquad+\Psi_{V_j,\lambda_j,\chi_{V_j}}(\zeta,x,w,\tilde{h}_1)\left(\hat{E}[V_je^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]-E[V_je^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]\right) \\
&\qquad\left.\left.+\Psi_{V_j,\lambda_j,f_W}(\zeta,x,w,\tilde{h}_1)\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)\right]d\zeta\right\}s_j(x,w)dwdx \\
&\quad+\lim_{\tilde{h}_2\to 0}\int\int\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)s_{J+1}(x,w)dwdx.
\end{aligned}$$

Because the assumption that $\bar{\Psi}_{V,\lambda,s}<\infty$ ensures the integrand is absolutely integrable for any given sample, integrals and limits can be interchanged:

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\sum_{j=1}^{J}\int\int L_{V_j,\lambda_j}(x,w,\tilde{h})s_j(x,w)dwdx \\
&+\lim_{\tilde{h}_2\to 0}\int\int\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)s_{J+1}(x,w)dwdx \\
&=\sum_{j=1}^{J}\int\left\{\lim_{\tilde{h}_1\to 0}\int\int\Psi_{V_j,\lambda_j,1}(\zeta,x,w,\tilde{h}_1)s_j(x,w)dwdx\left(\hat{E}[e^{i\zeta X_2}]-E[e^{i\zeta X_2}]\right)\right. \qquad (1.23)\\
&\qquad+\lim_{\tilde{h}_1\to 0}\int\int\Psi_{V_j,\lambda_j,X_1}(\zeta,x,w,\tilde{h}_1)s_j(x,w)dwdx\left(\hat{E}[X_1e^{i\zeta X_2}]-E[X_1e^{i\zeta X_2}]\right) \\
&\qquad+\lim_{\tilde{h}\to 0}\int\int\Psi_{V_j,\lambda_j,\chi_{V_j}}(\zeta,x,w,\tilde{h}_1)s_j(x,w) \\
&\qquad\quad\times\left(\hat{E}[V_je^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]-E[V_je^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]\right)dwdx \\
&\qquad+\lim_{\tilde{h}\to 0}\int\int\Psi_{V_j,\lambda_j,f_W}(\zeta,x,w,\tilde{h}_1)s_j(x,w) \\
&\qquad\quad\left.\times\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)dwdx\right\}d\zeta \\
&\quad+\int\lim_{\tilde{h}_2\to 0}\int\left(\hat{E}[k_{\tilde{h}_2}(W-w)]-E[k_{\tilde{h}_2}(W-w)]\right)s_{J+1}(x,w)dwdx.
\end{aligned}$$

For the first term in the integrand of eqn.(1.23), we have

$$\begin{aligned}
&\lim_{\tilde{h}_1\to 0}\int\int \Psi_{V,\lambda,1}(\zeta,x,w,\tilde{h}_1)s(x,w)dwdx \\
&= \lim_{\tilde{h}_1\to 0}\int\int\Big\{-\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(\tilde{h}_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi \\
&\qquad -\frac{1}{2\pi}(-i\zeta)^\lambda\kappa(\tilde{h}_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\theta_1(\zeta)}\Big\}s(x,w)dwdx \\
&= -\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}\left(\int\exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^\lambda\lim_{\tilde{h}_1\to 0}\kappa(\tilde{h}_1\xi)d\xi \\
&\qquad -\frac{1}{2\pi}\frac{(-i\zeta)^\lambda}{\theta_1(\zeta)}\left(\int\exp(-i\zeta x)\int s(x,w)\phi_V(\zeta,w)dwdx\right)\lim_{\tilde{h}_1\to 0}\kappa(\tilde{h}_1\zeta) \\
&= -\frac{1}{2\pi}\frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2}\int_\zeta^{\pm\infty}\left(\int\exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^\lambda d\xi \\
&\qquad -\frac{1}{2\pi}\frac{(-i\zeta)^\lambda}{\theta_1(\zeta)}\left(\int\exp(-i\zeta x)\int s(x,w)\phi_V(\zeta,w)dwdx\right) \\
&\equiv \Psi_{V,\lambda,1,s}(\zeta).
\end{aligned}$$

Similarly, for the second term, we have

$$\begin{aligned}
&\lim_{\tilde{h}_1\to 0}\int\int \Psi_{V,\lambda,X_1}(\zeta,x,w,\tilde{h}_1)s(x,w)dwdx \\
&= \lim_{\tilde{h}_1\to 0}\int\int\left\{\frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_\zeta^{\pm\infty}(-i\xi)^\lambda\kappa(\tilde{h}_1\xi)\exp(-i\xi x)\phi_V(\xi,w)d\xi\right\}s(x,w)dwdx \\
&= \frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_\zeta^{\pm\infty}\left(\int\exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^\lambda\lim_{\tilde{h}_1\to 0}\kappa(\tilde{h}_1\xi)d\xi \\
&= \frac{1}{2\pi}\frac{i}{\theta_1(\zeta)}\int_\zeta^{\pm\infty}\left(\int\exp(-i\xi x)\int s(x,w)\phi_V(\xi,w)dwdx\right)(-i\xi)^\lambda d\xi \\
&\equiv \Psi_{V,\lambda,X_1,s}(\zeta).
\end{aligned}$$

We also note that for the third term,

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\int\int \Psi_{V,\lambda,\chi_V}(\zeta,x,w,\tilde{h}_1)s(x,w) \\
&\qquad\qquad \times\left(\hat{E}[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]-E[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]\right)dwdx \\
&= \lim_{\tilde{h}\to 0}\int\int\left\{\frac{1}{2\pi}(-i\zeta)^\lambda\kappa(\tilde{h}_1\zeta)\exp(-i\zeta x)\frac{\phi_V(\zeta,w)}{\chi_V(\zeta,w)}\right\}s(x,w) \\
&\qquad\qquad \times\left(\hat{E}[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]-E[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]\right)dwdx \\
&= \frac{1}{2\pi}(-i\zeta)^\lambda\int\exp(-i\zeta x)\lim_{\tilde{h}_2\to 0}\int\frac{1}{\chi_V(\zeta,w)}s(x,w)\phi_V(\zeta,w) \\
&\qquad\qquad \times\left(\hat{E}[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]-E[Ve^{i\zeta X_2}k_{\tilde{h}_2}(W-w)]\right)dwdx\lim_{\tilde{h}_1\to 0}\kappa(\tilde{h}_1\zeta)
\end{aligned}$$

$$= \frac{1}{2\pi}(-i\zeta)^{\lambda} \int \exp(-i\zeta x) \lim_{\tilde{h}_2 \to 0} \int \frac{1}{\chi_V(\zeta, w)} s(x,w)\phi_V(\zeta, w) \\ \times \left(\hat{E}[Ve^{i\zeta X_2} k_{\tilde{h}_2}(W-w)] - E[Ve^{i\zeta X_2} k_{\tilde{h}_2}(W-w)] \right) dwdx \\ \equiv \hat{E}[\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta; V, X_2, W)] - E[\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta; V, X_2, W)],$$

and for the fourth term,

$$\lim_{\tilde{h} \to 0} \int \int \Psi_{V,\lambda,f_W}(\zeta, x, w, \tilde{h}_1) s(x,w) \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) dwdx \\ = \lim_{\tilde{h} \to 0} \int \int \left\{ \frac{-1}{2\pi}(-i\zeta)^{\lambda} \kappa(\tilde{h}_1 \zeta) \exp(-i\zeta x) \frac{\phi_V(\zeta, w)}{f_W(w)} \right\} s(x,w) \\ \times \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) dwdx \\ = -\frac{1}{2\pi}(-i\zeta)^{\lambda} \int \exp(-i\zeta x) \lim_{\tilde{h}_2 \to 0} \int \frac{1}{f_W(w)} s(x,w)\phi_V(\zeta, w) \\ \times \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) dwdx \lim_{\tilde{h}_1 \to 0} \kappa(\tilde{h}_1 \zeta) \\ = -\frac{1}{2\pi}(-i\zeta)^{\lambda} \int \exp(-i\zeta x) \lim_{\tilde{h}_2 \to 0} \int \frac{1}{f_W(w)} s(x,w)\phi_V(\zeta, w) \\ \times \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) dwdx \\ \equiv \hat{E}[\mathcal{Z}_{V,\lambda,f_W}(s,\zeta; W)] - E[\mathcal{Z}_{V,\lambda,f_W}(s,\zeta; W)],$$

where $\mathcal{Z}_{V,\lambda,\chi_V}(s,\zeta; V, X_2, W)$ and $\mathcal{Z}_{V,\lambda,f_W}(s,\zeta; W)$ are defined in the statement of the theorem.

Thus it follows that

$$\lim_{\tilde{h} \to 0} \sum_{j=1}^{J} \int \int L_{V_j,\lambda_j}(x, w, \tilde{h}) s_j(x,w) dwdx \\ + \lim_{\tilde{h}_2 \to 0} \int \int \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) s_{J+1}(x,w) dwdx \\ = \sum_{j=1}^{J} \int \Bigg\{ \Psi_{V_j,\lambda_j,1,s_j}(\zeta) \left(\hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right) \\ + \Psi_{V_j,\lambda_j,X_1,s_j}(\zeta) \left(\hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}] \right) \\ + \left(\hat{E}[\mathcal{Z}_{V_j,\lambda_j,\chi_{V_j}}(s_j,\zeta; V_j, X_2, W)] - E[\mathcal{Z}_{V_j,\lambda_j,\chi_{V_j}}(s_j,\zeta; V_j, X_2, W)] \right) \\ + \left(\hat{E}[\mathcal{Z}_{V_j,\lambda_j,f_W}(s_j,\zeta; W)] - E[\mathcal{Z}_{V_j,\lambda_j,f_W}(s_j,\zeta; W)] \right) \Bigg\} d\zeta \\ + \int \lim_{\tilde{h}_2 \to 0} \int \left(\hat{E}[k_{\tilde{h}_2}(W-w)] - E[k_{\tilde{h}_2}(W-w)] \right) s_{J+1}(x,w) dwdx$$

$$= \hat{E}\left[\sum_{j=1}^{J} \psi_{V_j,\lambda_j}(s_j; V_j, X_1, X_2, W) + \psi_f(s_{J+1}; W)\right]$$
$$= \hat{E}\left[\psi_s(V, X_1, X_2, W)\right],$$

as defined in the statement of the theorem. The assumption that $\bar{\Psi}_{V,\lambda,s} < \infty$ ensures that for some $C < \infty$,

$$|\psi_s(v, x_1, x_2, \tilde{w})| \leq C \max\{1, |x_1|\} \bar{\Psi}_{V,\lambda,s}.$$

Since $E[X_1^2] < \infty$ by Assumption 4.2, and $E\left[|\psi_s(V, X_1, X_2, W)|^2\right] < \infty$, the Lindeberg-Levy central limit theorem gives that $\hat{E}\left[\psi_s(V, X_1, X_2, W)\right]$ is $\sqrt{n}$ consistent and asymptotically normal.

The second term of eqn.(1.22) can be shown to be $o_p(n^{-1/2})$ because it can be written as an h_n-dependent sample average $\hat{E}\left[\bar{\psi}_s(V, X_1, X_2, W, h_n)\right]$ where $\bar{\psi}_s(V, X_1, X_2, W, h_n)$ is such that $\lim_{h_n \to 0} E\left[\left|\bar{\psi}_s(V, X_1, X_2, W, h_n)\right|^2\right] = 0$. The similar procedure to the case of $\hat{E}\left[\psi_s(V, X_1, X_2, W)\right]$ is used just by replacing $\kappa\left(\tilde{h}_1\xi\right)$ by $\left(\kappa\left(h_{1n}\xi\right) - \kappa\left(\tilde{h}_1\xi\right)\right)$ and $k_{\tilde{h}_2}(\cdot)$ by $\left(k_{h_{2n}}(\cdot) - k_{\tilde{h}_2}(\cdot)\right)$, and taking the limit as $h_n \equiv (h_{1n}, h_{2n}) \to 0$ and $\tilde{h} \equiv (\tilde{h}_1, \tilde{h}_2) \to 0$. $\square$

Proof of Theorem 4.9 From a first-order Taylor expansion of $\hat{\beta}(x, w, h_n) - \beta(x, w)$ in $\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)$, we get

$$\hat{\beta}(x, w, h_n) - \beta(x, w) = \sum_{V=1,Y} \sum_{\lambda=0,1} s_{V,\lambda}(x, w)\left(\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)\right) \tag{1.24}$$
$$+ R_{V,\lambda}\left(\bar{g}_{V,\lambda}(x, w, h_n), \left(\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)\right)\right),$$

where $R_{V,\lambda}\left(\bar{g}_{V,\lambda}(x, w, h_n), \left(\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)\right)\right)$ is a remainder term in which $\bar{g}_{V,\lambda}(x, w, h_n)$ lies between $\hat{g}_{V,\lambda}(x, w, h_n)$ and $g_{V,\lambda}(x, w)$ for each (x, w, h_n), and the $s_{V,\lambda}(x, w)$ are given in the statement of Theorem 4.10.

We note that by Corollary 4.5,

$$\max_{V=1,Y} \max_{\lambda=0,1} \sup_{(x,w)\in(X,W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)| = O_p(\varepsilon_n),$$
$$\varepsilon_n \equiv \left(h_{1n}^{-1}\right)^{\gamma_{1,B}} \exp\left(\alpha_B \left(h_{1n}^{-1}\right)^{\beta_B}\right)$$
$$+ n^{-1/2}(\max\{(h_{1n}^{-1})^{\delta_L}, h_{2n}^{-1}\}) \left(h_{1n}^{-1}\right)^{\gamma_{1,L}} \exp\left(\alpha_L \left(h_{1n}^{-1}\right)^{\beta_L}\right) \to 0.$$

The first terms in the Taylor expansion of $\hat{\beta}(x, w, h_n) - \beta(x, w)$ can be shown to be $O_p(\varepsilon_n/\tau_n^3)$ uniformly for $(x, w) \in \Gamma_\tau$. Each term of $s_{V,\lambda}(x, w)$ consists of products of functions of the form $g_{V,\lambda}(x,w)$ divided by products of at most 3 functions of the form $g_{1,0}(x,w)$. Because $g_{V,\lambda}(x,w)$ are uniformly bounded over $\mathbb{R}$ by assumption and $g_{1,0}(x,w)$ are bounded below by τ_n uniformly for $(x,w) \in \Gamma_\tau$ by construction, we have that $\sup_{(x,w)\in\Gamma_\tau} |s_{V,\lambda}(x,w)\,(\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w))\,| = O(1)O_p(\tau_n^{-3})O_p(\varepsilon_n) = O_p(\varepsilon_n/\tau_n^3)$.

The remainder terms in the Taylor expansion of $\hat{\beta}(x,w,h_n) - \beta(x,w)$ can be shown to be $o_p(\varepsilon_n/\tau_n^3)$ uniformly for $(x,w) \in \Gamma_\tau$. These terms involve a finite sum of (i) finite products of the functions $\bar{g}_{V,\lambda}(x,w,h_n)$ for $V = 1, Y$ and $\lambda = 0, 1$; (ii) division by a product of at most 4 functions of the form $\bar{g}_{1,0}(x,w,h_n)$; and (iii) pairwise products of functions of the form $\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)$ for $V = 1, Y$ and $\lambda = 0, 1$. First, the contribution of (i) is bounded in probability uniformly for $(x,w) \in \Gamma_\tau$ because

$$
\begin{aligned}
|\bar{g}_{V,\lambda}(x,w,h_n)| &\le |g_{V,\lambda}(x,w)| + |\bar{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)| \\
&\le |g_{V,\lambda}(x,w)| + |\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)| \\
&\le |g_{V,\lambda}(x,w)| + \max_{V=1,Y}\max_{\lambda=0,1}\sup_{(x,w)\in(X,W)} |\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)| \\
&= O_p(1) + o_p(1) \\
&= O_p(1).
\end{aligned}
$$

Second, the contribution of (ii) is bounded as well. We note that for $(x,w) \in \Gamma_\tau$

$$
\begin{aligned}
\bar{g}_{1,0}(x,w,h_n) &= g_{1,0}(x,w)\left(1 + \frac{\bar{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)}{g_{1,0}(x,w)}\right) \\
&= f_{X|W}(x \mid w)\left(1 + \frac{\bar{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)}{f_{X|W}(x \mid w)}\right) \\
&= f_{X|W}(x \mid w)\left(1 + O_p\left(\frac{\varepsilon_n}{\tau_n}\right)\right).
\end{aligned}
$$

By selecting $\{\tau_n\}$ such that $\tau_n > 0$, $\tau_n \to 0$ as $n \to \infty$, and $\varepsilon_n/\tau_n^3 \to 0$ we also have $\varepsilon_n/\tau_n \to 0$. Thus we get for $(x,w) \in \Gamma_\tau$

$$
\bar{g}_{1,0}(x,w,h_n) = f_{X|W}(x \mid w)\,(1 + o_p(1)) \ge \tau_n/2
$$

with probability approaching one since $f_{X|W}(x \mid w) \ge \tau_n$ for $(x,w) \in \Gamma_\tau$ by construction. Therefore we have that the contribution of (ii) is $\bar{g}_{1,0}^{-4}(x,w,h_n) = O_p(\tau_n^{-4})$. Finally, the contribution of (iii) is $O_p(\varepsilon_n^2)$. Putting all together, we have

$$
\begin{aligned}
&R_{V,\lambda}\left(\bar{g}_{V,\lambda}(x,w,h_n), (\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w))\right) \\
&\quad = O_p(1)O_p(\tau_n^{-4})O_p(\varepsilon_n^2) = O_p\left(\frac{\varepsilon_n}{\tau_n^3}\right)O_p\left(\frac{\varepsilon_n}{\tau_n}\right) = o_p\left(\frac{\varepsilon_n}{\tau_n^3}\right)
\end{aligned}
$$

so that

$$\sup_{(x,w)\in\Gamma_\tau}\left|\hat{\beta}(x,w,h_n)-\beta(x,w)\right| = O_p\left(\frac{\varepsilon_n}{\tau_n^3}\right)+o_p\left(\frac{\varepsilon_n}{\tau_n^3}\right) = o_p(1). \ \square$$

Proof of Theorem 4.10 We have established the asymptotic normality of $\hat{g}_{V,\lambda}(x,w,h_n)-g_{V,\lambda}(x,w)$ in Corollary 4.6 and we have the Taylor expansion in eqn.(1.24). Thus the result is immediate from the delta method. $\square$

Proof of Theorem 4.11 We prove the theorem by applying Theorem 4.7 and straightforward Talyor expansions. (i) From the definitions of $\hat{\beta}_m(x)$ and $\beta_m(x)$, we have

$$\begin{aligned}
&\sup_{x\in\mathbb{M}}\left|\hat{\beta}_m(x)-\beta_m(x)\right| \\
&=\sup_{x\in\mathbb{M}}\left|\int_{S^w_{\hat{\beta}(\cdot,h_n)}}\left(\hat{\beta}(x,w,h_n)-\beta(x,w)\right)m(w)dw\right| \\
&=\sup_{x\in\mathbb{M}}\left|\int_{S^w_{\hat{\beta}(\cdot,h_n)}}\sum_{V=1,Y}\sum_{\lambda=0,1}m(w)s_{V,\lambda}(x,w)\left(\hat{g}_{V,\lambda}(x,w,h_n)-g_{V,\lambda}(x,w)\right)dw\right|+o_p(1) \\
&=\sup_{x\in\mathbb{M}}\left|\sum_{V=1,Y}\sum_{\lambda=0,1}\int_{S^w_{\hat{\beta}(\cdot,h_n)}}m(w)s_{V,\lambda}(x,w)\left(\hat{g}_{V,\lambda}(x,w,h_n)-g_{V,\lambda}(x,w)\right)dw\right|+o_p(1) \\
&=O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}}\exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad+O_p\left(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}},h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}}\exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\right),
\end{aligned}$$

where the last equality is attained by Theorem 4.7.

(ii) From the definitions of $\hat{\beta}_m(x)$ and $\beta_m(x)$, we have

$$\begin{aligned}
&\sup_{x\in\mathbb{M}}\left|\hat{\beta}_{mf_W}(x)-\beta_{mf_W}(x)\right| \\
&=\sup_{x\in\mathbb{M}}\left|\int_{S^w_{\hat{\beta}(\cdot,h_n)}}\left(\hat{\beta}(x,w,h_n)m(w)\hat{f}_W(w)-\beta(x,w)m(w)f_W(w)\right)dw\right| \\
&=\sup_{x\in\mathbb{M}}\left|\int_{S^w_{\hat{\beta}(\cdot,h_n)}}m(w)f_W(w)\left(\hat{\beta}(x,w,h_n)-\beta(x,w)\right)dw\right. \\
&\quad\left.+\int_{S^w_{\hat{\beta}(\cdot,h_n)}}\beta(x,w)m(w)\left(\hat{f}_W(w)-f_W(w)\right)dw\right|+o_p(1)
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{x \in \mathbb{M}} \Bigg| \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \sum_{V=1,Y} \sum_{\lambda=0,1} m(w) f_W(w) s_{V,\lambda}(x,w) \left(\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right) dw \\
&\quad + \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w) m(w) \left(\hat{f}_W(w) - f_W(w)\right) dw \Bigg| + o_p(1) \\
&= \sup_{x \in \mathbb{M}} \Bigg| \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} m(w) f_W(w) s_{V,\lambda}(x,w) \left(\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right) dw \\
&\quad + \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w) m(w) \left(\hat{f}_W(w) - f_W(w)\right) dw \Bigg| + o_p(1) \\
&= O\left(\tau^{-3} \left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}} \exp\left(\alpha_{B,s} \left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&\quad + O_p\Bigg(\tau^{-3} n^{-1/2} \left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right) \left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}} \exp\left(\alpha_{L,s} \left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\Bigg).
\end{aligned}
$$

(iii) From the definitions of $\hat{\beta}_{mf_{W|X}}(x)$ and $\beta_{mf_{W|X}}(x)$, we have

$$
\begin{aligned}
&\sup_{x \in \mathbb{M}} \left| \hat{\beta}_{mf_{W|X}}(x) - \beta_{mf_{W|X}}(x) \right| \\
&= \sup_{x \in \mathbb{M}} \Bigg| \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \Bigg(\hat{\beta}(x,w,h_n) m(w) \frac{\hat{g}_{1,0}(x,w,h_n) \hat{f}_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n) dw} \\
&\qquad - \beta(x,w) m(w) \frac{g_{1,0}(x,w) f_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w) dw} \Bigg) dw \Bigg| \\
&= \sup_{x \in \mathbb{M}} \Bigg| \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \Bigg[m(w) \frac{g_{1,0}(x,w) f_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)} \left(\hat{\beta}(x,w,h_n) - \beta(x,w)\right) \\
&\qquad + \beta(x,w) m(w) \frac{f_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w) dw} \left(\hat{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)\right) \\
&\qquad + \beta(x,w) m(w) \frac{g_{1,0}(x,w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w) dw} \left(\hat{f}_W(w) - f_W(w)\right) \\
&\qquad - \beta(x,w) m(w) \frac{g_{1,0}(x,w) f_W(w)}{\left(\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w) dw\right)^2} \\
&\qquad\quad \times \left(\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n) dw - \int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w) dw \right) \Bigg] dw \Bigg| + o_p(1) \\
&= \sup_{x \in \mathbb{M}} \Bigg| \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} m(w) f_{W|X}(w \mid x) s_{V,\lambda}(x,w) \left(\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right) dw \\
&\qquad + \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w) m(w) \frac{f_W(w)}{f_X(x)} \left(\hat{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)\right) dw
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_{S^w_{\hat\beta(\cdot,h_n)}} \beta(x,w)m(w)\frac{f_{X|W}(x \mid w)}{f_X(x)} \left(\hat f_W(w) - f_W(w)\right) dw \\
&- \int_{S^w_{\hat\beta(\cdot,h_n)}} \beta(x,w)m(w)\frac{f_{W|X}(w \mid x)}{f_X(x)} \\
&\quad \times \left(\int_{S^w_{\hat\beta(\cdot,h_n)}} \hat g_{1,0}(x,w,h_n)dw - \int_{S^w_{\hat\beta(\cdot,h_n)}} g_{1,0}(x,w)dw \right) dw \Bigg| + o_p(1) \\
= {} & O\left(\tau^{-3}\left(h_{1n}^{-1}\right)^{\gamma_{1,B,s}} \exp\left(\alpha_{B,s}\left(h_{1n}^{-1}\right)^{\beta_{B,s}}\right)\right) \\
&+ O_p\Bigg(\tau^{-3}n^{-1/2}\left(\max\{(h_{1n}^{-1})^{\delta_{L,s}}, h_{2n}^{-1}\}\right)\left(h_{1n}^{-1}\right)^{\gamma_{1,L,s}} \exp\left(\alpha_{L,s}\left(h_{1n}^{-1}\right)^{\beta_{L,s}}\right)\Bigg). \ \square
\end{aligned}
$$

Proof of Theorem 4.12 Similarly, Theorem 4.8 and Talyor expansions are used for the proof. (i) From the definition of $\hat\beta_{\tilde m}$ and $\beta_{\tilde m}$, we have

$$
\begin{aligned}
&\hat\beta_{\tilde m} - \beta_{\tilde m} \\
&= \int_{S^x_{\hat\beta(\cdot,h_n)}} \int_{S^w_{\hat\beta(\cdot,h_n)}} \left(\hat\beta(x,w,h_n) - \beta(x,w)\right) \tilde m(x,w)dwdx \\
&= \int_{S^x_{\hat\beta(\cdot,h_n)}} \int_{S^w_{\hat\beta(\cdot,h_n)}} \sum_{V=1,Y}\sum_{\lambda=0,1} \tilde m(x,w)s_{V,\lambda}(x,w)\left(\hat g_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right)dwdx \\
&\quad + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1} \int_{S^x_{\hat\beta(\cdot,h_n)}} \int_{S^w_{\hat\beta(\cdot,h_n)}} \tilde m(x,w)s_{V,\lambda}(x,w)\left(\hat g_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right)dwdx \\
&\quad + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1} \hat E\left[\psi_{V,\lambda}(\tilde m s_{V,\lambda}; V, X_1, X_2, W)\right] + o_p(n^{-1/2}) \\
&= \hat E\left[\sum_{V=1,Y}\sum_{\lambda=0,1} \psi_{V,\lambda}(\tilde m s_{V,\lambda}; V, X_1, X_2, W)\right] + o_p(n^{-1/2}).
\end{aligned}
$$

Let $\psi_{\beta_{\tilde m}}(v,x_1,x_2,\tilde w) \equiv \sum_{V=1,Y}\sum_{\lambda=0,1}\psi_{V,\lambda}(\tilde m s_{V,\lambda}; v, x_1, x_2, \tilde w)$. The result is immediate from the application of Theorem 4.8.

(ii) From the definitions of $\hat\beta_{\tilde m f_{W|X}}$ and $\beta_{\tilde m f_{W|X}}$, we have

$$
\begin{aligned}
&\hat\beta_{\tilde m f_{W|X}} - \beta_{\tilde m f_{W|X}} \\
&= \int_{S^x_{\hat\beta(\cdot,h_n)}} \int_{S^w_{\hat\beta(\cdot,h_n)}} \Bigg[\hat\beta(x,w,h_n)\tilde m(x,w)\frac{\hat g_{1,0}(x,w,h_n)\hat f_W(w)}{\int_{S_{\hat\beta(\cdot,h_n)}} \hat g_{1,0}(x,w,h_n)dw}
\end{aligned}
$$

$$
\begin{aligned}
&- \beta(x,w)\tilde{m}(x,w)\frac{g_{1,0}(x,w)f_W(w)}{\int_{S_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw}\Bigg]dwdx \\
&= \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \Bigg[\tilde{m}(x,w)\frac{g_{1,0}(x,w)f_W(w)}{\int_{S_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw}\left(\hat{\beta}(x,w,h_n)-\beta(x,w)\right) \\
&\quad + \beta(x,w)\tilde{m}(x,w)\frac{f_W(w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw}\left(\hat{g}_{1,0}(x,w,h_n)-g_{1,0}(x,w)\right) \\
&\quad + \beta(x,w)\tilde{m}(x,w)\frac{g_{1,0}(x,w)}{\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw}\left(\hat{f}_W(w)-f_W(w)\right) \\
&\quad - \beta(x,w)\tilde{m}(x,w)\frac{g_{1,0}(x,w)f_W(w)}{\left(\int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw\right)^2} \\
&\qquad \times \left(\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n)dw - \int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw\right)\Bigg]dwdx + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1}\int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \tilde{m}(x,w)f_{W|X}(w\mid x)s_{V,\lambda}(x,w) \\
&\qquad \times \left(\hat{g}_{V,\lambda}(x,w,h_n)-g_{V,\lambda}(x,w)\right)dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)\frac{f_W(w)}{f_X(x)}\left(\hat{g}_{1,0}(x,w,h_n)-g_{1,0}(x,w)\right)dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)\frac{f_{X|W}(x\mid w)}{f_X(x)}\left(\hat{f}_W(w)-f_W(w)\right)dwdx \\
&\quad - \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)\frac{f_{W|X}(w\mid x)}{f_X(x)} \\
&\qquad \times \left(\int_{S^w_{\hat{\beta}(\cdot,h_n)}} \hat{g}_{1,0}(x,w,h_n)dw - \int_{S^w_{\hat{\beta}(\cdot,h_n)}} g_{1,0}(x,w)dw\right)dwdx + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1}\hat{E}\left[\psi_{V,\lambda}(\tilde{m}f_{W|X}s_{V,\lambda};V,X_1,X_2,W)\right] + \hat{E}\left[\psi_{1,0}(P_1;1,X_1,X_2,W)\right] \\
&\quad - \hat{E}\left[\psi_{1,0}(P_2;1,X_1,X_2,W)\right] + \hat{E}\left[\psi_f(P_3;W)\right] + o_p(n^{-1/2}) \\
&= \hat{E}\Bigg[\sum_{V=1,Y}\sum_{\lambda=0,1}\psi_{V,\lambda}(\tilde{m}f_{W|X}s_{V,\lambda};V,X_1,X_2,W) + \psi_{1,0}(P_1;1,X_1,X_2,W) \\
&\quad - \psi_{1,0}(P_2;1,X_1,X_2,W) + \psi_f(P_3;W)\Bigg] + o_p(n^{-1/2}).
\end{aligned}
$$

Let $\psi_{\beta_{\tilde{m}f_{W|X}}}(v,x_1,x_2,\tilde{w}) \equiv \sum_{V=1,Y}\sum_{\lambda=0,1}\psi_{V,\lambda}(\tilde{m}f_{W|X}s_{V,\lambda};v,x_1,x_2,\tilde{w})+\psi_{1,0}(P_1;1,x_1,x_2,\tilde{w})-\psi_{1,0}(P_2;1,x_1,x_2,\tilde{w})+\psi_f(P_3;\tilde{w})$ where P_1, P_2, and P_3 are defined in the statement of the theorem. The result is immediate from the application of Theorem 4.8.

(iii) From the definitions of $\hat{\beta}_{\tilde{m}f_{W,X}}$ and $\beta_{\tilde{m}f_{W,X}}$, we have

$$\begin{aligned}
&\hat{\beta}_{\tilde{m}f_{W,X}} - \beta_{\tilde{m}f_{W,X}} \\
&= \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \Big[\hat{\beta}(x,w,h_n)\tilde{m}(x,w)\hat{g}_{1,0}(x,w,h_n)\hat{f}_W(w) \\
&\qquad - \beta(x,w)\tilde{m}(x,w)g_{1,0}(x,w)f_W(w)\Big] dwdx \\
&= \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \tilde{m}(x,w)g_{1,0}(x,w)f_W(w)\left(\hat{\beta}(x,w,h_n) - \beta(x,w)\right) dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)f_W(w)\left(\hat{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)\right) dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)g_{1,0}(x,w)\left(\hat{f}_W(w) - f_W(w)\right) dwdx + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1} \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \tilde{m}(x,w)f_{W,X}(w,x)s_{V,\lambda}(x,w) \\
&\qquad \times \left(\hat{g}_{V,\lambda}(x,w,h_n) - g_{V,\lambda}(x,w)\right) dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)f_W(w)\left(\hat{g}_{1,0}(x,w,h_n) - g_{1,0}(x,w)\right) dwdx \\
&\quad + \int_{S^x_{\hat{\beta}(\cdot,h_n)}} \int_{S^w_{\hat{\beta}(\cdot,h_n)}} \beta(x,w)\tilde{m}(x,w)f_{X|W}(x \mid w)\left(\hat{f}_W(w) - f_W(w)\right) dwdx + o_p(n^{-1/2}) \\
&= \sum_{V=1,Y}\sum_{\lambda=0,1} \hat{E}\left[\psi_{V,\lambda}(\tilde{m}f_{W,X}s_{V,\lambda}; V, X_1, X_2, W)\right] + \hat{E}\left[\psi_{1,0}(\beta\tilde{m}f_W; 1, X_1, X_2, W)\right] \\
&\quad + \hat{E}\left[\psi_f(\beta\tilde{m}f_{X|W}; W)\right] + o_p(n^{-1/2}) \\
&= \hat{E}\Bigg[\sum_{V=1,Y}\sum_{\lambda=0,1} \psi_{V,\lambda}(\tilde{m}f_{W,X}s_{V,\lambda}; V, X_1, X_2, W) + \psi_{1,0}(\beta\tilde{m}f_W; 1, X_1, X_2, W) \\
&\qquad + \psi_f(\beta\tilde{m}f_{X|W}; W)\Bigg] + o_p(n^{-1/2}).
\end{aligned}$$

Let $\psi_{\beta_{\tilde{m}f_{W,X}}}(v,x_1,x_2,\tilde{w}) \equiv \sum_{V=1,Y}\sum_{\lambda=0,1}\psi_{V,\lambda}(\tilde{m}f_{W,X}s_{V,\lambda}; v, x_1, x_2, \tilde{w}) + \psi_{1,0}(\beta\tilde{m}f_W; 1, x_1, x_2, \tilde{w}) + \psi_f(\beta\tilde{m}f_{X|W}; \tilde{w})$. The result is immediate from the application of Theorem 4.8. $\square$

1.9 Tables and Figures

Table 1.1: Monte Carlo simulation designs

Ex	U_x	W	U_1, U_2	U_u	$f_1(X)$	$f_2(W)$
1	N(0, 0.5)	N(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
2	N(0, 0.5)	N(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
3	N(0, 0.5)	N(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
4	N(0, 0.5)	N(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
5	N(0, 0.5)	L(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
6	N(0, 0.5)	L(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
7	N(0, 0.5)	L(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
8	N(0, 0.5)	L(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
9	L(0, 0.5)	N(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
10	L(0, 0.5)	N(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
11	L(0, 0.5)	N(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
12	L(0, 0.5)	N(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
13	L(0, 0.5)	L(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
14	L(0, 0.5)	L(0, 1)	N(0, 0.25)	N(0, 0.09)	erf(x)	S(w)
15	L(0, 0.5)	L(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	erf(w)
16	L(0, 0.5)	L(0, 1)	L(0, 0.25)	N(0, 0.09)	erf(x)	S(w)

Table 1.2: Simulation results from Fourier 1

$h_1 \diagdown h_2$		4	4.25	4.5	4.75	5	5.25	5.5
	Bias^2	0.018	0.004	0.000	0.003	0.025	0.051	0.116
9.5	Variance	0.481	0.496	0.423	0.384	0.385	0.308	0.316
	RMSE	0.706	0.708	0.650	0.622	0.640	0.600	0.657
	Bias^2	0.007	0.002	0.000	0.008	0.032	0.075	0.136
9.75	Variance	0.462	0.406	0.372	0.348	0.322	0.305	0.279
	RMSE	0.684	0.639	0.610	0.597	0.595	0.617	0.644
	Bias^2	0.003	0.000	0.002	0.016	0.043	0.089	0.154
10	Variance	0.425	0.384	0.362	0.349	0.313	0.288	0.263
	RMSE	0.655	0.620	0.604	0.604	0.597	0.614	0.646
	Bias^2	0.001	0.000	0.005	0.024	0.057	0.108	0.173
10.25	Variance	0.404	0.374	0.354	0.340	0.315	0.290	0.257
	RMSE	0.636	0.612	0.599	0.603	0.610	0.630	0.656
	Bias^2	0.000	0.003	0.011	0.033	0.070	0.124	0.201
10.5	Variance	0.392	0.386	0.344	0.321	0.297	0.273	0.258
	RMSE	0.626	0.624	0.596	0.595	0.606	0.631	0.677
	Bias^2	0.001	0.008	0.021	0.048	0.090	0.150	0.231
10.75	Variance	0.368	0.363	0.332	0.310	0.288	0.264	0.249
	RMSE	0.608	0.608	0.594	0.598	0.614	0.643	0.693
	Bias^2	0.007	0.016	0.035	0.068	0.119	0.181	0.264
11	Variance	0.354	0.338	0.320	0.299	0.285	0.254	0.233
	RMSE	0.600	0.596	0.596	0.605	0.635	0.660	0.704
optimal		h_1	h_2	Bias^2	Variance	RMSE		
		9.75	4.9	0.020	0.332	0.594		

Table 1.3: Simulation results from local linear without correction

$h_1 \diagdown h_2$		2	2.25	2.5	2.75	3	3.25	3.5
	Bias^2	0.329	0.468	0.478	0.559	0.577	0.800	1.302
3	Variance	18.22	23.44	12.41	12.52	6.340	9.208	8.231
	RMSE	4.307	4.890	3.589	3.617	2.630	3.164	3.088
	Bias^2	0.178	0.131	0.386	0.599	0.881	0.879	1.145
3.25	Variance	13.90	8.145	2.364	2.238	4.678	0.960	1.132
	RMSE	3.752	2.877	1.658	1.684	2.358	1.356	1.509
	Bias^2	0.601	0.763	0.895	0.878	0.997	0.992	1.559
3.5	Variance	5.349	0.654	1.839	0.661	1.132	3.892	2.753
	RMSE	2.439	1.190	1.653	1.240	1.459	2.210	2.077
	Bias^2	1.037	0.824	1.016	0.994	1.223	1.414	1.552
3.75	Variance	2.284	2.117	0.434	0.563	0.280	0.566	1.049
	RMSE	1.822	1.715	1.204	1.248	1.226	1.407	1.613
	Bias^2	1.041	1.065	1.204	1.189	1.256	1.564	1.715
4	Variance	0.623	1.208	0.225	0.485	1.371	0.603	0.365
	RMSE	1.290	1.508	1.196	1.294	1.621	1.472	1.442
	Bias^2	1.167	1.273	1.142	1.467	1.415	1.393	1.871
4.25	Variance	0.291	0.166	2.474	1.109	0.638	2.815	1.672
	RMSE	1.207	1.200	1.902	1.605	1.433	2.051	1.882
	Bias^2	1.334	1.352	1.393	1.389	1.601	1.743	1.941
4.5	Variance	0.258	0.263	0.165	0.851	0.213	0.648	0.972
	RMSE	1.262	1.271	1.249	1.496	1.347	1.546	1.706
optimal		h_1	h_2	Bias^2	Variance	RMSE		
		3.5	2.55	0.787	0.450	1.112		

Table 1.4: Simulation results from local linear without errors

$h_1 \diagdown h_2$		2	2.25	2.5	2.75	3	3.25	3.5
	Bias^2	0.865	0.538	0.422	0.229	0.126	0.021	0.009
3	Variance	10.74	4.365	3.039	4.041	3.187	1.007	0.901
	RMSE	3.406	2.214	1.861	2.066	1.820	1.014	0.954
	Bias^2	0.039	0.083	0.045	0.015	0.001	0.026	0.127
3.25	Variance	5.378	1.135	0.710	0.586	1.618	0.436	0.860
	RMSE	2.327	1.103	0.869	0.775	1.273	0.680	0.994
	Bias^2	0.004	0.000	0.004	0.016	0.037	0.109	0.203
3.5	Variance	4.982	0.571	0.420	0.415	0.688	0.325	0.792
	RMSE	2.233	0.756	0.651	0.657	0.851	0.659	0.997
	Bias^2	0.036	0.044	0.071	0.087	0.134	0.213	0.324
3.75	Variance	0.549	0.418	0.683	0.362	0.310	0.279	0.454
	RMSE	0.765	0.680	0.869	0.670	0.666	0.701	0.883
	Bias^2	0.087	0.130	0.130	0.147	0.192	0.309	0.445
4	Variance	1.403	0.364	0.298	0.444	0.846	0.238	0.549
	RMSE	1.221	0.703	0.655	0.769	1.019	0.739	0.997
	Bias^2	0.190	0.201	0.231	0.243	0.243	0.401	0.427
4.25	Variance	0.467	0.429	0.341	0.251	2.378	0.214	1.719
	RMSE	0.810	0.793	0.756	0.703	1.619	0.785	1.465
	Bias^2	0.173	0.312	0.296	0.324	0.399	0.488	0.715
4.5	Variance	4.722	3.717	0.228	0.225	0.218	0.201	3.082
	RMSE	2.212	2.007	0.724	0.741	0.785	0.830	1.948
optimal		h_1	h_2	Bias^2	Variance	RMSE		
		3.7	2.55	0.046	0.336	0.618		

Table 1.5: Simulation results from Fourier 2

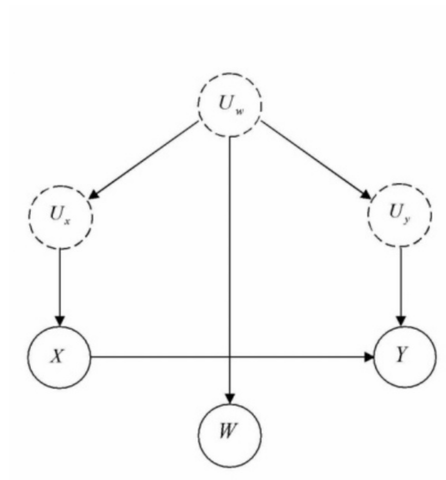
$h_1 \diagdown h_2$		4	4.25	4.5	4.75	5	5.25	5.5
	Bias^2	0.018	0.004	0.000	0.003	0.025	0.051	0.116
9.5	Variance	0.481	0.496	0.423	0.384	0.385	0.308	0.316
	RMSE	0.706	0.708	0.650	0.622	0.640	0.600	0.657
	Bias^2	0.007	0.002	0.000	0.008	0.032	0.075	0.136
9.75	Variance	0.462	0.406	0.372	0.348	0.322	0.305	0.279
	RMSE	0.684	0.639	0.610	0.597	0.595	0.617	0.644
	Bias^2	0.003	0.000	0.002	0.016	0.043	0.089	0.154
10	Variance	0.425	0.384	0.362	0.349	0.313	0.288	0.263
	RMSE	0.655	0.620	0.604	0.604	0.597	0.614	0.646
	Bias^2	0.001	0.000	0.005	0.024	0.057	0.108	0.173
10.25	Variance	0.404	0.374	0.354	0.340	0.315	0.290	0.257
	RMSE	0.636	0.612	0.599	0.603	0.610	0.630	0.656
	Bias^2	0.000	0.003	0.011	0.033	0.070	0.124	0.201
10.5	Variance	0.392	0.386	0.344	0.321	0.297	0.273	0.258
	RMSE	0.626	0.624	0.596	0.595	0.606	0.631	0.677
	Bias^2	0.001	0.008	0.021	0.048	0.090	0.150	0.231
10.75	Variance	0.368	0.363	0.332	0.310	0.288	0.264	0.249
	RMSE	0.608	0.608	0.594	0.598	0.614	0.643	0.693
	Bias^2	0.007	0.016	0.035	0.068	0.119	0.181	0.264
11	Variance	0.354	0.338	0.320	0.299	0.285	0.254	0.233
	RMSE	0.600	0.596	0.596	0.605	0.635	0.660	0.704
optimal		h_1	h_2	Bias^2	Variance	RMSE		
		9.7	4.95	0.023	0.329	0.594		

Table 1.6: Simulation results from Fourier 3

$h_1 \diagdown h_2$		4	4.25	4.5	4.75	5	5.25	5.5
	Bias^2	0.017	0.004	0.000	0.003	0.025	0.052	0.117
9.5	Variance	0.472	0.487	0.415	0.377	0.378	0.302	0.309
	RMSE	0.699	0.701	0.644	0.616	0.635	0.595	0.653
	Bias^2	0.006	0.002	0.000	0.009	0.032	0.077	0.138
9.75	Variance	0.453	0.398	0.365	0.341	0.315	0.299	0.273
	RMSE	0.677	0.633	0.604	0.591	0.590	0.613	0.641
	Bias^2	0.003	0.000	0.002	0.017	0.044	0.091	0.156
10	Variance	0.417	0.376	0.355	0.342	0.307	0.282	0.258
	RMSE	0.648	0.613	0.598	0.599	0.592	0.610	0.643
	Bias^2	0.001	0.000	0.006	0.024	0.058	0.109	0.175
10.25	Variance	0.395	0.367	0.346	0.334	0.309	0.284	0.252
	RMSE	0.630	0.606	0.593	0.598	0.605	0.627	0.653
	Bias^2	0.000	0.003	0.012	0.034	0.071	0.126	0.203
10.5	Variance	0.384	0.378	0.337	0.315	0.291	0.268	0.252
	RMSE	0.620	0.618	0.590	0.590	0.602	0.627	0.674
	Bias^2	0.001	0.008	0.021	0.048	0.091	0.151	0.233
10.75	Variance	0.360	0.355	0.326	0.304	0.282	0.259	0.244
	RMSE	0.601	0.603	0.589	0.594	0.610	0.640	0.690
	Bias^2	0.007	0.017	0.036	0.069	0.120	0.183	0.266
11	Variance	0.347	0.331	0.313	0.292	0.279	0.249	0.227
	RMSE	0.594	0.590	0.591	0.601	0.632	0.657	0.702
optimal		h_1	h_2	Bias^2	Variance	RMSE		
		9.75	4.9	0.021	0.325	0.589		

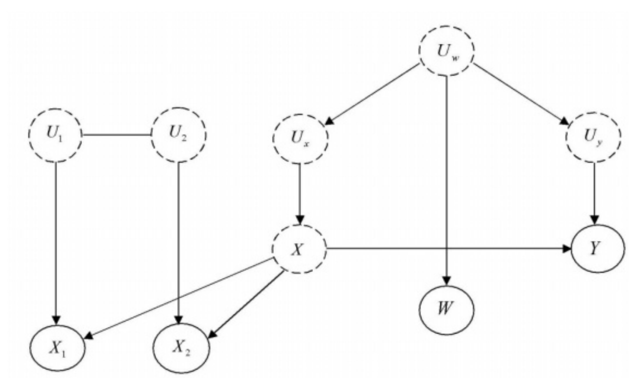
Table 1.7: Monte Carlo simulation results as a function of sample size

Ex ╲	Size	1,000	2,000	8,000	Ex ╲	Size	1,000	2,000	8,000
	h_1	9.75	9.55	9.85		h_1	9.4	9.65	9.9
	h_2	4.9	4.9	4.85		h_2	5.05	4.95	5
1	B^2	0.020	0.001	0.000	9	B^2	0.009	0.003	0.000
	V	0.332	0.090	0.014		V	0.222	0.121	0.012
	R	0.594	0.303	0.120		R	0.481	0.352	0.112
	h_1	11.7	11.35	11.05		h_1	8.8	8.65	8.85
	h_2	4.75	4.85	5		h_2	5.05	5.05	5
2	B^2	0.010	0.001	0.000	10	B^2	0.015	0.003	0.001
	V	0.205	0.053	0.047		V	0.278	0.155	0.065
	R	0.464	0.232	0.219		R	0.541	0.398	0.257
	h_1	11.2	11.55	11.25		h_1	9.55	9.55	9.8
	h_2	5.05	4.9	4.9		h_2	5.1	5.05	5
3	B^2	0.010	0.000	0.000	11	B^2	0.006	0.001	0.000
	V	0.197	0.042	0.029		V	0.176	0.072	0.013
	R	0.455	0.207	0.172		R	0.427	0.271	0.114
	h_1	9.5	9.55	9.7		h_1	8.85	8.75	8.55
	h_2	5	5	4.4		h_2	4.95	5.2	5.15
4	B^2	0.012	0.002	0.000	12	B^2	0.020	0.002	0.001
	V	0.251	0.100	0.014		V	0.328	0.122	0.093
	R	0.513	0.318	0.119		R	0.590	0.353	0.307
	h_1	9.75	9.45	9.3		h_1	8.45	8.45	8.35
	h_2	4.35	4.4	4.4		h_2	4.7	4.75	4.8
5	B^2	0.018	0.001	0.001	13	B^2	0.017	0.003	0.000
	V	0.283	0.091	0.055		V	0.284	0.117	0.025
	R	0.548	0.303	0.236		R	0.549	0.346	0.158
	h_1	8.4	8.6	8.75		h_1	7.25	7.65	7.2
	h_2	4.6	4.4	4.35		h_2	4.7	4.6	4.8
6	B^2	0.023	0.005	0.000	14	B^2	0.028	0.001	0.001
	V	0.344	0.182	0.042		V	0.407	0.092	0.075
	R	0.606	0.432	0.207		R	0.660	0.305	0.276
	h_1	9.1	9.25	9.4		h_1	8.4	8.6	8.3
	h_2	4.6	4.5	4.45		h_2	4.7	4.6	4.7
7	B^2	0.017	0.001	0.000	15	B^2	0.020	0.001	0.000
	V	0.281	0.062	0.023		V	0.306	0.078	0.048
	R	0.546	0.251	0.152		R	0.571	0.282	0.221
	h_1	8.25	8.25	8.25		h_1	7.4	7.65	7.65
	h_2	4.75	4.5	4.7		h_2	4.8	4.65	4.6
8	B^2	0.015	0.001	0.000	16	B^2	0.017	0.002	0.001
	V	0.297	0.099	0.017		V	0.322	0.101	0.074
	R	0.558	0.316	0.130		R	0.582	0.321	0.274



Notes: Arrows denote direct causal relationships. Dashed circles denote unobservables and complete circles denote observables. W, a proxy for common cause U_w could be used conditioning instrument ensuring conditional independence between X and U_y.

Figure 1.1: Causal effects - conditioning instrument



Notes: A line without an arrow denotes dependence arising from a causal relation in either direction or the existence of an underlying common cause. Because true X is unobservable, it becomes a dashed circle. However, error-laden measurements of X help recovering identification of causal relationship.

Figure 1.2: Causal effects - conditioning instrument and measurement error

Chapter 2

Semiparametric Estimation of Models with Conditional Moment Restrictions in the Presence of Nonclassical Measurement Errors

2.1 Introduction

We consider the following models defined by conditional moment restrictions,

$$E[\rho(Z, \theta_0, h_0(\cdot)) \mid X] = 0, \tag{2.1}$$

where $Z \equiv (Y', X_1')'$, $Y \equiv (Y_1, Y_2')'$ is a vector of endogenous (or dependent) variables, X_1 is a subset of conditioning variables $X \equiv (X_1', X_2')'$, $\rho()$ is a vector of generalized residual functions whose functional forms are known up to the unknown vector of finite dimensional parameters θ_0 and the unknown functions ($h_0 \equiv (h_{01}(\cdot), ..., h_{0q}(\cdot))$), where the arguments of each function $h_{0\ell}(\cdot)$, $\ell = 1, ..., q$, may depend on different arguments, and, in particular, may depend on Y. $E[\rho(Z, \theta_0, h_0) \mid X]$ is the conditional expectation of $\rho(Z, \theta_0, h_0)$ given X. Classical model of conditional moment restrictions without the unknown functions h_0 has been exploited considerably in the literature on nonlinear parametric models (see, for instance, Hansen (1982), Chamberlain (1987), Newey (1990, 1993)). There has also been a lot of work on more general frameworks including the unknown function h_0 in the literature on nonparametric and semiparametric models (see, for instance, Robinson (1988), Powell, Stock, and Stoker (1989), Chamberlain (1992), Ichimura (1993)). In their seminal papers, Newey and Powell (2003), and Ai and Chen (2003) study method of sieves when the unknown functions h_0 depend on the endogenous variables. To be specific, they approximate the unkown fuctions h_0 by sieves, and apply the method of minimum distance to estimate parameters of interest. Ai and Chen find that an estimator of h_0 is consistent with a rate faster than $n^{-1/4}$, and that an estimator of the parametric components θ_0 is $\sqrt{n}$ consistent, asymptotically normally distributed, and efficient, while Newey and Powell characterize sufficient identification conditions and propose a consistent estimator for the parameters of interest.

The main contribution of our setup to the literature is that the model (2.1) encompasses the case where the true Y_2, causes of interest, are unobserved due to nonclassical measurement errors on the true Y_2. There have been few works which simultaneously resolve both endogeneity and measurement errors imposed on the same variable of interest in nonparametric and semiparametric models, despite there being a number of empirical observations where endogenous variables are also measurement error-laden. In the returns-to-education literature, for instance, education, the cause of interest, is endogenous in that it is correlated with unobserved ability which is an unobservable driver of income, dependent variable. Moreover, there is often erreneous reporting due to the nature of survey data. In the linear parametric models, the use of valid instruments could resolve issues of identificaiton and estimation associated with measurement errors. However, the existence of valid instruments is not sufficient for the identification and estimation of parameters in nonlinear models, as demonstrated by Amemiya (1985) and Hsiao (1989). As a result, accounting for both endogeneity and measurement errors in nonparametric and semiparametric models is not straightforward.

In this paper, we propose a two-step estimation addressing the aforementioned issues. In the first step, a consistent estimate of the true conditional density of endogenous variables given conditioning variables, which are masked by the nonclassical measurement errors, is obtained. In the second step, a consistent estimate of parameters of interest, $\alpha_0 \equiv (\theta_0, h_0)$, is obtained. For the first-step estimation, we make use of a method proposed by Hu and Schennach (2008), which relies on the eigenvalue-eigenfunction decomposition of an integral operator associated with joint densities of observables, and extend their method to allow for the presence of a vector of additional observable regressors. We also propose a sieve maximum likelihood estimator of conditional densities associated with the unobserved regressors of interest. We then propose a sieve minimum distance estimator of parameters, α_0. Interestingly, we find that one instrument is sufficient to identify and estimate parameters of interest, even when one regressor of interest is endogenous and measurement error-laden. We also show that the sieve estimator of the infinite dimensional unknown functions is consistent with a rate faster than $n^{-1/4}$ under certain metrics, and the sieve estimator of the finite dimensional unknown parameters is $\sqrt{n}$ consistent and asymptotically normally distributed.

The rest of the paper is organized as follows. We describe the proposed two-step estimation in section 2.2. Issues of identification and estimation of distributions in presence of nonclassical measurement errors are discussed in section 2.3. In section 2.4, we prove consistency and $n^{-1/4}$ convergence rates of the parameters from both steps. Asymptotic normality of finite-dimentional parameters of both steps is analyzed in section 2.5. In section 2.6, the finite-sample properties of the estimator are investigated via Monte Carlo studies. Section 2.7 briefly concludes. All technical proofs are included in the Mathematical Appendix.

2.2 Two-Stage Estimation

Let $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_2^*, \mathcal{X}_1$, and $\mathcal{X}_2$ denote the support of the distribution of the random variables Y_1, Y_2, Y_2^*, X_1, and X_2, respectively. Let $Y \equiv (Y_1, Y_2')' \in \mathcal{Y} \equiv \mathcal{Y}_1 \times \mathcal{Y}_2$, $Y^* \equiv (Y_1, Y_2^{*'})' \in \mathcal{Y}^* \equiv \mathcal{Y}_1 \times \mathcal{Y}_2^*$, $X \equiv (X_1', X_2')' \in \mathcal{X} \equiv \mathcal{X}_1 \times \mathcal{X}_2$. Suppose that the true observations $\{(Y_i, X_i) : i = 1, 2, ..., n\}$ are drawn independently from the distribution of (Y, X) with support $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y}$ is a subset of $\mathcal{R}^{d_y}$ and $\mathcal{X}$ is a compact subset of $\mathcal{R}^{d_x}$. Suppose that the unknown distribution of (Y, X) satisfies the conditional moment restriction given by eqn. 2.1, where $\rho : \mathcal{Z} \times \mathcal{A} \to \mathcal{R}^{d_\rho}$ is a known mapping, up to an unknown vector of parameters, $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$. We assume that $\Theta \subseteq \mathcal{R}^{d_\theta}$ is compact with a nonempty interior, and that $\mathcal{H} \equiv \mathcal{H}^1 \times \cdots \times \mathcal{H}^q$ is a space of continuous functions. We further assume that $Z \equiv (Y', X_1')' \in \mathcal{Z} \equiv \mathcal{Y} \times \mathcal{X}_1$ and $\mathcal{X}_1 \subseteq \mathcal{X}$. We use the notation $f_{R_1}(r_1)$, $f_{R_1|R_2}(r_1 \mid r_2)$, and $F_{R_1|R_2}(r_1 \mid r_2)$ to denote the marginal density of variable R_1, the conditional density of R_1 conditional on R_2, and the cumulative density of R_1 conditional on R_2, respectively.

Let $m(x, \alpha) \equiv \int \rho(y, x_1, \theta, h(\cdot)) dF_{Y|X}(y \mid x)$ denote the conditional mean function of the residuals, $\rho(Y, X_1, \theta, h(\cdot))$, given X. Under the assumption that model (1) identifies α_0, one can solve for α_0 as follows:

$$\alpha_0 = \arg \inf_{\alpha=(\theta,h)\in\Theta\times\mathcal{H}} E\left[m(X,\alpha)'[\Sigma(X)]^{-1}m(X,\alpha)\right] \tag{2.2}$$

where $\Sigma(X)$ is a positive-definite matrix for any given X. Because the conditional distribution $F_{Y|X}(y \mid x)$ and conditional mean function $m(x, \alpha)$ are not specified, Newey and Powell (2003) and Ai and Chen (2003) propose a sieve minimum distance (hereafter SMD) estimator that replaces $m(X,\alpha)$ with a consistent nonparametric estimator $\hat{m}(X,\alpha)$ and the function space $\mathcal{H}$ with a sieve space $\mathcal{H}_n \equiv \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$ (Grenander, 1981). However, the method is infeasible in our setup because elements of the true Y (i.e., Y_2) are unobserved so that the empirical distribution of (Y_i, X_i) cannot be used to estimate $m(X,\alpha)$. Instead, we base an estimate of $m(X,\alpha)$ on a sieve maximum likelihood (hereafter SML) estimator of $F_{Y|X}(y \mid x)$. For this, we adapt a method proposed by Hu and Schennach (2008). Let $F_{Y|X}(y \mid x)$ be absolutely continuous with respect to Lebesque measure. To be specific, the conditional mean function can be rewritten as follows: for true values $(\phi_0, \eta_0) \in \Phi \times \mathcal{M}$

$$
\begin{aligned}
m(x,\alpha) &\equiv \int_{\mathcal{Y}} \rho(y, x_1, \theta, h(\cdot)) dF_{Y|X}(y \mid x; \phi_0, \eta_0) \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} \rho(y, x_1, \theta, h(\cdot)) dF_{Y_1|Y_2X}(y_1 \mid y_2, x; \phi_0, \eta_0) \right] dF_{Y_2|X}(y_2 \mid x) \quad (2.3) \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} \rho(y, x_1, \theta, h(\cdot)) f_{Y_1|Y_2X}(y_1 \mid y_2, x; \phi_0, \eta_0) dy_1 \right] f_{Y_2|X}(y_2 \mid x) dy_2 \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} \rho(y, x_1, \theta, h(\cdot)) f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0) dy_1 \right] \\
&\qquad \times f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1) dy_2,
\end{aligned}
$$

where the last equality holds by the exclusion restriction specified in assumption 3.2 in the next section. Note that $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ are the second stage parameters, and $\beta_0 \equiv (\psi_0, f_1, f_2) \in \mathcal{B} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ are the first stage parameters where $\psi_0 \equiv (\phi_0, \eta_0) \in \Psi \equiv \Phi \times \mathcal{M}$ is a vector of parameters of $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0)$, $f_1 \equiv f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1)$, and $f_2 \equiv f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)$. In the first step, we use a SML estimation to estimate $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \phi_0, \eta_0)$ and $f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)$ needed for eqn. 2.3. Then in the second step, the SMD estimator of $\alpha_0 \equiv (\theta_0, h_0)$ minimizes the sample analog of a nonparametric version of (2) with a sieve space $\mathcal{H}_n \equiv \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$ in place of $\mathcal{H}$:

$$
\hat{\alpha}_n = \arg \min_{\alpha=(\theta,h)\in\Theta\times\mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} \hat{m}(X_i,\alpha)' [\hat{\Sigma}(X_i)]^{-1} \hat{m}(X_i,\alpha), \qquad (2.4)
$$

where $\mathcal{H}_n$ is some finite-dimensional approximation space that becomes dense in $\mathcal{H}$ as sample size $n \to \infty$ (e.g., Fourier series, power series, splines, wavelets, etc.),

$\hat{\Sigma}(X)$ is a consistent estimator of $\Sigma(X)$, and $\hat{m}(X,\alpha)$ is the plug-in SML estimator of $m(X,\alpha)$ for any fixed $\alpha = (\theta, h_n)$:

$$\hat{m}(X,\alpha) \equiv \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} \rho(y, x_1, \theta, h_n(\cdot)) \hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \hat{\phi}_n, \hat{\eta}_n) dy_1 \right] \hat{f}_{Y_2|X_2X_1}(y_2 \mid x_2, x_1) dy_2. \tag{2.5}$$

We now introduce useful spaces of smooth functions to analyze how well a sieve can approximate either $\mathcal{H}$ or $\mathcal{M}$. Let $\xi \in \mathcal{V} \subset \mathcal{R}^{d_\xi}$, $\|\cdot\|_E$ denote the Euclidean norm, and

$$\nabla^{\mathbf{a}} \mathrm{g}(\xi) \equiv \frac{\partial^{a_1+a_2+\cdots+a_{d_\xi}} g(\xi)}{\partial \xi_1^{a_1} \cdots \partial \xi_{d_\xi}^{a_{d_\xi}}}$$

denote the $\sum_{i=1}^{d_\xi} a_i$-th derivative where $\mathbf{a} = (a_1, a_2, \ldots, a_{d_\xi})'$ is a vector of non-negative integers. Let $\underline{\gamma}$ denote the largest integer satisfying $\underline{\gamma} < \gamma$. The Hölder space $\Lambda^\gamma(\mathcal{V})$ of order $\gamma > 0$ is a space of functions $\mathrm{g} : \mathcal{V} \mapsto \mathcal{R}$ such that the first $\underline{\gamma}$ derivative is bounded and the $\underline{\gamma}$-th derivatives are Hölder continuous with the exponent $\gamma - \underline{\gamma} \in (0, 1]$, i.e., for all $\xi, \xi' \in \mathcal{V}$ and some constant c

$$\max_{\sum_{i=1}^{d_\xi} a_i = \underline{\gamma}} |\nabla^{\mathbf{a}} \mathrm{g}(\xi) - \nabla^{\mathbf{a}} \mathrm{g}(\xi')| \leq c(\|\xi - \xi'\|_E)^{\gamma - \underline{\gamma}}.$$

The space $\Lambda^\gamma(\mathcal{V})$ becomes a Banach space under the Hölder norm:

$$\|\mathrm{g}\|_{\Lambda^\gamma} = \sup_\xi |\mathrm{g}(\xi)| + \max_{\sum_{i=1}^{d_\xi} a_i = \underline{\gamma}} \sup_{\xi \neq \xi'} \frac{|\nabla^{\mathbf{a}} \mathrm{g}(\xi) - \nabla^{\mathbf{a}} \mathrm{g}(\xi')|}{\|\xi - \xi'\|_E)^{\gamma - \underline{\gamma}}} < \infty.$$

A Hölder ball (of radius c) is defined as $\Lambda_c^\gamma(\mathcal{V}) \equiv \{\mathrm{g} \in \Lambda^\gamma(\mathcal{V}) : \|\mathrm{g}\|_{\Lambda^\gamma} \leq c < \infty\}$. Let $\omega(\cdot)$ be a positive continuous weight function on $\mathcal{V}$ where $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}, \varsigma > \gamma > 0$. Denote $\Lambda_c^{\gamma,\omega}(\mathcal{V})$ as the weighted Hölder space with a weighted Hölder norm $\|\mathrm{g}\|_{\Lambda^{\gamma,\omega}} \equiv \|\tilde{\mathrm{g}}\|_{\Lambda^\gamma}$ for $\tilde{\mathrm{g}}(\xi) \equiv \mathrm{g}(\xi)\omega(\xi)$. Also define a weighted Hölder ball $\Lambda_c^{\gamma,\omega}(\mathcal{V}) \equiv \{\mathrm{g} \in \Lambda^{\gamma,\omega}(\mathcal{V}) : \|\mathrm{g}\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$.

2.3 Identification and Estimation of Distribution

2.3.1 Identification of Distributions

In this section, we consider the identification of two densities, $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \psi)$ and $f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)$. Hu and Schennach (2008) show that the joint distribution of y_1 and y_2 is identified from knowledge of the distribution of all observed variables. For our case, we straightforwardly extend the treatment in Hu and Schennach (2008) to allow for the presence of a vector X_1 of additional observable regressors. We consider Y_2, Y_2^*, and X_2 to be jointly continuously distributed, while Y_1 and X_1 can be either continuous or discrete. We first state a useful note that a function of three variables can be associated with an integral operator.

Definition 3.1 *Let R_1, R_2, R_3 and R_4 denote four random variables with respective supports $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$, and $\mathcal{R}_4$, distributed according to the joint density $f_{R_1R_2R_3R_4}(r_1, r_2, r_3, r_4)$. Given four corresponding spaces $\mathcal{G}(\mathcal{R}_1), \mathcal{G}(\mathcal{R}_2), \mathcal{G}(\mathcal{R}_3)$,* and *$\mathcal{G}(\mathcal{R}_4)$ of functions with domains $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$, and $\mathcal{R}_4$, respectively, let (i) $L_{R_1|R_2r_3}$ denote an integral operator mapping $g \in \mathcal{G}(\mathcal{R}_2)$ to $L_{R_1|R_2r_3}g \in \mathcal{G}(\mathcal{R}_1)$ for a given r_3 defined by*

$$[L_{R_1|R_2r_3}g](r_1) \equiv \int_{\mathcal{R}_2} f_{R_1|R_2R_3}(r_1 \mid r_2, r_3)g(r_2)dr_2; \tag{2.6}$$

(ii) $L_{r_1R_2|R_3r_4}$ denote an integral operator mapping $g \in \mathcal{G}(\mathcal{R}_3)$ to $L_{r_1R_2|R_3r_4}g$ $\in \mathcal{G}(\mathcal{R}_2)$ for a given (r_1, r_4) defined by

$$[L_{r_1R_2|R_3r_4}g](r_2) \equiv \int_{\mathcal{R}_3} f_{R_1R_2|R_3R_4}(r_1, r_2 \mid r_3, r_4)g(r_3)dr_3; \tag{2.7}$$

(iii) $\triangle_{r_1|R_2r_3}$ denote a diagonal operator mapping $g \in \mathcal{G}(\mathcal{R}_2)$ to $\triangle_{r_1|R_2r_3}g \in \mathcal{G}(\mathcal{R}_2)$ for a given (r_1, r_3) defined by

$$\triangle_{r_1|R_2r_3}g \equiv f_{R_1|R_2R_3}(r_1 \mid r_2, r_3)g(r_2). \tag{2.8}$$

For the identification of distributions, we assume following hypotheses. Note that the absence of correctly measured regressors, X_1, draws on simliar assumptions to those in Hu and Schennach (2008).

Assumption 3.1 *(i) The joint density of Y_1 and Y_2, Y_2^*, X_1, X_2 admits a bounded density with respect to the product measure of some dominating measure μ (defined on $\mathcal{Y}_1$) and the Lebesque measure on $\mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{X}_1 \times \mathcal{X}_2$. (ii) All marginal and conditional densities are also bounded.*

Assumption 3.2 *(i) $f_{Y_1|Y_2Y_2^*X_1X_2}(y_1 \mid y_2, y_2^*, x_1, x_2) = f_{Y_1|Y_2X_1X_2}(y_1 \mid y_2, x_1, x_2) = f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1)$ for all $(Y_1, Y_2, Y_2^*, X_1, X_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{X}_1 \times \mathcal{X}_2$ and (ii) $f_{Y_2^*|Y_2X_1X_2}(y_2^* \mid y_2, x_1, x_2) = f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1)$ for all $(Y_2, Y_2^*, X_1, X_2) \in \mathcal{Y}_2 \times \mathcal{Y}_2^* \times \mathcal{X}_1 \times \mathcal{X}_2$.*

Assumption 3.3 *The operators $L_{Y_2^*|Y_2x_1}$ and $L_{Y_2^*|X_2x_1}$ are one-to-one (for either $\mathcal{G} = \mathcal{L}^1$ or $\mathcal{G} = \mathcal{L}^1_{bnd}$ where $\mathcal{L}^1(\mathcal{A})$ is the set of all absolutely integrable functions with domain $\mathcal{A}$ endowed with the norm $||g||_1 = \int_{\mathcal{A}} |g(a)|da$ and where $\mathcal{G} = \mathcal{L}^1_{bnd}$ is the set of functions in $\mathcal{L}^1(\mathcal{A})$ that are also bounded such that $\sup_{a \in \mathcal{A}} |g(a)| < \infty$).*

Assumption 3.4 *For any $x_1 \in \mathcal{X}_1$ and any $\tilde{y}_2, \bar{y}_2 \in \mathcal{Y}_2$, the set $\{y_1 : f_{Y_1|Y_2X_1}(y_1 \mid \tilde{y}_2, x_1) \neq f_{Y_1|Y_2X_1}(y_1 \mid \bar{y}_2, x_1)\}$ has positive probability (under the marginal of Y_1) whenever $\tilde{y}_2 \neq \bar{y}_2$.*

Assumption 3.5 *For any given $x_1 \in \mathcal{X}_1$, there exists a known functional M such that $M[f_{Y_2^*|Y_2X_1}(\cdot \mid y_2, x_1)] = y_2$ for all $y_2 \in \mathcal{Y}_2$.*

A few remarks are in order. Assumption 3.1 restricts all densities to regular bounded densities. Assumption 3.2 states conditional independence restrictions which have been imposed by Altonji and Matzkin (2005), White and Chalak (2006), Chalak and White (2007), and Hoderlein and Mammen (2007), among others. To be specific, Assumption 3.2(i) states that Y_2^*, X_2 do not provide further information on Y_1, given Y_2, X_1. Similarly, Assumption 3.2(ii) indicates that X_2 does not provide further information on Y_2^*, given Y_2, X_1. Assumption 3.3 is associated with restrictions on the relationships between Y_2, Y_2^*, X_2, and X_1, which have been phrased as singular value decompositions with nonzero singular values (Darolles, Florens, and Renault (2002)), nonsingularity (Hall and Horowitz (2005), Horowitz (2006)), and completeness (or bounded completeness) (Newey and Powell (2003), Blundell, Chen, and Kristensen (2007)). Assumption 3.4 states a fairly weak condition which is only violated if the distribution of Y_1 conditional on Y_2, X_1 is identical at different values of Y_2. Assumption 3.5 places restrictions on some measure of the location of a density, denoted by M. The assumption is essential in that it enables the model to include nonclassical measurement errors as well as classical measurement errors.[1] It is invoked by the observation that, even though the measurement error may have a nonzero mean conditional on the true value of the variable, other measures of location, such as the median, mode, or quantile, could be zero. The next theorem provides identificaiton results of unknown distributions.

Theorem 3.1 *Under Assumptions 3.1 – 3.5, given the true observed density* $f_{Y^*|X}(y^* \mid x) \equiv f_{Y_1 Y_2^* | X_1 X_2}(y_1, y_2^* \mid x_1, x_2)$, *the equation*

$$f_{Y^*|X}(y^* \mid x) = \int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1) f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1) f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1) dy_2 \tag{2.9}$$

admits a unique solution $(f_{Y_1|Y_2X_1}, f_{Y_2^*|Y_2X_1}, f_{Y_2|X_2X_1})$ *for all* $y_1 \in \mathcal{Y}_1, y_2^* \in \mathcal{Y}_2^*, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$.

The result is parallel to eqn. 5 in Theorem 1 of Hu and Schennach (2008). The integral equation relates the joint densities of the observables to the product of the joint densities of the unobservables. The identification of unobserved densities enables us to propose the first-stage estimation procedure, and, in turn, to estimate the prameters of interest in the second-stage via estimates of $f_{Y_1|Y_2X_1}$ and $f_{Y_2|X_2X_1}$.

2.3.2 Estimation Using Sieve Maximum Likelihood

Theorem 3.1 implies that β_0 is obtained by the maximization problem:

[1]For instance, M could be the mean, τ quantile, and mode: $M[f] = \int_{\mathcal{Y}_2^*} y_2^* f_{Y_2^*}(y_2^*) dy_2^*$, $M[f] = \inf\left\{y_2 \in \mathcal{Y}_2 : \int 1\{y_2^* \le y_2\} f_{Y_2^*}(y_2^*) dy_2^* \ge \tau\right\}$, and $M[f] = \arg\max_{y_2^* \in \mathcal{Y}_2^*} f_{Y_2^*}(y_2^*)$ respectively.

$$\begin{aligned}\beta_0 &= (\psi_0, f_{Y_2^*|Y_2X_1}, f_{Y_2|X_2X_1})' \\ &= \arg\max_{\beta=(\psi,f_1,f_2)'\in\mathcal{B}} E\bigg(\ln\int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1;\psi) f_1(y_2^* \mid y_2, x_1) \\ &\qquad\qquad\qquad\qquad \times f_2(y_2 \mid x_2, x_1)dy_2\bigg),\end{aligned} \tag{2.10}$$

where $\mathcal{B} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ with $\Psi \equiv \Phi \times \mathcal{M}$. We impose some restrictions on the sets $\mathcal{M}, \mathcal{F}_1$, and $\mathcal{F}_2$ to which the functions η, $f_{Y_2^*|Y_2X_1}$, and $f_{Y_2|X_2X_1}$ belong, respectively, in the following assumptions.

Assumption 3.6 *$\eta \in \Lambda_c^{\gamma_1,\omega}(\mathcal{U})$ where $\gamma_1 > 1$.*

Assumption 3.7 *$f_1 \in \Lambda_c^{\gamma_1,\omega}(\mathcal{Y}_2^* \times \mathcal{Y}_2 \times \mathcal{X}_1)$ where $\gamma_1 > 1$ and $\int_{\mathcal{Y}_2^*} f_1(y_2^* \mid y_2, x_1)dy_2^* = 1$ for all $y_2 \in \mathcal{Y}_2, x_1 \in \mathcal{X}_1)$.*

Assumption 3.8 *$f_2 \in \Lambda_c^{\gamma_1,\omega}(\mathcal{Y}_2 \times \mathcal{X}_2 \times \mathcal{X}_1)$ where $\gamma_1 > 1$ and $\int_{\mathcal{Y}_2} f_2(y_2 \mid x_2, x_1)dy_2 = 1$ for all $x_2 \in \mathcal{X}_2, x_1 \in \mathcal{X}_1)$.*

Then we define three sets as follows:

$$\begin{aligned}\mathcal{M} &= \{\eta(\cdot,\cdot,\cdot) : \text{Assumption 3.6 holds}\}, \\ \mathcal{F}_1 &= \{f_1(\cdot \mid \cdot,\cdot) : \text{Assumption 3.3, 3.5, and 3.7 hold}\}, \\ \mathcal{F}_2 &= \{f_2(\cdot \mid \cdot,\cdot) : \text{Assumption 3.3 and 3.8 hold}\}.\end{aligned}$$

As in eqn. 2.2, the optimization method provides an inconsistent estimator for β_0 or a consisent estimator which converges slowly when the function spaces $\mathcal{M}, \mathcal{F}_1$, and $\mathcal{F}_2$ are large. Thus, we replace $\mathcal{M}, \mathcal{F}_1$, and $\mathcal{F}_2$ with finite-dimentional compact parameter spaces $\mathcal{M}_n, \mathcal{F}_{1n}$, and $\mathcal{F}_{2n}$, respectively, where

$$\begin{aligned}\mathcal{M}_n &= \{\eta(\xi_1,\xi_2,\xi_3) = p^{k_n}(\xi_1,\xi_2,\xi_3)'\delta \text{ for all } \delta \text{ s.t. Assumption 3.6 holds}\}, \\ \mathcal{F}_{1n} &= \{f(y_2^* \mid y_2, x_1) = p^{k_n}(y_2^*, y_2, x_1)'\rho \text{ for all } \rho \text{ s.t. Assumption 3.3, 3.5,} \\ &\qquad \text{and 3.7 hold}\}, \\ \mathcal{F}_{2n} &= \{f(y_2 \mid x_2, x_1) = p^{k_n}(y_2, x_2, x_1)'\pi \text{ for all } \pi \text{ s.t. Assumption 3.3} \\ &\qquad \text{and 3.8 hold}\}.\end{aligned}$$

Let the projection of the true parameter β_0 onto the space $\mathcal{B}_n$ where $\mathcal{B}_n = \Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$ with $\Psi_n = \Phi \times \mathcal{M}_n$:

$$\begin{aligned}\Pi_n\beta &\equiv \beta_n \\ &= \arg \max_{\beta_n=(\psi,f_1,f_2)'\in\mathcal{B}_n} E\Big(\ln \int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \psi) f_1(y_2^* \mid y_2, x_1) \\ &\qquad\qquad \times f_2(y_2 \mid x_2, x_1) dy_2 \Big).\end{aligned}$$

Then a corresponding measurement-error robust sieve maximum likelihood estimator of β_0 maximizes the sample analog of eqn. 2.10 with $\Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ restricted to the sieve space $\Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$:

$$\begin{aligned}\hat{\beta}_n &= (\hat{\psi}_n, \hat{f}_{1n}, \hat{f}_{2n})' \qquad (2.11)\\ &= \arg \max_{(\psi,f_1,f_2)'\in\mathcal{B}_n} \frac{1}{n}\sum_{i=1}^{n} \ln \int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_{1i} \mid y_2, x_{1i}; \psi) f_1(y_{2i}^* \mid y_2, x_{1i}) \\ &\qquad\qquad \times f_2(y_2 \mid x_{2i}, x_{1i}) dy_2.\end{aligned}$$

2.4 Consistency and Convergence Rates

2.4.1 Consistency

In this section, we first obtain consistency of the SML estimator $\hat{\beta}$ for $\beta_0 \equiv (\psi_0, f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1), f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1))$ under a strong metric $\|\cdot\|_{s,\beta}$ and the SMD estimator $\hat{\alpha}$ for $\alpha_0 \equiv (\theta_0, h_0)$ under a strong metric $\|\cdot\|_{s,\alpha}$ by applying the results in Newey and Powell (2003). Following Ai and Chen (2003), we then establish that $\hat{\beta}$ and $\hat{\alpha}$ converge to β_0 and α_0 at a rate faster than $n^{-1/4}$ under suitably constructed weaker metrics $\|\cdot\|_\beta$ and $\|\cdot\|_\alpha$, respectively. Let $(Y^{*'}, X')'$ be a vector of observed variables for $Y^* \in \mathcal{Y}^*, X \in \mathcal{X}$. Define $\|\beta\|_{s,\beta} \equiv \|\phi\|_E + \|\eta\|_{\infty,\omega} + \|f_1\|_{\infty,\omega} + \|f_2\|_{\infty,\omega}$ where $\|g\|_{\infty,\omega} \equiv \sup_\xi |g(\xi)\omega(\xi)|$ with weight function $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}, \varsigma > \gamma_1 > 0$. Note that the meaning of ξ depend on the domain of g (e.g., when $g = f_2$, $\xi = (y_2, x_2, x_1)$).

Assumption 4.1 *(i) The data $\{(Y_i^*, X_i)_{i=1}^n\}$ are i.i.d. (ii) The density of $(Y^{*'}, X')'$, f_{Y^*X}, satisfies $\int \omega(Y^*, X)^{-2} f_{Y^*X}(y^*, x) d(y^*, x) < \infty$.*

Assumption 4.2 *$\alpha_0 \in \mathcal{A}$ is the only $\alpha \in \mathcal{A}$ satisfying $m(X, \alpha) = 0$.*

Assumption 4.3 *(i) $\hat{\Sigma}(X) = \Sigma(X) + o_p(1)$ uniformly over $X \in \mathcal{X}$. (ii) $\Sigma(X)$ is finite positive-definite uniformly over $X \in \mathcal{X}$.*

Assumption 4.4 *(i) There is a metric $\|\cdot\|_{s,\alpha}$ such that $\mathcal{A} \equiv \Theta \times \mathcal{H}$ is compact under $\|\cdot\|_{s,\alpha}$. (ii) For any $\alpha \in \mathcal{A}$, there exists $\Pi_n \alpha \in \mathcal{A}_n \equiv \Theta \times \mathcal{H}_n$ such that $\|\Pi_n \alpha - \alpha\|_{s,\alpha} = o(1)$.*

Assumption 4.5 *(i) There is a metric $\|\cdot\|_{s,\beta}$ such that $\mathcal{B} \equiv \Psi \times \mathcal{F}_1 \times \mathcal{F}_2$ is compact under $\|\cdot\|_{s,\beta}$. (ii) For any $\beta \in \mathcal{B}$, there exists $\Pi_n \beta \in \mathcal{B}_n \equiv \Psi_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$ with $\Psi_n \equiv \Phi \times \mathcal{M}_n$ such that $\|\Pi_n \beta - \beta\|_{s,\beta} = o(1)$.*

Assumption 4.6 *(i) $E[|\rho(Z, \alpha_0)|^2 \mid X]$ is bounded. (ii) $\rho(Z, \alpha)$ is Hölder continuous in $\alpha \in \mathcal{A}$.*

Assumption 4.7 *(i) $E[|\ln f_{Y^*|X}(y^* \mid x)|^2]$ is bounded. (ii) There exists a measurable function $h_1(y^*, x)$ with $E[|h_1(y^*, x)|^2] < \infty$ such that for any $\bar{\beta} = (\bar{\psi}, \bar{f}_1, \bar{f}_2)' \in \mathcal{B}$,*

$$\left| \frac{f^{|1|}_{Y^*|X}(y^* \mid x; \bar{\beta}, \bar{\omega})}{f_{Y^*|X}(y^* \mid x; \bar{\beta})} \right| \leq h_1(y^*, x),$$

where $f^{|1|}_{Y^*|X}(y^* \mid x; \bar{\beta}, \bar{\omega})$ is defined as $(\frac{d}{dt} f_{Y^*|X}(y^* \mid x; \bar{\beta} + t\bar{\omega})|_{t=0}$ with each linear term, that is, $\frac{d}{d\psi} f_{Y_1|Y_2 X_1}$, $\bar{f}_1$, and $\bar{f}_2$, replaced by its absolute value, and $\bar{\omega}(\xi, y_2^*, y_2, x_2.x_1) = [1, \omega^{-1}(\xi), \omega^{-1}((y_2^*, y_2, x_1)'), \omega^{-1}((y_2, x_2, x_1)')]'$ with $\xi \in \mathcal{U}$.

Assumption 4.8 *(i) $k_{1n} \to +\infty$.*

Assumption 4.9 *(i) $k_n / n \to 0$.*

Theorem 4.1 *(i) Under Assumptions 3.1-3.8, 4.5 (i) and (ii), 4.7 (i) and (ii), and 4.9, we have $\|\hat{\beta}_n - \beta_0\|_{s,\beta} = o_p(1)$.*

(ii) Under Assumptions 3.1-3.8, 4.1 (i), 4.2, 4.3 (i) and (ii), 4.4 (i) and (ii), 4.5 (i) and (ii), 4.6 (i) and (ii), 4.7 (i) and (ii), 4.8 (i), and 4.9 (i), we have $\|\hat{\alpha}_n - \alpha_0\|_{s,\alpha} = o_p(1)$.

Theorem 4.1 provides consistency results under the metrics $\|\cdot\|_{s,\beta}$ and $\|\cdot\|_{s,\alpha}$, which are stepping stones to establishing the asymptotic normality of $\hat{\phi}$ and $\hat{\theta}$, respectively.

2.4.2 Convergence Rates

As in Ai and Chen (2003) and Hu and Schennach (2008), we now consider $n^{-1/4}$ convergence rates of $\hat{\beta}_n$ and $\hat{\alpha}_n$ under weaker metrics, which are sufficient to establish the asymptotic normality and $\sqrt{n}$-consistency results. First, we recall the weaker metric introduced by Ai and Chen (2003).

Suppose that the parameter space $\mathcal{B}$ is connected in the sense that for any two points $\beta_1, \beta_2 \in \mathcal{B}$, there exists a continuous path $\{\beta(t) : t \in [0,1]\}$ in $\mathcal{B}$ such that $\beta(0) = \beta_1$ and $\beta(1) = \beta_2$. And suppose that $\mathcal{B}$ is convex at the true value β_0 in the sense that, for any $\beta \in \mathcal{B}, (1-t)\beta_0 + t\beta \in \mathcal{B}$ for small $t > 0$. Furthermore, suppose that for almost all D and any $\beta \in \mathcal{B}$, $\ln f_{Y^*|X}(D, (1-t)\beta_0 + t\beta)$ is continuously differentiable at $t = 0$. Similarly, suppose that for any two points $\alpha_1, \alpha_2 \in \mathcal{A}$, there exists a continuous path $\{\alpha(\tau) : \tau \in [0,1]\}$ in $\mathcal{A}$ such that $\alpha(0) = \alpha_1$ and $\alpha(1) = \alpha_2$. Also, suppose that $\mathcal{A}$ is convex at the true value α_0, and suppose that for almost all X, $m(X, (1-\tau)\alpha_0 + \tau\alpha)$ is continuously differentiable at $\tau = 0$.

Denote the first pathwise derivative of $\ln f_{Y^*|X}(y^* \mid x; \beta_0)$ at the direction $[\beta - \beta_0]$ evaluated at β_0 by:

$$\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[\beta - \beta_0] \equiv \left. \frac{d \ln f_{Y^*|X}(y^* \mid x; (1-t)\beta_0 + t\beta)}{dt} \right|_{t=0}$$

almost everywhere (under the probability measure of (Y^*, X)) and for $\beta_1, \beta_2 \in \mathcal{B}$ denote

$$\begin{aligned} &\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[\beta_1 - \beta_2] \\ &\equiv \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[\beta_1 - \beta_0] - \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[\beta_2 - \beta_0]. \end{aligned}$$

Specifically, the pathwise derivative is denoted by:

$$
\begin{aligned}
&\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\beta}[\beta-\beta_0] \\
&= \frac{1}{f_{Y^*|X}(y^* \mid x;\beta_0)}\left\{ \int_{\mathcal{Y}_2} \frac{d}{d\psi} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1;\psi_0)[\psi-\psi_0] \right. \\
&\qquad\qquad \times f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1) f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1) dy_2 \\
&\quad + \int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1;\psi_0)[f_1(y_2^* \mid y_2, x_1) - f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1)] \\
&\qquad \times f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1) dy_2 \\
&\quad + \int_{\mathcal{Y}_2} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1;\psi_0) f_{Y_2^*|Y_2X_1}(y_2^* \mid y_2, x_1) \\
&\qquad \left. \times [f_2(y_2 \mid x_2, x_1) - f_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)] dy_2 \right\}.
\end{aligned}
$$

For any $\beta_1, \beta_2 \in \mathcal{B}$, the metric is defined as

$$
\|\beta_1-\beta_2\|_\beta \equiv \sqrt{E\left\{\left(\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\beta}[\beta_1-\beta_2]\right)^2\right\}}.
$$

Similarly, denote the first pathwise derivative of $\rho(Z,\alpha_0)$ at the direction $[\alpha-\alpha_0]$ evaluated at α_0 by:

$$
\frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha-\alpha_0] \equiv \left.\frac{d\rho(Z,(1-\tau)\alpha_0+\tau\alpha)}{d\tau}\right|_{\tau=0}
$$

almost everywhere (under the probability measure of Z) and for any $\alpha_1, \alpha_2 \in \mathcal{A}$ denote

$$
\begin{aligned}
\frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2] &\equiv \frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_0] - \frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha_2-\alpha_0], \\
\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2] &\equiv E\left\{\frac{d\rho(Z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\middle| X\right\}.
\end{aligned}
$$

Also, for any $\alpha_1, \alpha_2 \in \mathcal{A}$, the metric $\|\cdot\|_\alpha$ is defined as

$$
\begin{aligned}
&\|\alpha_1-\alpha_2\|_\alpha \\
&\equiv \sqrt{E\left\{\left(\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)' \Sigma(X)^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right\}}.
\end{aligned}
$$

Assumption 4.3 *(iii)* $\hat{\Sigma}(X) = \Sigma(X) + o_p(n^{-1/4})$ *uniformly over* $X \in \mathcal{X}$.

Assumption 4.4 *(iii) There is a constant* $\mu_1 > 0$ *such that for any* $\alpha \in \mathcal{A}$, *there exists* $\Pi_n\alpha \in \mathcal{A}_n$ *satisfying* $\|\Pi_n\alpha - \alpha\|_\alpha = O(k_{1n}^{-\mu_1})$, *and* $k_{1n}^{-\mu_1} = o(n^{-1/4})$.

Assumption 4.5 *(iii) There is a constant* $\gamma_1 > 1$ *as in Assumptions 6-8 such that for any* $\beta \in \mathcal{B}$, *there exists* $\Pi_n\beta \in \mathcal{B}_n$ *satisfying* $\|\Pi_n\beta - \beta\|_\beta = O(k_n^{-\gamma_1/d_1})$, *and* $k_n^{-\gamma_1/d_1} = o(n^{-1/4})$.

Assumption 4.6 *(iii) Each element of* $\rho(Z, \alpha)$ *satisfies an envelope condition in* $\alpha \in \mathcal{A}_n$; *(iv) each element of* $m(\cdot, \alpha) \in \Lambda_c^\gamma(\mathcal{X})$ *with* $\gamma > d_x/2$, *for all* $\alpha \in \mathcal{A}_n$.

Assumption 4.7 *(iii)* $\ln f_{Y^*|X}(y^* \mid x; \beta)$ *satisfies an envelope condition in* $\beta \in \mathcal{B}_n$; *(iv)* $\ln f_{Y^*|X}(y^* \mid x; \beta) \in \Lambda_c^{\gamma,\omega}(\mathcal{Y}^* \times \mathcal{X})$ *for some constant* $c > 0$ *with* $\gamma > d_{(Y^*,X)}/2$, *for all* $\beta \in \mathcal{B}_n$, *where* $d_{(Y^*,X)}$ *is the dimension of* (Y^*, X).

Denote $\xi_{0n} \equiv \sup_{(\xi_1,\xi_2,\xi_3)\in(\mathcal{U}\cup(\mathcal{Y}_2^*\times\mathcal{Y}_2\times\mathcal{X}_1)\cup(\mathcal{Y}_2\times\mathcal{X}_2\times\mathcal{X}_1))} \|p^{k_n}(\xi_1, \xi_2, \xi_3)\|_E^2$, which is nondecresasing in k_n. Let $N(\varepsilon, \mathcal{B}_n, \|\cdot\|_{s,\beta})$ and $N(\delta, \mathcal{A}_n, \|\cdot\|_{s,\alpha})$ denote the minimal number of radius ε covering balls of $\mathcal{B}_n$ under the $\|\cdot\|_{s,\beta}$ metric, and the minimal number of radius δ covering balls of $\mathcal{A}_n$ under the $\|\cdot\|_{s,\alpha}$ metric, respectively.

Assumption 4.8 *(ii)* $k_{1n} \times \ln n \times \xi_{0n}^2 \times n^{-1/2} = o(1)$; *(iii)* $\ln[N(\delta^{1/\kappa}, \mathcal{A}_n, \|\cdot\|_{s,\alpha})] \leq \text{const.} \times k_{1n} \times \ln(k_{1n}/\delta)$.

Assumption 4.9 *(ii)* $k_n \times \ln n \times \xi_{0n}^2 \times n^{-1/2} = o(1)$; *(iii)* $\ln[N(\varepsilon, \mathcal{B}_n, \|\cdot\|_{s,\beta})] \leq \text{const.} \times k_n \times \ln(k_n/\varepsilon)$.

Assumption 4.10 *(i)* $\mathcal{A}$ *is convex in* α_0, *and* $\rho(Z, \alpha)$ *is pathwise differentiable at* α_0; *(ii) for some* $c_1, c_2 > 0$,

$$c_1 E\{m(X,\alpha)'\Sigma(X)^{-1}m(X,\alpha)\} \leq \|\alpha - \alpha_0\|_\alpha^2 \leq c_2 E\{m(X,\alpha)'\Sigma(X)^{-1}m(X,\alpha)\}$$

holds for all $\alpha \in \mathcal{A}_n$ *with* $\|\alpha - \alpha_0\|_{s,\alpha} = o(1)$.

Assumption 4.11 *(i)* $\mathcal{B}$ *is convex in* β_0 *and* $f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \psi)$ *is pathwise differentiable at* ψ_0; *(ii) for some* $c_1, c_2 > 0$,

$$c_1 E\left\{\ln \frac{f_{Y^*|X}(y^* \mid x; \beta_0)}{f_{Y^*|X}(y^* \mid x; \beta)}\right\} \leq \|\beta - \beta_0\|_\beta^2 \leq c_2 E\left\{\ln \frac{f_{Y^*|X}(y^* \mid x; \beta_0)}{f_{Y^*|X}(y^* \mid x; \beta)}\right\}$$

holds for all $\beta \in \mathcal{B}_n$ *with* $\|\beta - \beta_0\|_{s,\beta} = o(1)$.

Theorem 4.2 *(i) Under Assumptions 3.1-3.8, 4.1, 4.5, 4.7, 4.9 and 4.11, we have* $\|\hat{\beta}_n - \beta_0\|_\beta = o_p(n^{-1/4})$.

(ii) Under Assumptions 3.1-3.8 and 4.1-4.11, we have $\|\hat{\alpha}_n - \alpha_0\|_\alpha = o_p(n^{-1/4})$.

2.5 Asymptotic Normality and Efficiency

2.5.1 Asymptotic Normality and Efficiency

We consider the asymptotic normality of $\hat{\phi}_n$ and $\hat{\theta}_n$, and efficiency of a three-step estimation of $\hat{\theta}_n$. We first introduce important notation aligning with that of Ai and Chen (2003) and Hu and Schennach (2008). Let $\overline{\mathbf{V}}_1$ denote the closure of the linear span of $\mathcal{B} - \{\beta_0\}$ under the metric $\|\cdot\|_\beta$ (i.e., $\overline{\mathbf{V}}_1 = \mathcal{R}^{d_\phi} \times \overline{\mathcal{W}}_1$ with $\overline{\mathcal{W}}_1 \equiv \overline{\mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2} - \{(\eta_0, f_{Y_2^*|Y_2X_1}, f_{Y_2|X_2X_1})'\})$ and $(\overline{\mathbf{V}}_1, \|\cdot\|_\beta)$ is a Hilbert space with the inner product:

$$\langle v_{11}, v_{12}\rangle_\beta = E\left\{\left(\frac{d\ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[v_{11}]\right)\left(\frac{d\ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[v_{12}]\right)\right\}.$$

Similarly, let $\overline{\mathbf{V}}_2$ denote the closure of the linear span of $\mathcal{A} - \{\alpha_0\}$ under the metric $\|\cdot\|_\alpha$ (i.e., $\overline{\mathbf{V}}_2 = \mathcal{R}^{d_\theta} \times \overline{\mathcal{W}}_2$ with $\overline{\mathcal{W}}_2 \equiv \overline{\mathcal{H}} - \{h_0\}$). Then $(\overline{\mathbf{V}}_2, \|\cdot\|_\alpha)$ is a Hilbert space with the inner product:

$$\langle v_{21}, v_{22}\rangle_\alpha = E\left\{\left(\frac{dm(X, \alpha_0)}{d\alpha}[v_{21}]\right)' \Sigma(X)^{-1}\left(\frac{dm(X, \alpha_0)}{d\alpha}[v_{22}]\right)\right\}.$$

The pathwise derivative at β_0 is defined as

$$\begin{aligned}
&\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\beta}[\beta-\beta_0] \\
&= \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\phi'}(\phi-\phi_0) + \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\eta}[\eta-\eta_0] \\
&\quad + \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_1}[f_1 - f_{Y_2^*|Y_2X_1}] + \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_2}[f_2 - f_{Y_2|X_2X_1}].
\end{aligned}$$

For each component ϕ_j of ϕ, $j = 1, 2, \ldots, d_\phi$, we define $w_{1j}^* \in \overline{\mathcal{W}}_1$ as

$$\begin{aligned}
w_{1j}^* &\equiv (\eta_j^*, f_{1j}^*, f_{2j}^*)' \\
&= \arg\min_{(\eta,f_1,f_2)'\in\overline{\mathcal{W}}_1} E\Bigg\{\Bigg(\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\phi_j} - \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\eta}[\eta_j] \\
&\qquad - \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_1}[f_{1j}] - \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_2}[f_{2j}]\Bigg)^2\Bigg\}.
\end{aligned}$$

Define

$$\begin{aligned}
&w_1^* = (w_{11}^*, w_{12}^*, \ldots, w_{1d_\phi}^*), \\
&\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df}[w_{1j}^*] \\
&= \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\eta}[\eta_j] + \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_1}[f_{1j}] + \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df_2}[f_{2j}] \\
&\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df}[w_1^*] \\
&= \left(\frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df}[w_{11}^*], \ldots, \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df}[w_{1d_\phi}^*]\right),
\end{aligned}$$

and the row vector

$$G_{w_1^*}(Y^*, X, \beta_0) \equiv \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{d\phi'} - \frac{d\ln f_{Y^*|X}(y^* \mid x;\beta_0)}{df}[w_1^*].$$

We also introduce some notation for the second stage parameters, θ_0. As shown before, the pathwise derivative at α_0 is

$$\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha_0] \equiv \frac{dm(X,\alpha_0)}{d\theta'}(\theta-\theta_0) + \frac{dm(X,\alpha_0)}{dh}[h-h_0].$$

For each component θ_j of θ, $j = 1, 2, \ldots, d_\theta$, we define $w_{2j}^* \in \overline{\mathcal{W}}_2$ as

$$\begin{aligned}
w_{2j}^* \equiv \arg\min_{w_{2j}\in\overline{\mathcal{W}}_2} E\Bigg\{&\left(\frac{dm(X,\alpha_0)}{d\theta_j} - \frac{dm(X,\alpha_0)}{dh}[w_{2j}]\right)'\Sigma(X)^{-1} \\
&\times \left(\frac{dm(X,\alpha_0)}{d\theta_j} - \frac{dm(X,\alpha_0)}{dh}[w_{2j}]\right)\Bigg\}.
\end{aligned}$$

Define

$$w_2^* = (w_{21}^*, w_{22}^*, \ldots, w_{2d_\theta}^*),$$
$$\frac{dm(X,\alpha_0)}{dh}[w_2^*] = \left(\frac{dm(X,\alpha_0)}{dh}[w_{21}^*], \ldots, \frac{dm(X,\alpha_0)}{dh}[w_{2d_\theta}^*]\right),$$

and the row vector

$$G_{w_2^*}(X,\alpha_0) \equiv \frac{dm(X,\alpha_0)}{d\theta_2'} - \frac{dm(X,\alpha_0)}{dh}[w_2^*].$$

Define $s_1(\beta) \equiv \lambda_1'\phi$ for $\lambda_1 \in \mathcal{R}^{d_\phi}$ and $\lambda_1 \neq 0$, and define $s_2(\alpha) \equiv \lambda_2'\theta$ for $\lambda_2 \in \mathcal{R}^{d_\theta}$ and $\lambda_2 \neq 0$. As mentioned in Ai and Chen (2003), $s_1(\beta) \equiv \lambda_1'\phi$ is bounded if and only if $E[G_{w_1^*}(Y^*, X, \beta_0)' G_{w_1^*}(Y^*, X, \beta_0)]$ is finite positive-definite. The Riesz representation theorem then implies that there exists a representor v_1^* such that

$$s_1(\beta) - s_1(\beta_0) \equiv \lambda_1'(\phi - \phi_0) = \langle v_1^*, \beta - \beta_0\rangle_\beta$$

for all $\beta \in \mathcal{B}$ where $v_1^* \equiv (v_\phi^*, v_f^*) \in \overline{\mathbf{V}}_1$, $v_\phi^* = J_1^{-1}\lambda_1$, $v_f^* = -w_1^* \times v_\phi^*$ with $J_1 = E[G_{w_1^*}(Y^*, X, \beta_0)' G_{w_1^*}(Y^*, X, \beta_0)]$. Similarly, because of the fact that $s_2(\alpha) \equiv \lambda_2'\theta$ is bounded if and only if $E[G_{w_2^*}(X,\alpha_0)'\Sigma(X)^{-1}G_{w_2^*}(X,\alpha_0)]$ is finite positive-definite, we have

$$s_2(\alpha) - s_2(\alpha_0) \equiv \lambda_2'(\theta - \theta_0) = \langle v_2^*, \alpha - \alpha_0\rangle_\alpha$$

for all $\alpha \in \mathcal{A}$ where $v_2^* \equiv (v_\theta^*, v_h^*) \in \overline{\mathbf{V}}_2$, $v_\theta^* = J_2^{-1}\lambda_2$, $v_h^* = -w_2^* \times v_\theta^*$ with $J_2 = E[G_{w_2^*}(X,\alpha_0)'\Sigma(X)^{-1}G_{w_2^*}(X,\alpha_0)]$.

We now state the sufficient conditions for the $\sqrt{n}$-normality of $\hat{\phi}_n$ and $\hat{\theta}_n$.

Assumption 5.1 *(i) $E[G_{w_2^*}(X,\alpha_0)'\Sigma(X)^{-1}G_{w_2^*}(X,\alpha_0)]$ exists, is bounded, and is positive-definite; (ii) $\theta_0 \in \mathrm{int}(\Theta)$; (iii) $\Sigma_0(X) \equiv \mathrm{var}[\rho(Z,\alpha_0) \mid X]$ is positive-definite for all $X \in \mathcal{X}$.*

Assumption 5.2 *(i) $E[G_{w_1^*}(Y^*, X, \beta_0)' G_{w_1^*}(Y^*, X, \beta_0)]$ exists, is bounded, and is positive-definite; (ii) $\phi_0 \in \mathrm{int}(\Phi)$.*

Assumption 5.3 *There is a $v_{2n}^* = (v_\theta, -\Pi_n w_2^* \times v_\theta) \in \mathcal{A}_n - \{\Pi_n\alpha_0\}$ such that $\|v_{2n}^* - v_2^*\|_\alpha = O(n^{-1/4})$.*

Assumption 5.4 *There is a* $v_{1n}^* = (v_\phi, -\Pi_n w_1^* \times v_\phi) \in \mathcal{B}_n - \{\Pi_n \beta_0\}$ *such that* $\|v_{1n}^* - v_1^*\|_\beta = O(n^{-1/4})$.

Define $\mathcal{N}_{01n} \equiv \{\beta \in \mathcal{B}_n : \|\beta - \beta_0\|_{s,\beta} \le v_{1n}, \|\beta - \beta_0\|_\beta \le v_{1n} n^{-1/4}\}$ with $v_{1n} = o(1)$ and define $\mathcal{N}_{01}$ the same way with $\mathcal{B}_n$ replaced by $\mathcal{B}$. Define $\mathcal{N}_{02n} \equiv \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_{s,\alpha} \le v_{2n}, \|\alpha - \alpha_0\|_\alpha \le v_{2n} n^{-1/4}\}$ with $v_{2n} = o(1)$ and define $\mathcal{N}_{02}$ the same way with $\mathcal{A}_n$ replaced by $\mathcal{A}$. For $\beta \in \mathcal{N}_{01n}$, we denote a local alternative $\beta^*(\beta, \varepsilon_n) = (1 - \varepsilon_n)\beta + \varepsilon_n(v_1^* + \beta_0)$ with $\varepsilon_n = o(n^{-1/2})$. Let $\Pi_n \beta^*(\beta, \varepsilon_n)$ be the projection of $\beta^*(\beta, \varepsilon_n)$ onto $\mathcal{B}_n$. We denote

$$\frac{d\rho(Z, \alpha)}{d\alpha}[v_2] \equiv \left.\frac{d\rho(Z, \alpha + \tau v_2)}{d\tau}\right|_{\tau=0} \quad \text{a.s. } Z,$$

and

$$\frac{dm(X, \alpha)}{d\alpha}[v_2] \equiv \frac{dm(X, \alpha + \tau v_2)}{d\alpha}[v_2] \quad \text{a.s. } X,$$

for any $v_2 \in \overline{\mathbf{V}}_2$. Also for any $v_1 \in \overline{\mathbf{V}}_1$, we denote

$$\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta)}{d\beta}[v_1] \equiv \left.\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta + t v_1)}{dt}\right|_{t=0} \quad \text{a.s. } (Y^*, X).$$

Assumption 5.5 *For all* $\alpha \in \mathcal{N}_{02}$, *the pathwise first derivative* $(d\rho(Z, \alpha(\tau))/d\alpha)[v_2]$ *exists a.s.* $Z \in \mathcal{Z}$. *Moreover,* (i) *each element of the pathwise first derivative evaluated at* v_{2n}^*, $(d\rho(Z, \alpha)/d\alpha)[v_{2n}^*]$, *satisfies an envelope condition and is Hölder continuous in* $\alpha \in \mathcal{N}_{02n}$; (ii) *each element of* $(dm(X, \alpha)/d\alpha)[v_{2n}^*]$ *is in* $\Lambda_c^\gamma(\mathcal{X}), \gamma > d_x/2$ *for all* $\alpha \in \mathcal{N}_{02}$.

Assumption 5.6 *Uniformly over* $\alpha \in \mathcal{N}_{02n}$, *we have*

$$E\left(\left\|\frac{dm(X, \alpha)}{d\alpha}[v_{2n}^*] - \frac{dm(X, \alpha_0)}{d\alpha}[v_{2n}^*]\right\|_E^2\right) = o(n^{-1/2}).$$

Assumption 5.7 *Uniformly over* $\alpha \in \mathcal{N}_{02}$, $\bar{\alpha} \in \mathcal{N}_{02n}$, *we have*

$$E\left(\left\{\frac{dm(X, \alpha_0)}{d\alpha}[v_2^*]\right\} \Sigma(X)^{-1} \left\{\frac{dm(X, \alpha)}{d\alpha}[\bar{\alpha} - \alpha_0] - \frac{dm(X, \alpha_0)}{d\alpha}[\bar{\alpha} - \alpha_0]\right\}\right)$$
$$= o(n^{-1/2}).$$

Assumption 5.8 *For all $\alpha \in \mathcal{N}_{02n}$, the pathwise second derivative $d^2\rho(Z, \alpha+\tau v_{2n}^*)/d\tau^2|_{\tau=0}$ exists a.s. $Z \in \mathcal{Z}$, and is bounded by a measurable function $c_5(Z)$ with $E[c_5(Z)^2] < \infty$.*

Assumption 5.9 *There exists a measurable function $h_2(Y^*, X)$ with $E[h_2(Y^*, X)^2] < \infty$ such that for any $\bar{\beta} = (\bar{\psi}, \bar{f}_1, \bar{f}_2)' \in \mathcal{N}_{01}$,*

$$\left| \frac{f^{|1|}_{Y^*|X}(y^* \mid x; \bar{\beta}, \bar{\omega})}{f_{Y^*|X}(y^* \mid x; \bar{\beta})} \right|^2 + \left| \frac{f^{|2|}_{Y^*|X}(y^* \mid x; \bar{\beta}, \bar{\omega})}{f_{Y^*|X}(y^* \mid x; \bar{\beta})} \right| \leq h_2(Y^*, X),$$

where $f^{|2|}_{Y^*|X}(y^* \mid x; \bar{\beta}, \bar{\omega})$ is defined as $\frac{d^2}{dt^2} f_{Y^*|X}(y^* \mid x; \bar{\beta} + t\bar{\omega})|_{t=0}$ with each linear term, that is, $\frac{d}{d\psi} f_{Y_1|Y_2X_1}$, $\frac{d^2}{d\psi^2} f_{Y_1|Y_2X_1}$, $\bar{f}_1$, and $\bar{f}_2$, replaced by its absolute value.

Following Hu and Schennach (2008), we write the following notations for the next assumption:

$$\begin{aligned}
&\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[p^{k_n}] \\
&= \Bigg(\left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\phi} \right)', \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\eta}[p^{k_n}] \right)', \\
&\qquad \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{df_1}[p^{k_n}] \right)', \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{df_2}[p^{k_n}] \right)' \Bigg)',
\end{aligned}$$

where for $\tilde{f} = \eta, f_1$, or f_2,

$$\begin{aligned}
&\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\tilde{f}}[p^{k_n}] \\
&= \Bigg(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\tilde{f}}[p_1^{k_n}], \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\tilde{f}}[p_2^{k_n}], \ldots, \\
&\qquad \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\tilde{f}}[p_{k_n}^{k_n}] \Bigg)', \\
&\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\phi} \\
&= \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\phi_1}, \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\phi_2}, \ldots, \frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\phi_{d_\phi}} \right)',
\end{aligned}$$

and

$$\Omega_{k_n} = E\left\{ \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[p^{k_n}] \right) \left(\frac{d \ln f_{Y^*|X}(y^* \mid x; \beta_0)}{d\beta}[p^{k_n}] \right)' \right\}.$$

Assumption 5.10 *The smallest eigenvalue of the matrix* Ω_{k_n} *is bounded away from zero, and* $\|p_j^{k_n}\|_{\infty,\omega} < \infty$ *for* $j = 1, 2, \ldots, k_n$ *uniformly in* k_n.

Assumption 5.11 *For all* $\beta \in \mathcal{N}_{01n}$, *there exists a measurable function* $h_4(Y^*, X)$ *with* $E|h_4(Y^*, X)| < \infty$ *such that*

$$\left|\frac{d^4}{dt^4} \ln f_{Y^*|X}(y^* \mid x; \bar{\beta} + t(\beta - \beta_0))\right|_{t=0} \le h_4(Y^*, X)\|\beta - \beta_0\|_{s,\beta}^4.$$

Theorem 5.1 *(i) Under Assumptions 3.1-3.8, 4.1, 4.5, 4.7, 4.11, 5.2, 5.4, 5.9-5.11,* $\sqrt{n}(\hat{\phi}_n - \phi_0) \xrightarrow{d} N(0, J_1^{-1})$,
where $J_1 = E[G_{w_1^*}(Y^*, X, \beta_0)' G_{w_1^*}(Y^*, X, \beta_0)]$.

(ii) Under Assumptions 3.1-3.8, 4.1-4.11, 5.1-5.11,
$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, J_2^{-1})$, *where*

$$\begin{aligned} J_2 = {} & E[G_{w_2^*}(X, \alpha_0)' \Sigma(X)^{-1} G_{w_2^*}(X, \alpha_0)] \\ & \times (E[G_{w_2^*}(X, \alpha_0)' \Sigma(X)^{-1} \Sigma_0(X) \Sigma(X)^{-1} G_{w_2^*}(X, \alpha_0)])^{-1} \\ & \times E[G_{w_2^*}(X, \alpha_0)' \Sigma(X)^{-1} G_{w_2^*}(X, \alpha_0)]. \end{aligned}$$

2.5.2 Consistent Covariance Estimator

We now establish a consistent estimator $\tilde{J}_{02}$ of the covariance matrix J_{02}, which is needed to perform any statistical inference using the semiparametrically efficient estimator $\tilde{\theta}_n$.

Let $\hat{G}_{w_{2j}}(X, \tilde{\alpha})$ be a consistent estimator of $G_{w_{2j}}(X, \alpha_0)$ as follows:

$$\hat{G}_{w_{2j}}(X, \tilde{\alpha}) = \frac{d\hat{m}(X, \tilde{\alpha})}{d\theta_j} - \frac{d\hat{m}(X, \tilde{\alpha})}{dh}[w_{2j}].$$

We estimate w_{02j} by $\tilde{w}_{2j}$, which is the solution to the minimization problem:

$$\min_{w_{2j} \in \mathcal{W}_{2n}} \frac{1}{n} \sum_{i=1}^{n} \hat{G}_{w_{2j}}(X_i, \tilde{\alpha})' \hat{\Sigma}_0(X_i)^{-1} \hat{G}_{w_{2j}}(X_i, \tilde{\alpha}).$$

If we let $w_{02} = (w_{021}, \ldots, w_{02d_\theta})$ and $\tilde{w}_2 = (\tilde{w}_{21}, \ldots, \tilde{w}_{2d_\theta})$, then $\hat{G}_{\tilde{w}_2}(X, \tilde{\alpha})$ is a consistent estimator of $G_{w_{02}}(X, \alpha_0)$. Therefore, the estimator of J_{02} is $\tilde{J}_{02} = \frac{1}{n} \sum_{i=1}^{n} \hat{G}_{\tilde{w}_2}(X_i, \tilde{\alpha})' \hat{\Sigma}_0(X_i)^{-1} \hat{G}_{\tilde{w}_2}(X_i, \tilde{\alpha})$.

Theorem 5.2 *Under the conditions of Theorem 5.1 (ii)*, $\tilde{J}_{02} = J_{02} + o_p(1)$.

Theorem 5.2 states that the estimator $\tilde{J}_{02}$ of the covariance matrix J_{02} is consistent.

2.6 Simulation

We assess the finite sample performance of the proposed estimator in this section. The simulation is based on a nonparametric regression

$$Y_1 = h_0(Y_2) + U$$

where $h_0(Y_2) = \exp(Y_2)/(1 + \exp(Y_2))$. We assume that Y_2 is generated as $Y_2 = aX_2 + R(U + c) + b\varepsilon$. X_2, ε, and U are independent and distributions of those are $X_2 \sim N(1, \sigma^2)$, $\varepsilon \sim N(1, \sigma^2)$, and $U \sim N(0, \sigma^2)$ with $(a, b, c, R, \sigma) = (0.6, 0.2, 1, 0.2, 0.7)$. The distributions of X_2 and ε are truncated on $[0, 2]$ and the distribution of U is truncated on $[-1, 1]$. Thus the support of Y_2 is $[0, 2]$. As in Ai and Chen (2003), we approximate the unknown $h_0(Y_2)$ by a power series of fourth order multiplied by the cumulative distribution function of a standard normal since $h_0(Y_2)$ is bounded between zero and one. So the approximate regression model is

$$Y_1 \approx \pi_0\Phi(Y_2) + \pi_1\Phi(Y_2)Y_2 + \pi_2\Phi(Y_2)Y_2^2 + \pi_3\Phi(Y_2)Y_2^3 + U$$

where $\Phi(Y_2)$ denotes the standard normal cumulative distribution function.

We also use the general form of generating processes for the measurement error which is similar to those in Hu and Schennach (2008)

$$f_{Y_2^*|Y_2}(y_2^* \mid y_2) = \frac{1}{\sigma(y_2)} f_\nu\left(\frac{y_2^* - y_2}{\sigma(y_2)}\right),$$

where $\sigma(y_2) = 1.5\exp(-y_2)$ and f_ν is a density function to be specified below for three models: heteroskedastic measurement error with zero mean, nonadditive measurement error with zero mode, and nonadditive measurement error with zero median.

(1) Heteroskedastic Measurement Error with Zero Mean: a measurement error is

$$Y_2^* = Y_2 + \sigma(y_2)\nu$$

with $Y_2 \perp \nu$. The error structure in the simulation is $F_\nu(\nu) = \Phi(\nu)$.

(2) Nonadditive Measurement Error with Zero Mode: let

$$f_{Y_2^*|Y_2}(y_2^* \mid y_2) = \frac{g(y_2^*, y_2)}{\int_{-\infty}^{\infty} g(y_2^*, y_2) dy_2^*},$$
$$g(y_2^*, y_2) = \exp\left\{ h(y_2) \left[\left(\frac{y_2^* - y_2}{\sigma(y_2)} \right) - \exp\left(\frac{y_2^* - y_2}{\sigma(y_2)} \right) \right] \right\}$$

with $h(y_2) = \exp(-0.1y_2)$. Then $f_{Y_2^*|Y_2}(y_2^* \mid y_2)$ has the unique mode at y_2 for any $h(y_2) > 0$.

(3) Nonadditive Measurement Error with Zero Median: let the corresponding cumulative distribution function be

$$F_{Y_2^*|Y_2}(y_2^* \mid y_2)$$
$$= \frac{1}{\pi} \arctan\left\{ h(y_2) \left[\frac{1}{2} + \frac{1}{2} \exp\left(\frac{y_2^* - y_2}{\sigma(y_2)} \right) - \left(-\frac{y_2^* - y_2}{\sigma(y_2)} \right) \right] \right\} + \frac{1}{2}$$

with $h(y_2) = \exp(-0.1y_2)$. Then $F_{Y_2^*|Y_2}(y_2 \mid y_2) = \frac{1}{2}$ for any $h(y_2) > 0$.

We consider three estimators: (i) the (inconsistent) SMD estimator from Ai and Chen (2003) which is obtained using error-laden data, (ii) the (infeasible) SMD estimator from Ai and Chen (2003) which is obtained using error-free data, and (iii) the proposed two-stage SML-SMD estimator. We construct sieves for functions of two variables using tensor product bases of univariate trigonometric series in our estimator. In both SMD estimators, we use a tensor product polynomial sieve to approximate the conditional mean function which is the set of instruments: $\{1, X_2, X_2^2, \ldots, X_2^{k_n}\}$ for $k_n \geq 3$. The sample size is $1,000$ and the procedures are repeated 100 times to obtain the root integrated mean squared error (RIMSE) according to the following discrete expression: $((200)^{-1} \sum_{j=0}^{199} \text{mean}\{[h_0(0+0.01j) - \hat{h}(0 + 0.01j)]^2\})^{1/2}$, where mean$\{\cdot\}$ denotes the average over all 100 estimators $\hat{h}$ for each procedure.

Table 2.1 reports estimation results. RIMSE from our proposed estimator is smaller than that from the SMD estimator obtained using error-laden data for all cases of identification conditions for measurement error.

2.7 Summary and Concluding Remarks

We consider semiparametric estimation of models defined by conditional moment restrictions, which contain finite dimensional unknown parameters and infinite dimensional unknown functions. We extend these models to include the case where the unknown functions depend on endogenous variables which are contaminated by nonclassical measurement errors. A two-stage estimation procedure is proposed to recover the true conditional density of endogenous variables given conditioning variables masked by the nonclassical measurement errors, and to rectify the difficulty associated with endogeneity of the unknown functions. Specifically, we estimate conditional density of endogenous variables given conditioning variables in the first stage using sieve maximum likelihood estimation, and then estimate parameters of interest in the second stage using sieve minimum distance estimation. We show that the proposed estimator of the infinite dimensional unknown functions is consistent with a rate faster than $n^{-1/4}$ under a certain metric, and the proposed estimator of the finite dimensional unknown parameters obtains root-n asymptotic normality. Monte Carlo evidence illustrates the usefulness of our method.

2.8 Mathematical Appendix

Proof of Theorem 3.1 Since $Y^* \equiv (Y_1, Y_2^{*\prime})'$ and $X \equiv (X_1', X_2')'$, eqn. 2.9 follows by the fact that

$$
\begin{aligned}
& f_{Y^*|X}(y \mid x) \\
& \equiv f_{Y_1 Y_2^* | X_2 X_1}(y_1, y_2^* \mid x_2, x_1) \\
& = \int_{\mathcal{Y}_2} f_{Y_1 Y_2 Y_2^* | X_2 X_1}(y_1, y_2, y_2^* \mid x_2, x_1) dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 Y_2^* X_2 X_1}(y_1 \mid y_2, y_2^*, x_2, x_1) f_{Y_2 Y_2^* | X_2 X_1}(y_2, y_2^* \mid x_2, x_1) dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) f_{Y_2 Y_2^* | X_2 X_1}(y_2, y_2^* \mid x_2, x_1) dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) f_{Y_2^* | Y_2 X_2 X_1}(y_2^* \mid y_2, x_2, x_1) f_{Y_2 | X_2 X_1}(y_2 \mid x_2, x_1) dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) f_{Y_2 | X_2 X_1}(y_2 \mid x_2, x_1) dy_2,
\end{aligned}
$$

where the fourth equality and the sixth equality are obtained by Assumption 3.2 (i) and (ii), respectively. The equation above relates the joint densities of the observable variables to those of unobservable variables. We need to show the solution to the equation is unique. By the definition 3.1 and the eqn. 2.9, we get an operator equivalence relationship: for an arbitrary $g \in \mathcal{G}(\mathcal{X}_2)$

$$
\begin{aligned}
& \left[L_{y_1 Y_2^* | X_2 x_1} g \right] (y_2^*) \\
& = \int_{\mathcal{X}_2} f_{Y_1 Y_2^* | X_2 X_1}(y_1, y_2^* \mid x_2, x_1) g(x_2) dx_2 \\
& = \int_{\mathcal{X}_2} \int_{\mathcal{Y}_2} f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) f_{Y_2 | X_2 X_1}(y_2 \mid x_2, x_1) dy_2 g(x_2) dx_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) \int_{\mathcal{X}_2} f_{Y_2 | X_2 X_1}(y_2 \mid x_2, x_1) g(x_2) dx_2 dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) f_{Y_1 | Y_2 X_1}(y_1 \mid y_2, x_1) \left[L_{Y_2 | X_2 x_1} g \right] (y_2) dy_2 \\
& = \int_{\mathcal{Y}_2} f_{Y_2^* | Y_2 X_1}(y_2^* \mid y_2, x_1) \left[\triangle_{y_1 | Y_2 x_1} L_{Y_2 | X_2 x_1} g \right] (y_2) dy_2 \\
& = \left[L_{Y_2^* | Y_2 x_1} \triangle_{y_1 | Y_2 x_1} L_{Y_2 | X_2 x_1} g \right] (y_2^*),
\end{aligned}
$$

where the third equality is obtained by an interchange of the order of integration. Thus eqn. 2.9 defines the operator equivalence over the domain $g \in \mathcal{G}(\mathcal{X}_2)$:

$$L_{y_1 Y_2^*|X_2 x_1} = L_{Y_2^*|Y_2 x_1} \triangle_{y_1|Y_2 x_1} L_{Y_2|X_2 x_1}. \tag{2.12}$$

Next, we note that integration of eqn. 2.13 over all $y_1 \in \mathcal{Y}_1$ yields

$$L_{Y_2^*|X_2 x_1} = L_{Y_2^*|Y_2 x_1} L_{Y_2|X_2 x_1},$$

since integration of $\triangle_{y_1|Y_2 x_1}$ becomes the identity operator. Since $L_{Y_2^*|Y_2 x_1}$ is one-to-one from Assumption 3.3, isolating $L_{Y_2|X_2 x_1}$ yields

$$L_{Y_2|X_2 x_1} = L^{-1}_{Y_2^*|Y_2 x_1} L_{Y_2^*|X_2 x_1}.$$

Substitution of the expression into eqn. 2.13 yields

$$L_{y_1 Y_2^*|X_2 x_1} = L_{Y_2^*|Y_2 x_1} \triangle_{y_1|Y_2 x_1} L^{-1}_{Y_2^*|Y_2 x_1} L_{Y_2^*|X_2 x_1}.$$

Since $L_{Y_2^*|X_2 x_1}$ is one-to-one from Assumption 3.3, by rearranging, we get the operator equivalence defined over a dense subset of $\mathcal{G}(\mathcal{Y}_2^*)$

$$L_{y_1 Y_2^*|X_2 x_1} L^{-1}_{Y_2^*|X_2 x_1} = L_{Y_2^*|Y_2 x_1} \triangle_{y_1|Y_2 x_1} L^{-1}_{Y_2^*|Y_2 x_1}.$$

Thus the known operator $L_{y_1 Y_2^*|X_2 x_1} L^{-1}_{Y_2^*|X_2 x_1}$ defined in terms of densties of the observable variables (Y^*, X) admits a spectral decomposition (an eigenvalue-eigenfunction decomposition). The eigenvalues of the known operator (the diagonal elements of the $\triangle_{y_1|Y_2 x_1}$ operator, i.e., $\{f_{Y_1|Y_2 X_1}$
$(y_1 \mid y_2, x_1)\}$ for a given (y_1, x_1) and for all Y_2) and the eigenfunctions of the known operator (the kernel of the integral operator $L_{Y_2^*|Y_2 x_1}$, i.e., $\{f_{Y_2^*|Y_2 X_1}(\cdot \mid y_2, x_1)\}$ for a given x_1 and for all Y_2) provide the unobserved densities of interest. For the uniqueness of the spectral decomposition, we use similar arguments in Theorem 1 of Hu and Schennach (2008) [2]. $\square$

[2] To ensure uniqueness of the spectral decomposition, they show four techniques: First, Theorem XV.4.5 in Dunford and Schwartz (1971) guarantees uniqueness up to some normalizations. Second, the a priory arbitrary scale of the eigenfunctions is fixed by the requirement that densities must integrate to 1. Third, Assumption 3.4 and the fact that the eigenfunctions (which do not depend on Y_1, unlike the eigenvalues) must be consistent across different values of the dependent variable Y_1 are employed to avoid any ambiquity in the definition of the eigenfunctions when there is an eigenvalue degeneracy that involves two eigenfunctions $f_{Y_2^*|Y_2 X_1}(\cdot \mid y_2^a, x_1)$ and $f_{Y_2^*|Y_2 X_1}(\cdot \mid y_2^b, x_1)$ for some value of Y_1. Fourth, Assumption 3.5 is used to uniquely determine the ordering and indexing of the eigenvalues and eigenfunctions.

Let

$$\hat{Q}_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}(X_i,\alpha)'[\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\alpha),$$
$$Q_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} m(X_i,\alpha)'[\Sigma(X_i)]^{-1}m(X_i,\alpha),$$
$$Q(\alpha) = E\left[m(X,\alpha)'[\Sigma(X)]^{-1}m(X,\alpha)\right].$$

We use the following result to prove Theorem 4.1.

Lemma A.1 *Suppose that* $\mathcal{A}$ and $\mathcal{B}$ *are compact subsets of a space with norm* $\|\alpha\|_{s,\alpha}$ *and a space with norm* $\|\beta\|_{s,\beta}$*, respectively, and* Z_t $(t = 1, 2, \cdots)$ *are i.i.d. Also suppose that (i)* $Var[\rho(Z,\alpha) \mid X]$ *is bounded for each* $\alpha \in \mathcal{A}$*; (ii)* $\|\hat{\beta} - \beta_0\|_{s,\beta} = o(1)$*; (iii) there is* $b(Z)$ *and* $\nu > 0$ *with* $|\rho(Z,\tilde{\alpha}) - \rho(Z,\alpha)| \le b(Z)\|\tilde{\alpha}-\alpha\|^{\nu}_{s,\alpha}$ *and* $E[\hat{q}(x_1)^2] < \infty$ *where* $\hat{q} = (\hat{q}(x_1),\cdots,\hat{q}(x_n))' = \left(\int b(Z)\hat{f}_{Y|X}(y \mid x_1;\hat{\psi})dy, \cdots, \int b(Z)\hat{f}_{Y|X}(y \mid x_n;\hat{\psi})dy\right)'$. *Then* $\sup_{\alpha\in\mathcal{A}} |\hat{Q}_n(\alpha) - Q(\alpha)| = o_p(1)$ *and* $Q(\alpha)$ *is continuous.*

Proof of Lemma A.1 The proof will proceed by verifying the hypotheses of Lemma A.2 of Newey and Powell (2003). Their compactness of a parameter space is assumed directly in our hypothesis (i). To show that hypothesis (ii) holds (pointwise convergence in α), let $\hat{g}(\alpha) = (\hat{m}(X_1,\alpha),\cdots,\hat{m}(X_n,\alpha))'$, and $g(\alpha) = (m(X_1,\alpha),\cdots,m(X_n,\alpha))'$. We use the notation $\lesssim$ for " smaller than up to a generic constant." Note that for some subsequence $\{n_j\}$ a.s.,

$$\left|\hat{Q}_n(\alpha) - Q_n(\alpha)\right| \lesssim \left|\|\hat{g}(\alpha)\|_E^2 - \|g(\alpha)\|_E^2\right|/n$$
$$\le \left(\|\hat{g}(\alpha) - g(\alpha)\|_E^2 + 2\|g(\alpha)\|_E \cdot \|\hat{g}(\alpha) - g(\alpha)\|_E\right)/n.$$

Strictly speaking, the first inequality above holds almost surely for some subsequence $\{n_{j_k}\}$ of an arbitrary subsequence $\{n_j\}$ of $\{n\}$ as a consequence of Assumption 4.3, ensuring that $\hat{\Sigma}(X) = O_p(1)$ uniformly over $X \in \mathcal{X}$. For clarity and convenience, we will continue to use the notation above without explicit reference to sub-subsequences or probability zero concepts.

Also note that $\|g(\alpha)\|_E^2/n = O_p(1)$ by the Markov inequality from $Var(\rho(Z,\alpha) \mid X) < \infty$. Thus, it suffices to show $\|\hat{g}(\alpha) - g(\alpha)\|_E^2/n = o_p(1)$ to show $\left|\hat{Q}_n(\alpha) - Q_n(\alpha)\right| =$

$o_p(1)$.

$$
\begin{aligned}
&E\left[\|\hat{g}(\alpha)-g(\alpha)\|_E^2/n\right]\\
&=E\left[\left(\hat{g}(\alpha)-g(\alpha)\right)'\left(\hat{g}(\alpha)-g(\alpha)\right)/n\right]\\
&=E\Bigg[\frac{1}{n}\sum_{i=1}^{n}\left(\int\rho(Z,\alpha)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy-\int\rho(Z,\alpha)f_{Y|X}(y\mid x_i;\psi_0)dy\right)'\\
&\qquad\qquad\times\left(\int\rho(Z,\alpha)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy-\int\rho(Z,\alpha)f_{Y|X}(y\mid x_i;\psi_0)dy\right)\Bigg]\\
&=E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int\rho_j(Z,\alpha)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy-\int\rho_j(Z,\alpha)f_{Y|X}(y\mid x_i;\psi_0)dy\right)^2\right]\\
&=E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int\rho_j(Z,\alpha)\left(\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})-f_{Y|X}(y\mid x_i;\psi_0)\right)dy\right)^2\right]\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}E\left[\left(\int\rho_j(Z,\alpha)\left(\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})-f_{Y|X}(y\mid x_i;\psi_0)\right)dy\right)^2\right]\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}E\Bigg[\Bigg(\int\int\rho_j(Z,\alpha)\\
&\qquad\times\Big\{\hat{f}_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\hat{\psi})\left(\hat{f}_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})-f_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})\right)\\
&\qquad+f_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})\\
&\qquad\quad\times\left(\hat{f}_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\hat{\psi})-f_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\psi_0)\right)\Big\}dy_1dy_2\Bigg)^2\Bigg]\\
&\le\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}E\Bigg[\Bigg(\int\int|\rho_j(Z,\alpha)|\\
&\qquad\times\Big\{|\hat{f}_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\hat{\psi})\omega^{-1}(y_2,x_{2i},x_{1i})|\|\hat{f}_2-f_2\|_{\infty,\omega}\\
&\qquad+\left|f_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})\frac{df_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\psi_0)}{d\psi}\omega^{-1}(\xi)\right|\|\hat{\psi}-\psi_0\|_{\infty,\omega}\Big\}dy_1dy_2\Bigg)^2\Bigg]\\
&\le\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}E\Bigg[\Bigg(\int\int|\rho_j(Z,\alpha)|\Big\{|\hat{f}_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\hat{\psi})\omega^{-1}(y_2,x_{2i},x_{1i})|\\
&\qquad+\left|f_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})\frac{df_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\psi_0)}{d\psi}\omega^{-1}(\xi)\right|\Big\}dy_1dy_2\|\hat{\beta}-\beta_0\|_{s,\beta}\Bigg)^2\Bigg]\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}E\Bigg[\Bigg(\int\int|\rho_j(Z,\alpha)|\Big\{|\hat{f}_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\hat{\psi})\omega^{-1}(y_2,x_{2i},x_{1i})|\\
&\qquad+\left|f_{Y_2|X_2X_1}(y_2\mid x_{2i},x_{1i})\frac{df_{Y_1|Y_2X_1}(y_1\mid y_2,x_{1i};\psi_0)}{d\psi}\omega^{-1}(\xi)\right|\Big\}dy_1dy_2\Bigg)^2\Bigg]\|\hat{\beta}-\beta_0\|_{s,\beta}^2\\
&=o(1),
\end{aligned}
$$

since $\|\hat{\beta}-\beta_0\|_{s,\beta}=o_p(1)$. Therefore, we get $\|\hat{g}(\alpha)-g(\alpha)\|_E^2/n=o_p(1)$ by the Markov inequality. Since $Q_n(\alpha)=Q(\alpha)+o_p(1)$ by the weak law of large numbers, the triangle inequality gives hypothesis (ii) of Newey and Powell (2003). To show hypothesis (iii), let $\hat{q}=(\hat{q}(x_1),\cdots,\hat{q}(x_n))'$, and $\tilde{B}_n=\left[\|\hat{q}\|_E^2+2\|\hat{q}\|_E\cdot\|\hat{g}(\alpha_0)\|_E\right]/n$. Note that $\|\hat{q}\|_E^2/n=O_p(1)$ and $\|\rho(Z,\alpha_0)\|_E^2/n=O_p(1)$ so that $\tilde{B}_n=O_p(1)$. Since $\|\cdot\|_{s,\alpha}^{\nu}$ is bounded on $\mathcal{A}\times\mathcal{A}$ by the compactness of the parameter space, there is a constant C such that

$$
\begin{aligned}
&\left|\hat{Q}_n(\tilde{\alpha})-\hat{Q}_n(\alpha)\right| \\
&\lesssim \left|\|\hat{g}(\tilde{\alpha})\|_E^2-\|\hat{g}(\alpha)\|_E^2\right|/n \\
&\leq \left(\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E^2+2\|\hat{g}(\alpha)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E\right)/n \\
&\leq \Bigg(\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E^2+2\|\hat{g}(\alpha)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E \\
&\quad +2\|\hat{g}(\alpha_0)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E-2\|\hat{g}(\alpha_0)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E\Bigg)/n \\
&\leq \Bigg(\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E^2+2\|\hat{g}(\alpha)-\hat{g}(\alpha_0)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E \\
&\quad +2\|\hat{g}(\alpha_0)\|_E\cdot\|\hat{g}(\tilde{\alpha})-\hat{g}(\alpha)\|_E\Bigg)/n \\
&=\Bigg\{\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int(\rho_j(Z,\tilde{\alpha})-\rho_j(Z,\alpha))\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2 \\
&\quad +2\Bigg[\left(\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int(\rho_j(Z,\alpha)-\rho_j(Z,\alpha_0))\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\right) \\
&\qquad\quad \times\left(\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int(\rho_j(Z,\tilde{\alpha})-\rho_j(Z,\alpha))\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\right)\Bigg]^{1/2} \\
&\quad +2\Bigg[\left(\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int\rho_j(Z,\alpha_0)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\right) \\
&\qquad\quad \times\left(\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int(\rho_j(Z,\tilde{\alpha})-\rho_j(Z,\alpha))\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\right)\Bigg]^{1/2}\Bigg\}/n \\
&\leq\Bigg\{\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int b_j(Z)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\|\tilde{\alpha}-\alpha\|_{s,\alpha}^{2\nu} \\
&\quad +2\Bigg[\left(\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\left(\int b_j(Z)\hat{f}_{Y|X}(y\mid x_i;\hat{\psi})dy\right)^2\|\alpha-\alpha_0\|_{s,\alpha}^{2\nu}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\qquad \times \left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \|\tilde{\alpha} - \alpha\|_{s,\alpha}^{2\nu} \right) \Bigg]^{1/2} \\
&+ 2 \Bigg[\left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int \rho_j(Z, \alpha_0) \right)^2 \right) \\
&\qquad \times \left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \|\tilde{\alpha} - \alpha\|_{s,\alpha}^{2\nu} \right) \Bigg]^{1/2} \Bigg\} / n \\
= &\Bigg\{ \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \|\tilde{\alpha} - \alpha\|_{s,\alpha}^{\nu} \\
&+ 2 \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \|\alpha - \alpha_0\|_{s,\alpha}^{\nu} \\
&+ 2 \Bigg[\left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int \rho_j(Z, \alpha_0) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \\
&\qquad \times \left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \Bigg]^{1/2} \Bigg\} \|\tilde{\alpha} - \alpha\|_{s,\alpha}^{\nu} / n \\
\le &B_n \|\tilde{\alpha} - \alpha\|_{s,\alpha}^{\nu},
\end{aligned}
$$

where $B_n = C\tilde{B_n}$ for some constant C and

$$
\begin{aligned}
\tilde{B_n} = &\Bigg\{ \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \\
&+ 2 \Bigg[\left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int (\rho_j(Z, \alpha_0)) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \\
&\qquad \times \left(\sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \left(\int b_j(Z) \hat{f}_{Y|X}(y \mid x_i; \hat{\psi}) dy \right)^2 \right) \Bigg]^{1/2} \Bigg\} / n.
\end{aligned}
$$

Then hypothesis (iii) follows by $B_n = C \cdot O_p(1)$. $\square$

Proof of Theorem 4.1 (i) See Lemma 2 in Hu and Schennach (2008).

(ii) We prove the results by verifying the hypotheses of Lemma A.1 of Newey and Powell (2003). Hypothesis (i) follows by Theorem 4.1 of Newey and Powell (2003). Hypothesis (ii) follows by Lemma A.1. Note that hypotheses (i) and (iii) of Lemma A.1 are satisfied by Assumption 4.6 and hypothesis (ii) of Lemma A.1 is satisfied by the result in Theorem 4.1 (i). Finally, we verify hypothesis (iii) by choosing $\Pi_n \alpha \in \mathcal{A}_n$ such that $\|\Pi_n \alpha - \alpha\|_{s,\alpha} = o(1)$. $\square$

Lemma A.2 *Suppose that* $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. *Then we have* (i) *under Assumptions 4.1, 4.6, 4.8 and 4.9,* $\|\hat{g}(\alpha) - g(\alpha)\|_E^2/n = o_p(n^{-1/2})$ *uniformly over* $\alpha \in \mathcal{A}$; (ii) *under Assumptions 4.1-4.2 and 5.1,* $\|\hat{g}(\alpha_0)\|_E^2/n = O_p(\delta_{1n})$ *such that* $\delta_{1n} = o(1)$.

Proof of Lemma A.2 (i) From the proof of Lemma A.1, we have

$$
\begin{aligned}
&E\left[\|\hat{g}(\alpha) - g(\alpha)\|_E^2/n\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\left[\left(\int \rho_j(Z,\alpha)\left(\hat{f}_{Y|X}(y \mid x_i;\hat{\psi}) - f_{Y|X}(y \mid x_i;\psi_0)\right) dy\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\Bigg[\Bigg(\int\int \rho_j(Z,\alpha)\Bigg(\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi})\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) \\
&\qquad - f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\psi_0)f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})\Bigg)dy_1dy_2\Bigg)^2\Bigg] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\Bigg[\Bigg(\int\int \rho_j(Z,\alpha)\Bigg(\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi})\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) \\
&\qquad - \hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi})f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) \\
&\qquad + \hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi})f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) \\
&\qquad - f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\psi_0)f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})\Bigg)dy_1dy_2\Bigg)^2\Bigg] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\Bigg[\Bigg(\int\int \rho_j(Z,\alpha) \\
&\qquad \times \Bigg\{f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})\left(\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi}) - f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\psi_0)\right) \\
&\qquad + \hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\hat{\psi}) \\
&\qquad\quad \times \left(\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})\right)\Bigg\}dy_1dy_2\Bigg)^2\Bigg] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\Bigg[\Bigg(\int\int \rho_j(Z,\alpha) \\
&\qquad \times \Bigg\{\left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})}\right)\frac{d}{d\psi}f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i};\psi_0)[\hat{\psi} - \psi_0] \\
&\qquad\quad \times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})
\end{aligned}
$$

$$+\left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})})\right) f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})$$

$$\times [\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})]\Big\} dy_1 dy_2\Big)^2\Big]$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\Big[\int\Big(\int \rho_j(Z,\alpha)$$

$$\times\Big\{\left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})}\right)\frac{d}{d\psi} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi}-\psi_0]$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})$$

$$+\left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})})\right) f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})$$

$$\times [\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})]\Big\} dy_2\Big)^2 dy_1\Big]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\int E\Big[\Big(f_{Y^*|X}(y_i^* \mid x_i; \beta_0)\frac{1}{f_{Y^*|X}(y_i^* \mid x_i; \beta_0)}\int \rho_j(Z,\alpha)$$

$$\times\Big\{\left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})}\right)\frac{d}{d\psi} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi}-\psi_0]$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})$$

$$+\left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})})\right) f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})$$

$$\times [\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})]\Big\} dy_2\Big)^2\Big] dy_1$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho}\int\left(\sup_x f_{Y^*|X}(y_i^* \mid x_i; \beta_0)\right)\left(\sup_x \sup_{y_2} \rho_j(Z,\alpha)\right) E\Big[\Big(\frac{1}{f_{Y^*|X}(y_i^* \mid x_i; \beta_0)}$$

$$\times\int\Big\{\left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})}\right)\frac{d}{d\psi} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0)[\hat{\psi}-\psi_0]$$

$$\times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})$$

$$
\begin{aligned}
&+ \left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})}) \right) f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) \\
&\qquad \times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) \\
&\qquad \times [\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})] \Big\} dy_2 \Big)^2 \Big] dy_1 \\
&\le \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \int \left(\sup_x f_{Y^*|X}(y_i^* \mid x_i; \beta_0) \right) \left(\sup_x \sup_{y_2} \rho_j(Z, \alpha) \right) \\
&\quad \times \left(\max \left\{ \sup_x \sup_{y_2} \left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right), \right. \right. \\
&\qquad \left. \left. \sup_x \sup_{y_2} \left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right) \right\} \right) \\
&\times E \Big[\Big(\frac{1}{f_{Y^*|X}(y_i^* \mid x_i; \beta_0)} \Big\{ \int \frac{d}{d\psi} f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) [\hat{\psi} - \psi_0] \\
&\qquad \times f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) dy_2 \\
&\quad + \int f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i}) \\
&\qquad \times [\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i}) - f_{Y_2|X_2X_1}(y_2 \mid x_{2i}, x_{1i})] dy_2 \Big\} \Big)^2 \Big] dy_1 \\
&\le \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \int \left(\sup_x f_{Y^*|X}(y_i^* \mid x_i; \beta_0) \right) \left(\sup_x \sup_{y_2} \rho_j(Z, \alpha) \right) \\
&\quad \times \left(\max \left\{ \sup_x \sup_{y_2} \left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right), \right. \right. \\
&\qquad \left. \left. \sup_x \sup_{y_2} \left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right) \right\} \right) \\
&\quad \times E \left\{ \left(\frac{d \ln f_{Y^*|X}(y_i^* \mid x_i; \beta_0)}{d\beta} [\hat{\beta} - \beta_0] \right)^2 \right\} dy_1 \\
&\le \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{d_\rho} \int \left(\sup_x f_{Y^*|X}(y_i^* \mid x_i; \beta_0) \right) \left(\sup_x \sup_{y_2} \rho_j(Z, \alpha) \right) \\
&\quad \times \left(\max \left\{ \sup_x \sup_{y_2} \left(\frac{1}{f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right), \right. \right. \\
&\qquad \left. \left. \sup_x \sup_{y_2} \left(\frac{\hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \hat{\psi})}{f_{Y_1|Y_2X_1}(y_1 \mid y_2, x_{1i}; \psi_0) f_{Y_2^*|Y_2X_1}(y_{2i}^* \mid y_2, x_{1i})} \right) \right\} \right) dy_1 \|\hat{\beta} - \beta_0\|_\beta^2 \\
&= o(n^{-1/2}),
\end{aligned}
$$

since $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. Thus the result follows by the Markov inequality.

(ii) See Corollary A.1 (ii) in Ai and Chen (2003). $\square$

Lemma A.3 *(i) Under Assumptions 4.1-4.2, 4.3(ii), 4.6(iii), and 4.10, we obtain uniformly over $\alpha \in \{\mathcal{A}_n : \|\alpha - \alpha_0\|_\alpha = o(1)\}$: $(1/n)\sum_{i=1}^n \|m(X_i, \alpha)\|_E^2 - E[\|m(X,\alpha)\|_E^2] = o_p(n^{-1/2})$.*

(ii) Suppose that $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. Under Assumptions 4.1-4.3, 4.6, 4.8, 4.10, we obtain uniformly over $\alpha \in \mathcal{A}$, $\|\alpha - \alpha_0\|_\alpha = o(\eta_n)$: $(1/n)\sum_{i=1}^n \|\hat{m}(X_i, \alpha)\|_E^2 = o_p(\eta_n^2)$ and $(1/n)\sum_{i=1}^n \|m(X_i, \alpha)\|_E^2 = o_p(\eta_n^2)$, where $\eta_n = n^{-\tau}$ with $\tau \le 1/4$.

Proof of Lemma A.3 (i) See Corollary A.2(i) of Ai and Chen (2003).

(ii) The result follows from applying Lemma A.2(i) and A.3(i), and $E[\|m(X,\alpha)\|_E^2] = o(\eta_n^2)$ by Assumptions 4.3(ii) and 4.9. $\square$

Lemma A.4 *Suppose that $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. Assumptions 4.1-4.3, 4.6, and 4.8-4.9 imply: (i) $\hat{Q}_n(\alpha) - Q_n(\alpha) = o_p(n^{-1/4})$ uniformly over $\alpha \in \mathcal{A}_n$; and (ii) $\hat{Q}_n(\alpha) - \hat{Q}_n(\alpha_0) - \{Q_n(\alpha) - Q_n(\alpha_0)\} = o_p(\eta_n n^{-1/4})$ uniformly over $\alpha \in \mathcal{A}_n$ with $\|\alpha - \alpha_0\|_\alpha \le o(\eta_n)$, where $\eta_n = n^{-\tau}$ with $\tau \le 1/4$.*

Proof of Lemma A.4 (i) The result follows from Lemma A.2(i) and Assumption 4.3.

(ii) The result follows from Lemma A.3 and Lemma A.4(ii). $\square$

Proof of Theorem 4.2 (i) See Theorem 2 in Hu and Schennach (2008).

(ii) It follows from a similar argument of Theorem 3.1 in Ai and Chen (2003). $\square$

Let

$$\frac{d\hat{g}(\alpha)}{d\alpha}[v_{2n}^*] = \left(\frac{d\hat{m}(X_1,\alpha)}{d\alpha}[v_{2n}^*], \cdots, \frac{d\hat{m}(X_n,\alpha)}{d\alpha}[v_{2n}^*]\right)',$$
$$\frac{dg(\alpha)}{d\alpha}[v_{2n}^*] = \left(\frac{dm(X_1,\alpha)}{d\alpha}[v_{2n}^*], \cdots, \frac{d\tilde{m}(X_n,\alpha)}{d\alpha}[v_{2n}^*]\right)',$$

where

$$\frac{d\hat{m}(X,\alpha)}{d\alpha}[v_{2n}^*] = \int \frac{d\rho(Z,\alpha)}{d\alpha}[v_{2n}^*]\hat{f}_{Y|X}(y \mid x; \hat{\psi})dy,$$
$$\frac{dm(X,\alpha)}{d\alpha}[v_{2n}^*] = \int \frac{d\rho(Z,\alpha)}{d\alpha}[v_{2n}^*]f_{Y|X}(y \mid x; \psi_0)dy,$$

by the interchangability of integral and derivative. Recall the definition of neighborhoods $\mathcal{N}_{02n}$ and $\mathcal{N}_{02}$ introduced in Section 5.

Lemma A.5 *(i) Assumptions 4.1, 4.8 and 5.1, 5.3, 5.5-5.6 imply*:

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{02n}} \frac{1}{n}\left\| \frac{dg(\tilde{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dg(\alpha_0)}{d\alpha}[v_{2n}^*] \right\|_E^2 = o_p(n^{-1/2}).$$

(ii) In addition, if $\|\hat{\beta}-\beta_0\|_\beta = o_p(n^{-1/4})$ *holds, then*

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{02n}} \frac{1}{n}\left\| \frac{d\hat{g}(\tilde{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dg(\tilde{\alpha})}{d\alpha}[v_{2n}^*] \right\|_E^2 = o_p(n^{-1/2}).$$

Proof of Lemma A.5 (i) The result can be proved by the same argument of Corollary C.1 (ii) of Ai and Chen (2003).

(ii) We have

$$\begin{aligned}
&E\left[\frac{1}{n}\left\| \frac{d\hat{g}(\tilde{\alpha})}{d\alpha}[v_{2n}^*] - \frac{dg(\tilde{\alpha})}{d\alpha}[v_{2n}^*] \right\|_E^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_\rho} E\left[\left(\int \frac{d\rho_j(Z,\tilde{\alpha})}{d\alpha}[v_{2n}^*]\left(\hat{f}_{Y|X}(y \mid x_i;\hat{\psi}) - f_{Y|X}(y \mid x_i;\psi_0)\right)\right)^2\right] \\
&= o(n^{-1/2}),
\end{aligned}$$

uniformly over $\tilde{\alpha}\in\mathcal{N}_{02n}$ since $\|\hat{\beta}-\beta_0\|_\beta = o_p(n^{-1/4})$. Thus the result follows by the Markov inequality. $\square$

Let

$$\begin{aligned}
\frac{d\hat{Q}_n(\alpha)}{d\alpha}[v_{2n}^*] &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{m}(X_i,\alpha)}{d\alpha}[v_{2n}^*]\right\}'[\hat{\Sigma}(X_i)]^{-1}\left\{\frac{d\hat{m}(X_i,\alpha)}{d\alpha}[v_{2n}^*]\right\}, \\
\frac{dQ_n(\alpha)}{d\alpha}[v_{2n}^*] &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha)}{d\alpha}[v_{2n}^*]\right\}'[\Sigma(X_i)]^{-1}\left\{\frac{dm(X_i,\alpha)}{d\alpha}[v_{2n}^*]\right\}
\end{aligned}$$

and

$$\begin{aligned}
\frac{d^2\hat{g}(\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*] &= \left(\frac{d^2\hat{m}(X_1,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*],\cdots,\frac{d^2\hat{m}(X_n,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*]\right)', \\
\frac{d^2 g(\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*] &= \left(\frac{d^2 m(X_1,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*],\cdots,\frac{d^2 m(X_n,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*]\right)',
\end{aligned}$$

where

$$\begin{aligned}
\frac{d^2\hat{m}(X,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*] &= \int \frac{d^2\rho(Z,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*]\hat{f}_{Y|X}(y \mid x;\hat{\psi})dy, \\
\frac{d^2 m(X,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*] &= \int \frac{d^2\rho(Z,\alpha)}{d\alpha d\alpha}[v_{2n}^*,v_{2n}^*]f_{Y|X}(y \mid x;\psi_0)dy.
\end{aligned}$$

by the interchangability of integral and derivative.

Lemma A.6 *Suppose that* $\|\hat{\beta} - \beta_0\|_\beta = o_p(n^{-1/4})$. *(i) Under Assumptions 4.1, 4.3-4.4 4.6, 4.8, 4.10, and 5.8, we have*

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{02n}} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d^2\hat{m}(X_i,\tilde{\alpha})}{d\alpha d\alpha}[v_{2n}^*, v_{2n}^*]\right\}' [\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha}) = o_p(n^{-1/4}).$$

(ii) Under Assumptions 4.1, 4.3, 4.8, 5.1(ii), 5.3, 5.5-5.6, we have

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{02n}} \frac{d\hat{Q}_n(\tilde{\alpha})}{d\alpha}[v_{2n}^*] = \frac{dQ_n(\alpha_0)}{d\alpha}[v_{2n}^*] + o_p(n^{-1/4}).$$

Proof of Lemma A.6 Proof is similar to Ai and Chen (2003). (i) For some constant C, Assumption 4.3 implies

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d^2\hat{m}(X_i,\tilde{\alpha})}{d\alpha d\alpha}[v_{2n}^*, v_{2n}^*]\right\}' [\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\right| \\ &\le C\sqrt{\left\|\frac{d^2\hat{g}(\alpha)}{d\alpha d\alpha}[v_{2n}^*, v_{2n}^*]\right\|_E^2/n}\sqrt{\|\hat{g}(\alpha)\|_E^2/n}.\end{aligned}$$

Then the result follows from Lemma A.3(ii) because we have that uniformly over $\tilde{\alpha} \in \mathcal{N}_{02n}$,

$$\left\|\frac{d^2\hat{g}(\alpha)}{d\alpha d\alpha}[v_{2n}^*, v_{2n}^*]\right\|_E^2/n \le c_1(Z)^2 = O_p(1)$$

by Assumption 5.8.

(ii) Uniformly over $\tilde{\alpha} \in \mathcal{N}_{02n}$,

$$\begin{aligned}&\frac{d\hat{Q}_n(\tilde{\alpha})}{d\alpha}[v_{2n}^*] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{m}(X_i,\alpha)}{d\alpha}[v_{2n}^*] - \frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\}' [\hat{\Sigma}(X_i)]^{-1}\left\{\frac{d\hat{m}(X_i,\alpha)}{d\alpha}[v_{2n}^*]\right\} \\ &\quad+ \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\}' [\hat{\Sigma}(X_i)]^{-1}\left\{\frac{d\hat{m}(X_i,\alpha)}{d\alpha}[v_{2n}^*] - \frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\} \\ &\quad+ \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\}' \{[\hat{\Sigma}(X_i)]^{-1} - [\Sigma(X_i)]^{-1}\}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\} \\ &\quad+ \frac{dQ_n(\alpha_0)}{d\alpha}[v_{2n}^*].\end{aligned}$$

The result follows from Assumption 4.3 and Lemma A.5. □

Lemma A.7 *(i) Under Assumptions 4.1 4.3-4.4, 4.6, 4.8, 4.10, 5.1(ii), 5.3, 5.5-5.6, we have uniformly over* $\tilde{\alpha} \in \mathcal{N}_{02n}$:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{m}(X_i,\tilde{\alpha})}{d\alpha}[v_{2n}^*]\right\}'[\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\\&=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m(X_i,\alpha_0)}{d\alpha}[v_2^*]\right\}'[\Sigma(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})+o_p(n^{-1/2}).\end{aligned}$$

(ii) Under Assumptions 4.1, 5.1(ii), 5.3, 5.5-5.7, we have uniformly over $\tilde{\alpha} \in \mathcal{N}_{02n}$:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v_2^*]\right\}'[\Sigma(X_i)]^{-1}\{\hat{m}(X_i,\tilde{\alpha})-\hat{m}(X_i,\alpha_0)\}=\langle v_2^*,\tilde{\alpha}-\alpha_0\rangle_\alpha+o_p(n^{-1/2}).$$

(iii) Under Assumptions 4.1, 4.3(ii), 4.8, 5.1(iii), 5.3, we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v_2^*]\right\}'[\Sigma(X_i)]^{-1}\hat{m}(X_i,\alpha_0)\\&=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m(X_i,\alpha_0)}{d\alpha}[v_2^*]\right\}'[\Sigma(X_i)]^{-1}\rho(X_i,\alpha_0)+o_p(n^{-1/2}).\end{aligned}$$

Proof of Lemma A.7 (i) Uniformly over $\tilde{\alpha} \in \mathcal{N}_{02n}$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{m}(X_i,\tilde{\alpha})}{d\alpha}[v_{2n}^*]\right\}'[\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\\&-\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m(X_i,\alpha_0)}{d\alpha}[v_2^*]\right\}'[\Sigma(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\\&=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\hat{m}(X_i,\tilde{\alpha})}{d\alpha}[v_{2n}^*]-\frac{dm(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\}'[\hat{\Sigma}(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m(X_i,\alpha_0)}{d\alpha}[v_{2n}^*]\right\}'\left\{[\hat{\Sigma}(X_i)]^{-1}-[\Sigma(X_i)]^{-1}\right\}\hat{m}(X_i,\tilde{\alpha})\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m(X_i,\alpha_0)}{d\alpha}[v_{2n}^*-v_2^*]\right\}'[\Sigma(X_i)]^{-1}\hat{m}(X_i,\tilde{\alpha})\\&\equiv A_1+A_2+A_3.\end{aligned}$$

Then the result follows from the fact that $A_1 = o_p(n^{-1/2})$ by Lemma A.3(ii), A.5(ii), and Assumption 4.3(ii); $A_2 = o_p(n^{-1/2})$ by Assumption 4.3(iii) and Lemma A.3(ii); $A_3 = o_p(n^{-1/2})$ by Assumption 5.3 and Lemma A.3(ii).

(ii) Let $\varphi(X, v_2^*) = \left(\frac{dm(X,\alpha_0)}{d\alpha}[v_2^*]\right)' \Sigma(X)^{-1}$ and let

$$\tilde{\mathcal{F}} = \Big\{\varphi(X, v^*)\tilde{m}(X, \alpha) : \alpha \in \mathcal{N}_{02n}, \tilde{m} \in \Lambda_c^\gamma(\mathcal{X}) \text{ s.t.}$$
$$\sup_{x\in\mathcal{X},\alpha\in\mathcal{N}_{02n}} |\tilde{m}(x, \alpha) - m(x, \alpha)| = o(1)\Big\},$$
$$\mathcal{F} = \Big\{\varphi(X, v^*)m(X, \alpha) : \alpha \in \mathcal{N}_{02n}\Big\}.$$

By a similar argument to Corollary C.3(ii) of Ai and Chen (2003), $\tilde{\mathcal{F}}$ and $\mathcal{F}$ are Donsker classes, and we have uniformly over $\alpha \in \mathcal{N}_{02n}$,

$$\frac{1}{n}\sum_{i=1}^n \varphi(X_i, v_2^*)\{\hat{m}(X_i, \alpha) - m(X_i, \alpha)\} - E\left[\varphi(X_i, v_2^*)\{\hat{m}(X_i, \alpha) - m(X_i, \alpha)\}\right]$$
$$= o_p(n^{-1/2}), \tag{2.13}$$
$$\frac{1}{n}\sum_{i=1}^n \varphi(X_i, v_2^*)\{\hat{m}(X_i, \alpha_0) - m(X_i, \alpha_0)\} - E\left[\varphi(X_i, v_2^*)\{\hat{m}(X_i, \alpha_0) - m(X_i, \alpha_0)\}\right]$$
$$= o_p(n^{-1/2}), \tag{2.14}$$
$$\frac{1}{n}\sum_{i=1}^n \varphi(X_i, v_2^*)\{\tilde{m}(X_i, \alpha) - m(X_i, \alpha_0)\} - E\left[\varphi(X_i, v^*)\{m(X_i, \alpha) - m(X_i, \alpha_0)\}\right]$$
$$= o_p(n^{-1/2}). \tag{2.15}$$

From eqns. 2.14 and 2.15,

$$\frac{1}{n}\sum_{i=1}^n \varphi(X_i, v_2^*)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\} \tag{2.16}$$
$$= \frac{1}{n}\sum_{i=1}^n \varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\} + E\left[\varphi(X_i, v_2^*)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\}\right]$$
$$- E\left[\varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}\right] + o_p(n^{-1/2}).$$

Let $\tilde{\varphi}(X_i, v_2^*) = \int[\int \varphi \hat{f}_{Y_1|Y_2X_1}(y_1 \mid y_2, x_1; \hat{\phi}, \hat{\eta})dy_1]\hat{f}_{Y_2|X_2X_1}(y_2 \mid x_2, x_1)dy_2$. Then we have

$$E\left[\varphi(X_i, v_2^*)\{\hat{m}(X_i, \hat{\alpha}) - \hat{m}(X_i, \alpha_0)\}\right] = E\left[\tilde{\varphi}(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}\right], \tag{2.17}$$
$$E\left[\tilde{\varphi}(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}\right] - E\left[\varphi(X_i, v_2^*)\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}\right]$$
$$= E\left[\{\tilde{\varphi}(X_i, v_2^*) - \varphi(X_i, v_2^*)\}\{m(X_i, \hat{\alpha}) - m(X_i, \alpha_0)\}\right] \tag{2.18}$$
$$= o_p(n^{-1/2}).$$

Plugging eqns. 2.16, 2.18 and 2.19 into 2.17 gives for some $\bar{\alpha} \in \mathcal{N}_{02}$, a convex combination of $\hat{\alpha}$ and α_0 that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i, v_2^*)\{\hat{m}(X_i,\hat{\alpha}) - \hat{m}(X_i,\alpha_0)\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\varphi(X_i, v_2^*)\{m(X_i,\hat{\alpha}) - m(X_i,\alpha_0)\} + o_p(n^{-1/2})\\
&= E\left[\varphi(X_i, v_2^*)\{m(X_i,\hat{\alpha}) - m(X_i,\alpha_0)\}\right] + o_p(n^{-1/2})\\
&= E\left[\varphi(X_i, v_2^*)\frac{dm(X_i,\bar{\alpha})}{d\alpha}[\hat{\alpha}-\alpha_0]\right] + E\left[\varphi(X_i, v_2^*)\frac{dm(X_i,\alpha_0)}{d\alpha}[\hat{\alpha}-\alpha_0]\right]\\
&\quad - E\left[\varphi(X_i, v_2^*)\frac{dm(X_i,\alpha_0)}{d\alpha}[\hat{\alpha}-\alpha_0]\right] + o_p(n^{-1/2})\\
&= \langle v_2^*, \hat{\alpha}-\alpha_0\rangle_\alpha + E\left[\varphi(X_i, v_2^*)\left(\frac{dm(X_i,\bar{\alpha})}{d\alpha}[\hat{\alpha}-\alpha_0] - \frac{dm(X_i,\alpha_0)}{d\alpha}[\hat{\alpha}-\alpha_0]\right)\right]\\
&\quad + o_p(n^{-1/2})\\
&= \langle v_2^*, \hat{\alpha}-\alpha_0\rangle_\alpha + o_p(n^{-1/2}),
\end{aligned}$$

where the third, fourth and fifth equalities follow from the mean value theorem, the definition of $\langle v_2^*, \hat{\alpha}-\alpha_0\rangle$, and Assumption 5.1($ii$) and 5.7, respectively.

(iii) From the definition of $\tilde{\varphi}(X_i, v_2^*)$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i, v_2^*)\{\hat{m}(X_i,\alpha_0) - \frac{1}{n}\sum_{i=1}^{n}\varphi(X_i, v_2^*)\rho(Z_i,\alpha_0)\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varphi}(X_i, v_2^*) - \varphi(X_i, v_2^*)\right)\rho(Z_i,\alpha_0).
\end{aligned}$$

Then the result follows from the same argument of Corollary C.3(iii) in Ai and Chen (2003). □

Proof of Theorem 5.1 (i) See Theorem 3 in Hu and Schennach (2008).

(ii) It follows from a similar argument of Theorem 4.1 in Ai and Chen (2003). □

Proof of Theorem 5.2

See Theorem 5.1 in Ai and Chen (2003). □

2.9 Tables

Table 2.1: Monte Carlo simulation results

Estimator	Zero Mode	Zero Mean	Zero Median
Infeasible SMD	0.14334	0.15796	0.14059
SML-SMD	0.14683	0.17255	0.14683
Inconsistent SMD	0.23990	0.18691	0.15668

Chapter 3

Empirical Analysis of Endogeneity and Measurement Error in Nonparametric and Semiparametric Models

3.1 Introduction

In 1970's, many of studies show the importance of controlling for both measurement error and endogeneity due to simultaneous equation models. Most of the studies consider the case that exogenous variables are mismeasured. It is because of the property of linear models. As pointed out by Geraci (1997) and Hausman (1997) among others, consideration of additional measurement error in endogenous variables is not interesting since it is observationally equivalent to measurement error in exogenous variables as long as both measurement errors are uncorrelated. Indeed, if only an endogenous variable is mismeasured in the linear parametric model, one instrumental variable is sufficient to control for both endogeneity and measurement error unless the instrumental variable is correlated with measurement error and unobserved drivers of dependent variable.

However, it is not easy to identify parameters in nonlinear models in the presence of both endogeneity and measurement error because measurement error problem becomes a problem associated with the knowledge on distributions of measurement errors. To illustrate this point, consider the following nonlinear parametric model:

$$
\begin{aligned}
y_i &= f(x_i; \theta_0) + u_i \\
x_i^* &= x_i + e_i,
\end{aligned}
$$

where f is a known real-valued function and θ_0 is a vector of unknown parameters, and where x_i^* is a mismeasured version of true x_i and e_i is measurement error. So there is no endogeneity issue by construction. By plugging in x_i, the first equation becomes $y_i = f(x_i^* + e_i; \theta_0) + u_i$. If the function f is linear, it is simply $y_i = \theta_0 x_i^* + \epsilon_i$, where $\epsilon_i = u_i + \theta_0 e_i$, so that one instrumental variable which is uncorrelated with ϵ_i could control for the measurement error. Adding endogeneity to the equation causes no extra cost on the problem. However, if the function f is nonlinear, the problem deviates from the method of standard instrumental variables because x_i^* is not additively separable with the measurement error e_i anymore. This is one of the reasons why measurement error is differentiated from endogeneity issue in nonlinear models and only single issue among them has been considered in nonlinear models.

Since economic theory provides no general guideline in model specification and distribution of economic variables for econometricians, general nonparametric or semiparametric models become more popular. Chapter 1 and 2 consider both endogeneity and measurement error in nonparametric or semiparametric regression models. They contribute to the literature since there is no published work prior to them on the identification and estimation of nonparametric or semiparametric models in the presence of endogenous and mismeasured variables. Nevertheless, each chapter has its own distinct properties. Chapter 1 considers very general additively nonseparable models between regressors and unobserved drivers of dependent variable and shows the identification and estimation of covariate-conditioned average marginal effects. Chapter 2 restricts the model on additively separable one but allows nonclassical measurement error. It develops semiparametric estimation of models with conditional moment restricts, and shows that one instrumental variable is sufficient to identify and estimate parameters of interest, even when one regressor of interest is endogenous and mismeasured.

The purpose of the present chapter is to support the effectiveness of both methods in the previous chapters for empirical analysis. The structure of the paper is as follows. Section 1 uses the proposed method in chapter 1 to estimate the impact of family income on children's achievement. In section 2, we apply the proposed method in chapter 2 to the estimation of Engel curves. Section 3 concludes.

3.2 The Impact of Family Income on Children's Achievement

This section applies the proposed estimator in chapter 1 to study the causal effect of family income on child achievement. We also discuss how to choose optimal bandwiths since estimation results highly depend on the choice of the smoothing parameters.

3.2.1 Overview

The association between family income and child development is a contentious issue in economics, sociology, and developmental psychology. Even though it has been examined in a number of studies, there is no consensus on the relative effectiveness of income transfers and direct intervention in augmenting the human capital of children. Income transfers could have a significant impact on the economic well-being of children growing up in poor families if family income plays a substantial role in child development. If not, then direct interventions, such as the Head Start program, to improve child health, education, and parenting may be more effective.

Using data from the Panel Study of Income Dynamics (PSID), Duncan, Yeung, Brooks-Gunn, and Smith (1998) find that family income in early childhood has the greatest impact on completed schooling, especially among children in families with low incomes, regardless of whether they control for fixed family effects or not. Blau (1999) uses the matched mother-child subsample of the National Longitudinal Survey of Youth (NLSY) to estimate the impact of parental income on children's cognitive, social, and emotional development. He finds that OLS estimates of income effects are generally statistically significant and positive, but that they are smaller and insignificant when he uses either random- or fixed-effect strategies. In addition, his findings indicate that the effect of permanent income is much larger, but not large enough to make income transfer a feasible approach to achieving substantial improvements in child outcomes. He also find that there is no evidence for any systematic indication of diminishing returns to income, i.e., income effects that are larger at lower levels of income.

Aughinbaugh and Gittleman (2003) examine the relationship between child development and income in Great Britain and compare it with that in the United States. Using the NLSY and Great Britain's National Child Development Study, they find that the relationship between income and child development is quite similar in the two countries. Income tends to improve cognitive test scores, but the magnitude of the impact is small. Using participants from the National Institute of Child Health and Human Development (NICHD) study of Early Child Care, Taylor, Dearing, and McCartney (2004) estimate the impact of family economic resources on developmental outcomes in early childhood. They find that economic resources are important when properly compared with other important variables, such as maternal verbal intelligence, and when compared with established interventions, such as Early Head Start. Their findings also indicate that there are significant nonlinear effects of permanent (but not current) income, implying that income effects are larger for children living in poor families.

Dahl and Lochner (2005) address both omitted variables bias and attenuation bias due to measurement error on family income using fixed-effect (parametric) instrumental variables estimation. They use panel data on over $6,000$ children matched to their mothers in the NLSY data. They find that estimates from the fixed-effect instrumental variables approach are larger than cross-section OLS or standard fixed-effects estimates, so that current income has a significant effect on a child's math and reading test scores.

Here we examine the effect of family income on child achievement, as measured by scores on math and reading assessments. We address measurement errors, endogeneity of family income, and nonlinearity of income effects, by considering a data generating process of the form

$$Y = r(X, U_y),$$

where Y is child scholastic achievement, X is family income, and U_y represents other unobserved drivers of child achievement; r is an unknown measurable scalar-valued function. Because unobserved parents' ability could be a common cause of both family incomes and child achievement, the explanatory variable X is generally correlated with the error term U_y. Moreover, income is noisily measured in most surveys, and the data used here are no exception.

Figure 3.1 depicts the causal relation postulated to operate here. Mother's cognitive ability is a common cause for family earning potential and child ability. The fact that earning potential and child ability share a common cause induces a correlation between family income and child ability. Nevertheless, the conditional independence assumption makes it possible to recover features of the causal relationship. Because AFQT scores, a proxy for mother's cognitive ability, are observable, they serve as conditioning instruments to ensure the conditional independence between family income and unobserved child ability. Moreover, true family income is unobservable because income is noisily measured in survey data. Without correcting for the measurement error, estimates would be biased towards zero. Fortunately, we observe two error-laden measurements of true family income. This permits us to recover the desired effect measures using our estimator.

We also use the matched mother-child subsample of the NLSY from Dahl and Lochner (2005) in the cross-sectional nonparametric model.[1] The dependent variables, i.e., child scholastic achievement (Y) are measures of achievement in math and reading based on standardized scores of the Peabody Individual Achievement Tests (PIAT). Math achievement is measured by mathemathics scores, and reading achievement is measured by a simple average of the reading recognition and reading comprehension scores. We use measures of both current income and permanent income in different estimation equations. Our error-laden measurement of current family income (X_1) is the natural logarithm (log) of family income in 1998. The error-laden measurement of permanent family income (X_1) is the log of the average of family incomes in 1994, 1996, and 1998. The log of family income in year 2000 is commonly used as additional error-laden measurement of family income (X_2) for both current and permanent family income. Income in each year is after-tax and after-transfer. The conditioning instrument (W) is the mother's Armed Forces Qualifying Test (AFQT) score; see Dahl and Lochner (2005) for further details. We assume true family incomes and unobserved drivers of child achievement are independent, conditional on AFQT scores (i.e., $X \perp U_y \mid W$). We create standardized test scores, AFQT scores, and family incomes having mean zero and standard deviation one.

[1] We thank Gordon Dahl for providing the NLSY data.

3.2.2 Bandwidth Selection

We consider leaving-one-out cross-validation to estimate the optimal bandwidths. Let $h^{[k,\lambda]}$ be the minimizer of

$$Q_\lambda^k(h) = \frac{1}{n}\sum_{i=1,i\neq k}^{n}(y_i - f_h(x_i))^2 + \lambda\frac{1}{n}\sum_{i=1,i\neq k}^{n}(D_x^2 f_h(u_i))^2,$$

where $D_x^2 f_h(u_i)$ is the second derivative of $f_h(x)$ with respect to x which is evaluated at u_i. Let $f_{h^{[k,\lambda]}}$ be a value of f_h evaluated at $h^{[k,\lambda]}$. Then the cross-validation function $V_0(\lambda)$ is

$$V_0(\lambda) = \frac{1}{n}\sum_{k=1}^{n}(y_k - f_{h^{[k,\lambda]}}(x_k))^2.$$

We obtain the cross-validation estimate of the smoothing parameter λ by minimizing the cross-validation function $V_0(\lambda)$. From the optimal λ, we also obtain the cross-validation estimate of the bandwidths $h \equiv (h_1, h_2)$. This procedure is similar to the ordinary cross validation in Wahba (1990) except that here h are additional smoothing parameters to be estimated and instead of the integral, a sample average is used in the second term of $Q_\lambda^k(h)$. In the language of her book, $Q_\lambda^k(h)$ represents a tradeoff between fidelity to the data and smoothness of the solution. The first is represented by the mean square of residuals and the second is represented by the mean square of the second derivative. Thanks to the smoothing parameter λ controlling the tradeoff between fidelity and smoothness, one can choose optimal bandwidths even with noisy data.

Table 3.1 reports optimal choices of the smoothing parameters, λ and h. For the local linear estimator, we use the 2nd-order local polynomial estimator to obtain the smoothing parameters because it automatically estimates the second derivatives of $f_h(x)$ and both local linear and local polynomial estimators are first-order identical. Since Fourier estimator has more roughness in the estimated function, it obtains smaller λ, which means that more penalties are given to the term for smoothness.

3.2.3 Estimation Results

Tables 3.2 and 3.3 show estimation results obtained by our new estimator and a local linear estimator ignoring the family income measurement error. Each estimate is evaluated at given values of standardized family income (X) and mother's AFQT score (W) ranging from -0.8 to 0.8 in increments of 0.1. Estimates from only a subset of the covariates are reported for conciseness. Estimated smoothing parameters in Table 3.1 are used for each estimator. All standard errors of the estimates are obtained by bootstrap methods. As Gonçalves and White (2005) remarked, one must formally justify using the bootstrap to compute standard errors because the consistency of the bootstrap distribution does not guarantee the consistency of the variance of the bootstrap distribution as an estimator of the asymptotic variance. Nevertheless, the bootstrap gives us standard errors with first-order accuracy, which should be sufficient for our purposes.

Table 3.2 reports the estimated impact of family income on children's math achievement. The covariate-conditioned average marginal effects of family income on children's math achievement from our estimator are positive and significantly large over all ranges of x and w. The average marginal effect is about 8.764 at $x = -0.8$ and $w = 0.4$, which means that the effect of a one standard deviation increase in log of family income is to increase a child's math score by about 8.764 of a standard deviation. For given mother's AFQT scores, w, effects decrease as family income, x, increases toward about 0.2 but increase again when family income is above 0.2. Interestingly, the covariate-conditioned average marginal effects from the local linear estimator are much smaller than those from our estimator for all (x, w) values. Notice that the average marginal effect from the local linear estimator is about 0.079 at $x = -0.8$ and $w = 0.4$, whereas that from our estimator is 8.764, a difference of about 8.6. It follows measurement errors in family income have an important impact on estimated effects, and that use of our new estimator is critical to obtain accurate estimates here. Note that due to high standard errors, parts of effects from Fourier estimator are not statistically significant. Nevertheless, for the family in which mother's AFQT scores is positive, effects of family income on children's math scores, who are in poor families (range of family income is from

-0.8 to -0.6), are statistically significant and large.

Table 3.3 shows the impact of family income on children's reading achievement. The covariate-conditioned average marginal effects of family income on children's reading achievement from our estimator are also positive and much larger than those from the local linear estimator in all ranges of (x, w). The average marginal effect from our estimator, for instance, is about 4.718 at $x = -0.8$ and $w = 0.8$, which means that the effect of a one standard deviation increase in log of family income is to increase a child's reading score by about 4.718 of a standard deviation, while that from the local linear estimator is 0.055. As observed in math achievement, even though part of effects from Fourier estimator are not statistically significant, children in low family incomes have large and statistically significant effects of family income on their reading scores.

Figure 3.2 shows a graph of the covariate-conditioned average marginal effect (top) and average counterfactual response (bottom) of family income on children's math scores at various values of family income and mother's AFQT, ranging from -0.8 to 0.8, obtained using our estimator with bandwidths in Table 3.1. All estimates of the average marginal effect are positive over the ranges of both family income and AFQT score. In general, the impact of family income at a given AFQT increases as family income moves from 0.2 to -0.8 or 0.8, making a broad U-shape. It attains a minimum of 0.5823 at $x = 0.2$ and $w = -0.4$. As a result, one can find slightly increasing returns to family income for children in high family incomes. However, diminishing returns to family income are apparently observed at income levels below $x = 0.2$. We note that the shape of income effect is varying over different levels of mother's AFQT. For instance, at 0.4 of mother's AFQT, the average marginal effect is very dynamic, while that at -0.6 of mother's AFQT is flat. Thus, the average marginal effect depends on the level of mother's AFQT, which means the nonseparable model is appropriate for this example.

Figure 3.3 shows a graph of the apparent causal effect (top) and average counterfactual response (bottom) of family income on children's math scores obtained using the local linear estimator. It shows much smaller marginal effects than those from our estimator. And the average counterfactual response is more flat than that from our estimator. Moreover, it is interesting to note that the results from the local linear estimator indicate increasing returns to income, i.e., income effects that are larger at higher levels of family income, which is unexpected by the economic theory.

Figure 3.4 shows the covariate-conditioned average marginal effect (top) and average counterfactual response (bottom) of family income on children's reading scores at various points of family income and AFQT ranging from -0.8 to 0.8, obtained by our estimator. The same bandwidths are used as in Table 3.1 . The effects are always positive over the ranges of both family income and AFQT score as well. Children in poor families are likely to have higher effect of family income at a given value of AFQT. However, for children in families with income above 0 and whose mothers have low AFQT scores, the effect of family income on reading scores increases with family income. The effect attains a minimum value of 0.5057 at $x = 0$ and $w = 0.2$. We also observe the dependence of the average marginal effect on mother's AFQT.

Figure 3.5 depicts the apparent causal effect (top) and average counterfactual response (bottom) of family income on children's reading scores obtained using the local linear estimator. The results indicate much smaller income effects than those from our estimator. Family income shows increasing returns to income.

Taken as a whole, these results suggest that our estimator effectively accounts for the measurement errors of family income, compared to the local linear estimator, which ignores measurement errors. We find that the effects of family income on both math and reading scores from our estimator are positive and that the magnitudes of the income effects are substantially larger, whereas those apparent from the local linear estimator are statistically significant, but rather modest, as seen in previous studies. Because these results hold for family income, it follows that income transfers could have a significant impact on the development of children growing up in poor families. Our findings indicate nonlinearity in income effects over ranges of family income, specifically diminishing returns to income for families with income levels below $x = 0.2$ but a wide U-shape overall. Moreover, we observe that the income effect depends on the level of mother's AFQT scores, which supports the use of the nonseparable model for this application.

3.3 Instrumental Variables Estimation of Engel Curves

We apply the proposed estimator in chapter 2 to the estimation of Engel curves (or consumer demand models) using the British Family Expenditure Survey (FES) data. Findings confirm that correcting for both endogeneity and measurement error is necessary to identify the economically meaningful structural Engel curves.

3.3.1 Overview

Demand models play an important role in the welfare analysis. One of the reason is that the evaluation of indirect tax policy reform needs the accurate specification of demand models which is consistent with consumer theory. Because of that, the study of the Engel curves, the relationship between expenditure (or income) and budget shares, has been an area of interest among econometricans since the early studies of Engel (1895), Working (1943), Leser (1963). Many of previous studies exploits the best model specification for the Engel curves and 'Leser-Working' specification of Engel curve in which budget shares are a linear function of the log of income or expenditure, has been the most popular one. However, economic theory provides almost no general guidance in specification of Engel curves and recent empirical studies show that linear specification of the Engel curves is far from an accurate feature of consumer behavior. Some empirical analysis of consumer behavior suggest that nonlinear parametric or semiparametric and nonparametric models are more favorable in the specification of the Engel curves. Along with the model specification, there have been two directions in the analysis of Engel curves: endogeneity and measurement errors.

A group of studies estimate Engel curves based on that budget shares and expenditure are endogenous to the consumer and are determined simultaneously, as pointed out by Summers (1959). Using a nonparametric method and correcting for the endogeneity of the log-total expenditure, Banks, Blundell, and Lewbel (1997) suggest that Engel curves require quadratic terms in the log-total expenditure. They also find that models failing to account for nonlinearity of the Engel curves could distort the patterns of welfare losses associated with a tax increase. Blundell, Duncan, and Pendakur (1998) allow for endogeneity of the log-total expenditure by adopting a parametric additive control function approach to the partially linear regression context and find that taking accound of endogeneity has an important impact on the shape of the Engel curve relationship, while Blundell, Browning, and Crawford (2003) use a nonparametric control function technique to adjust for endogeneity. Base on a nonparametric method, Lyssiotou, Pashardes and Stengos (1999) find that controlling for endogeneity tends to be more supportive of the rank 3 hypothesis. Blundell, Chen, and Kristensen (2007) (BCK) studies a shape-invariant Engel curve with endogenous log-total expenditure by applying the sieve minimum distance estimation of conditional moment restrictions and find the importance of correcting for endogeneity. Chen and Pouzo (2008a, b) studies nonparametric or semiparametric estimation of conditional moment models with possibly nonsmooth residuals, respectively and applied their methods to estimate quantile Engel curves with endogenous log-total expenditure.

Another issue on the estimation of the Engel curves is measurement error in total expenditure. Measurement error would be because of survey errors or a form of errors which come from the discrepancy between purchases and consumption due to storage or waste. In a linear parametric model, Liviatan (1961) applies the the method of instrumental variables to the Engel curves, with income serving as the instrumental variable. Aasness, Biorn, and Skjerpen (1993) model measurement error in total expenditure to estimate Engel curves with panel data. Hausman, Newey, and Powell (1995) propose consistent estimators for nonlinear regression framework in the presence of measurement error. In their application to the Engel curves, they find that measurement error in income should be accounted for and 'Lesser-Working' specification should be generalized to higher-order terms in log income. Lewbel (1996) develops a consistent estimator of nonlinear Engel curves to correct for measurement errors in total expenditures on the left and right hand side since an observed budget share has expenditure in its denominator. Newey (2001) studies the estimation of nonlinear errors-in-variables models using simulated moments and a flexible disturbance distribution, and applies the models to Engel curves with expenditures measurement errors on the left and right hand side. Hasegawa and Kozumi (2001) correct for expenditure measurement errors on both the left and right sides in the 'Lesser-Working' specification. They propose the Bayesian estimation procedure in both models without an instrument variable and with an instrument variable. Schennach (2004b) proposes a general solution to measurement error in general nonlinear models when one repeated observation is available for each mismeasured variable and applies it to the estimation of Engel curves. She finds that the impact of measurement error in total expenditure can not be neglected.

Even though there are plenty of evidences that total expenditure is endogenous as well as mismeasured, there has been no study which corrects for both endogeneity and measurement error in nonlinear parametric, nonparametric, or semiparametric models. As discussed by Amemiya (1985) and Hsiao (1989), it is because nonlinear regression models with measurement error are difficult to estimate with standard linear instrumental variables approach, due to the lack of additive separability between true regressor and measurement error. The present study employs the method which is proposed in chapter 2, in order to fill this gap. So our target is to control for both endogeneity and measurement error in the nonparametric shapes of the Engel curves.

The nonparametric specification of Engel curves we consider is

$$E[Y_{1i,l} - h_l(Y_{2i}) \mid X_i] = 0, \quad l = 1, \cdots, 7, \tag{3.1}$$

where Y_{1il} is the budget share of household i on good l (e.g., 1: food-out, 2: food-in, 3: alcohol, 4: fares, 5: fuel, 6: leisure goods, and 7: travel). Y_{2i} is the log-total expenditure of household i that is endogenous and unobservable, and X_i is gross earnings of the head of household, which is the instrumental variable. We consider the no kids sample that consists of 628 observations. BCK have used the same data set as well as a subset of married couples with one or two children in their study of a shape-invariant system of IV Engel curves.[2] Table 3.4 summarizes descriptive statistics for the main variables in the data set. We see that budget shares on food-in, leisure, and travel are large, while food-out, alcohol, fares, and fuel are relatively small. Leisure goods have a large standard deviation. The mean and standard deviation for log nondurable expenditure are similar to those for log gross earnings. As shown in BCK, log-total expenditure and log earnings have a strong positive correlation, which is 0.5111. We also assume that log earnings are independent of the residual, $(Y_{1i} - h(Y_{2i}))$. So the log gross earnings would be a proper instrumental variable to analyze the conditional moment restriction model.

BCK assume that the log of total expenditure on nondurables and services is endogenous but measurement error-free. However, their approach is infeasible if the true log-total expenditure suffers from measurement errors so that only a mis-measured version is observed.[3] As reviewed above, indeed, many empirical papers on the estimation of Engel curves show that measurement errors on the log-total expenditure is considerable. As a result, failure of controlling for measurement errors makes it difficult to estimate the economically meaningful Engel curves.

3.3.2 Two-step SML-SMD Procedure

In order to use the two-step sieve maximum likelihood and sieve minimum distance (SML-SMD) estimator, we specify the conditional mean function as fol-

[2]We thank Richard Blundell for providing the UK Family Expenditure Survey data.

[3]We assume that there is no measurement error on the left-hand variable in Eqn. (12) to ease the argument. It could be possible because both expenditure on good l and total expenditure might have measurement errors but the budget share could be correctly reported one if proportion of error-laden expenditure on good l to error-laden total expenditure is the same as true budget share. For instance, assume there are multiplicative measurement errors on expenditure on good l and total expenditure such that $Y^*_{0i,l} = Y_{0i,l}e_{0i,l}$ and $Y^*_{2i} = Y_{2i}e_{2i}$ where $Y^*_{0i,l}$, $Y_{0i,l}$, and $e_{0i,l}$ are measurement error-laden expenditure of household i on good l, true expenditure of household i on good l, and its mesurement error, respectively, and where Y^*_{2i}, Y_{2i}, and e_{2i} are measurement error-laden total expenditure of household i, true total expenditure of household i, and its mesurement error, respectively. If $e_{0i,l} = e_{2i}$, we can get $\frac{Y^*_{0i,l}}{Y^*_{2i}} = \frac{Y_{0i,l}e_{0i,l}}{Y_{2i}e_{2i}} = \frac{Y_{0i,l}}{Y_{2i}}$.

lows

$$
\begin{aligned}
m(x,h) &\equiv \int_{\mathcal{Y}} (y_1 - h(y_2)) dF_{Y|X}(y \mid x; \phi_0, \eta_0) \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} (y_1 - h(y_2)) dF_{Y_1|Y_2X}(y_1 \mid y_2, x) \right] dF_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0) \qquad (3.2) \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} (y_1 - h(y_2)) f_{Y_1|Y_2X}(y_1 \mid y_2, x) dy_1 \right] f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0) dy_2 \\
&= \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} (y_1 - h(y_2)) f_{Y_1|Y_2}(y_1 \mid y_2) dy_1 \right] f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0) dy_2,
\end{aligned}
$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are the support of the distribution of Y_1 and Y_2, respectively. In the empirical application, $\mathcal{Y}_1 = [0, 0.350]$ and $\mathcal{Y}_2 = [3.609, 6.947]$. Since partially parameterizing distributions eases nonparametric estimation of densities, we allow $f_{Y_2|X}(y_2 \mid x)$ to be parameterized. In fact, the conditional distribution of log-total expenditure given log gross earnings is close to normal (see BCK), we specify $f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0)$ as normal distribution. This is one of useful properties of the two-step SML-SMD estimator, which the sieve minimum distance procedure can not utilize because of its nature of the estimation.

In the first step, we estimate the population conditional mean function $m(x,h)$ semiparametrically by $\hat{m}(x,h)$. To do this, we use a SML estimation to estimate $f_{Y_1|Y_2}(y_1 \mid y_2)$ and $f_{Y_2|X}(y_2 \mid x; \phi_0, \eta_0)$.

$$
\begin{aligned}
\beta_0 &= (\psi_0, f_{Y_1|Y_2}, f_{Y_2^*|Y_2})' \qquad (3.3) \\
&= \arg \max_{\beta=(\psi, f_0, f_1)' \in \mathcal{B}} E\left(\ln \int_{\mathcal{Y}_2} f_0(y_1 \mid y_2) f_1(y_2^* \mid y_2) f_{Y_2|X}(y_2 \mid x; \psi) dy_2 \right),
\end{aligned}
$$

where $\mathcal{B} \equiv \Psi \times \mathcal{F}_0 \times \mathcal{F}_1$ with $\Psi \equiv \Phi \times \mathcal{M}$ and $\psi_0 = (\phi_0, \eta_0)$.

We also approximate the unknown function $h \in \mathcal{H}$ by $h_n \in \mathcal{H}_n \equiv \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$ where $\mathcal{H}_n$ is some finite-dimensional approximation space that becomes dense in $\mathcal{H}$ as sample size $n \to \infty$. In the second step, the SMD estimator of unknown sieve coefficients of h_0 is obtained by applying the SMD procedure

$$
\hat{h}_n = \arg \min_{h_n \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} \hat{m}(X_i, h_n)' \hat{m}(X_i, h_n), \qquad (3.4)
$$

where $\hat{m}(X,h)$ is the plug-in SML estimator of $m(X, \alpha)$ for any fixed $h_n = (h_{1,n}, \ldots,$

$h_{q,n}$):

$$\hat{m}(x, h_n) \tag{3.5}$$
$$\equiv \int_{\mathcal{Y}_2} \left[\int_{\mathcal{Y}_1} (y_1 - h_n(y_2)) \hat{f}_{Y_1|Y_2}(y_1 \mid y_2) dy_1 \right] \hat{f}_{Y_2|X}(y_2 \mid x; \hat{\phi}_n, \hat{\eta}_n) dy_2.$$

For the purpose of comparison, we also estimate the Engel curves using SMD estimator from BCK, which does not control for measurement errors of log-total expenditure. Both SMD and SML-SMD estimators are constructed without smoothness constraints for simplicity. We use a power series of fourth order multiplied by the cumulative distribution function of a standard normal to approximate $h_0(Y_2)$ for both estimators. In the SMD estimator, a set of instruments, $\{1, X_2, X_2^2, \ldots, X_2^{k_n}\}$ for $k_n \geq 3$ is used to approximate the conditional mean function.

3.3.3 Estimation Results

Figures 3.6 $\sim$ 3.7 show estimated Engel curves for four of the goods in the system. We plot curves over a set of log-total expenditures ranging from 4.5 to 6.5. Engel curves from our SML-SMD estimator which controls for both endogeneity and measurement errors in the log-total expenditure are plotted by real curves, while those from SMD estimator which only control for endogeneity in the log-total expenditure are plotted by dashed curves.

We note several interesting features. For households with low log-total expenditure, shares of food-in from our SML-SMD estimator are bigger than those from SMD estimator. Food-out from SMD estimator is a reverse U-shape and values are similar over different level of log-total expenditure. But Food-out from SML-SMD estimator dramatically decreases as log-total expenditure increases. As a result, for households with low log-total expenditure, the estimated shares of food from our estimator, which is sum of food-in and food-out, are much bigger than those from SMD estimator, even though food shares of households with high log-total expenditure from both estimators look similar. The Engel curve for fuel from SMD estimator shows a reverse S-shape and is close to that from SML-SMD estimator. However, the estimated Engel curves for leisure from both estimator show huge gaps. For example, the estimated shares of leisure for households with high log-total from SML-SMD estimator are around 0.7 bigger than those from SMD estimator. Thus measurement errors in log-total expenditure can make it difficult to estimate the Engel curves and controlling for the measurement errors are necessary to get correct estimates of the Engel curves.

Our empirical results can be extended in several directions. First, as in BCK, we could consider shape invariant Engel curves and compare the shapes of the estimated Engel curves to theirs. Then corresponding semiparametric model is

$$E[Y_{1i,l} - h_l(Y_{2i} - \phi(X_{1i}^{'}\theta_1)) - X_{1i}^{'}\theta_{2,l} \mid X_i] = 0, \quad l = 1, \cdots, 7, \tag{3.6}$$

where $\phi(X_{1i}^{'}\theta_1)$ is a known function up to a finite set of unknown parameters θ_1 and can be interpreted as the log of a general equivalence scale for household i. X_{1i} is a vector of demographic variables that represent different household types and θ_2 is the vector of corresponding equivalence scales (see, e.g., Pendakur (1998) and Blundell, Browning, and Crawford (2003)) and $X_i = (X_{1i}, X_{2i})$. Second, we could consider smoothness constraints in the second-step of our estimation procedure and compare the shapes of the estimated Engel curves to theirs. The penalized SMD estimation is

$$\hat{h}_n = \arg \min_{h_n \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} \hat{m}(X_i, h_n)^{'} \hat{m}(X_i, h_n) + \lambda_n \hat{P}_n(h_n), \tag{3.7}$$

where $\hat{P}_n(h_n)$ is the penalization function on the smoothness and λ_n is the smoothing parameter. Third, our empirical analysis needs to carry out the robustness check of the estimated Engel curves with respect to the selection of sieve basis functions and the smoothing parameters in the smoothness constraints. Two approximations are required to proceed our SML-SMD estimation: one to approximate h and the other to approximate unknown densities. So it would be useful to examine how the choice of sieve basis and the smoothing parameter affect the shapes of the estimated Engel curves. Fourth, a semiparametric Hausman-test on measurement errors could be developed. Let $\hat{\theta}_{SML-SMD}$ and $\hat{\theta}_{SMD}$ denote the semiparametric esimate of θ under H_0 : Y_2 measurement error-free and H_1 :

Y_2 measurement error-laden, respectively and let $\hat{V}_{SML-SMD}$ and $\hat{V}_{SMD}$ denote the estimates of their respective variances. It then follows that $n(\hat{\theta}'_{SML-SMD} - \hat{\theta}'_{SMD})\hat{V}^{-1}(\hat{\theta}_{SML-SMD} - \hat{\theta}_{SMD}) \overset{\text{asy.}}{\sim} \chi^2_{q+1}$ under the null, where $\hat{V} \xrightarrow{p} V$ with V the asymptotic covariance matrix of $n(\hat{\theta}_{SML-SMD} - \hat{\theta}_{SMD})$.

3.4 Summary and Concluding Remarks

In the article, we study empirical importance of endogeneity and measurement error in economic examples. To do this, we apply the proposed methods in chapter 1 and 2 to topics of interest among (applied) econometricians, the impact of family income on children's achievement and the estimation of Engel curves, respectively. The application to the impact of family income on children's achievement finds that the effects of family income on both math and reading scores from the proposed estimator are positive and that the magnitudes of the income effects are substantially larger. We also observe that the income effect depends on the level of mother's AFQT scores, which supports the use of the nonseparable model for this application. From the application to the estimation of Engel curves, our findings indicate that correcting for both endogeneity and measurement error obtains significantly different shapes of Engel curves, compared to the method which ignores measurement error on total expenditure.

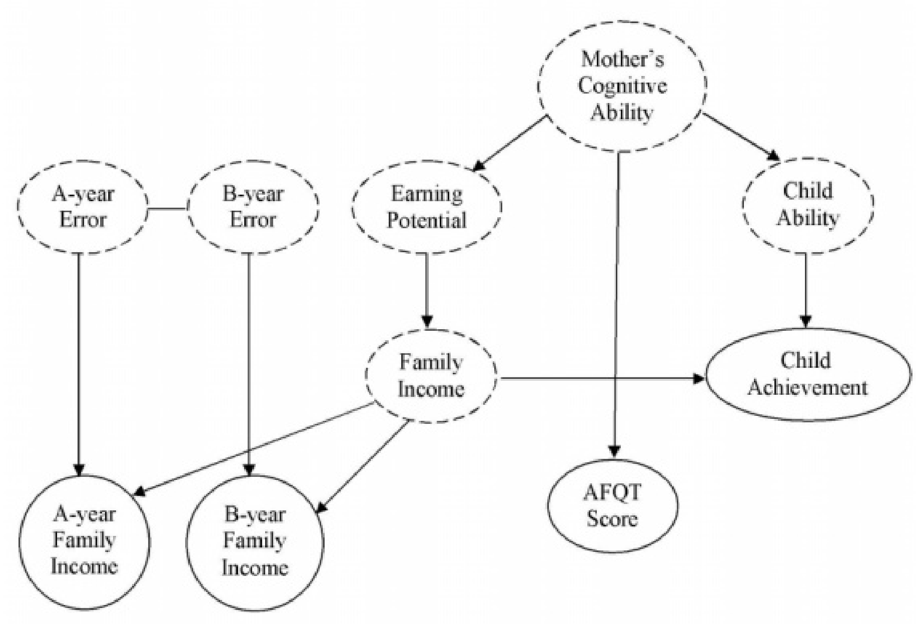
3.5 Acknowledgements

Part 2 of Chapter 3, the application to the impact of family income on children's achievement, is coauthored with Susanne M. Schennach and Halbert L. White.

3.6 Tables and Figures

Table 3.1: Optimal choice of smoothing parameters

		λ	h_1	h_2
Math	Fourier	10^{-15}	1.95	1.2
	Local linear	10^{-2}	6	6.7
Reading	Fourier	10^{-22}	2.15	1.9
	Local linear	10^{-4}	6.3	5.35



Notes: Due to common cause mother's cognitive ability, family income and child ability are correlated. AFQT score, a proxy for the common cause plays a key role as conditioning instrument ensuring conditional independence between family income and child ability. Two error-laden measurements of family income are used to get rid of attenuation bias due to measurement errors of family income.

Figure 3.1: Causal effects - impact of family income on child achievement

Table 3.2: Impact of family income on children's math achievement

w/x		-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
	F	2.386	1.845	1.570	1.456	1.465	1.603	1.927	2.599	4.103
-0.8		1.267	1.123	1.033	0.977	0.935	0.910	0.902	0.937	1.084
	L	0.140	0.159	0.178	0.199	0.222	0.248	0.279	0.319	0.379
		0.048	0.052	0.051	0.054	0.056	0.063	0.072	0.086	0.108
	F	1.226	0.888	0.717	0.635	0.612	0.641	0.735	0.936	1.361
-0.6		1.117	1.128	1.127	1.123	1.117	1.111	1.109	1.109	1.124
	L	0.131	0.154	0.176	0.199	0.224	0.252	0.285	0.325	0.380
		0.050	0.047	0.053	0.051	0.055	0.062	0.066	0.075	0.098
	F	1.477	0.999	0.759	0.636	0.583	0.582	0.637	0.771	1.059
-0.4		1.148	1.153	1.150	1.145	1.138	1.132	1.128	1.126	1.132
	L	0.123	0.148	0.173	0.200	0.227	0.257	0.291	0.331	0.381
		0.049	0.050	0.050	0.052	0.055	0.057	0.066	0.078	0.087
	F	1.510	1.117	0.902	0.789	0.745	0.757	0.835	1.008	1.360
-0.2		1.133	1.141	1.140	1.135	1.130	1.123	1.118	1.114	1.114
	L	0.113	0.142	0.171	0.200	0.230	0.262	0.297	0.336	0.382
		0.048	0.049	0.050	0.050	0.051	0.056	0.063	0.072	0.078
	F	2.133	1.650	1.363	1.207	1.146	1.169	1.287	1.543	2.047
0		1.166	1.191	1.197	1.195	1.191	1.187	1.182	1.179	1.179
	L	0.103	0.136	0.169	0.201	0.234	0.268	0.303	0.341	0.382
		0.049	0.048	0.049	0.050	0.051	0.056	0.063	0.070	0.082
	F	6.672	3.958	2.798	2.242	1.990	1.934	2.052	2.383	3.063
0.2		1.296	1.323	1.326	1.324	1.319	1.314	1.309	1.305	1.302
	L	0.091	0.130	0.167	0.203	0.239	0.274	0.310	0.345	0.383
		0.051	0.049	0.053	0.050	0.053	0.055	0.059	0.065	0.072
	F	8.764	4.769	3.218	2.511	2.196	2.117	2.239	2.603	3.356
0.4		1.366	1.382	1.385	1.383	1.380	1.377	1.373	1.369	1.367
	L	0.079	0.124	0.166	0.206	0.244	0.281	0.316	0.350	0.383
		0.055	0.050	0.053	0.052	0.053	0.056	0.063	0.062	0.074
	F	5.332	3.069	2.149	1.718	1.524	1.480	1.568	1.818	2.328
0.6		1.444	1.465	1.468	1.467	1.464	1.461	1.457	1.454	1.451
	L	0.065	0.118	0.165	0.210	0.251	0.289	0.323	0.355	0.383
		0.058	0.052	0.054	0.055	0.058	0.057	0.064	0.067	0.069
	F	2.998	1.955	1.433	1.160	1.025	0.980	1.013	1.136	1.391
0.8		1.589	1.628	1.633	1.631	1.627	1.623	1.618	1.614	1.610
	L	0.048	0.111	0.166	0.215	0.258	0.297	0.331	0.359	0.383
		0.067	0.057	0.059	0.056	0.059	0.061	0.062	0.069	0.070
	N	1544								

Notes: F and L refer to our Fourier estimator and local linear estimator, respectively. Standard errors obtained by bootstrap methods are in the second row of each results.

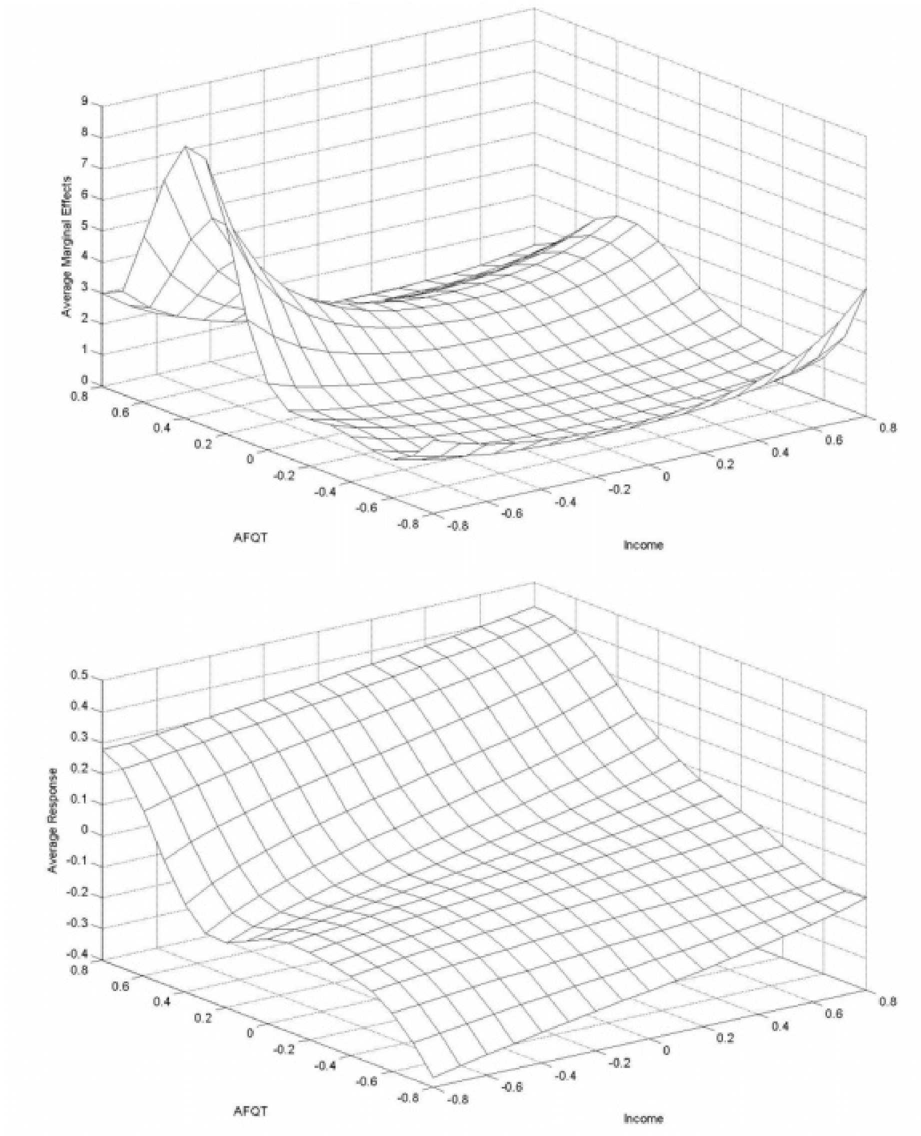
Table 3.3: Impact of family income on children's reading achievement

w/x		-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
	F	3.196	2.570	2.243	2.108	2.128	2.313	2.727	3.543	5.228
-0.8		0.767	0.721	0.688	0.657	0.633	0.618	0.619	0.648	0.774
	L	0.217	0.228	0.237	0.243	0.247	0.250	0.249	0.241	0.209
		0.063	0.063	0.061	0.067	0.074	0.085	0.097	0.123	0.159
	F	2.874	2.258	1.936	1.792	1.783	1.907	2.206	2.792	3.959
-0.6		0.762	0.762	0.757	0.753	0.750	0.746	0.741	0.739	0.752
	L	0.214	0.229	0.242	0.253	0.264	0.275	0.287	0.301	0.320
		0.057	0.059	0.062	0.064	0.069	0.075	0.087	0.100	0.124
	F	2.188	1.728	1.483	1.369	1.353	1.431	1.629	2.014	2.753
-0.4		0.765	0.771	0.771	0.770	0.770	0.769	0.768	0.766	0.769
	L	0.206	0.223	0.237	0.251	0.263	0.275	0.287	0.300	0.316
		0.059	0.055	0.061	0.063	0.064	0.066	0.080	0.091	0.098
	F	1.400	1.132	0.984	0.913	0.903	0.952	1.074	1.308	1.745
-0.2		0.797	0.803	0.805	0.805	0.805	0.804	0.803	0.803	0.803
	L	0.195	0.213	0.229	0.243	0.256	0.269	0.281	0.292	0.305
		0.057	0.056	0.058	0.061	0.066	0.069	0.074	0.083	0.103
	F	0.881	0.711	0.620	0.579	0.576	0.612	0.695	0.849	1.131
0		0.834	0.843	0.844	0.843	0.842	0.841	0.841	0.840	0.840
	L	0.181	0.201	0.218	0.233	0.247	0.259	0.271	0.282	0.292
		0.058	0.059	0.060	0.056	0.063	0.072	0.072	0.078	0.093
	F	0.907	0.675	0.563	0.514	0.506	0.535	0.606	0.740	0.981
0.2		0.872	0.882	0.884	0.884	0.883	0.882	0.880	0.879	0.879
	L	0.163	0.185	0.204	0.220	0.235	0.248	0.259	0.270	0.279
		0.059	0.054	0.058	0.063	0.066	0.066	0.075	0.075	0.088
	F	1.675	1.108	0.848	0.723	0.673	0.679	0.737	0.865	1.101
0.4		0.917	0.929	0.931	0.930	0.929	0.927	0.926	0.924	0.923
	L	0.141	0.165	0.186	0.204	0.220	0.234	0.246	0.256	0.264
		0.070	0.062	0.062	0.065	0.070	0.069	0.079	0.079	0.088
	F	3.143	1.830	1.273	1.004	0.876	0.833	0.858	0.957	1.160
0.6		0.973	0.987	0.989	0.988	0.986	0.985	0.983	0.981	0.979
	L	0.110	0.139	0.163	0.184	0.202	0.217	0.230	0.241	0.249
		0.074	0.072	0.066	0.067	0.069	0.070	0.079	0.079	0.084
	F	4.718	2.378	1.499	1.098	0.901	0.812	0.798	0.853	0.994
0.8		1.046	1.063	1.065	1.064	1.063	1.061	1.059	1.057	1.055
	L	0.055	0.096	0.129	0.156	0.178	0.197	0.212	0.223	0.232
		0.086	0.083	0.075	0.079	0.080	0.078	0.082	0.088	0.090
	N	1274								

Notes: F and L refer to our Fourier estimator and local linear estimator, respectively. Standard errors obtained by bootstrap methods are in the second row of each results.

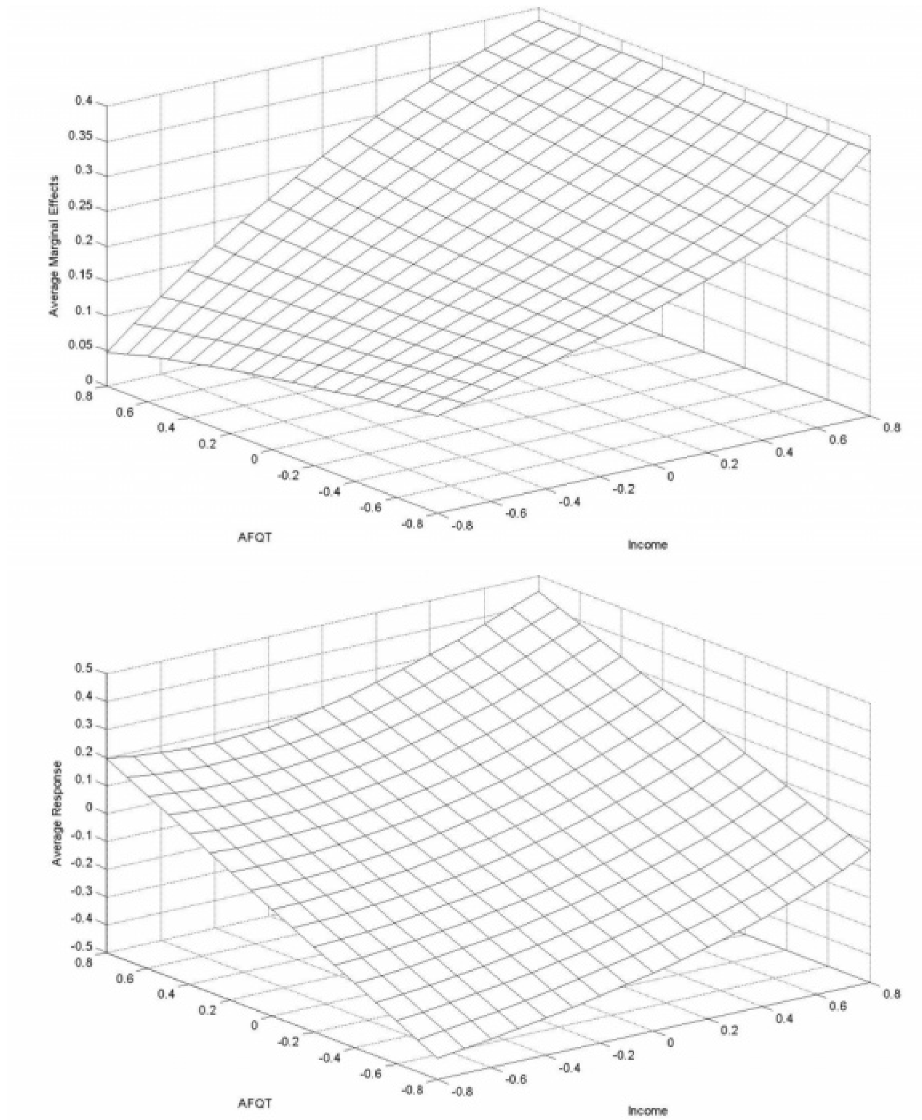
Table 3.4: Data descriptives

	Mean	Std.
Budget shares:		
Food-in	0.1776	0.0950
Food-out	0.0829	0.0591
Alcohol	0.0712	0.0791
Fuel	0.0612	0.0385
Travel	0.1488	0.0985
Fares	0.0216	0.0499
Leisure goods	0.1357	0.1456
Expenditure and income:		
log nondurable expenditure	5.3744	0.4864
log gross earnings	5.7712	0.5389
Sample size	628	



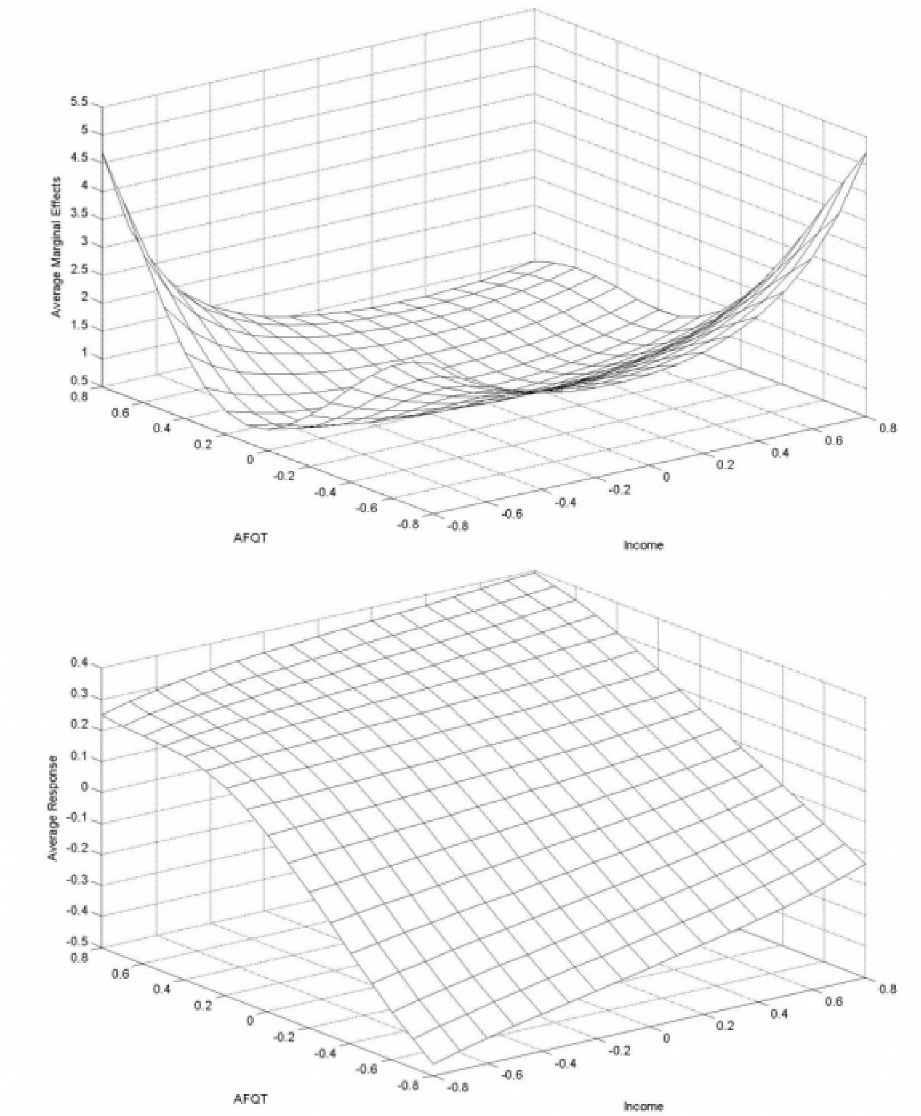
Notes: Our estimator is used for covariate-conditioned average marginal effect (top) and average counterfactual response (bottom). Error-laden measurement of family income is family income in 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 3.2: Impact of family income on children's math scores (Fourier)



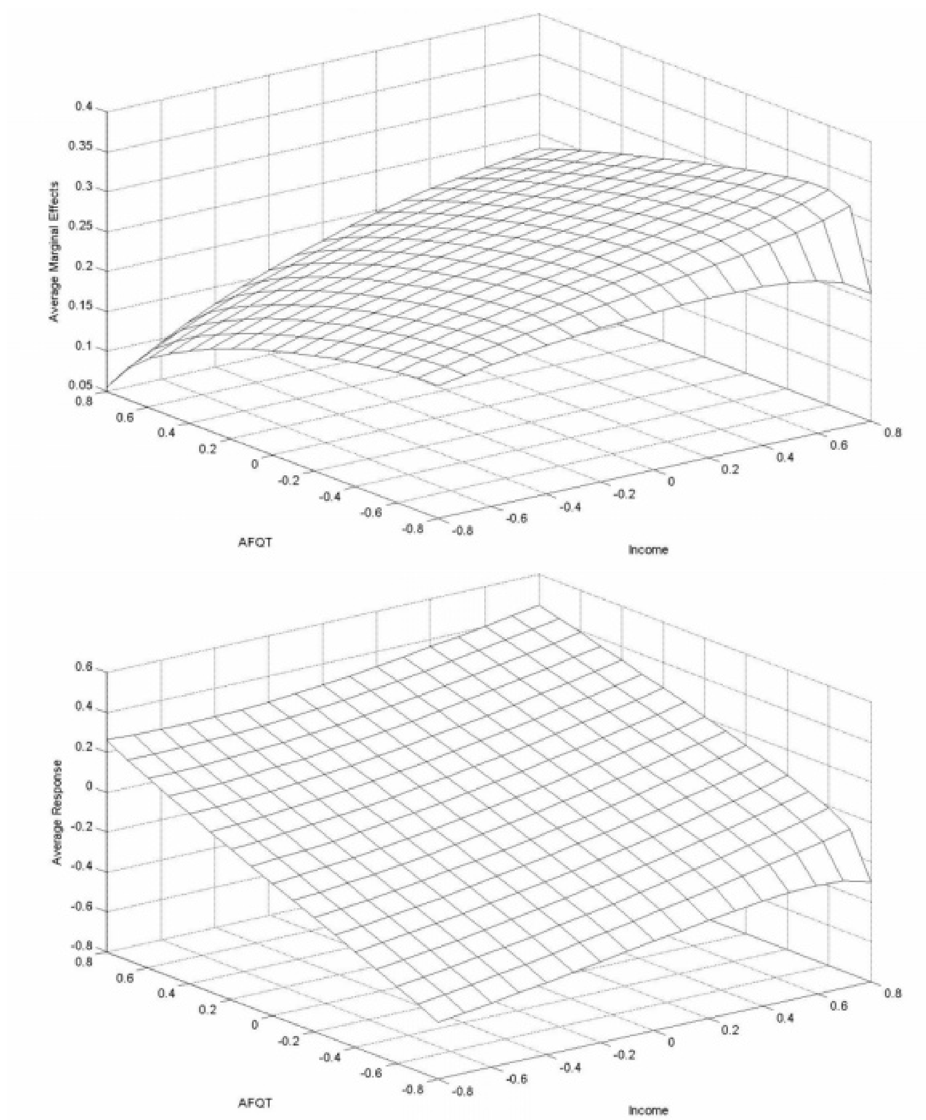
Notes: Local linear estimator is used for covariate-conditioned average marginal effect (top) and average counterfactual response (bottom). Error-laden measurement of family income is family income in 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 3.3: Impact of family income on children's math scores (Local linear)



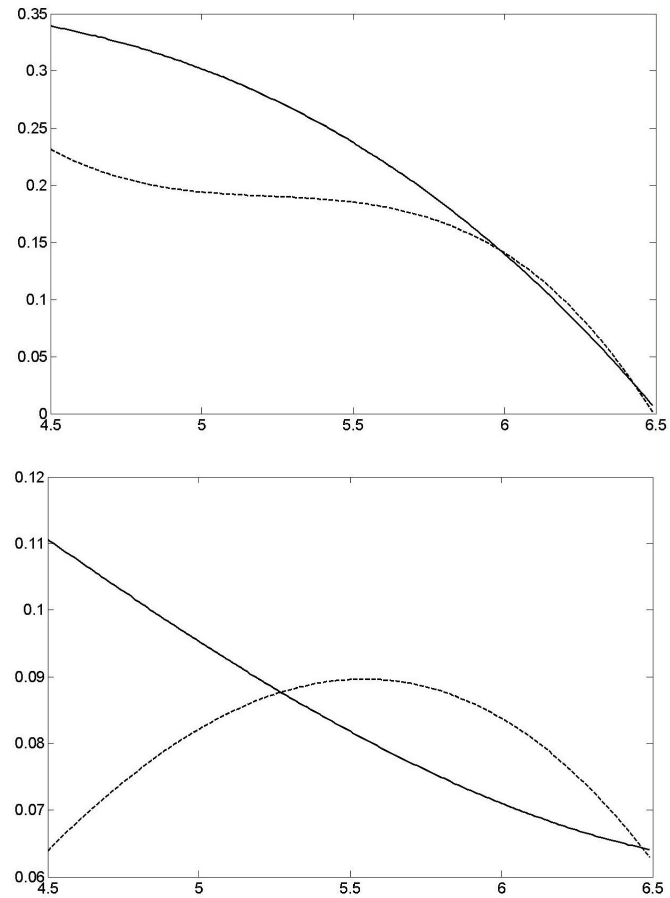
Notes: Our estimator is used for covariate-conditioned average marginal effect (top) and average counterfactual response (bottom). Error-laden measurement of current family income is family income in 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 3.4: Impact of family income on children's reading scores (Fourier)



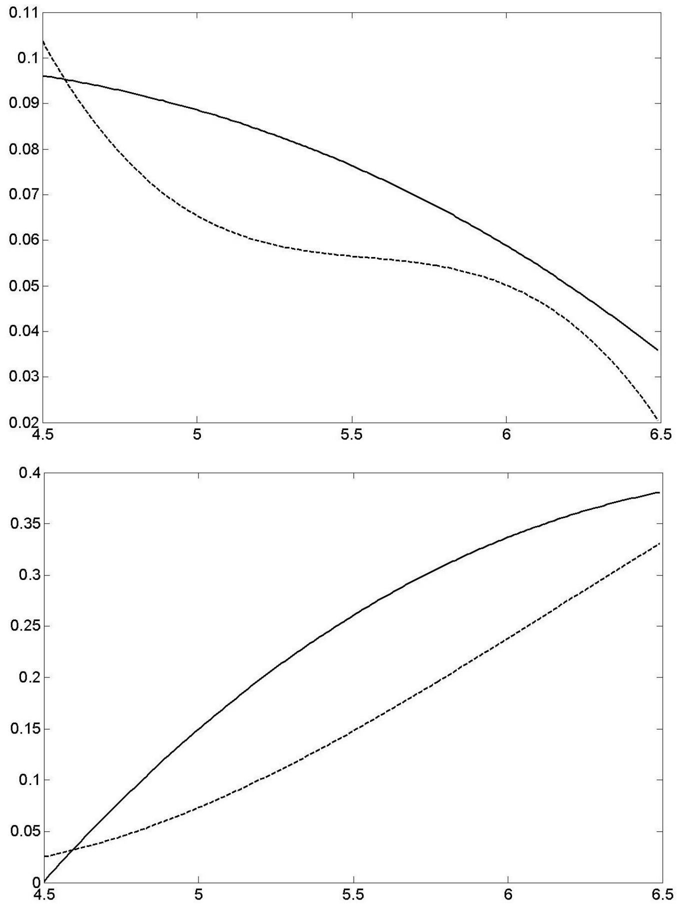
Notes: Local linear estimator is used for covariate-conditioned average marginal effect (top) and average counterfactual response (bottom). Error-laden measurement of current family income is family income in 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 3.5: Impact of family income on children's reading scores (Local linear)



Notes: Top figure is food-in and bottom figure is food-out. Our SML-SMD is the solid curve and SMD is dashed curve.

Figure 3.6: Engel curves for food-in and food-out



Notes: Top figure is fuel and bottom figure is leisure. Our SML-SMD is the solid curve and SMD is dashed curve.

Figure 3.7: Engel curves for fuel and leisure

Bibliography

Aasness, Jørgen and Biørn, Erik and Skjerpen, Terje (1993): "Engel Functions, Panel Data, and Latent Variables," *Econometrica*, 61, 6, 1395–1422.

Ai, Chunrong and Chen, Xiaohong (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 6, 1795–1843.

Altonji, Joseph G. and Matzkin, Rosa L. (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 4, 1053–1102.

Amemiya, Yasuo (1985): "Instrumental Variable Estimator for the Nonlinear Error-in-variables Model," *Journal of Econometrics*, 28, 273–289.

Aughinbaugh, Alison and Gittleman, Maury (2003): "Does Money Matter? A Comparison of the Effect of Income on Child Development in the United States and Great Britain," *Journal of Human Resources*, 38, 2, 416–440.

Banks, James, Blundell, Richard and Lewbel, Arthur (1997): "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79, 527–539.

David M. Blau (1999): "The effect of Income on Child Development," *Review of Economics and Statistics*, 81, 2, 261–276.

Blundell, Richard and Browning, Martin and Crawford, Ian (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71, 1, 205–240.

Blundell, Richard and Chen, Xiaohong and Kristensen, Dennis (2007): "Semi-nonparametric IV Estimation of Shape-invariant Engel Curves," *Econometrica*, 75, 6, 1613–1669.

Blundell, Richard and Duncan, Alan and Pendakur, Krishna (1998): "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13, 435–461.

Blundell, Richard W. and Powell, James L. (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655–679.

Blundell, Richard W. and Powell, James L. (2007): "Censored Quantile Regression with Endogenous Regressors," *Journal of Econometrics*, 141, 65–83.

Carroll, Raymond J. and Ruppert, David and Stefanski, Leonard A. and Crainiceanu, Ciprian M. (2006): "Measurement Error in Nonlinear Models," Chapman and Hall/CRC.

Chalak, Karim and White, Halbert (2007a): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," UCSD, Department of Economics Discussion Paper.

Chalak, Karim and White, Halbert (2007b): "Independence and Conditional Independence in Causal Systems," UCSD, Department of Economics Discussion Paper.

Chamberlain, Gary (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

Chamberlain, Gary (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 3, 567–596.

Chen, Xiaohong and Pouzo, Demian (2008a): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," Yale University, Cowles Foundation Discussion Paper.

Chen, Xiaohong and Pouzo, Demian (2008b): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," Yale University, Cowles Foundation Discussion Paper.

Chernozhukov, Victor and Imbens, Guido W. and Newey, Whitney K. (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

Chernozhukov, Victor and Hansen, Christian (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 1, 245–261.

Chesher, Andrew (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 5, 1405–1441.

Chesher, Andrew (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 5, 1525–1550.

Darolles, Serge and Florens, Jean-Pierre and Renault, Eric (2003): "Nonparametric Instrumental Regression," University of Toulouse, Working paper.

Das, M. (2005): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335–361.

Dawid, A.P. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B*, 41, 1–31.

Duncan, Greg J. and Yeung, W. Jean and Brooks-Gunn, Jeanne and Smith, Judith R. (1998): "How Much Does Childhood Poverty Affect the Life Chances of Children?," *American Sociological Review*, 63, 406–423.

Dunford, Nelson and Schwartz, Jacob T. (1971): *Linear Operators*. New York: Wiley Series.

Engel, Ernst (1895): "Die Lebenskosten Belgischer Arbeiter-Familien Früher and Jetzt," *International Statistical Institute Bulletin*, 9, 1–74.

Fan, Jianqing (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of Statistics*, 19, 3, 1257–1272.

Fan, Jianqing and Truong, Young K. (1993): "Nonparametric Regression with Errors in Variables," *Annals of Statistics*, 21, 4, 1900–1925.

Florens, Jean-Pierre and Heckman, James and Meghir, Costas and Vytlacil, Edward (2008): "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 5, 1191–1206.

Gonçalves, Sílvia and White, Halbert (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Jouranl of the American Statistical Association*, 100, 471, 970–979.

Grenander, U. (1981): *Abstract Inference*. New York: Wiley Series.

Hahn, Jinyong and Ridder, Geert (2007): "Conditional Moment Restrictions and Triangular Simultaneous Equations," University of Southern California, IEPR Working paper.

Hall, Peter and Horowitz, Joel L. (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 6, 2904–2929.

Hansen, Lars P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 4, 1029–1054.

Hasegawa, Hikaru and Kozumi, Hideo (2001): "Bayesian Analysis on Engel Curves Estimation with Measurement Errors and an Instrumental Variable," *Journal of Business and Economic Statistics*, 19, 3, 292–298.

Hausman, Jerry A. and Newey, Whitney K. and Ichimura, Hidehiko and Powell, James L. (1991): "Identification and Estimation of Polynomial Errors-in-variables Models ," *Journal of Econometrics*, 65, 205–233.

Hausman, Jerry A. and Newey, Whitney K. and Powell, James L. (1995): "Nonlinear Errors In Variables Estimation of Some Engel Curves ," *Journal of Econometrics*, 50, 273–295.

Hoderlein, Stefan (2007): "How Many Consumers are Rational?," Mannheim University, Department of Economics Working Paper.

Hoderlein, Stefan and Mammen, Enno (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 5, 1513–1518.

Horowitz, Joel L. (2006): "Testing a Parametric Model Against a Nonparametric Alternative with Identificaiton Through Instrumental Variables," *Econometrica*, 74, 2, 521–538.

Hsiao, Cheng (1989): "Consistent Estimation for Some Nonlinear Errors-in-variables Models," *Journal of Econometrics*, 41, 159–185.

Hu, Yingyao and Schennach, Susanne M. (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 1, 195–216.

Ichimura, Hidehiko (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71–120.

Imbens, Guido W. (2006): "Nonadditive Models with Endogenous Regressors," Harvard University, Department of Economics Working Paper.

Imbens, Guido W. and Newey, Whitney K. (2006): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Harvard University, Department of Economics Working Paper.

Leser, C. E. V. (1963): "Forms of Engel Functions," *Econometrica*, 31, 4, 694–703.

LEWBEL, ARTHUR (1996): "Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side," *Review of Economics and Statistics*, 78, 4, 718–725.

LI, TONG (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-varialbes Models," *Journal of Econometrics*, 110, 1–26.

LI, TONG AND VUONG, QUANG (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

LIVIATAN, NISSAN (1963): "Errors in Variables and Engel Curve Analysis," *Econometrica*, 29, 3, 336–362.

LYSSIOTOU, PANAYIOTA AND PASHARDES, PANOS AND STENGOS, THANASIS (1999): "Testing the Rank of Engel Curves with Endogenous Expenditure," *Economics Letters*, 64, 61–65.

MATZKIN, ROSA L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 5, 1339–1375.

NEWEY, WHITNEY K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 4, 809–837.

NEWEY, WHITNEY K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics*, Vol. 11, ed. by G.S. Maddala, C.R. Rao, and H.D. Vinod. Amsterdam: North-Holland, Chapter 16.

NEWEY, WHITNEY K. (2001): "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, 83, 4, 616–627.

NEWEY, WHITNEY K. AND POWELL, JAMES L. (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 5, 1565–1578.

NEWEY, WHITNEY K. AND POWELL, JAMES L. AND VELLA, FRANCIS (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 3, 565–603.

PAGAN, ADRIAN AND ULLAH, AMAN (1999): *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

PENDAKUR, KRISHNA (1998): "Semiparametric Estimates and Tests of Base-independent Equivalence Scales," *Journal of Econometrics*, 88, 1–40.

POLITIS, DIMITRIS N. AND ROMANO, JOSEPH P. (1999): "Multivariate Density Estimation with General Flat-Top Kernels of Infinite Order," *Journal of Multivariate Analysis*, 68, 1–25.

Powell, James L. and Stock, James H. and Stoker, Thomas M. (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 6, 1403–1430.

Rao, Prakasa (1992): *Identifiability in Stochastic Models*. Academic Press.

Robinson, Peter M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 4, 931–954.

Schennach, Susanne M. (2004a): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

Schennach, Susanne M. (2004b): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 1, 33–75.

Schennach, Susanne M. and White, Halbert and Chalak, Karim (2007): "Estimating Average Marginal Effects in Nonseparable Structural Systems," UCSD, Department of Economics Discussion Paper.

Severini, Thomas A. and Tripathi, Gautam (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory*, 22, 258–278.

Strotz, Robert H. and Wold, H. O. A. (1960): "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," *Econometrica*, 28, 2, 417–427.

Taylor, Beck A. and Dearing, Eric and McCartney, Kathleen (2004): "Incomes and Outcomes in Early Childhood," *Journal of Human Resources*, 39, 4, 980–1007.

Wahba, Grace (1992): *Spline Models for Observational Data*. SIAM.

White, Halbert and Chalak, Karim (2006): "A Unified Framework for Defining and Identifying Causal Effects," UCSD, Department of Economics Discussion Paper.

White, Halbert and Chalak, Karim (2008): "Identifying Structural Effects in Nonseparable Systems Using Covariates," UCSD, Department of Economics Discussion Paper.

Working, Holbrook (1943): "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 221, 43–56.