Identification and Estimation of Nonseparable Models with Measurement Errors

(Job Market Paper)

SUYONG SONG* SUSANNE M. SCHENNACH† HALBERT WHITE‡

October 20, 2009

Abstract

We study 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. Specifically, we show that two error-laden measurements of the unobserved true causes are sufficient to identify objects of interest and to deliver consistent estimators. 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 instruments, and root-n consistency and asymptotic normality for estimators of their weighted averages over instruments and regressors. We investigate their finite sample behaviors through Monte Carlo simulations and apply our new methods to study the impact of family income on child achievement. Our findings suggest that these effects are considerably larger than previously recognized.

JEL Classification: C13, C14, C31.

Keywords: causal effects; child development; endogeneity; measurement error; nonparametric estimation; nonseparable structural equation.

*Department of Economics, University of California, San Diego. Email: s7song@ucsd.edu
†Department of Economics, University of Chicago. Email: smschem@uchicago.edu
‡Department of Economics, University of California, San Diego. Email: hwhite@ucsd.edu
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.


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 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 3, and asymptotic properties of the estimator are analyzed in Section 4. The practical usefulness of the proposed estimator is illustrated by Monte Carlo experiments in Section 5. We apply our new methods to study the impact of family income on child achievement in Section 6. Section 7 concludes. All technical proofs are included in the Mathematical Appendix.

2 Data Generation and Identification

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 $U_y \equiv v_y(U)$ 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 $W$.

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).
Figure 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 $U_1 \equiv v_1(U)$ and $U_2 \equiv v_2(U)$ for measurable functions $v_1$ and $v_2$.

Figure 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.

### 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 $W$,

$$
\mu(X, W) \equiv E(Y \mid X, W).
$$

(1)

The function $\mu$ 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 $Y$. 

4
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 \( \mu \) acquires causal meaning from \( \rho \). We call this a “structural identification” result because it identifies an aspect of the causal structure, \( \rho \), with \( \mu \), a standard stochastic object. When \( \mu(x, w) \) is differentiable, let \( \beta(x, w) \equiv D_x \mu(x, w) \). With \( \mu \) 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).
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

$$\beta_m^*(x) \equiv \int \beta^*(x \mid w)m(w)dw,$$

(2)

$$\beta_{mf}^*(x) \equiv \int \beta^*(x \mid w)m(w)f_W(w)dw,$$

(3)

$$\beta_{mf|X}^*(x) \equiv \int \beta^*(x \mid w)m(w)f_{W|X}(w \mid x)dw,$$

(4)

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

(5)

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

(6)

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

(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, $\beta_{mf}^*(x)$ 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 $m(w) = 1$, $\beta_{mf|X}^*(x)$ 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 $\tilde{m}(x, w) = m(w)$, $\beta_{\tilde{mf}|X}^*$ corresponds to the weighted average of the local average response (Altonji and Matzkin, 2005).

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

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 $D_x^\lambda \equiv (\partial^\lambda / \partial x^\lambda)$ denotes the derivative operator of degree $\lambda$, $V$ 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

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

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 $\text{supp}(\cdot)$. By convention, we take the value of any referenced function to be zero except when the indicated random variable lies in $\text{supp}(\cdot)$. 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**
1. \(E[U_1 \mid X, U_2] = 0\);
2. \(U_2 \perp (X, W)\);
3. \(E[Y \mid X, U_2, W] = E[Y \mid X, W]\).

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

**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 $g_{Y,0}(x, w)$ 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 \(N \equiv \{0, 1, \ldots\}\) and \(\bar{N} \equiv N \cup \{\infty\}\).

**Assumption 3.5** For \(V = 1, Y, g_{V,0}(\cdot, w)\) is continuously differentiable of order \(\Lambda \in \bar{N}\) 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 distri-
bution of observable variables \(V, X_1,\) and \(X_2,\) we say it is “stochastically identified.” This is shown in the
next lemma.\(^1\)

**Lemma 3.1** Suppose Assumptions 2.1(i), 2.3, and 3.1 - 3.5 hold. Then for \(V = 1, Y\) and for each
\(\lambda \in \{0, \ldots, \Lambda\}\) 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 \(\zeta,\)

\[
\phi_V(\zeta, w) \equiv E[V e^{i\zeta X} \mid W = w] = \frac{E[V e^{i\zeta X_2} \mid W = w]}{E[e^{i\zeta X_2}]} \exp\left(\int_0^\zeta iE[X_1 e^{i\xi X_2}] d\xi\right).
\]

Note that \(g_{V,\lambda}\) is a empirically accessible when it involves only observable variables. Thus, knowledge
of \(E[V e^{i\zeta X_2} \mid W = w], E[e^{i\zeta X_2}],\) and \(E[X_1 e^{i\xi X_2}]\) is sufficient to obtain stochastic identification of \(g_{V,\lambda}\).

### 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) \(\kappa\) 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 \(|\xi| < \bar{\xi}\).

\(^1\)Derivation of a part of the expression for \(\phi_V\) 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 above assumption is similar to that used in SWC. The fact that the kernel is continuously differentiable to any order is ensured by the requirement of Assumption 3.6(i) that the Fourier transform of the kernel is compactly supported. The assumption of compact support of $\kappa$ 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 $\kappa$ from the original Fourier transform $\tilde{\kappa}$ of $\tilde{k}$ put

$$\kappa(\xi) = \mathcal{W}(\xi)\tilde{\kappa}(\xi),$$

$$\mathcal{W}(\xi) = \begin{cases} 
1 & \text{if } |\xi| \leq \bar{\xi} \\
(1 + \exp((1 - \bar{\xi})((1 - |\xi|)-1 - (|\xi| - \bar{\xi})-1)))^{-1} & \text{if } \bar{\xi} < |\xi| \leq 1 \\
0 & \text{if } 1 < |\xi| 
\end{cases}$$

(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 $\lambda \in \{0, ..., \Lambda\}$, $(x, w) \in \text{supp}(X, W)$, and $h > 0$, let

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

Then

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

We let $h$ denote the kernel bandwidth or smoothing parameter. Because $\lim_{h \to 0} g_{V,\lambda}(x, w, h) = g_{V,\lambda}(x, w)$ 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 The estimator for $g_{V,\lambda}(x, w)$ is defined as

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

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

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

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

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

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, $\hat{\phi}_V(\zeta, w)$. $\hat{E}[Ve^{i\zeta X_2} \mid W = w]$ is a kernel estimator of $E[Ve^{i\zeta X_2} \mid W = w]$.

4 Asymptotics

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 $h > 0$,

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

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

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

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

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

where $\tilde{g}_{V,\lambda}(x, w, h)$ is the linearization of $\hat{g}_{V,\lambda}(x, w, h)$ in terms of $\hat{E}[e^{i\xi X_2}] - E[e^{i\xi X_2}]$,

$$(\hat{E}[X_1e^{i\xi X_2}] - E[X_1e^{i\xi X_2}]), (\hat{E}[Ve^{i\xi X_2k_h(W - w)}] - E[Ve^{i\xi X_2k_h(W - w)})], and (\hat{E}[k_h(W - w)] - E[k_h(W - w)]),$$

where

$$\ell_{V,\lambda}(x, w; v, x_1, x_2, \tilde{w}) \equiv \int \Psi_{V,\lambda,1}(\zeta, x, w, h) \left( e^{i\xi x_2} - E \left[ e^{i\xi X_2} \right] \right) d\zeta$$

$$+ \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h) \left( x_1 e^{i\xi x_2} - E \left[ X_1 e^{i\xi X_2} \right] \right) d\zeta$$

$$+ \int \Psi_{V,\lambda,X_2}(\zeta, x, w, h) \left( ve^{i\xi x_2}k_h(\tilde{w} - w) - E \left[ Ve^{i\xi X_2k_h(W - w)} \right] \right) d\zeta$$

$$+ \int \Psi_{V,\lambda,fW}(\zeta, x, w, h) \left( k_h(\tilde{w} - w) - E \left[ k_h(W - w) \right] \right) d\zeta,$$

and where, letting $\theta_A(\zeta) \equiv E \left[ Ae^{i\xi X_2} \right]$ for $A = 1, X_1$ and $\chi_V(\zeta, w) \equiv \int \int ve^{i\xi x_2} f_{V,X_2,W}(v, x_2, w) dv dx_2$, we define
where for a given function $\xi \to f(\xi)$, we write \( \int_{-\infty}^{\infty} f(\xi) d\xi \equiv \lim_{c \to +\infty} \int_{-c}^{c} f(\xi) d\xi \); 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 \( \hat{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h) \) is equivalent to that of \( L_{V,\lambda}(x, w, h) \), as long as \( B_{V,\lambda}(x, w, h) \) 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, \( |\xi|^\gamma, \gamma < 0 \), as the frequency, \( |\xi| \), goes to zero (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 \( \exp(\alpha|\xi|^{\beta}), \alpha < 0, \beta > 0 \) as the frequency goes to zero (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 \( (1 + |\xi|)^\gamma \exp(\alpha|\xi|^{\beta}) \).

Assumption 4.1 Let \( \phi_1(\xi) \equiv E[e^{i\xi X}] \).
\[(i) \text{ There exist constants } C_1 > 0 \text{ and } \gamma_1 \geq 0 \text{ such that} \]
\[
|D_\zeta \ln \phi_1(\zeta)| = \left| \frac{D_\zeta \phi_1(\zeta)}{\phi_1(\zeta)} \right| \leq C_1 (1 + |\zeta|)^{\gamma_1};
\]

\[(ii) \text{ There exist constants } C_\phi > 0, \alpha_0 \leq 0, \beta_0 \geq 0, \text{ and } \gamma_\phi \in \mathbb{R} \text{ such that } \beta_0 \gamma_\phi \geq 0 \text{ and for } V = 1, Y
\]
\[
\sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \leq C_\phi (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_0|\zeta|^{\beta_0}),
\]
and if \( \alpha_0 = 0, \text{ then } \gamma_\phi < -\lambda - 1 \text{ for given } \lambda \in \{0, \ldots, \Lambda\};
\]

\[(iii) \text{ There exist constants } C_\theta > 0, \alpha_0 \leq 0, \beta_0 \geq 0, \text{ and } \gamma_\theta \in \mathbb{R} \text{ such that } \beta_0 \gamma_\theta \geq 0 \text{ and for } V = 1, Y
\]
\[
\min\left\{ \inf_{w \in \text{supp}(W)} |\chi_V(\zeta, w)|, |\theta_1(\zeta)| \right\} \geq C_\theta (1 + |\zeta|)^{\gamma_\theta} \exp(\alpha_0|\zeta|^{\beta_0}).
\]

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 \( \zeta \) for large \( \zeta \), 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 \( \bar{g}_{V:1}(x, w, h) \). 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 \( \beta_0 \geq \beta_\phi \) because

\[
C_\phi (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_0|\zeta|^{\beta_0}) \geq \sup_{w \in \text{supp}(W)} |\phi_1(\zeta, w)| = \sup_{w \in \text{supp}(W)} |E[e^{i\zeta X} \mid W = w]|
\]
\[
\geq \left| \int E[e^{i\zeta X} \mid W = w] f_W(w) dw \right| = |E[e^{i\zeta X}]| \geq |E[e^{i\zeta X}]| |E[e^{i\zeta U_2}]| \geq |E[e^{i\zeta X_2}]|
\]
\[
= |\theta_1(\zeta)| \geq C_\theta (1 + |\zeta|)^{\gamma_\theta} \exp(\alpha_0|\zeta|^{\beta_0}).
\]

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 \( \lambda \in \{0, \ldots, \Lambda\} \) and \( h > 0, \)

\[
\sup_{(x, w) \in \text{supp}(X,W)} |B_{V,\lambda}(x, w, h)| = O \left( (h^{-1})^{\gamma_{\lambda,B}} \exp \left( \alpha_B (h^{-1})^{\beta_B} \right) \right),
\]
where \( \alpha_B \equiv \alpha \beta_B \), \( \beta_B \equiv \beta \), 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 \( B_{V, \lambda}(x, w, h) \) only involves the kernel and error-free variables.\(^2\)

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 \( \beta_0 \neq 0 \) in Assumption 4.1(iii), then \( h_n^{-1} = O \left( (\ln n)^{1/\beta_0 - \eta} \right) \) for some \( \eta > 0 \); otherwise, for each \( \lambda \in \{0, \ldots, \Lambda\} \), \( h_n^{-1} = O \left( n^{-\eta}(3/2)/(\gamma_0 + \lambda + \gamma_1 - \gamma_0 + 3) \right) \) 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 in quantities in the denominator \( \chi_V(\zeta, w) \) and \( \theta_1 \) 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)).

---

\(^2\)When \( X \) is perfectly observed, one can propose an estimator of \( g_{V, \lambda} \) using a similar Fourier transform as

\[
\hat{g}_{V, \lambda}(x, w, h_n) \equiv \frac{1}{2\pi} \int (-i\zeta)^\lambda \chi_V(\zeta, w) \exp(-i\zeta x) d\zeta,
\]

for \( h_n \to 0 \) as \( n \to \infty \), where

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

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.
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 \( \lambda \in \{0, ..., \Lambda\} \), \((x, w) \in \text{supp}(X, W)\), and \( h > 0 \), \( E[L_{V,\lambda}(x, w, h)] = 0 \), and if Assumption 4.2 also holds, then

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

where

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

is finite. Further, if Assumption 4.1 holds, then

\[
\sqrt{n} \sup_{(x,w) \in \text{supp}(X,W)} \Omega_{V,\lambda}(x, w, h) = O \left( (h^{-1})^{\gamma_{\lambda,L}} \exp \left( \alpha_L \left( h^{-1}\right)^{\beta_L} \right) \right),
\]

with \( \alpha_L \equiv \alpha_\phi 1_{\beta_\phi = \beta_\theta} \) \( - \alpha_\theta \), \( \beta_L \equiv \beta_\theta \), and \( \gamma_{\lambda,L} \equiv 2 + \gamma_\phi - \gamma_\theta + \gamma_1 + \lambda \). We also have

\[
\sup_{(x,w) \in \text{supp}(X,W)} |L_{V,\lambda}(x, w, h)| = O_P \left( n^{-1/2} (h^{-1})^{\gamma_{\lambda,L}} \exp \left( \alpha_L \left( h^{-1}\right)^{\beta_L} \right) \right);
\]

(ii) If Assumptions 4.3 and 4.4 also hold, and if for \( V = 1, Y \) and for each \( \lambda \in \{0, ..., \Lambda\} \), \((x, w) \in \text{supp}(X, W)\), \( \Omega_{V,\lambda}(x, w, h_n) > 0 \) 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 \( \gamma_\phi, \alpha_\theta, \beta_\theta \)) 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 \( \gamma_1 \)), 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.3 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 assump-

---

3With perfectly observed \( X \), the order of the variance term of the estimator in footnote 2 can be derived as
\( n^{-1/2} (h^{-1})^{2 + \gamma_\phi + \lambda} \exp \left( \alpha_\phi \left( h^{-1}\right)^{\beta_\phi} \right) \). Thus if \( \beta_\phi > 0 \), \( \beta_L \equiv \beta_\theta \geq \beta_\phi \) by construction, and if \( \beta_L \equiv \beta_\phi = \beta_\theta = 0 \), \( \gamma_{\lambda,L} \equiv 2 + \gamma_\phi + \lambda - \gamma_\theta + \gamma_1 \geq 2 + \gamma_\phi + \lambda \) by the fact that \( \gamma_1 - \gamma_\theta > 0 \). Then the order of the variance term in Theorem 4.3 is greater than that of the kernel estimator with perfectly observed variables.
tion 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** \(E[|X_2|] < \infty, E[|X_1X_2|] < \infty, \text{ and } E[|YX_2|] < \infty\).

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** \[
\sup_{w \in \text{supp}(W)} |\hat{f}_W(w) - f_W(w)| = O_p \left( \frac{\sqrt{\ln n}}{nh} + h^2 \right).
\]

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 \(\beta_0 \neq 0\) in Assumption 4.1, then \(h_n^{-1} = O \left((\ln n)^{1/\beta_0 - \eta}\right)\) for some \(\eta > 0\); otherwise \(h_n^{-1} = O \left(n^{-\eta}n^{1/(2\gamma_1-2\gamma_\theta+6)}\right)\) 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 \(\lambda \in \{0, \ldots, \Lambda\}\), and some \(\epsilon > 0\),

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{V,\lambda}(x, w, h_n)| = O_p \left( n^{-1/2+\epsilon}(1 + h_n^{-1})^{3+\gamma_1-\gamma_\theta} \exp \left(-\alpha_\theta(h_n^{-1})^{\beta_\theta}\right) \right) \\
\times O_p \left( n^{-1/2}(h_n^{-1})^{\gamma_L} \exp \left(\alpha_L(h_n^{-1})^{\beta_L}\right) \right);
\]

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

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{V,\lambda}(x, w, h_n)| = o_p \left( n^{-1/2}(h_n^{-1})^{\gamma_L} \exp \left(\alpha_L(h_n^{-1})^{\beta_L}\right) \right).
\]
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 \( \lambda \in \{0, ..., \Lambda\} \),

\[
\sup_{(x,w) \in \text{supp}(X,W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w, 0)| = O \left( \left( \frac{1}{n} \right)^{\gamma_{\lambda,B}} \exp \left( \alpha_B \left( \frac{1}{n} \right)^{\beta_B} \right) \right) + O_p \left( n^{-1/2} \left( \frac{1}{n} \right)^{\gamma_{\lambda,L}} \exp \left( \alpha_L \left( \frac{1}{n} \right)^{\beta_L} \right) \right).
\]

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 \( (x, w) \in \text{supp}(X,W) \) we have: (i) \( \Omega_{V,\lambda}(x, w, h_n) > 0 \) for all \( n \) sufficiently large; (ii) \( n^{1/2} (\Omega_{V,\lambda}(x, w, h_n))^{-1/2} |B_{V,\lambda}(x, w, h_n)| \to 0 \); 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 \( B_{V,\lambda}(x, w, h_n) \) 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(ii).

**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 \( (x, w) \in \text{supp}(X,W) \), we have

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

### 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 we 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)) -
The first of the following theorems is relevant to estimating $\beta_m(x)$, $\beta_{mW}(x)$, and $\beta_{mfW|x}(x)$. 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 $\beta_\tilde{m}$, $\beta_\tilde{mW,X}$, and $\beta_\tilde{mfW|x}$ 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 $x \in \text{supp}(X)$, $\sup_{x \in \text{supp}(X)} \int |s(x, w)|dw < \infty$. Then for $V = 1, Y$, there exist constants $C_{\phi_s} > 0$, $\alpha_{\phi_s} \leq 0$, $\beta_{\phi_s} \geq 0$, 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

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

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

$$
\sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \geq \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) \left( \sup_{x \in \text{supp}(X)} \left| \int s(x, w) dw \right| \right).
$$

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 $\sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)|$ and on $\sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) dw \right|$ 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 \( \lambda \in \{0, \ldots, \Lambda\} \), we have: (i) if \( \beta_{\phi} = \beta_{\phi} > 0 \) or \( \gamma_{\phi} = \gamma_{\phi} \) for \( \beta_{\phi} = \beta_{\phi} = 0 \), \( \sup_{(x,w) \in \text{supp}(X,W)} |B_{V,\lambda}(x,w,h_n)| = o(\alpha_{1n}) \), \( \sup_{(x,w) \in \text{supp}(X,W)} |V,\lambda(x,w,h_n)| = o(\alpha_{1n}) \) where \( \alpha_{1n} \equiv (h_n^{-1})^{\gamma_{\lambda,B}} \exp\left(\alpha_B (h_n^{-1})^{\beta_B}\right) + n^{-1/2} (h_n^{-1})^{\gamma_{\lambda,L}} \exp\left(\alpha_L (h_n^{-1})^{\beta_L}\right) \) and where \( \alpha_B, \beta_B, \gamma_{\lambda,B}, \alpha_L, \beta_L, \) and \( \gamma_{\lambda,L} \) are as defined in Theorem 4.2 and 4.3.

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

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 \( \beta_m(x), \beta_{mW}(x) \), and \( \beta_{mW|X}(x) \).

Theorem 4.7  For given \( \Lambda, J \in \mathbb{N} \), let \( \lambda_1, \ldots, \lambda_J \) belong to \( \{0, \ldots, \Lambda\} \), let \( V_1, \ldots, V_J \) 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 \( \bar{g}_x \equiv (\bar{g}_{V_1,\lambda_1}(x,\cdot), \ldots, \bar{g}_{V_J,\lambda_J}(x,\cdot)) \) in an \( L_\infty \) neighborhood of the \( J \)-vector \( g_x \equiv (g_{V_1,\lambda_1}(x,\cdot), \ldots, g_{V_J,\lambda_J}(x,\cdot)) \),

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

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

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

\[
\sup_{x \in \text{supp}(X)} |b(\bar{g}_x(h_n)) - b(g_x)| = O\left( (h_n^{-1})^{\gamma_{\lambda,B}} \exp\left(\alpha_B (h_n^{-1})^{\beta_B}\right) \right)
+ O_p\left( n^{-1/2} (h_n^{-1})^{\gamma_{\lambda,L}} \exp\left(\alpha_L (h_n^{-1})^{\beta_L}\right) \right);
\]

(ii) If Assumption 4.10(ii) holds, then
\[
\sup_{x \in \text{supp}(X)} |b(\tilde{g}_x(h_n)) - b(g_x)| = O \left( (h_n^{-1})^{\gamma_L, s} \exp \left( \alpha_{B, s} (h_n^{-1})^{\beta_{B, s}} \right) \right) + O_p \left( n^{-1/2} (h_n^{-1})^{\gamma_L, s} \exp \left( \alpha_{L, s} (h_n^{-1})^{\beta_{L, s}} \right) \right). 
\]

Note that Eqn. (9) of this result is Fréchet differentiability of \( b(\tilde{g}_x) \) with respect to \( \tilde{g}_x \) in the norm \( \| \tilde{g}_{V, \lambda}(x, \cdot) \|_\infty^2 \), 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 \( \sup_{(x,w) \in \text{supp}(X,W)} |B_{V, \lambda}(x, w, h_n)| = o \left( n^{-1/2}, \sup_{(x,w) \in \text{supp}(X,W)} |L_{V, \lambda}(x, w, h_n)| = o_p \left( n^{-1/4} \right) \right), \sup_{(x,w) \in \text{supp}(X,W)} |R_{V, \lambda}(x, w, h_n)| = o_p \left( n^{-1/2} \right), \text{ and sup}_{w \in \text{supp}(W)} |\tilde{f}_W(w) - f_W(w)| = o_p \left( n^{-1/4} \right) \).

The following theorem gives a convenient asymptotic normality and \( \sqrt{n} \)- consistency result useful for analyzing \( \beta_{\hat{m}}, \beta_{\hat{m}f_{V,W}}, \text{ and } \beta_{\hat{m}f_{W|x}} \).

**Theorem 4.8** For given \( \Lambda, J \in \mathbb{N} \), 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 \( L_\infty \) 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 \( f \equiv f_W \),

\[
b(\tilde{g}, \tilde{f}) - b(g, f) = \sum_{j=1}^{J} \int \int (\tilde{g}_{V_j, \lambda_j}(x, w) - g_{V_j, \lambda_j}(x, w)) s_j(x, w) dwdx \\
+ \int \int (\tilde{f}_W(w) - f_W(w)) s_{J+1}(x, w) dwdx \\
+ \sum_{j=1}^{J} O \left( \| \tilde{g}_{V_j, \lambda_j} - g_{V_j, \lambda_j} \|_\infty^2 \right) + O \left( \| \tilde{f}_W - f_W \|_\infty^2 \right)
\]

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

\[
\Psi_{V, \lambda, s}(\xi) = \frac{1}{|\theta_1(\xi)|} \left( 1 + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|} \right) \int_{|\xi|}^{\infty} |\sigma_{V_1, s}(\xi)||\xi|^\lambda d\xi \\
+ |\xi|^\lambda \left( \frac{|\sigma_{V, X_1, s}(\xi)|}{|\theta_1(\xi)|} + |\sigma_{V, X_1, s}(\xi)| + |\sigma_{V, f_W, s}(\xi)| \right)
\]
\[
\sigma_{V,1,s}(\zeta) \equiv \int \exp(i\zeta x) \int s(x, w) \phi_V(\zeta, w) dw dx
\]

\[
\sigma_{V,\chi V,s}(\zeta) \equiv \int \exp(i\zeta x) \lim_{h \to 0} \int \frac{1}{\chi_V(\zeta, w)} s(x, w) \phi_V(\zeta, w) ve^{i\zeta x_2} k_h(\tilde{w} - w) dw dx
\]

\[
\sigma_{V,f_w,s}(\zeta) \equiv \int \exp(i\zeta x) \lim_{h \to 0} \int \frac{1}{f_w(w)} s(x, w) \phi_V(\zeta, w) k_h(\tilde{w} - w) dw dx
\]

\[
\sigma_{f_w,s} \equiv \int \lim_{h \to 0} \int s_{J+1}(x, w) k_h(\tilde{w} - w) dw dx,
\]

then, letting \( \hat{g}(h_n) \equiv (\hat{g}_{V,1,\lambda}(\cdot, h_n), \ldots, \hat{g}_{V,j,\lambda}(\cdot, h_n)) \) and \( \hat{f}(h_n) \equiv \hat{E}[k_{h_n}(\cdot)] \),

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

where

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

and where

\[
\psi_{V,\lambda}(s; v, x_1, x_2, \tilde{w}) \equiv \int \left\{ \Psi_{V,\lambda,1,s}(\zeta) \left( e^{i\zeta x_2} - E[e^{i\zeta X_2}] \right) + \Psi_{V,\lambda,X_1,s}(\zeta) \left( x_1 e^{i\zeta x_2} - E[X_1 e^{i\zeta X_2}] \right) \\
+ (Z_{V,\chi V}(s, \zeta; v, x_2, \tilde{w}) - E[Z_{V,\chi V}(s, \zeta; V, X_2, W)]) \\
+ (Z_{V,\lambda,f_w}(s, \zeta; \tilde{w}) - E[Z_{V,\lambda,f_w}(s, \zeta; W)]) \right\} d\zeta
\]

\[
\psi_f(s_{J+1}; \tilde{w}) \equiv \int \lim_{h \to 0} \int s_{J+1}(x, w) (k_h(\tilde{w} - w) - E[k_h(W - w)]) dw dx,
\]

with

\[
\Psi_{V,\lambda,1,s}(\zeta) \equiv -\frac{1}{2\pi} \frac{i\theta_{X_1}(\zeta)}{(\theta_1(\zeta))^2} \int_{-\infty}^{\infty} \left( \int \exp(-i\eta x) \int s(x, w) \phi_V(\zeta, w) dw dx \right) (-i\zeta)^\lambda \\
- \frac{1}{2\pi} \frac{(-i\zeta)^\lambda}{\theta_1(\zeta)} \left( \int \exp(-i\eta x) \int s(x, w) \phi_V(\zeta, w) dw dx \right)
\]

\[
\Psi_{V,\lambda,X_1,s}(\zeta) \equiv \frac{1}{2\pi} \frac{i}{\theta_1(\zeta)} \int_{-\infty}^{\infty} \left( \int \exp(-i\eta x) \int s(x, w) \phi_V(\zeta, w) dw dx \right) (-i\zeta)^\lambda d\eta
\]

\[
Z_{V,\chi V}(s, \zeta; v, x_2, \tilde{w}) \equiv \frac{1}{2\pi} (-i\zeta)^\lambda \int \exp(-i\eta x) \lim_{h \to 0} \int \frac{1}{\chi_V(\zeta, w)} s(x, w) \phi_V(\zeta, w) \\
\times ve^{i\zeta x_2} k_h(\tilde{w} - w) dw dx
\]

\[
Z_{V,\lambda,f_w}(s, \zeta; \tilde{w}) \equiv -\frac{1}{2\pi} (-i\zeta)^\lambda \int \exp(-i\eta x) \lim_{h \to 0} \int \frac{1}{f_w(w)} s(x, w) \phi_V(\zeta, w) k_h(\tilde{w} - w) dw dx.
\]
Moreover,
\[
n^{1/2}(b(\hat{g}(h_n), \hat{f}(h_n)) - b(g, f)) \xrightarrow{d} N(0, \Omega_b),
\]
where
\[
\Omega_b \equiv E \left[ (\psi_s(V, X_1, X_2, W))^2 \right] < \infty.
\]

4.3 Asymptotics for Covariate-Conditioned 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,
\[
\hat{\beta}(x, w, h_n) \equiv \frac{\hat{g}_{Y,1}(x, w, h_n) - \hat{g}_{Y,0}(x, w, h_n)}{\hat{g}_{1,0}(x, w, h_n) - \hat{g}_{1,0}(x, w, h_n)}
\]
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 \(\max V, Y, \max \lambda = 0, 1\)
\[
\sup_{(x, w) \in \text{supp}(X, W)} |g_{V,\lambda}(x, w)| < \infty. \text{ Further, for } \tau = \tau_n > 0, \text{ define}
\]
\[
\Gamma_\tau \equiv \{(x, w) \in \mathbb{R}^2 : f_{X|W}(x | w) \geq \tau_n\}.
\]

Then we have
\[
\sup_{(x, w) \in \Gamma_\tau} \left| \hat{\beta}(x, w, h_n) - \beta(x, w) \right| = O \left( \tau^{-3} (h_n^{-1})^{\gamma_1} \exp \left( \alpha_B (h_n^{-1})^{\beta_B} \right) \right) + O_p \left( \tau^{-3} n^{-1/2} (h_n^{-1})^{\gamma_1} \exp \left( \alpha_L (h_n^{-1})^{\beta_L} \right) \right),
\]
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 \(\max V, Y, \max \lambda = 0, 1\)
the asymptotic properties of the following estimators of the weighted averages in eqns. (2)

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

provided that

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

is finite and positive for all \( n \) sufficiently large, where

\[
\ell_\beta(x, w, h; v, x_1, x_2, \tilde{w}) = s_{Y,1}(x, w)\ell_{Y,1}(x, w, h; y, x_1, x_2, \tilde{w}) + s_{Y,0}(x, w)\ell_{Y,0}(x, w, h; y, x_1, x_2, \tilde{w}) \\
+ s_{1,1}(x, w)\ell_{1,1}(x, w, h; 1, x_1, x_2, \tilde{w}) + s_{1,0}(x, w)\ell_{1,0}(x, w, h; 1, x_1, x_2, \tilde{w}),
\]

where \( \ell_{V,\lambda} \) is as defined in Lemma 4.1, and

\[
s_{Y,1}(x, w) = \frac{1}{g_{1,0}(x, w)}, \\
s_{Y,0}(x, w) = -\frac{g_{1,1}(x, w)}{g_{1,0}(x, w)} \frac{1}{g_{1,0}(x, w)}, \\
s_{1,1}(x, w) = -\frac{g_{Y,0}(x, w)}{g_{1,0}(x, w)} \frac{1}{g_{1,0}(x, w)}, \\
s_{1,0}(x, w) = \left( 2g_{Y,0}(x, w)g_{1,1}(x, w) - g_{Y,1}(x, w) \right) \frac{1}{g_{1,0}(x, w)}
\]

Because we are interested in weighted averages of \( \hat{\beta}(x, w) \) as well as \( \beta(x, w) \) itself, we now consider
the asymptotic properties of the following estimators of the weighted averages in eqns. (2)~(7):

\[
\hat{\beta}_m(x, h_n) = \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)m(w)dw, \\
\hat{\beta}_{m,W}(x, h_n) = \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)m(w)\hat{f}_W(w)dw, \\
\hat{\beta}_{m,W,X}(x, h_n) = \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)m(w)\hat{f}_{W|X}(w | x)dw \\
= \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)m(w) \frac{\hat{g}_{1,0}(x, w, h_n)\hat{f}_{W}(w)}{\int_{S_{\beta,0}(w, h_n)} \hat{g}_{1,0}(x, w, h_n)dw}dw, \\
\hat{\beta}_m(h_n) = \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)\tilde{m}(x, w)dw dx, \\
\hat{\beta}_{m,W,X}(h_n) = \int_{S_{\beta,0}(w, h_n)} \hat{\beta}(x, w, h_n)\tilde{m}(x, w)\hat{f}_{W|X}(w | x)dw dx
\]
\[ = \int_{S_{\beta(\cdot, h_n)}^w} \tilde{\beta}(x, w, h_n) \tilde{m}(x, w) \frac{\hat{g}_{1,0}(x, w, h_n) \hat{f}_W(w)}{\hat{g}_{1,0}(x, w, h_n) dw} dx, \]

\[ \tilde{\beta}_{m,W,X}(h_n) = \int_{S_{\beta(\cdot, h_n)}^w} \tilde{\beta}(x, w, h_n) \tilde{m}(x, w) \hat{f}_{W,X}(w, x) dwdx \]

\[ = \int_{S_{\beta(\cdot, h_n)}^w} \tilde{\beta}(x, w, h_n) \tilde{m}(x, w) \hat{g}_{1,0}(x, w, h_n) \hat{f}_W(w) dwdx, \]

where \( S_{\beta(\cdot, h_n)}^w \equiv \{ w \in \mathbb{R} : \hat{g}_{1,0}(x, w, h_n) > 0 \} \), \( S_{\beta(\cdot, h_n)}^{x,w} \equiv \{ (x, w) \in \mathbb{R}^2 : \hat{g}_{1,0}(x, w, h_n) > 0 \} \), 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 \( \mathcal{M} \) and \( \tilde{\mathcal{M}} \) 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 \( \mathcal{M} \) and \( \tilde{\mathcal{M}} \), respectively; (ii) \( \inf_{(x,w)\in\mathcal{M}} f_{X|W}(x | w) > 0 \); (iii) \( \max_{\gamma, \lambda=1,1} \max_{\beta,\lambda=0,1} \sup_{(x,w)\in\tilde{\mathcal{M}}} |g_{V,\lambda}(x, w)| < \infty. \)

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. (11)~(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)

\[ \sup_{x \in \mathcal{M}} |\tilde{\beta}_m(x, h_n) - \beta_m(x)| = O \left( \tau^{-3} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{B,s} (h_n^{-1})^{\beta_{B,s}} \right) \right) \]

\[ + O_p \left( \tau^{-3} n^{-1/2} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{L,s} (h_n^{-1})^{\beta_{L,s}} \right) \right), \]

(ii)

\[ \sup_{x \in \mathcal{M}} |\tilde{\beta}_{m,W}(x, h_n) - \beta_{m,W}(x)| = O \left( \tau^{-3} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{B,s} (h_n^{-1})^{\beta_{B,s}} \right) \right) \]

\[ + O_p \left( \tau^{-3} n^{-1/2} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{L,s} (h_n^{-1})^{\beta_{L,s}} \right) \right), \]

and (iii)

\[ \sup_{x \in \mathcal{M}} |\tilde{\beta}_{m,W,X}(x, h_n) - \beta_{m,W,X}(x)| = O \left( \tau^{-3} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{B,s} (h_n^{-1})^{\beta_{B,s}} \right) \right) \]

\[ + O_p \left( \tau^{-3} n^{-1/2} (h_n^{-1})^{\gamma_1,\beta,s} \exp \left( \alpha_{L,s} (h_n^{-1})^{\beta_{L,s}} \right) \right). \]
where $\alpha_{B,s}, \beta_{B,s}, \gamma_{B,s}, \alpha_{L,s}, \beta_{L,s},$ and $\gamma_{L,s}$ are as defined in Theorem 4.7.

The following theorem establishes asymptotic results for the semiparametric functionals taking the forms of eqns. (14)~(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)

$$n^{1/2} (\Omega_m)^{-1/2} \left( \hat{\beta}_m(h_n) - \beta_m \right) \xrightarrow{d} N(0, 1),$$

provided that

$$\Omega_m \equiv E \left[ (\psi_{\beta_m}(V, X_1, X_2, W))^2 \right]$$

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

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

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)

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

provided that

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

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

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

$$+ \psi_{1,0}(P_1; 1, x_1, x_2, \tilde{w}) - \psi_{1,0}(P_2; 1, x_1, x_2, \tilde{w}) + \psi_f(P_3; \tilde{w}),$$

and where $\tilde{m}_{fW|X} s_{V,\lambda}, P_1, P_2,$ and $P_3$ denote the functions mapping $(x, w)$ to $\tilde{m}(x, w)f_{W|X}(w \mid x)$

$s_{V,\lambda}(x, w), \beta(x, w)\tilde{m}(x, w)f_{W}(w) / f_X(x), \int s_{V,\lambda}(x, w, \tilde{w}) \beta(x, w)\tilde{m}(x, w)f_{W|X}(w \mid x) dw / f_X(x),$ and $\beta(x, w)$

$\tilde{m}(x, w)f_{W|X}(x \mid w) / f_X(x),$ respectively; (iii)$$n^{1/2} (\Omega_{\tilde{m}_{fW|X}})^{-1/2} \left( \hat{\beta}_{\tilde{m}_{fW|X}}(h_n) - \beta_{\tilde{m}_{fW|X}} \right) \xrightarrow{d} N(0, 1),$$

25
provided that
\[ \Omega \tilde{m}_{fW,X} \equiv E\left[ (\psi_{\beta \tilde{m}_{fW,X}} (V, X_1, X_2, W))^2 \right] \]
is finite and positive for all \( n \) sufficiently large, where
\[
\psi_{\beta \tilde{m}_{fW,X}} (v, x_1, x_2, \tilde{w}) \equiv \sum_{V=1}^{Y} \sum_{\lambda=0,1} \psi_{V,\lambda} (\tilde{m}_{fW,X} s_{V,\lambda}; v, x_1, x_2, \tilde{w}) + \psi_{1,0} (\beta \tilde{m}_{fW}; 1, x_1, x_2, \tilde{w}) + \psi_f (\beta \tilde{m}_{fX|W}; \tilde{w}),
\]
where \( \tilde{m}_{fW,X} s_{V,\lambda}, \beta \tilde{m}_{fW}, \) and \( \beta \tilde{m}_{fX|W} \) 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 \( \beta(x, w) \tilde{m}(x, w) f_{X|W}(x | w), \) respectively.

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 \( X \perp U_y \mid W. \)\(^4\) 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 for purposes of comparison. For those estimators, we consider a flat-top kernel of infinite order, where the Fourier transform is given in eqn.(8) with \( \bar{\xi} = .5. \) All estimates are constructed at values \( x = 1 \) and \( w = 1. \) For our estimator, we scan a set of bandwidths\(^5\) ranging from 6.5 to 8.0 in increments of 0.05 in order to find the optimal bandwidth minimizing the mean square error (MSE). For both local linear estimators, we scan a set of bandwidths ranging from 2 to 3.5, with the same increments. All simulations draw 500 samples of 1,000, 2,000, or 8,000 observations.

\(^4\)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 \( X \perp U_y \mid W. \)

\(^5\)Note that the flat-top kernel has a very narrow central peak, so that even moderately large bandwidths result in highly local smoothing.
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. As 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 $\mu$ and variance $\sigma^2$. Its characteristic function has a tail of the form $|\zeta|^{-2}$. The normal distribution with variance $\sigma^2$ is used as an example of a supersmooth distribution. The tail of the characteristic function of the normal distribution is of the form $\exp(-\frac{\sigma^2}{2} |\zeta|^2)$. Our example of an ordinarily smooth function for $f_2(W)$ is a piecewise linear continuous function with a discontinuous first derivative

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

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 $|\zeta|^{-1} \exp\left(-\frac{1}{4} |\zeta|^2\right)$ as $|\zeta| \to \infty$ for $V = X$ or $W$.

Table 2 reports the bias squared, the variance, and the MSE of the three estimators, which are functions of bandwidth for a sample size of 1,000. We show results from only a subset of the bandwidths for conciseness. The results from the optimal bandwidth are reported at the last column of each estimator. 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 gives larger variance than the infeasible local linear estimator and the local linear estimator based on error-contaminated covariates. Nevertheless, our estimator attains lower MSE than the local linear estimator ignoring measurement error, as the bias reduction of the former is enough to compensate for the increase in variance. As a result, our estimator outperforms the local linear estimator in terms of MSE. Table 3 reports Monte Carlo simulation results for the convergence rate as a function of sample size for each example. MSEs in all examples decrease as sample size increases, corroborating our theoretical results.
6 Application: The Impact of Family Income on Child Achievement

This section applies our estimator to study the causal effect of family income on child achievement. 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 depicts the causal relationships 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.\(^6\) 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 mathematics scores, and reading achievement is measured by a simple average of the reading recognition and reading comprehension scores. We use

\(^6\)We thank Gordon Dahl for providing the NLSY data.
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.

Tables 4.1 ∼ 5.2 show estimation results obtained by our new estimator and a local linear estimator ignoring the family income measurement error. We scan a set of bandwidths ranging from 14 to 21 in increments of 0.1 for both estimators. Estimates from only a subset of the bandwidths are reported for conciseness. The practical implementation of both methods requires the selection of the smoothing parameters. Because a formal decision rule for the optimal bandwidth selection is beyond the scope of the present paper, we consider an informal selection rule. It is well known that the estimates are not very sensitive to small changes of the bandwidth in the vicinity of a valid bandwidth. Thus we conjecture that a proper bandwidth could be obtained by carefully examining a region where the estimates are stable. Each estimate is evaluated at given values of standardized family income ($X$) and mother’s AFQT score ($W$) ranging from $-1$ to $1$ in increments of 0.25. 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 4.1 reports the estimated impact of current family income on children’s math achievement. Because the estimates are very stable over the bandwidth range $15 \sim 18$, we take valid bandwidths to lie in this range. The covariate-conditioned average marginal effects of current family income on children’s math achievement from our estimator are positive and statistically significant over all ranges of $x$ and $w$. The average marginal effect is about 0.664 at $x = -1$ and $w = -1$, which means that the effect of a one standard deviation increase in log of current family income is to increase a child’s math score by about 0.664 of a standard deviation. Effects are decreasing as $x$ and $w$ increase toward 0.25 but increase again when $x$ and $w$ are above 0.25. 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 0.139 at \(x = 0.25\) and \(w = 0.25\), whereas that from our estimator is 0.647, a difference of a factor of about five. 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.

Table 4.2 reports the covariate-conditioned average marginal effects of permanent family income on children’s math achievement. The average marginal effects from our estimator are slightly less than those in Table 4.1 over all ranges of \((x, w)\), and the differences are statistically significant, but the magnitude of the differences are tiny. Thus there is little evidence that the effect of permanent income on children’s math scores is much larger than the effect of current income. As we expect, the local linear estimator shows a larger effect for permanent income than for current income because averaging usually mitigates the attenuation bias from measurement errors.

Table 5.1 shows the impact of current family income on children’s reading achievement. We observe that the estimates from both estimators are not sensitive to bandwidths ranging from 16 to 18. The covariate-conditioned average marginal effects of current family income on children’s reading achievement from our estimator are close to those in Table 4.1. Moreover, they are also positive, statistically significant, 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 0.684 at \(x = −1\) and \(w = −1\), which means that the effect of a one standard deviation increase in log of current family income is to increase a child’s reading score by about 0.684 of a standard deviation, while that from the local linear estimator is 0.100.

Table 5.2 reports the impact of permanent family income on children’s reading achievement. The covariate-conditioned average marginal effects from our estimator are slightly less than those in Table 5.1 over all ranges of \((x, w)\), and the differences are statistically significant, but the magnitudes of the differences are very small. Thus, the results indicate that the effect of permanent income on children’s reading scores is slightly less than the effect of current income. However, the local linear estimator obtains a larger apparent effect of permanent family income than that for current family income over all ranges of \((x, w)\) except those from 0.75 to 1.

Figure 4 shows a graph of the causal effects of current (top) and permanent (bottom) family income on children’s math scores at various values of family income and mother’s AFQT, ranging from −1 to 1, obtained using our estimator with bandwidth 16.5. All estimates are positive over the ranges of both family income and AFQT score. In general, the impact of current family income at a given AFQT increases as family income moves from 0.1 to −1 or 1, making a broad \(U\)-shape. It attains a minimum of 0.645 at
\(x = 0.1\) and \(w = 1\). As a result, diminishing returns to income are observed only at income levels below \(x = 0.1\). Nevertheless, the difference between the income effect of family with \(x = 0.1\) and that of a family with \(x = 1\) is economically meaningless because the difference is negligible. At any given value for family income, the effects are rather stable across AFQT scores. Impacts of current and permanent family income generally have similar magnitudes.

Figure 5 shows a graph of the apparent causal effects of current (top) and permanent (bottom) family income on children’s math scores obtained using the local linear estimator. It shows much smaller effects than those 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.

Figure 6 shows the impacts of current (top) and permanent (bottom) family income on children’s reading scores at various points of family income and AFQT ranging from \(-1\) to \(1\), obtained by our estimator. The same bandwidth is used as before. 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 \(0.1\), the effect of family income on reading scores increases with family income. The effect attains a minimum value of \(0.658\) at \(x = 0.1\) and \(w = 1\). As mentioned before, the difference between the income effect of a family with \(x = 0.1\) and that of a family with \(x = 1\) is economically meaningless. Both current and permanent incomes also have similar effects on children’s reading scores. Thus there is little evidence that the effect of permanent income on reading score is much larger than the effect of current income.

Figure 7 depicts the apparent causal effects of current (top) and permanent (bottom) 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. The two figures show quite different effect shapes. Current 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 current 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 $x = 0.1$ but a wide $U$-shape overall. Nevertheless, the differences between the high and low income effects are economically meaningless because the differences are negligible. Further, we find that there is little evidence that the effect of permanent family income on children’s achievement is much larger than the effect of current family income.

7 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. We investigate the finite-sample behavior of our estimators using Monte Carlo simulations, and we apply our new methods to study the impact of family income on child achievement. There we find interesting new results, suggesting that these effects are considerably larger than previously recognized.
A 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

$$
\frac{iE[X_1 e^{i\xi X_2}]}{E[e^{i\xi X_2}]} = \frac{iE[X e^{i\xi(X+U_2)}] + iE[U_1 e^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]}
$$

$$
= \frac{iE[X e^{i\xi(X+U_2)}] + iE[U_1 e^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]}
$$

$$
= \frac{iE[X e^{i\xi(X+U_2)}] + iE[U_1 | X, U_2] e^{i\xi(X+U_2)}]}{E[e^{i\xi(X+U_2)}]}
$$

as considered by SWC. We use $E[U_1 | X, U_2] = 0$ 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 $f(U_2, X, W) = f(U_2, X \mid W)f(W) = f(U_2)f(X \mid W)f(W) = f(U_2)f(X, W)$. 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 $\zeta$, we have

$$
\phi_V(\zeta, W) \equiv E[V e^{i\xi X} \mid W]
$$

$$
= \frac{E[V e^{i\xi X} \mid W] E[e^{i\xi U_2}]}{E[e^{i\xi X}] E[e^{i\xi U_2}]}
$$

$$
= \frac{E[V e^{i\xi X} \mid X, W] e^{i\xi X} e^{i\xi U_2} \mid W]}{E[e^{i\xi X}] E[e^{i\xi U_2}]}
$$

$$
= \frac{E[V \mid X, U_2, W] e^{i\xi X} \mid X, U_2, W] E[e^{i\xi U_2} \mid W]}{E[e^{i\xi X}] E[e^{i\xi U_2}]}
$$

$$
= \frac{E[V \mid X, U_2, W] e^{i\xi X} e^{i\xi U_2} \mid X, U_2, W] E[e^{i\xi X}]}{E[e^{i\xi X}] E[e^{i\xi U_2}]}
$$

as needed.

34
\[
\begin{align*}
    & = \frac{E[V e^{i\zeta X}] W}{E[e^{i\zeta X}] E[e^{i\zeta U_2}] E[e^{i\zeta X}]}
    \quad \frac{E[e^{i\zeta X}] E[e^{i\zeta X}]}{E[e^{i\zeta X}]}
    \quad \exp \left( \ln(E[e^{i\zeta X}]) - \ln 1 \right)
    \quad \exp \left( \int_0^\zeta D_\zeta \ln(E[e^{i\zeta X}]) d\zeta \right)
    \quad \exp \left( \int_0^\zeta iE[X x, \zeta] / E[e^{i\zeta X}] d\zeta \right).
\end{align*}
\]

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

\[
(-i\zeta)^\lambda E[V e^{i\zeta X} \mid W = w] = (-i\zeta)^\lambda \int E[V \mid W = w, X = x] f_{X \mid W}(x \mid w) e^{i\zeta x} dx
= (-1)^\lambda \int E[V \mid W = w, X = x] f_{X \mid W}(x \mid w) D^\lambda_x e^{i\zeta x} dx
= \int D^\lambda_x (E[V \mid W = w, X = x] f_{X \mid W}(x \mid w)) e^{i\zeta x} dx
= \int g_{V,\lambda}(x, w) e^{i\zeta x} dx.
\]

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

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

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

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

\[
g_{V,\lambda}(x, w, h) \equiv \int \frac{1}{h} k \left( \frac{\tilde{x} - x}{h} \right) g_{V,\lambda}(\tilde{x}, w) d\tilde{x}
= \int \frac{1}{h} k \left( \frac{\tilde{x} - x}{h} \right) D^\lambda_x (E[V \mid X = \tilde{x}, W = w] f_{X \mid W}(\tilde{x} \mid w)) d\tilde{x}.
\]

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

\[
g_{V,\lambda}(x, w, h) = \frac{1}{2\pi} \int \kappa(h\zeta) \left( (-i\zeta)^\lambda E[V e^{i\zeta X} \mid W = w] \right) \exp(-i\zeta x) d\zeta
= \frac{1}{2\pi} \int (-i\zeta)^\lambda \kappa(h\zeta) \phi_V(\zeta, w) \exp(-i\zeta x) d\zeta. \quad \square
\]
Proof of Lemma 4.1 For $A = 1, X_1$, we let $\theta_A(\zeta) \equiv E \left[ A e^{i\xi X_2} \right]$ and for $V = 1, Y$,

$$\theta_V(\zeta, w) \equiv E \left[ V e^{i\xi X_2} \mid W = w \right]$$

$$= \int \int ve^{i\xi x_2} f_{V, X_2 | W}(v, x_2 \mid w) dv dx_2$$

$$= \frac{\chi_V(\zeta, w)}{f_W(w)},$$

where $\chi_V(\zeta, w) \equiv \int \int ve^{i\xi x_2} f_{V, X_2, W}(v, x_2, w) dv dx_2$, $f_{V, X_2 | W}(v, x_2 \mid w)$ 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 $\hat{\theta}_A(\zeta) \equiv \hat{E} \left[ A e^{i\xi X_2} \right]$ and $\delta \hat{\theta}_A(\zeta) \equiv \hat{\theta}_A(\zeta) - \theta_A(\zeta)$. Similarly $\hat{\theta}_V(\zeta, w) \equiv \hat{E} \left[ V e^{i\xi X_2} \mid W = w \right] = \hat{\chi}_V(\zeta, w) / \hat{f}_W(w)$, where

$$\hat{\chi}_V(\zeta, w) = \frac{1}{n} \sum_{j=1}^{n} k_h(W_j - w) V_j e^{i\xi X_2 j} = \hat{E} \left[ V e^{i\xi X_2} k_h(W - w) \right]$$

$$\hat{f}_W(w) = \frac{1}{n} \sum_{j=1}^{n} k_h(W_j - w) = \hat{E} \left[ k_h(W - w) \right]$$

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 $\hat{\theta}_{X_1}(\zeta) / \hat{\theta}_1(\zeta)$:

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

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

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

(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
\[
\begin{align*}
\dot{q}_V(\zeta, w) &\equiv \frac{\dot{X}_V(\zeta, w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta, w)\dot{\theta}_1(\zeta)}{(\theta_1(\zeta))^2} \\
\dot{q}_1(\zeta, w) &\equiv \frac{\chi_V(\zeta, w)}{\theta_1(\zeta)} \left( \frac{\dot{\theta}_1(\zeta)}{\theta_1(\zeta)} \right)^2 \left( 1 + \frac{\delta \dot{\theta}_1(\zeta)}{\theta_1(\zeta)} \right)^{-1} - \frac{\delta \dot{X}_V(\zeta, w)}{\theta_1(\zeta)} \left( 1 + \frac{\delta \dot{\theta}_1(\zeta)}{\theta_1(\zeta)} \right)^{-1}.
\end{align*}
\]

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

\[
\frac{1}{\dot{f}_W(w)} = \frac{1}{f_W(w) + \delta \dot{f}_W(w)} = q_1(w) + \delta q_1(w)
\]

where \(q_1(w) \equiv 1/f_W(w)\) and where \(\delta q_1(w)\) can be written as either

\[
\delta q_1(w) = \left( -\frac{\delta \dot{f}_W(w)}{(f_W(w))^2} \right) \left( 1 + \frac{\delta \dot{f}_W(w)}{f_W(w)} \right)^{-1}
\]

or \(\delta q_1(w) = \delta q_1(w) + \delta q_1(w)\) with

\[
\delta q_1(w) = -\frac{\delta \dot{f}_W(w)}{(f_W(w))^2}
\]

\[
\delta q_1(w) = \frac{1}{f_W(w)} \left( \frac{\delta \dot{f}_W(w)}{f_W(w)} \right)^2 \left( 1 + \frac{\delta \dot{f}_W(w)}{f_W(w)} \right)^{-1}.
\]

For \(Q_{X_1}(\zeta) \equiv \int_0^\zeta (i\theta X_1(\xi)/\theta_1(\xi))d\xi\), \(\delta Q_{X_1}(\zeta) \equiv \int_0^\zeta (i\theta X_1(\xi)/\theta_1(\xi))d\xi - Q_{X_1}(\zeta)\) and some random function \(\delta Q_{X_1}(\zeta)\) such that \(|\delta Q_{X_1}(\zeta)| \leq |\delta Q_{X_1}(\zeta)|\) for all \(\zeta\),

\[
\exp\left( Q_{X_1}(\zeta) + \delta Q_{X_1}(\zeta) \right) = \exp(Q_{X_1}(\zeta)) \left( 1 + \delta Q_{X_1}(\zeta) + \frac{1}{2} \left[ \exp(\delta Q_{X_1}(\zeta)) \right] (\delta Q_{X_1}(\zeta))^2 \right).
\]

By substituting eqn.(18)~(21) into

\[
\dot{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h)
\]

\[
= \frac{1}{2\pi} \int (-i\zeta)^\lambda \kappa(h\zeta) \exp(-i\zeta x) \left[ \frac{\dot{X}_V(\zeta, w)}{\theta_1(\zeta)} \exp\left( \int_0^\zeta i\theta X_1(\xi)/\theta_1(\xi) d\xi \right) - \frac{\theta_V(\zeta, w)}{\theta_1(\zeta)} \exp\left( \int_0^\zeta i\theta X_1(\xi)/\theta_1(\xi) d\xi \right) \right] d\zeta,
\]

we have

\[
\dot{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h)
\]

\[
= \frac{1}{2\pi} \int (-i\zeta)^\lambda \kappa(h\zeta) \exp(-i\zeta x) \left[ -\frac{\theta_V(\zeta, w)}{\theta_1(\zeta)} \exp\left( \int_0^\zeta i\theta X_1(\xi)/\theta_1(\xi) d\xi \right) \right. \\
+ \left\{ \frac{\chi_V(\zeta, w)}{\theta_1(\zeta)} + \frac{\delta \dot{X}_V(\zeta, w)}{\theta_1(\zeta)} - \frac{\chi_V(\zeta, w)\delta \dot{\theta}_1(\zeta)}{(\theta_1(\zeta))^2} + \delta \dot{q}_V(\zeta, w) \right\} \\
\times \left\{ \frac{1}{f_W(w)} - \frac{\delta \dot{f}_W(w)}{(f_W(w))^2} + \delta q_1(w) \right\} \times \exp(Q_{X_1}(\zeta)).
\]
Using the identity

\[
\bar{\chi} \left( -L^2 V, \lambda \int [\int \left( V, \lambda \pi V, \lambda \pi \left( \frac{1}{\pi} \theta_1 \right) ^2 \right) \xi \right)
\]

\[
\times \left\{ 1 + \int_0^\zeta i \delta_1 q_{X_1}(\xi) d\xi + \int_0^\zeta i \delta_2 q_{X_1}(\xi) d\xi + \frac{1}{2} \exp(\delta \bar{Q}_{X_1}(\zeta)) \left( \int_0^\zeta i \delta_1 q_{X_1}(\xi) d\xi \right)^2 \right\} d\zeta.
\]

Keeping the terms linear in \( \delta \bar{\theta}_1(\zeta), \delta \bar{\theta}_2(\zeta), \delta \bar{\chi}_1(\zeta, w), \) and \( \delta \bar{f}_W(w) \) gives the linearization of \( \bar{g}_{V,\lambda}(x, w, h) \), denoted \( \bar{g}_{V,\lambda}(x, w, h) \):

\[
\bar{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h) = \frac{1}{2\pi} \int (-i\zeta)^2 (h(x) \exp(-i\zeta x) \times \left( \frac{\theta_1(\zeta, w)}{\theta_1(\zeta)} \exp(Q_{X_1}) \int_0^\zeta \left( \frac{i \delta X_1(\xi)}{\theta_1(\xi)} - \frac{i \theta_1(\xi) \delta \bar{\theta}_1(\xi)}{(\theta_1(\xi))^2} \right) d\xi 
\]

\[
- \exp(Q_{X_1}(\zeta)) \frac{\chi_1(\zeta, w)}{\theta_1(\zeta)} \delta \bar{f}_W(w)) + \exp(Q_{X_1}(\zeta)) \frac{1}{f_W(w)} \left( \frac{\delta \bar{\chi}_1(\zeta, w)}{\theta_1(\zeta)} - \frac{\chi_1(\zeta, w) \delta \bar{\theta}_1(\zeta)}{(\theta_1(\zeta))^2} \right) \right) d\zeta
\]

\[
+ \frac{1}{2\pi} \int (-i\zeta)^2 (h(x) \exp(-i\zeta x) \phi_1(\zeta, w) \left( - \frac{\delta \bar{f}_W(w)}{f_W(w)} + \frac{\delta \bar{\chi}_1(\zeta, w)}{\chi_1(\zeta, w)} - \frac{\delta \bar{\theta}_1(\zeta)}{(\theta_1(\zeta))} \right) d\zeta.
\]

Using the identity

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\zeta, \xi) d\zeta d\xi = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\zeta, \xi) d\zeta d\xi + \int_{-\infty}^{0} \int_{-\infty}^{\infty} f(\zeta, \xi) d\zeta d\xi = \int \int_{-\infty}^{\infty} f(\zeta, \xi) d\zeta d\xi
\]

for any absolutely integrable function \( f \), we get

\[
L_{V,\lambda}(x, w, h) \equiv \bar{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h)
\]

\[
= \frac{1}{2\pi} \int \left\{ \left\{ - \frac{1}{2\pi (\theta_1(\zeta))^2} \right\} \int_{-\infty}^{\infty} \left( -i\zeta \right)^2 (h(x) \exp(-i\zeta x) \phi_1(\zeta, w) \right) d\zeta + \left\{ - \frac{1}{2\pi (\theta_1(\zeta))^2} \right\} \int_{-\infty}^{\infty} \left( -i\zeta \right)^2 (h(x) \exp(-i\zeta x) \phi_1(\zeta, w) \right) d\zeta
\]

\[
= \left[ \Psi_{V,\lambda,1}(\zeta, x, w, h) \left( \hat{E}[e^{i\xi X_1}] - E[e^{i\xi X_1}] \right) + \Psi_{V,\lambda,2}(\zeta, x, w, h) \right. \left( \hat{E}[X_1 e^{i\xi X_1}] - E[X_1 e^{i\xi X_1}] \right)
\]

\[
+ \Psi_{V,\lambda,3}(\zeta, x, w, h) \left( \hat{E}[V e^{i\xi X_1} k_h(W - w)] - E[V e^{i\xi X_1} k_h(W - w)] \right)
\]

38
+ \Psi_{V,\lambda,f_w}(\zeta, x, w, h) \left( \bar{E}[h_k(W - w)] - E[h_k(W - w)] \right) d\zeta

= \bar{E} \left[ \int \Psi_{V,\lambda,1}(\zeta, x, w, h) \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) \left( X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}] \right) d\zeta + \int \Psi_{V,\lambda,X_2}(\zeta, x, w, h) \left( V e^{i\zeta X_2} k_h(W - w) - E[V e^{i\zeta X_2} k_h(W - w)] \right) d\zeta + \int \Psi_{V,\lambda,f_w}(\zeta, x, w, h) \left( k_h(W - w) - E[k_h(W - w)] \right) d\zeta \right]

= \bar{E} \left[ \ell_{V,\lambda}(x, w, h; V, X_1, X_2, W) \right]

where \Psi_{V,\lambda,A}(\zeta, x, w, h) \text{ and } \ell_{V,\lambda}(x, w, h; V, X_1, X_2, W) \text{ are defined in the statement of the Lemma 4.1.}\n
We define the following convenient notation as employed in SWC.

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

**Proof of Theorem 4.2** Using Parseval’s identity, we have

\[
|B_{V,\lambda}(x, w, h)| = |g_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, 0)|
\]

\[
= \left| \frac{1}{2\pi} \int \kappa(h\zeta)(-i\zeta)^{\lambda} \phi_V(\zeta, w) \exp(-i\zeta x) d\zeta - \frac{1}{2\pi} \int (-i\zeta)^{\lambda} \phi_V(\zeta, w) \exp(-i\zeta x) d\zeta \right|
\]

\[
= \left| \frac{1}{2\pi} \int \kappa(h\zeta) - 1(-i\zeta)^{\lambda} \phi_V(\zeta, w) \exp(-i\zeta x) d\zeta \right|
\]

\[
\leq \frac{1}{2\pi} \int \left| \kappa(h\zeta) - 1 \right| \left| \zeta \right|^\lambda \left| \phi_V(\zeta, w) \right| d\zeta
\]

\[
= \frac{1}{\pi} \int_{\xi/h}^{\infty} \left| \kappa(h\zeta) - 1 \right| \left| \zeta \right|^\lambda \left| \phi_V(\zeta, w) \right| d\zeta
\]

\[
\leq \int_{\xi/h}^{\infty} \left| \zeta \right|^\lambda \left| \phi_V(\zeta, w) \right| d\zeta
\]

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

\[
\sup_{(x, w) \in \text{supp}(X, W)} |B_{V,\lambda}(x, w, h)| \leq \int_{\xi/h}^{\infty} \left| \zeta \right|^\lambda C_0 (1 + |\zeta|)^{\gamma_0} \exp(\alpha_0 |\zeta|^{\beta_0}) d\zeta
\]

\[
\leq \int_{\xi/h}^{\infty} \left| \zeta \right|^\lambda (1 + |\zeta|)^{\gamma_0} \exp(\alpha_0 |\zeta|^{\beta_0}) d\zeta
\]

\[
= O \left( \left( \xi/h \right)^{\gamma_0 + \lambda + 1} \exp \left( \alpha_0 \left( \xi/h \right)^{\beta_0} \right) \right)
\]

\[
= O \left( \left( h^{-1} \right)^{\gamma_0 + \lambda} \exp \left( \alpha_B \left( h^{-1} \right)^{\beta_0} \right) \right).\]

**Lemma A.1** Suppose the conditions of Lemma 4.1 hold. For each \( \zeta \) and \( h \), and for \( A = 1, X_1, X_2, f_w \), let \( \Psi_{V,\lambda,A}(\zeta, x, w, h) = \sup_{(x, w) \in \text{supp}(X, W)} |\Psi_{V,\lambda,A}(\zeta, x, w, h)| \), and define

\[
\Psi_{V,\lambda}(h) = \sum_{A=1,X_1} \int \Psi_{V,\lambda,A}(\zeta, h) d\zeta + h^{-1} \sum_{B=\chi, f_w} \int \Psi_{V,\lambda,B}(\zeta, h) d\zeta.
\]
If Assumption 4.1 also holds, then for $h > 0$

$$
\Psi^+_{V, \lambda}(h) = O \left( (1 + h^{-1})^{\gamma_0 + \lambda + \gamma_1 - \gamma_0 + 2} \exp \left( (\alpha_0 1_{\{ \beta_\theta = \beta_\theta \}} - \alpha_\theta)(h^{-1})^{\beta_\theta} \right) \right).
$$

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

$$
\Psi^+_{V, \lambda, 1}(\zeta, h) = \sup_{(x, w) \in \text{supp}(X, W)} |\Psi_{V, \lambda, 1}(\zeta, x, w, h)|
= \sup_{(x, w) \in \text{supp}(X, W)} \left| \frac{i \theta_{X, 1}(\zeta)}{2\pi (\theta_1(\zeta))^2} \int_\zeta^{\pm \infty} (-i \xi) \kappa(h \xi) \exp(-i \xi x) \phi_V(\xi, w) d\xi - \frac{1}{2\pi} (-i \zeta) \kappa(h \zeta) \exp(-i \zeta x) \phi_V(\zeta, w) \right| \theta_1(\zeta)
\leq \sup_{(x, w) \in \text{supp}(X, W)} \left\{ \sup_{\xi \in \text{supp}(W)} |\theta_{X, 1}(\zeta)| \left| \frac{\theta_{X, 1}(\zeta)}{\theta_1(\zeta)} \right| \int_\zeta^{\pm \infty} |\xi|^\lambda |\kappa(h \xi)| \exp(-i \xi x) \phi_V(\xi, w) d\xi \right\} \frac{1}{|\theta_1(\zeta)|}
= \frac{1}{|\theta_1(\zeta)|} \left\{ |D_\zeta \ln \phi_1(\zeta)| \int_\zeta^{\pm \infty} |\xi|^\lambda |\kappa(h \xi)| \left( \sup_{\xi \in \text{supp}(W)} |\phi_V(\xi, w)| \right) d\xi \right\}
+ |\zeta|^\lambda |\kappa(h \zeta)| \left( \sup_{\xi \in \text{supp}(W)} |\phi_V(\zeta, w)| \right)
= \frac{1}{|\theta_1(\zeta)|} \left\{ |D_\zeta \ln \phi_1(\zeta)| \int_\zeta^{\pm \infty} |\xi|^\lambda |\kappa(h \xi)| \left( \sup_{\xi \in \text{supp}(W)} |\phi_V(\xi, w)| \right) d\xi \right\}
+ |\zeta|^\lambda |\kappa(h \zeta)| \left( \sup_{\xi \in \text{supp}(W)} |\phi_V(\zeta, w)| \right).
$$

because we have $\theta_{X, 1}(\zeta)/\theta_1(\zeta) = -i D_\zeta \ln \phi_1(\zeta)$ by eqn.(20) in the proof of Lemma 3.1. Then

$$
\Psi^+_{V, \lambda, 1}(\zeta, h) \leq \frac{1}{|\theta_1(\zeta)|} \left\{ |D_\zeta \ln \phi_1(\zeta)| \int_\zeta^{\pm \infty} |\xi|^\lambda \left( \sup_{\xi \in \text{supp}(w)} |\phi_V(\xi, w)| \right) d\xi \right\}
+ |\zeta|^\lambda \left( \sup_{\xi \in \text{supp}(w)} |\phi_V(\zeta, w)| \right)
\leq \frac{1}{|\theta_1(\zeta)|} \left\{ |D_\zeta \ln \phi_1(\zeta)| \int_\zeta^{h^{-1}} |\xi|^\lambda \left( \sup_{\xi \in \text{supp}(w)} |\phi_V(\xi, w)| \right) d\xi \right\}
+ |\zeta|^\lambda \left( \sup_{\xi \in \text{supp}(w)} |\phi_V(\zeta, w)| \right).
$$

By using Assumption 4.1 and integrating $\Psi^+_{V, \lambda, 1}(\zeta, h)$ with respect to $\zeta$, we obtain
\[ \int \Psi_{V, \lambda, 1}(\zeta, h) d\zeta \leq \int \frac{1}{|\theta_1(\zeta)|} 1_{\{|\zeta| \leq h^{-1}\}} \left[ |D\zeta \ln \phi_1(\zeta)| \int_{\zeta}^{h^{-1}} |\zeta|^\lambda \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) d\zeta + |\zeta|^\lambda \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) \right] d\zeta \]

\[ \leq \int (1 + |\zeta|)^{-\gamma_0} \exp \left( -\alpha_\theta |\zeta|^{\beta_\theta} \right) 1_{\{|\zeta| \leq h^{-1}\}} \left[ (1 + |\zeta|)^\gamma_1 \int_{0}^{h^{-1}} |\zeta|^\lambda (1 + |\zeta|)^{\gamma_0} \exp (\alpha_\phi |\zeta|^{\beta_\phi}) d\zeta + |\zeta|^\lambda (1 + |\zeta|)^{\gamma_0} \exp (\alpha_\phi |\zeta|^{\beta_\phi}) \right] d\zeta \]

\[ \leq (1 + h^{-1})^{1-\gamma_0} \exp \left( -\alpha_\theta (h^{-1})^{\beta_\theta} \right) \left[ (1 + h^{-1})^{\gamma_1} (1 + h^{-1})^{\lambda + \gamma_0 + 1} \exp (\alpha_\phi (h^{-1})^{\beta_\phi}) + (1 + h^{-1})^{\gamma_0 + \lambda} \exp (\alpha_\phi (h^{-1})^{\beta_\phi}) \right] \]

\[ \leq (1 + h^{-1})^{1-\gamma_0} \exp \left( -\alpha_\theta (h^{-1})^{\beta_\theta} \right) (1 + h^{-1})^{\lambda + \gamma_0} \exp (\alpha_\phi (h^{-1})^{\beta_\phi}) ((1 + h^{-1})^{\gamma_1 + 1} + 1) \]

\[ \leq (1 + h^{-1})^{\gamma_0 + \lambda + \gamma_1} \exp \left( -\alpha_\theta (h^{-1})^{\beta_\theta} \right) \exp (\alpha_\phi (h^{-1})^{\beta_\phi}) . \]

Second,

\[ \Psi_{V, \lambda, X_1}(\zeta, h) = \sup_{(x, w) \in \text{supp}(X, W)} |\Psi_{V, \lambda, X_1}(\zeta, x, w, h)| \]

\[ = \sup_{(x, w) \in \text{supp}(X, W)} \left| \frac{1}{2\pi i} \frac{1}{|\theta_1(\zeta)|} \int_{|h\xi| = \pm \infty} (-i\xi)^\lambda \kappa(h\xi) \exp(-i\xi x) \phi_V(\zeta, w) d\xi \right| \]

\[ \leq \sup_{(x, w) \in \text{supp}(X, W)} \left| \frac{1}{|\theta_1(\zeta)|} \int_{|h\xi| = \pm \infty} |\xi|^{\lambda \gamma_1} |\kappa(h\xi)| \exp(-i\xi x) |\phi_V(\zeta, w)| d\xi \right| \]

\[ = \frac{1}{|\theta_1(\zeta)|} \int_{|h\xi| = \pm \infty} |\xi|^{\lambda \gamma_1} \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) d\xi \]

\[ = \frac{1}{|\theta_1(\zeta)|} \int_{|h\xi| = \pm \infty} |\xi|^{\lambda \gamma_1} 1_{\{|\xi| \leq h^{-1}\}} \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) d\xi \]

so that

\[ \int \Psi_{V, \lambda, X_1}(\zeta, h) d\zeta \leq \int_{0}^{h^{-1}} (1 + |\zeta|)^{-\gamma_0} \exp \left( -\alpha_\theta |\zeta|^{\beta_\theta} \right) \left( \int_{0}^{h^{-1}} |\zeta|^{\lambda (1 + |\zeta|)^{\gamma_0} \exp (\alpha_\phi |\zeta|^{\beta_\phi})} d\zeta \right) d\zeta \]

\[ \leq (1 + h^{-1})^{1-\gamma_0} \exp \left( -\alpha_\theta (h^{-1})^{\beta_\theta} \right) (1 + h^{-1})^{\lambda + \gamma_0 + 1} \exp (\alpha_\phi (h^{-1})^{\beta_\phi}) \]

\[ \leq (1 + h^{-1})^{\gamma_0 + \lambda + \gamma_0 + 2} \exp \left( -\alpha_\theta (h^{-1})^{\beta_\theta} \right) \exp (\alpha_\phi (h^{-1})^{\beta_\phi}). \]
Third,

\[ \Psi_{V,\lambda,V}(\zeta, h) \equiv \sup_{(x,w) \in \text{supp}(X,W)} |\Psi_{V,\lambda,V}(\zeta, x, w, h)| = \frac{1}{2\pi} \left| \frac{1}{2\pi} (-i\zeta)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \frac{\phi_V(\zeta, w)}{\chi_V(\zeta, w)} \right| \leq |\zeta|^{\lambda} \left( \sup_{w \in \text{supp}(W)} \left| \frac{\phi_V(\zeta, w)}{\chi_V(\zeta, w)} \right| \right) \]

so that

\[ h^{-1} \int \Psi_{V,\lambda,V}(\zeta, h) d\zeta \leq h^{-1} \int_0^{h^{-1}} |\zeta|^{\lambda} (1 + |\zeta|)^{-\gamma_0} \exp(-\alpha_\phi |\zeta|^{\beta_0}) (1 + |\zeta|)^{\gamma_0} \exp(\alpha_\phi |\zeta|^{\beta_0}) d\zeta \leq (1 + h^{-1})^{\gamma_0 + \lambda - \gamma_0 + 2} \exp(-\alpha_\phi (h^{-1})^{\beta_0}) \exp(\alpha_\phi (h^{-1})^{\beta_0}). \]

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

\[ \Psi_{V,\lambda,f_W}(\zeta, h) \equiv \frac{1}{2\pi} \left| \frac{1}{2\pi} (-i\zeta)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \frac{\phi_V(\zeta, w)}{f_W(w)} \right| \leq |\zeta|^{\lambda} \left( \sup_{w \in \text{supp}(W)} \left| \frac{\phi_V(\zeta, w)}{f_W(\zeta)} \right| \right) \]

so that

\[ h^{-1} \int \Psi_{V,\lambda,f_W}(\zeta, h) d\zeta \leq h^{-1} \int_0^{h^{-1}} |\zeta|^{\lambda} (1 + |\zeta|)^{\gamma_0} \exp(\alpha_\phi |\zeta|^{\beta_0}) d\zeta \leq (1 + h^{-1})^{\gamma_0 + \lambda - \gamma_0 + 2} \exp(\alpha_\phi (h^{-1})^{\beta_0}). \]

Putting together these rates for each term of \( \Psi_{V,\lambda}(h) \) gives the desired result. \( \square \)

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

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

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

\[ \sigma_n^{-1} n^{1/2} \left( E \left[ \sum_{j=1}^J A_j P_{n,j}(X_2) + \sum_{k=1}^K B_k Q_{n,k}(W) \right] - E \left[ \sum_{j=1}^J A_j P_{n,j}(X_2) + \sum_{k=1}^K B_k Q_{n,k}(W) \right] \right) \xrightarrow{d} N(0, 1). \]
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 $E[L_{V,X}(x, w, h)] = 0$ by the definition of $L_{V,X}(x, w, h)$. Assumption 4.2 guarantees that $L_{V,X}(x, w, h)$ has a finite variance so that

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

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

$$\Omega_{V,X}(x, w, h) = n E \left[ (\tilde{g}_{V,X}(x, w, h) - g_{V,X}(x, w, h))^2 \right] = E \left[ \left( \int \Psi_{V,X,1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_1(\zeta) d\zeta + \int \Psi_{V,X,1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{X_1}(\zeta) d\zeta \right. \right.

\left. + \int \Psi_{V,X,V}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{V}(\zeta, w) d\zeta + \int \Psi_{V,X,fW}(\zeta, x, w, h) n^{1/2} \hat{f}_{W}(w) d\zeta \right)^2

\leq E \left[ \left( \int \Psi_{V,X,1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_1(\zeta) d\zeta \right)^2 + \int \Psi_{V,X,X_1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{X_1}(\zeta) d\zeta \right.

\left. + \int \Psi_{V,X,V}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{V}(\zeta, w) d\zeta + \int \Psi_{V,X,fW}(\zeta, x, w, h) n^{1/2} \hat{f}_{W}(w) d\zeta \right)^2

\leq \left[ E \left( \left( \int \Psi_{V,X,1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_1(\zeta) d\zeta \right)^2 \right)^{1/2} \right. \left. + E \left( \left( \int \Psi_{V,X,X_1}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{X_1}(\zeta) d\zeta \right)^2 \right)^{1/2} \right.

\left. + E \left( \left( \int \Psi_{V,X,V}(\zeta, x, w, h) n^{1/2} \hat{\delta}_{V}(\zeta, w) d\zeta \right)^2 \right)^{1/2} \right.

\left. + E \left( \left( \int \Psi_{V,X,fW}(\zeta, x, w, h) n^{1/2} \hat{f}_{W}(w) d\zeta \right)^2 \right)^{1/2} \right]^{1/2}

= \left[ \int \int \Psi_{V,X,1}(\zeta, x, w) E \left[ n \hat{\delta}_1(\zeta) \hat{\delta}_{X_1}^T(\xi) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \right. \left. + \int \int \Psi_{V,X,X_1}(\zeta, x, w) E \left[ n \hat{\delta}_{X_1}(\zeta) \hat{\delta}_{X_1}^T(\xi) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \right.

\left. + \int \int \Psi_{V,X,V}(\zeta, x, w) E \left[ n \hat{\delta}_{V}(\zeta, w) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \times (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right]^{1/2}

\left. + \int \int \Psi_{V,X,fW}(\zeta, x, w) E \left[ n \hat{\delta}_{W}(w) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \times (\Psi_{V,X,fW}(\xi, x, w, h))^T d\zeta d\xi \right]^{1/2}

\left. + \int \int \Psi_{V,X,fW}(\zeta, x, w) E \left[ n \hat{\delta}_{W}(w) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \times (\Psi_{V,X,fW}(\xi, x, w, h))^T d\zeta d\xi \right]^{1/2}

\left. + \int \int \Psi_{V,X,fW}(\zeta, x, w) E \left[ n \hat{\delta}_{W}(w) (\Psi_{V,X,X_1}(\xi, x, w, h))^T d\zeta d\xi \right] \times (\Psi_{V,X,fW}(\xi, x, w, h))^T d\zeta d\xi \right]^{1/2}.

43
Note that by Assumption 4.2

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

so that

\[ \left| E \left[ n \delta \hat{\theta}_1(\zeta) \delta \hat{\theta}_1^\dagger(\xi) \right] \right| = \left| \theta_1(\zeta - \xi) - \theta_1(\xi) \theta_1(-\xi) \right| 
\]
\[ \leq E \left[ e^{i(\zeta - \xi) X_2} \right] + E \left[ e^{i\xi X_2} \right] E \left[ e^{-i\xi X_2} \right] 
\]
\[ \leq 1; \]

\[ E \left[ n \delta \hat{\theta}_{X_1}(\zeta) \delta \hat{\theta}_{X_1}^\dagger(\xi) \right] = E \left[ n \left( \hat{\theta}_{X_1}(\zeta) - \theta_{X_1}(\zeta) \right) \left( \hat{\theta}_{X_1}^\dagger(\xi) - \theta_{X_1}^\dagger(\xi) \right) \right] 
\]
\[ = E \left[ \left( X_1 e^{i\xi X_2} - \theta_{X_1}(\zeta) \right) \left( X_1 e^{-i\xi X_2} - \theta_{X_1}^\dagger(\xi) \right) \right] 
\]
\[ = E \left[ X_1 e^{i\xi X_2} X_1 e^{-i\xi X_2} - \theta_{X_1}(\zeta) E \left[ X_1 e^{-i\xi X_2} \right] - E \left[ X_1 e^{i\xi X_2} \right] \theta_{X_1}^\dagger(\xi) + \theta_{X_1}(\zeta) \theta_{X_1}^\dagger(\xi) \right] 
\]
\[ = E \left[ X_1 X_1 e^{i(\zeta - \xi) X_2} - \theta_{X_1}(\zeta) \theta_{X_1}^\dagger(\xi) \right] 
\]

so that

\[ \left| E \left[ n \delta \hat{\theta}_{X_1}(\zeta) \delta \hat{\theta}_{X_1}^\dagger(\xi) \right] \right| = \left| E \left[ X_1 X_1 e^{i(\zeta - \xi) X_2} - \theta_{X_1}(\zeta) \theta_{X_1}^\dagger(\xi) \right] \right| 
\]
\[ \leq E \left[ \left[ X_1 X_1 \right] e^{i(\zeta - \xi) X_2} \right] + E \left[ \left[ X_1 \right] e^{i\xi X_2} \right] E \left[ \left[ X_1 \right] e^{-i\xi X_2} \right] 
\]
\[ \leq E \left[ \left[ X_1 X_1 \right] \right] + E \left[ \left[ X_1 \right] \right] E \left[ \left[ X_1 \right] \right] 
\]
\[ \leq 1; \]

\[ E \left[ n \left( \sup_{w \in \text{supp}(W)} h \hat{\chi}_V(\zeta, w) \right) \left( \sup_{w \in \text{supp}(W)} h \hat{\chi}_V^\dagger(\zeta, w) \right) \right] 
\]
\[ = E \left[ n \left( \sup_{w \in \text{supp}(W)} h (\hat{\chi}_V(\zeta, w) - \chi_V(\zeta, w)) \right) \left( \sup_{w \in \text{supp}(W)} h (\hat{\chi}_V^\dagger(\zeta, w) - \chi_V^\dagger(\zeta, w)) \right) \right] 
\]
\[ = E \left[ \sup_{w \in \text{supp}(W)} h (V e^{i\xi X_2} k_h(W - w) - \chi_V(\zeta, w)) \sup_{w \in \text{supp}(W)} h (V e^{-i\xi X_2} k_h(W - w) - \chi_V^\dagger(\zeta, w)) \right] 
\]

so that

\[ \left| E \left[ n \left( \sup_{w \in \text{supp}(W)} h \hat{\chi}_V(\zeta, w) \right) \left( \sup_{w \in \text{supp}(W)} h \hat{\chi}_V^\dagger(\zeta, w) \right) \right] \right| 
\]
\[
\leq E \left[ \sup_{w \in \text{supp}(W)} |V e^{i\xi X_2} h k_h(W - w) - h \chi V(\xi, w)| \right] \sup_{w \in \text{supp}(W)} |V e^{-i\xi X_2} h k_h(W - w) - h \chi_V(\xi, w)|
\]

\[
\leq E \left[ \left( \sup_{w \in \text{supp}(W)} |V e^{i\xi X_2} h k_h(W - w)| + |h \chi V(\xi, w)| \right) \sup_{w \in \text{supp}(W)} (|V e^{-i\xi X_2} h k_h(W - w)| + |h \chi_V(\xi, w)|) \right]
\]

\[
\leq E \left[ \left( \sup_{w \in \text{supp}(W)} |V e^{i\xi X_2} h k_h(W - w)| \right) + E \left[ \left( \sup_{w \in \text{supp}(W)} |V e^{-i\xi X_2} h k_h(W - w)| \right) \right] \right]
\]

\times \left( \sup_{w \in \text{supp}(W)} |V e^{i\xi X_2} h k_h(W - w)| + E \left[ \left( \sup_{w \in \text{supp}(W)} |V e^{-i\xi X_2} h k_h(W - w)| \right) \right] \right)
\]

\[
= E \left[ |V|^2 |e^{i(\xi - \bar{\zeta})X_2}| \sup_{w \in \text{supp}(W)} h k_h(W - w) \right]^2
\]

\[
+ 3E \left[ |V| |e^{i\xi X_2}| \sup_{w \in \text{supp}(W)} h k_h(W - w) \right] E \left[ |V| |e^{-i\xi X_2}| \sup_{w \in \text{supp}(W)} h k_h(W - w) \right]
\]

\[
\leq 1
\]

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

\[
\sup_{w \in \text{supp}(W)} |h k_h(w)| = \sup_{w \in \text{supp}(W)} \left| \frac{h}{2\pi} \int \kappa(h\zeta)e^{-i\xi\zeta}d\zeta \right| \leq \frac{h}{2\pi} \sup_{w \in \text{supp}(W)} \int |\kappa(h\zeta)||e^{-i\xi\zeta}|d\zeta
\]

\[
= \frac{h}{2\pi} \int |\kappa(h\zeta)|d\zeta = \frac{1}{2\pi} \int |\kappa(\zeta)|d\zeta = \frac{1}{2\pi} \int_{-1}^{1} |\kappa(\zeta)|d\zeta
\]

\[
\leq 1;
\]

Finally,

\[
E \left[ n \left( \sup_{w \in \text{supp}(W)} h \delta f_W(w) \right) \left( \sup_{w \in \text{supp}(W)} h \delta \tilde{f}_W(w) \right) \right]
\]

\[
= E \left[ n \left( \sup_{w \in \text{supp}(W)} h f_W(w) - f_W(w) \right) \left( \sup_{w \in \text{supp}(W)} h(\tilde{f}_W(w) - f_W(w)) \right) \right]
\]

\[
= E \left[ \left( \sup_{w \in \text{supp}(W)} h(k_h(W - w) - E[k_h(W - w)]) \right) \left( \sup_{w \in \text{supp}(W)} h(k_h(W - w) - E[k_h(W - w)]) \right) \right]
\]

so that

\[
\left| E \left[ n \left( \sup_{w \in \text{supp}(W)} h \delta f_W(w) \right) \left( \sup_{w \in \text{supp}(W)} h \delta \tilde{f}_W(w) \right) \right] \right|
\]

\[
\leq E \left[ \left( \sup_{w \in \text{supp}(W)} h k_h(W - w) \right)^2 \right] + E \left[ \left( \sup_{w \in \text{supp}(W)} h k_h(W - w) \right) \right] E \left[ \left( \sup_{w \in \text{supp}(W)} h k_h(W - w) \right) \right]
\]

\[
\leq 1.
\]

Thus we have
\[ \Omega_{V,\lambda}(x, w, h) \leq \left( \int \int |\Psi_{V,\lambda,1}(\zeta, x, w, h)| |\Psi_{V,\lambda,1}(\xi, x, w, h)| d\zeta d\xi \right)^{1/2} \\
+ \left\{ \int \int |\Psi_{V,\lambda,X_1}(\zeta, x, w, h)| |\Psi_{V,\lambda,X_1}(\xi, x, w, h)| d\zeta d\xi \right\}^{1/2} \\
+ \left\{ h^{-2} \int \int |\Psi_{V,\lambda,XV}(\zeta, x, w, h)| |\Psi_{V,\lambda,XV}(\xi, x, w, h)| d\zeta d\xi \right\}^{1/2} \\
+ \left\{ h^{-2} \int \int |\Psi_{V,\lambda,fw}(\zeta, x, w, h)| |\Psi_{V,\lambda,fw}(\xi, x, w, h)| d\zeta d\xi \right\}^{1/2} \right)^2 \\
= \left( \sum_{A=1,X_1} \int |\Psi_{V,\lambda,A}(\zeta, x, w, h)| d\zeta + h^{-1} \sum_{B=\chi_V,fw} \int |\Psi_{V,\lambda,B}(\zeta, x, w, h)| d\zeta \right)^2 \\
\leq \left( \sum_{A=1,X_1} \int \Psi_{V,\lambda,A}^+(\zeta, h) d\zeta + \sum_{B=\chi_V,fw} \int \Psi_{V,\lambda,B}^+(\zeta, h) d\zeta \right)^2 \\
= \left( \Psi_{V,\lambda}^+(h) \right)^2 ,
\]

where for \( A = 1, X_1, \chi_V, fw \)

\[
\Psi_{V,\lambda,A}^+(\zeta, h) \equiv \sup_{(x, w) \in \text{supp}(X, W)} |\Psi_{V,\lambda,A}(\zeta, x, w, h)| \\
\Psi_{V,\lambda}^+(h) \equiv \sum_{A=1,X_1} \int \Psi_{V,\lambda,A}^+(\zeta, h) d\zeta + h^{-1} \sum_{B=\chi_V,fw} \int \Psi_{V,\lambda,B}^+(\zeta, h) d\zeta \\
= O \left( (1 + h^{-1})^{\gamma_\phi + \lambda} \right) \exp \left( (\alpha_\phi 1_{\beta_\phi = \beta_\phi}) - \alpha_\phi (h^{-1})^{\beta_\phi} \right).
\]

Thus it follows that

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

with \( \alpha_L \equiv \alpha_\phi 1_{\{\beta_\phi = \beta_\phi\}} - \alpha_\phi, \beta_L \equiv \beta_\phi, \) and \( \gamma_{\lambda,L} \equiv 2 + \gamma_\phi - \gamma_\theta + \gamma_1 + \lambda. \)

To show uniform convergence,

\[
\sup_{(x, w) \in \text{supp}(X, W)} \left| \bar{\nu}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h) \right| \\
= \sup_{(x, w) \in \text{supp}(X, W)} \left\| \int \Psi_{V,\lambda,1}(\zeta, x, w, h) \left( \hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right) d\zeta \right\| \\
+ \Psi_{V,\lambda,X_1}(\zeta, x, w, h) \left( \hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}] \right) \\
+ \Psi_{V,\lambda,XV}(\zeta, x, w, h) \left( \hat{E}[V e^{i\zeta X_2} k_{h(W - w)}] - E[V e^{i\zeta X_2} k_{h(W - w)}] \right) \\
+ \Psi_{V,\lambda,fw}(\zeta, x, w, h) \left( \hat{E}[k_{h(W - w)}] - E[k_{h(W - w)}] \right) d\zeta \\
\leq \int \left\| \left( \sup_{(x, w) \in \text{supp}(X, W)} |\Psi_{V,\lambda,1}(\zeta, x, w, h)| \right) \left( \hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right) \right\| d\zeta,
\]

for all \( \zeta \in \mathbb{R}^d \) and \( \lambda, h > 0 \).
\[
\begin{align*}
&+ \left( \sup_{(x,w) \in \text{supp}(X,W)} |\Psi_{V,\lambda,X_1}(\zeta, x, w, h)| \right) \left| \hat{E}[X_1 e^{i \zeta} X_2] - E[X_1 e^{i \zeta} X_2] \right| \\
&+ \left( \sup_{(x,w) \in \text{supp}(X,W)} |\Psi_{V,\lambda,X_2}(\zeta, x, w, h)| \right) \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[Ve^{i \zeta} X_2 k_h(W - w)] - E[Ve^{i \zeta} X_2 k_h(W - w)] \right| \right) \\
&+ \left( \sup_{(x,w) \in \text{supp}(X,W)} |\Psi_{V,\lambda,f_w}(\zeta, x, w, h)| \right) \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right| \right) d\zeta \\
\end{align*}
\]

where the integrals are finite since \( \hat{E}[e^{i \zeta} X_2] - E[e^{i \zeta} X_2] \leq 1 \), \( \hat{E}[X_1 e^{i \zeta} X_2] - E[X_1 e^{i \zeta} X_2] \leq 1 \), and \( \sup_{w \in \text{supp}(W)} \left| \hat{E}[Ve^{i \zeta} X_2 k_h(W - w)] - E[Ve^{i \zeta} X_2 k_h(W - w)] \right| \leq 1 \), and since Lemma A.1 implies that \( \Psi_{V,\lambda}(h) < \infty \). Then we have

\[
E \left[ \sup_{(x,w) \in \text{supp}(X,W)} |\tilde{g}_{v,\lambda}(x, w, h) - g_{v,\lambda}(x, w, h)| \right] \\
\leq \int \left[ \Psi_{V,\lambda,1}(\zeta, h) E \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) E \left\{ \left( \hat{E}[X_1 e^{i \zeta} X_2] - E[X_1 e^{i \zeta} X_2] \right)^2 \right\}^{1/2} \\
+ h^{-1} \Psi_{V,\lambda,X_2}(\zeta, h) E \left\{ \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[Ve^{i \zeta} X_2 k_h(W - w)] - E[Ve^{i \zeta} X_2 k_h(W - w)] \right) \right)^2 \right\}^{1/2} \\
+ h^{-1} \Psi_{V,\lambda,f_w}(\zeta, h) E \left\{ \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) \right)^2 \right\}^{1/2} \right] d\zeta \\
\leq \int \left[ \Psi_{V,\lambda,1}(\zeta, h) E \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) E \left\{ \left( \hat{E}[X_1 e^{i \zeta} X_2] - E[X_1 e^{i \zeta} X_2] \right)^2 \right\}^{1/2} \\
+ h^{-1} \Psi_{V,\lambda,X_2}(\zeta, h) E \left\{ \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[Ve^{i \zeta} X_2 k_h(W - w)] - E[Ve^{i \zeta} X_2 k_h(W - w)] \right) \right)^2 \right\}^{1/2} \right] d\zeta \\
+ h^{-1} \Psi_{V,\lambda,f_w}(\zeta, h) E \left\{ \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) \right)^2 \right\}^{1/2} \right] d\zeta \\
\leq \int \left[ \Psi_{V,\lambda,1}(\zeta, h) \left\{ n^{-1} E \left( |e^{i \zeta} X_2 - E[e^{i \zeta} X_2]|^2 \right) \right\}^{1/2} \right] + \Psi_{V,\lambda,X_1}(\zeta, h) \left\{ n^{-1} E \left( |X_1 e^{i \zeta} X_2 - E[X_1 e^{i \zeta} X_2]|^2 \right) \right\}^{1/2} \\
+ h^{-1} \Psi_{V,\lambda,X_2}(\zeta, h) \left\{ n^{-1} E \left( \left| \sup_{w \in \text{supp}(W)} \left( Ve^{i \zeta} X_2 k_h(W - w) - E[Ve^{i \zeta} X_2 k_h(W - w)] \right) \right|^2 \right) \right\}^{1/2} \right] d\zeta \\
+ h^{-1} \Psi_{V,\lambda,f_w}(\zeta, h) \left\{ n^{-1} E \left( \left| \sup_{w \in \text{supp}(W)} \left( Ve^{i \zeta} X_2 k_h(W - w) - E[Ve^{i \zeta} X_2 k_h(W - w)] \right) \right|^2 \right) \right\}^{1/2} \right] d\zeta \\
\]
\[ + h^{-1} \Psi_{V,\lambda,f_{W}}(\zeta, h) \left\{ n^{-1} E \left( \sup_{w \in \text{supp}(W)} \left( h k_h(W - w) - E[h k_h(W - w)] \right)^2 \right) \right\}^{1/2} d\zeta \]

\[ = n^{-1/2} \int \left[ \Psi_{V,\lambda,1}(\zeta, h) \left\{ E \left( \left| e^{i\zeta X_2} - E[e^{i\zeta X_2}] \right|^2 \right)^{1/2} + \Psi_{V,\lambda,X_1}(\zeta, h) \left\{ E \left( \left| X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}] \right|^2 \right)^{1/2} \right\} \right. \right. \]

\[ + h^{-1} \Psi_{V,\lambda,X_1}(\zeta, h) \left\{ E \left( \sup_{w \in \text{supp}(W)} \left( V e^{i\zeta X_2} h k_h(W - w) - E[V e^{i\zeta X_2} h k_h(W - w)] \right)^2 \right)^{1/2} \right\} \]

\[ + h^{-1} \Psi_{V,\lambda,f_W}(\zeta, h) \left\{ E \left( \sup_{w \in \text{supp}(W)} \left( h k_h(W - w) - E[h k_h(W - w)] \right)^2 \right)^{1/2} \right\} \]

\[ \leq n^{-1/2} \left( \sum_{A=1,X_1} \int \Psi_{V,\lambda,A}(\zeta, h) d\zeta + h^{-1} \sum_{B=\chi_{V,f_{W}}} \int \Psi_{V,\lambda,B}(\zeta, h) d\zeta \right) \]

\[ = n^{-1/2} \Psi_{V,\lambda}(h), \]

where \( \Psi_{V,\lambda}(h) = O \left( (1 + h^{-1})^{\gamma_0 + \lambda + \gamma_1 - \gamma_0 + 2} \exp \left( (\alpha_0 1_{\{\beta_0 = \beta_0\}} - \alpha_0)(h^{-1})^{\beta_0} \right) \right) \). It follows that by Markov’s inequality

\[ \sup_{(x,w) \in \text{supp}(X,W)} |L_{V,\lambda}(x,w,h)| = \sup_{(x,w) \in \text{supp}(X,W)} |\hat{g}_{V,\lambda}(x,w,h) - g_{V,\lambda}(x,w,h)| \]

\[ = O_p \left( n^{-1/2} (1 + h^{-1})^{\gamma_0 + \lambda + \gamma_1 - \gamma_0 + 2} \exp \left( (\alpha_0 1_{\{\beta_0 = \beta_0\}} - \alpha_0)(h^{-1})^{\beta_0} \right) \right), \]

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

\[ \sum_{j=1}^{2} A_j P_{n,j}(X_2) + \sum_{k=1}^{2} B_k Q_{n,k}(W) \]

\[ \equiv \int \Psi_{V,\lambda,1}(\zeta, x, w, h) (e^{i\zeta X_2}) d\zeta + \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h) (X_1 e^{i\zeta X_2}) d\zeta \]

\[ + \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h) (V e^{i\zeta X_2} k_h(W - w)) d\zeta + \int \Psi_{V,\lambda,f_W}(\zeta, x, w, h) (k_h(W - w)) d\zeta; \]

with

\[ P_{n,1}(x_2) = \int \Psi_{V,\lambda,1}(\zeta, x, w, h) e^{i\zeta x_2} d\zeta, \]

\[ P_{n,2}(x_2) = \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h) e^{i\zeta x_2} d\zeta, \]

\[ Q_{n,1}(\tilde{w}) = \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h) e^{i\zeta X_2} k_h(\tilde{w} - w) d\zeta, \]

\[ Q_{n,2}(\tilde{w}) = \int \Psi_{V,\lambda,f_W}(\zeta, x, w, h) k_h(\tilde{w} - w) d\zeta, \]

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 \( N, \sup_{n>N} \Omega_{V,\lambda}(x,w,h) = \sup_{n>N} \text{var}[\{V_{V,\lambda}(x,w,h,V,X_1,X_2)\}] < \infty. \) 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 \( j = 1,2, \)
\[
\sup_{x_2 \in \text{supp}(X_2)} |D_{x_2} P_{n,j}(x_2)| = \sup_{x_2 \in \text{supp}(X_2)} \left| \int \zeta \Psi_{V,\lambda,j}(\zeta, x, w, h)e^{i\zeta x_2}d\zeta \right|
\leq \int \zeta \left| \Psi_{V,\lambda,j}(\zeta, x, w, h) \right|d\zeta
\leq h_n^{-1} \int_0^{h_n^{-1}} \Psi_{V,\lambda,j}(\zeta, h)d\zeta
\leq h_n^{-1} (1 + h_n^{-1})^{\gamma_0 + \gamma_1 - \gamma_0 + 3} \exp\left( (\alpha_0 1_{\{\beta_0 = \beta_0\}} - \theta_0) (h_n^{-1})^{\beta_0} \right).
\]

By Assumption 4.4, if \(\beta_0 \neq 0\), we have \(h_n^{-1} = O((\ln n)^{1/\beta_0 - \eta})\) for some \(\eta > 0\). Thus we have for \(j = 1, 2\)
\[
\sup_{x_2 \in \text{supp}(X_2)} |D_{x_2} P_{n,j}(x_2)| \leq (1 + (\ln n)^{1/\beta_0 - \eta})^{\gamma_0 + \gamma_1 - \gamma_0 + 3} \exp\left( (\alpha_0 1_{\{\beta_0 = \beta_0\}} - \theta_0) (\ln n)^{1/\beta_0 - \eta} \right).
\]

Because the right-hand side grows more slowly than any power of \(n\), we certainly have \(\sup_{x_2 \in \text{supp}(X_2)} |D_{x_2} P_{n,j}(x_2)| = O(n^{(3/2) - \eta})\) for \(j = 1, 2\). If \(\beta_0 = 0\), we have \(h_n^{-1} = O(n^{-\eta n^{(3/2)/(\gamma_0 + \gamma_1 - \gamma_0 + 3)}})\) for some \(\eta > 0\). Thus we have for \(j = 1, 2\)
\[
\sup_{x_2 \in \text{supp}(X_2)} |D_{x_2} P_{n,j}(x_2)| \leq (1 + n^{-\eta n^{(3/2)/(\gamma_0 + \gamma_1 - \gamma_0 + 3)}})^{\gamma_0 + \gamma_1 - \gamma_0 + 3}
\leq (1 + n^{-\eta n^{(3/2)}})
= O(n^{(3/2) - \eta}).
\]

Because the Fourier transform of \(D_x^k k_h(x)\) is \((-i\zeta)^k \kappa(h\zeta)\), we have
\[
|h^{\lambda+1}D_x^k k_h(x)| = \frac{h^{\lambda+1}}{2\pi} \left| \int (-i\zeta)^k \kappa(h\zeta)e^{-i\zeta x}d\zeta \right|
\leq \frac{h^{\lambda+1}}{2\pi} \int |\zeta|^k |\kappa(h\zeta)|d\zeta
= \frac{1}{2\pi} \int |\zeta|^k |\kappa(h\zeta)|d\zeta < \infty.
\]

Therefore we get
\[
\sup_{\tilde{w} \in \text{supp}(W)} |D_{\tilde{w}} Q_{n,1}(\tilde{w})| = \sup_{\tilde{w} \in \text{supp}(W)} \left| D_{\tilde{w}} \int \Psi_{V,\lambda,\chi,V}(\zeta, x, w, h)e^{i\zeta x_2}k_h(\tilde{w} - w)d\zeta \right|
= h^{-2} \int \left| \Psi_{V,\lambda,\chi,V}(\zeta, x, w, h) \right| |e^{i\zeta x_2}| \left( \sup_{\tilde{w} \in \text{supp}(W)} \left| h^2D_{\tilde{w}} k_h(\tilde{w} - w) \right| \right) d\zeta
\leq h^{-2} \int \left| \Psi_{V,\lambda,\chi,V}(\zeta, x, w, h) \right| d\zeta
= h^{-2} \int \left| \Psi_{V,\lambda,\chi,V}^+(\zeta, h) \right| d\zeta
= O\left( (1 + h^{-1})^{\gamma_0 + \gamma_1 - \gamma_0 + 3} \exp\left( (\alpha_0 1_{\{\beta_0 = \beta_0\}} - \theta_0) (h^{-1})^{\beta_0} \right) \right).
\]
Proof of Theorem 4.4

Lemma A.3 \textbf{Lemma A.3} \textit{Let }A\textit{ and }X_2\textit{ be random variables satisfying }E[|A|^2] < \infty \textit{ and }E[|A||X_2|] < \infty, \textit{ and let }
\{A_i, X_2,i\}_{i=1,..,n} \textit{ be a corresponding IID sample. Then for any }u, U \geq 0, \textit{ and }\epsilon > 0,
\sup_{\zeta \in [-U^n, U^n]} \left| E[A \exp(i\zeta X_2)] - E[A \exp(i\zeta X_2)] \right| = O_p(n^{-1/2+\epsilon}).

\textbf{Proof} \textbf{Proof} \textit{See Lemma 6 in Schennach (2004b).}

Proof of Theorem 4.4 \textbf{Proof of Theorem 4.4} \textit{By substituting eqn.(18)\textasciitilde(21) into }
\dot{g}_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, h)
= \frac{1}{2\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \left[ \frac{\dot{\theta}_V(\zeta, w)}{\theta_1(\zeta)} \exp \left( \int_0^\zeta \frac{i\dot{\theta} X_1(\xi)}{\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,

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

R_1 = \frac{1}{4\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) q_v(\zeta, w) q_1(w) \exp(Q_{X_1}(\zeta)) \exp \left( \delta \dot{Q}_{X_1}(\zeta) \right) \left( \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi \right)^2 d\zeta

R_2 = \frac{1}{4\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \delta \dot{q}_v(\zeta, w) q_1(w) \exp(Q_{X_1}(\zeta)) \exp \left( \delta \dot{Q}_{X_1}(\zeta) \right) \left( \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi \right)^2 d\zeta

R_3 = \frac{1}{4\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) q_v(\zeta, w) \delta \dot{q}_1(w) \exp(Q_{X_1}(\zeta)) \exp \left( \delta \dot{Q}_{X_1}(\zeta) \right) \left( \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi \right)^2 d\zeta

R_4 = \frac{1}{4\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \delta \dot{q}_v(\zeta, w) \delta \dot{q}_1(w) \exp(Q_{X_1}(\zeta)) \exp \left( \delta \dot{Q}_{X_1}(\zeta) \right) \left( \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi \right)^2 d\zeta

R_5 = \frac{1}{2\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) q_v(\zeta, w) q_1(w) \exp(Q_{X_1}(\zeta)) \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi d\zeta

R_6 = \frac{1}{2\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \delta \dot{q}_v(\zeta, w) q_1(w) \exp(Q_{X_1}(\zeta)) \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi d\zeta

R_7 = \frac{1}{2\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) q_v(\zeta, w) \delta \dot{q}_1(w) \exp(Q_{X_1}(\zeta)) \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi d\zeta

R_8 = \frac{1}{2\pi} \int \left( -i\zeta \right)^{\lambda} \kappa(h\zeta) \exp(-i\zeta x) \delta \dot{q}_v(\zeta, w) \delta \dot{q}_1(w) \exp(Q_{X_1}(\zeta)) \int_0^\zeta i\delta \dot{q}_{X_1}(\xi) d\xi d\zeta
We define $Y$ by assumption 4.2 and $\supp(W)$ by assumption 4.5, Lemma A.3 gives that for any $\epsilon > 0$,

$$
\sup_{w \in \supp(W)} \sup_{\zeta \in [h_n^{-1}, h_n^{-1}]} |\hat{\chi}_V(\zeta, w) - \chi_V(\zeta, w)|
$$

$$
= \sup_{w \in \supp(W)} \sup_{\zeta \in [h_n^{-1}, h_n^{-1}]} |\hat{E}[V_{k_n}(W - w) \exp(i\zeta X_2)] - E[V_{k_n}(W - w) \exp(i\zeta X_2)]|
$$

$$
= h_n^{-1} \sup_{w \in \supp(W)} |h_n k_n(W - w)| \sup_{\zeta \in [-\theta_n, \theta_n]} \sup_{\zeta \in [-h_n^{-1}, h_n^{-1}]} |\hat{E}[V \exp(i\zeta X_2)] - E[V \exp(i\zeta X_2)]|
$$

$$
= O_p(h_n^{-1/2 + \epsilon}).
$$

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

$$
\Upsilon(h_n) \equiv (1 + h_n^{-1}) \left( \sup_{\zeta \in [-h_n^{-1}, h_n^{-1}]} |D\ln \phi_1(\zeta)| \right)
$$

$$
\times \left( \max \left\{ \sup_{\zeta \in [-h_n^{-1}, h_n^{-1}]} \sup_{w \in \supp(W)} |\chi_V(\zeta, w)|^{-1}, \sup_{\zeta \in [-h_n^{-1}, h_n^{-1}]} |\theta_1(\zeta)|^{-1} \right\} \right)
$$

$$
= O \left( (1 + h_n^{-1})^{1+\gamma_1-\gamma_0} \exp \left( -\alpha \left( h_n^{-1}\right)^{\beta_0} \right) \right),
$$

51
\[ \Phi_n \equiv \max \left\{ \sup_{\zeta \in \left[ -h_n^{-1}, h_n^{-1} \right]} \left| \dot{\theta}_1(\zeta) - \theta_1(\zeta) \right|, \sup_{\zeta \in \left[ -h_n^{-1}, h_n^{-1} \right]} \left| \dot{\theta}_X(\zeta) - \theta_X(\zeta) \right|, \right. \\
\left. \sup_{w \in \text{supp}(W)} \sup_{\zeta \in \left[ -h_n^{-1}, h_n^{-1} \right]} \left| \dot{\chi}_V(\zeta, w) - \chi_V(\zeta, w) \right|, \sup_{w \in \text{supp}(W)} \left| \dot{f}_W(w) - f_W(w) \right| \right\} \\
= O_p \left( h_n^{-1} n^{-1/2 + \epsilon} \right) \\
\] for any \( \epsilon > 0 \). Note that the supremums associated with \( \zeta \) can be taken over \( \left[ -h_n^{-1}, h_n^{-1} \right] \) since \( \kappa(h_n, \zeta) \) vanishes outside the interval by Assumption 3.6 (ii). The second order of magnitude follows from Lemma A.3 and Assumption 4.6 since \( h_n^{-1} n^{-1/2 + \epsilon} = h_n^{-1/2} n^{-1/2} (n')^{-1/2} > h_n^{-1/2} n^{-1/2} (\ln n)^{1/2} + h_n^2 \) for any choices of \( h_n \) from Assumption 4.4 and 4.7. Then those terms in the nonlinear remainder can be bounded in terms of \( \Psi_{\nu, \lambda}^+(h_n), \Upsilon(h_n), \text{ and } \Phi_n \). We note that

\[ \Phi_n \times \left( \max \left\{ \sup_{\zeta \in \left[ -h_n^{-1}, h_n^{-1} \right]} \left| \dot{\chi}_V(\zeta, w) \right|^{-1}, \sup_{\zeta \in \left[ -h_n^{-1}, h_n^{-1} \right]} \left| \dot{\theta}_1(\zeta) \right|^{-1} \right\} \right) \\
\leq \Phi_n T(h_n) \\
= O_p \left( h_n^{-1} n^{-1/2 + \epsilon} \right) O \left( \left( 1 + h_n^{-1} \right)^{1 + \gamma_1 - \gamma_0} \exp \left( -\alpha_0 \left( h_n^{-1} \right)^{\beta_0} \right) \right) \\
= o_p(1). \]

We find upper bounds for each term, \( R_i, i = 1, ..., 22. \)
\[
\begin{align*}
&= \exp(o_p(1)) \Phi(h) \Phi_n^2 |1 + o_p(1)|^{-2} \int_0^\infty \left[ \int_\xi^\infty |\zeta| \kappa(h) \right] \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) d\zeta \frac{1}{|\theta_1(\xi)|} \\
&\quad + \int_\xi^\infty |\zeta| \kappa(h) \left( \sup_{w \in \text{supp}(W)} |\phi_V(\zeta, w)| \right) d\zeta \frac{1}{|\theta_1(\xi)|^2} d\xi \\
&= O_p(1) \Phi(h) \Phi_n^2 \Psi_{V,\lambda}^+(h) \\
&\leq \Psi(h) \Phi_n^2 \Psi_{V,\lambda}^+(h).
\end{align*}
\]

When the conditions of Theorem 4.3 hold, we have

\[
\sup_{(x,u) \in \text{supp}(X,W)} |R_1| = O_p \left( (1 + h_n^{-1})^{1+\gamma_\theta-\gamma_\eta} \exp \left( -\alpha_\theta (h_n^{-1})^{\beta_\theta} \right) h_n^{-1-2\epsilon} (h_n^{-1})^{\gamma_\lambda+\epsilon} \exp \left( \alpha_L (h_n^{-1})^{\beta_L} \right) \right)
\]

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

\[
\Psi(h) \Phi_n^2 \Psi_{V,\lambda}^+(h) = \left( \Psi(h) \Phi_n^2 n^{1/2} \right) n^{-1/2} \Psi_{V,\lambda}^+(h),
\]

\[
n^{-1/2} \Psi_{V,\lambda}^+(h) = O_p \left( n^{-1/2} (h_n^{1+\gamma_\lambda}) \exp \left( \alpha_L (h_n^{-1})^{\beta_L} \right) \right)
\]

where the second equality is obtained by Lemma A.1. Now we show that \(\Psi(h_n) \Phi_n^2 n^{1/2} = o_p(1)\). When \(\beta_\theta \neq 0\), we have \(h_n^{-1} \leq (\ln n)^{1/\beta_\theta-\eta} \) by the Assumption 4.7 so that

\[
\Psi(h_n) \Phi_n^2 n^{1/2} = \Psi(h_n) O_p \left( h_n^{-2} n^{-1+2\epsilon} \right) n^{1/2}
\]

\[
= O_p \left( (1 + h_n^{-1})^{3+\gamma_\lambda-\gamma_\eta} \exp \left( -\alpha_\theta (h_n^{-1})^{\beta_\theta} \right) n^{-1/2+2\epsilon} \right)
\]

\[
= O_p \left( (1 + (\ln n)^{1/\beta_\theta-\eta})^{3+\gamma_\lambda-\gamma_\eta} \exp \left( -\alpha_\theta (\ln n)^{1-\eta/\beta_\theta} \right) n^{-1/2+2\epsilon} \right)
\]

\[
= O_p \left( (\ln n)^{1/\beta_\theta-\eta} (3+\gamma_\lambda-I_{\text{dom}}) \exp \left( -\alpha_\theta (\ln n)^{1-\eta/\beta_\theta} \right) n^{-1/2+2\epsilon} \right)
\]

\[
= o_p(1),
\]

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 \(h_n^{-1} \leq n^{-\eta/2} \) so that

\[
\Psi(h_n) \Phi_n^2 n^{1/2} = \Psi(h_n) O_p \left( h_n^{-2} n^{-1+2\epsilon} \right) n^{1/2}
\]

\[
= O_p \left( (3+\gamma_\lambda) n^{-1/2+2\epsilon} \right)
\]

\[
= O_p \left( (n^{-\eta/2} n^{-2+2\epsilon}) \right)
\]

\[
= o_p(1),
\]

selecting \(\epsilon < \eta/2\).
Now we get the bounds for the remaining terms. Because they all contain the same leading term, $\Upsilon(h) \hat{\Phi}_n^\gamma \Psi_{\nu,\lambda}^\gamma(h)$, they can be similarly bounded:

\[
\sup_{(x, u) \in \text{supp}(X, W)} |R_2| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left| q_1(w) \right| \exp(Q_{X_1}(\zeta)) \exp \left( |\delta Q_{X_1}(\zeta)| \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left( 1 + \left| \frac{\chi_{\nu}(\zeta, w)}{\theta_1(\zeta)} \right| \right) \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| q_1(w) \exp(Q_{X_1}(\zeta)) \\
\times \exp \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left( 1 + \left| \frac{\chi_{\nu}(\zeta, w)}{\theta_1(\zeta)} \right| \right) q_1(w) \exp(Q_{X_1}(\zeta)) \\
\times \exp \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
+ \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left( \sup_{w \in \text{supp}(W)} |\phi_{\nu}(\zeta, w)| \right) \exp \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
\leq \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| \left( \sup_{(x, u) \in \text{supp}(X, W)} |R_1| \right) \\
= o_p(1) \left( \sup_{(x, u) \in \text{supp}(X, W)} |R_1| \right) ;
\]

\[
\sup_{(x, u) \in \text{supp}(X, W)} |R_3| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left| q_{\nu}(\zeta, w) \right| \delta q_1(w) \exp(Q_{X_1}(\zeta)) \exp \left( |\delta Q_{X_1}(\zeta)| \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left| q_{\nu}(\zeta, w) \right| \frac{1}{|f_{W}(w)|} |1 + o_p(1)|^{-1} \Upsilon(h) \hat{\Phi}_n \exp(Q_{X_1}(\zeta)) \\
\times \exp \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
= \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| \int_0^\infty |\zeta|^{\gamma(h\zeta)} \left( \sup_{w \in \text{supp}(W)} |\phi_{\nu}(\zeta, w)| \right) \frac{1}{2} \exp \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right) \left( \int_0^\zeta |\delta q_{X_1}(\xi)| d\xi \right)^2 d\zeta \\
\leq \Upsilon(h) \hat{\Phi}_n \left| 1 + o_p(1)^{-1} \right| \left( \sup_{(x, u) \in \text{supp}(X, W)} |R_1| \right) \\
= o_p(1) \left( \sup_{(x, u) \in \text{supp}(X, W)} |R_1| \right) ;
\]
\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_4| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^4 |\kappa(h\zeta)||\delta q_\nu(\zeta, w)||\delta \hat{q}_n(1 + o_p(1))|^{1 + o_p(1)}|q_1(w)| \exp(Q_X(\zeta)) \\
\left( \int_0^\zeta |\delta \hat{q}_n(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \frac{1}{\int_{[W(\zeta)]} |Q_1(\zeta)|} \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \\
\times \exp(Q_X(\zeta)) \frac{1}{2} \exp \left( \int_0^\zeta |\delta \hat{q}_n(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \left( \sup_{(x, w) \in \text{supp}(X, W)} |R_2| \right)
\]
\[
= o_p(1) \left( \sup_{(x, w) \in \text{supp}(X, W)} |R_2| \right);
\]
\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_5| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^4 |\kappa(h\zeta)||q_\nu(\zeta, w)||q_1(w)| \exp(Q_X(\zeta)) \left( \int_0^\zeta |\delta q_\nu(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
= \int_0^\infty |\zeta|^4 |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} \phi_\nu(\zeta, w) \right) \left( \int_0^\zeta |\delta q_\nu(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
\leq \int_0^\infty |\zeta|^4 |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} \phi_\nu(\zeta, w) \right) \left( \int_0^\zeta |\delta q_\nu(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
\leq \int_0^\infty |\zeta|^4 |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} \phi_\nu(\zeta, w) \right) \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \frac{1}{|\theta_1(\zeta)|^2} \left( \frac{|\theta_1(\zeta)|}{|\theta_1(\zeta)| + 1} \right) d\zeta
\]
\[
= \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \int_0^\infty |\zeta|^4 |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} \phi_\nu(\zeta, w) \right) \left( \frac{1}{|\theta_1(\zeta)|^2} \left( \frac{|\theta_1(\zeta)|}{|\theta_1(\zeta)| + 1} \right) \right) d\zeta
\]
\[
= \Upsilon(h) \hat{\Phi}_n y_n(h) (1 + o_p(1));
\]
\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_6| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^4 |\kappa(h\zeta)||\delta q_\nu(\zeta, w)||q_1(w)| \exp(Q_X(\zeta)) \left( \int_0^\zeta |\delta q_\nu(\zeta)| d\zeta \right)^2 d\zeta
\]
\[
= \sup_{w \in \text{supp}(W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( \frac{1}{|\theta_{1}(\zeta)|} + \frac{|\chi_{W}(\zeta, w)|}{|\theta_{1}(\zeta)|^2} \right) \Phi_{n}|1 + o_{p}(1)|^{-1} |q_{1}(w)| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
\leq \sup_{w \in \text{supp}(W)} \Psi(h, \Phi_{n}|1 + o_{p}(1)|^{-1} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( 1 + \frac{|\chi_{W}(\zeta, w)|}{|\theta_{1}(\zeta)|} \right) \left| q_{1}(w) \right| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
= \Psi(h, \Phi_{n}|1 + o_{p}(1)|^{-1} \left[ \sup_{w \in \text{supp}(W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( 1 + \frac{|\chi_{W}(\zeta, w)|}{|\theta_{1}(\zeta)|} \right) \left| q_{1}(w) \right| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
+ \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} |\phi_{W}(\zeta, w)| \right) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
\leq \Psi(h, \Phi_{n}|1 + o_{p}(1)| \left[ \sup_{(x, w) \in \text{supp}(X, W)} |R_{5}| \right]
\]
\[
= o_{p}(1) \left[ \sup_{(x, w) \in \text{supp}(X, W)} |R_{5}| \right]
\]
\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{7}|
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| |\chi_{W}(\zeta, w)| |\delta\dot{q}_{1}(w)| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
= \sup_{w \in \text{supp}(W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| |\chi_{W}(\zeta, w)| |\delta\dot{q}_{1}(w)| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
\Psi(h, \Phi_{n}|1 + o_{p}(1)|^{-1} \left[ \sup_{w \in \text{supp}(W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( 1 + \frac{|\chi_{W}(\zeta, w)|}{|\theta_{1}(\zeta)|} \right) \left| q_{1}(w) \right| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
\leq \Psi^{2}(h, \Phi_{n}|1 + o_{p}(1)|^{-2} \left[ \sup_{(x, w) \in \text{supp}(X, W)} \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( 1 + \frac{|\chi_{W}(\zeta, w)|}{|\theta_{1}(\zeta)|} \right) \left| q_{1}(w) \right| \exp(Q_{X_{1}}(\zeta)) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
+ \int_{0}^{\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| \left( \sup_{w \in \text{supp}(W)} |\phi_{W}(\zeta, w)| \right) \int_{0}^{\zeta} |\delta_{2}\dot{q}_{X_{1}}(\xi)| d\xi d\zeta
\]
\[
\leq \Psi^{2}(h, \Phi_{n}|1 + o_{p}(1)|^{-2} \left[ \sup_{(x, w) \in \text{supp}(X, W)} |R_{5}| \right]
\]
\[
= o_p(1) \left( \sup_{(x, w) \in \text{supp}(X, W)} |R_5| \right);
\]

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_9| 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| |\delta_1 \hat{g}_V(\zeta, w)| |q_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_{X_1}(\xi)| d\xi d\zeta 
= \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| \left( 1 + \frac{|\chi_V(\zeta, w)|}{|\theta_1(\zeta)|} \right) \Upsilon(h) \hat{\Phi}_n |q_1(w)| \exp(Q_{X_1}(\zeta))
\times \int_0^\zeta \left( 1 + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|} \right) \frac{1}{|\theta_1(\zeta)|} \hat{\Phi}_n d\xi d\zeta 
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \hat{\Phi}_n^2 \int_0^\infty \left[ \int_0^\zeta |\zeta|^\lambda |\kappa(h, \zeta)| \left( |q_1(w)| \exp(Q_{X_1}(\zeta)) + |\phi_V(\zeta, w)| \right) d\zeta \right] 
\times \left( 1 + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|} \right) \frac{1}{|\theta_1(\zeta)|} d\xi 
\leq \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V, \lambda}(h);
\]

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{10}| 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| |\delta_2 \hat{g}_V(\zeta, w)| |q_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_{X_1}(\xi)| d\xi d\zeta 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| \left( \frac{|\chi_V(\zeta, w)|}{|\theta_1(\zeta)|} \frac{1}{|\theta_1(\zeta)|^2} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} + \frac{1}{|\theta_1(\zeta)|^2} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \right) 
\times |q_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta \left( \frac{1}{|\theta_1(\zeta)|} + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|^2} \right) \hat{\Phi}_n d\xi d\zeta 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| \Upsilon(h) \hat{\Phi}_n [1 + o_p(1)]^{-1} \frac{1}{|\theta_1(\zeta)|} \left( \frac{|\chi_V(\zeta, w)|}{|\theta_1(\zeta)|} + 1 \right) |q_1(w)| \exp(Q_{X_1}(\zeta)) 
\times \int_0^\zeta \left( 1 + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|} \right) \Upsilon(h) \hat{\Phi}_n d\xi d\zeta 
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \hat{\Phi}_n \Upsilon(h) \hat{\Phi}_n^2 [1 + o_p(1)]^{-1} \left[ \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| \frac{1}{|\theta_1(\zeta)|} |\phi_V(\zeta, w)| \int_0^\zeta \left( 1 + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|} \right) d\xi d\zeta 
+ \int_0^\infty |\zeta|^\lambda |\kappa(h, \zeta)| \frac{|q_1(w)|}{|\theta_1(\zeta)|} \exp(Q_{X_1}(\zeta)) \int_0^\zeta \left( 1 + \frac{|\theta_X(\zeta)|}{|\theta_1(\zeta)|} \right) d\xi d\zeta \right] 
= o_p(1) \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V, \lambda}(h);
\]
\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{11}| 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) ||q_v(\xi, w)|| \delta_1 \dot{q}_1(w) |\exp(Q X_1(\xi))| \int_0^\xi |\delta_1 \dot{q}_1(\xi)| d\xi d\zeta 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) ||q_v(\xi, w)|| \frac{1}{|f_W(w)|} \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \exp(Q X_1(\xi)) 
\times \int_0^\xi \frac{1}{|\theta_1(\xi)|} \left(1 + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) \hat{\Phi}_n d\xi d\zeta 
= \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \int_0^\infty |\xi^\lambda| \kappa(h\xi) \left(\sup_{w \in \text{supp}(W)} |\phi_V(\xi, w)|\right) \int_0^\xi \frac{1}{|\theta_1(\xi)|} \left(1 + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta 
= \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V, \lambda}^+(h)(1 + o_p(1));
\]

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{12}| 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) ||\delta_1 \dot{q}_V(\xi, w)|| \delta_1 \dot{q}_1(w) |\exp(Q X_1(\xi))| \int_0^\xi |\delta_1 \dot{q}_1(\xi)| d\xi d\zeta 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\chi_V(\xi, w)|}{|\theta_1(\xi)|^2}\right) \hat{\Phi}_n \frac{1}{|f_W(w)|^2} \hat{\Phi}_n \exp(Q X_1(\xi)) 
\times \int_0^\xi \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|^2}\right) \hat{\Phi}_n d\xi d\zeta 
= \Upsilon^2(h) \hat{\Phi}_n^3 \left[\sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) \left(\frac{1}{|f_W(w)|}\right) \exp(Q X_1(\xi)) \int_0^\xi \frac{1}{|\theta_1(\xi)|} \left(1 + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta 
+ \int_0^\infty |\xi^\lambda| \kappa(h\xi) \left(\sup_{w \in \text{supp}(W)} |\phi_V(\xi, w)|\right) \int_0^\xi \frac{1}{|\theta_1(\xi)|} \left(1 + \frac{|\theta_{X_1}(\xi)|}{|\theta_1(\xi)|}\right) d\xi d\zeta\right] 
= o_p(1) \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V, \lambda}^+(h);
\]

\[
\sup_{(x, w) \in \text{supp}(X, W)} |R_{13}| 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) ||\delta_2 \dot{q}_V(\xi, w)|| \delta_1 \dot{q}_1(w) |\exp(Q X_1(\xi))| \int_0^\xi |\delta_1 \dot{q}_1(\xi)| d\xi d\zeta 
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\xi^\lambda| \kappa(h\xi) \left(\frac{|\chi_V(\xi, w)|}{|\theta_1(\xi)|^2}\right) \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1}
\]
\[
\begin{aligned}
&\quad + \frac{\vert \delta \hat{\chi}_v (\zeta, w) \vert}{\vert \hat{\Theta}_1 (\zeta) \vert} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \int_0^\zeta \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&= \sup_{w \in \text{supp}(W)} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \times \left( \frac{\vert \hat{\Theta}_1 (\zeta) \vert}{\vert \hat{\Theta}_1 (\zeta) \vert^2} + \frac{\vert \delta \hat{\chi}_v (\zeta, w) \vert}{\vert \hat{\Theta}_1 (\zeta) \vert} \right) \\
&\quad \times \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \leq \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{12} \right| \right) \\
&= o_p(1) \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{11} \right| \right) \\
&\quad \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{14} \right| \\
&\quad \leq \sup_{w \in \text{supp}(W)} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \leq \sup_{w \in \text{supp}(W)} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \frac{\hat{f}_W (w)}{\hat{f}_W (w)} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&= \sup_{w \in \text{supp}(W)} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \leq \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{11} \right| \right) \\
&= o_p(1) \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{11} \right| \right) \\
&\quad \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{15} \right| \\
&\quad \leq \sup_{w \in \text{supp}(W)} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \leq \sup_{w \in \text{supp}(W)} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \frac{\hat{f}_W (w)}{\hat{f}_W (w)} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&= \sup_{w \in \text{supp}(W)} \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert^{-1} \int_0^\infty \langle \chi (\zeta, w) \hat{\Phi}_n \rangle \exp (Q X_1 (\zeta)) \int_0^\zeta \delta_1 \hat{q}_1 (\zeta) \mathrm{d} \zeta \mathrm{d} \zeta \\
&\quad \leq \chi (h) \hat{\Phi}_n \vert 1 + o_p(1) \vert \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{12} \right| \right) \\
&= o_p(1) \left( \sup_{(x, w) \in \text{supp}(X, W)} \left| R_{12} \right| \right)
\end{aligned}
\]
\[
\sup_{(x,w) \in \text{supp}(X,W)} |R_{16}| \leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_X(\zeta)| d\xi d\zeta
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_X(\zeta)| d\xi d\zeta
\]
\[
+ \frac{|\delta_2 \hat{q}_V(\zeta,w)|}{|\theta_1(\zeta)|} Y(h) \hat{\Phi}_n[1 + o_p(1)]^{-1} |\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_X(\zeta)| d\xi d\zeta
\]
\[
= \sup_{w \in \text{supp}(W)} Y(h) \hat{\Phi}_n[1 + o_p(1)]^{-1} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) \int_0^\zeta |\delta_1 \hat{q}_X(\zeta)| d\xi d\zeta
\]
\[
\leq Y(h) \hat{\Phi}_n[1 + o_p(1)] \left( \sup_{(x,w) \in \text{supp}(X,W)} |R_{15}| \right)
\]
\[
o_p(1) \left( \sup_{(x,w) \in \text{supp}(X,W)} |R_{15}| \right);
\]

\[
\sup_{(x,w) \in \text{supp}(X,W)} |R_{17}|
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta
\]
\[
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta
\]
\[
= Y(h) \hat{\Phi}_n[1 + o_p(1)]^{-1} \left( \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \right)
\]
\[
+ \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)||\delta_2 \hat{q}_V(\zeta,w)||q_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta
\]
\[
\leq Y(h) \hat{\Phi}_n^2 \Psi_{V,\lambda}(h); \]
\[
\sup_{(x,w) \in \text{supp}(X,W)} |R_{19}| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_2 \hat{q}_V(\zeta,w)| |\delta_1 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| \left( \frac{|x_V(\zeta,w)| |\delta \hat{q}_1(\zeta)|}{|\theta_1(\zeta)|^2} \right) \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \\
+ \frac{|\delta x_V(\zeta,w)|}{|\theta_1(\zeta)|} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} |\delta_1 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_2 \hat{q}_V(\zeta,w)| |\delta_1 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
\leq \Upsilon(h) \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); \\
\sup_{(x,w) \in \text{supp}(X,W)} |R_{20}| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |q_V(\zeta,w)| |\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |q_V(\zeta,w)| \frac{1}{|f_W(\zeta)|^2} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \exp(Q_{X_1}(\zeta)) d\zeta \\
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \int_0^\infty |\zeta| |\kappa(h\zeta)| \frac{1}{|f_W(\zeta)|^2} |\delta q_V(\zeta,w)| d\zeta \\
\leq \Upsilon(h) \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); \\
\sup_{(x,w) \in \text{supp}(X,W)} |R_{21}| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_1 \hat{q}_V(\zeta,w)| |\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_1 \hat{q}_V(\zeta,w)| \frac{|\delta f_W(\zeta)|}{|f_W(\zeta)|^2} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \exp(Q_{X_1}(\zeta)) d\zeta \\
= \sup_{w \in \text{supp}(W)} \Upsilon(h) \Phi_n |1 + o_p(1)|^{-1} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_1 \hat{q}_V(\zeta,w)| |\delta_1 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta \\
\leq \Upsilon(h) \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); \\
\sup_{(x,w) \in \text{supp}(X,W)} |R_{22}| \\
\leq \sup_{w \in \text{supp}(W)} \int_0^\infty |\zeta| |\kappa(h\zeta)| |\delta_2 \hat{q}_V(\zeta,w)| |\delta_2 \hat{q}_1(w)| \exp(Q_{X_1}(\zeta)) d\zeta;
\]
If Assumption 4.9 also holds, then for Assumption 4.11, the result immediately follows from Theorem 4.3, Theorem 4.4(ii). We obtain rates for each term of Lemma A.4. Suppose the conditions of Lemma 4.1 hold. For each Proof of Corollary 4.6. Combining Theorem 4.2, Theorem 4.3 and Theorem 4.4(ii) immediately yields the result.

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(ii) and the fact that $\hat{g}_{V,\lambda}(x, w, h_{n}) - g_{V,\lambda}(x, w) = B_{V,\lambda}(x, w, h_{n}) + L_{V,\lambda}(x, w, h_{n}) + R_{V,\lambda}(x, w, h_{n})$. □

Lemma A.4 Suppose the conditions of Lemma 4.1 hold. For each $\zeta$ and $h$, and for $A = 1, X, \chi_{V}, f_{W}$, let $\Psi^{+}_{V,\lambda, A, s}(\zeta, h) \equiv \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda, A}(\zeta, x, w, h) s(x, w) dw \right|$, and define $\Psi^{+}_{V,\lambda, s}(h) \equiv \sum_{A=1, X} \int \Psi^{+}_{V,\lambda, A, s}(\zeta, h) d\zeta + h^{-1} \sum_{B=\chi_{V}, f_{W}} \int \Psi^{+}_{V,\lambda, B, s}(\zeta, h) d\zeta$.

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

$$\Psi^{+}_{V,\lambda, s}(h) = O \left( (1 + h^{-1})^{\gamma_{\phi_{s}} + \lambda + \gamma_{\gamma} - \gamma_{\theta} + 2} \exp \left( \alpha_{\phi_{s}} 1_{\{\beta_{\phi_{s}} \geq \beta_{\theta}\}} - \alpha_{\theta} 1_{\{\beta_{\theta} \leq \beta_{\phi_{s}}\}} (h^{-1})^{\max\{\beta_{\phi_{s}}, \beta_{\theta}\}} \right) \right).$$

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

$$\Psi^{+}_{V,\lambda, 1, s}(\zeta, h) \equiv \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda, 1}(\zeta, x, w, h) s(x, w) dw \right|$$

$$= \sup_{x \in \text{supp}(X)} \left| \int \left( -\frac{1}{2\pi} \frac{i\theta X_{1}(\zeta)}{\theta_{1}(\zeta)} \right)^{2} \int_{\zeta}^{\pm\infty} (-i\xi)^{\lambda} \kappa(h\xi) \exp(-i\xi x) \phi_{V}(\zeta, w) d\xi \right|$$

$$= \sup_{x \in \text{supp}(X)} \left| \frac{1}{2\pi} \frac{i\theta X_{1}(\zeta)}{\theta_{1}(\zeta)} \right|^{2} \int_{\zeta}^{\pm\infty} (-i\xi)^{\lambda} \kappa(h\xi) \exp(-i\xi x) \left( \int \phi_{V}(\zeta, w) s(x, w) dw \right) d\xi$$

$$= \frac{\theta X_{1}(\zeta)}{\theta_{1}(\zeta)} \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_{V}(\zeta, w) s(x, w) dw \right| \right)$$

$$+ \frac{1}{\theta_{1}(\zeta)} \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_{V}(\zeta, w) s(x, w) dw \right| \right)$$

62
By Assumption 4.1 and 4.9, we obtain

\[
\begin{align*}
&= \frac{1}{|\theta_1(\zeta)|} \left[ |D_\zeta \ln \phi_1(\zeta)| \left( |\xi|^\lambda |\kappa(h\xi)| \left( \sup_{x \in \text{supp}(X)} \left| \int X \phi_V(\xi, w)s(x, w)dw \right| \right) d\xi \\
&\quad + |\xi|^\lambda |\kappa(h\xi)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w)s(x, w)dw \right| \right) \right] \\
&\leq \frac{1}{|\theta_1(\zeta)|} \left[ |D_\zeta \ln \phi_1(\zeta)| \left( |\xi|^\lambda 1_{|\xi| \leq \tilde{h}^{-1}} \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w)s(x, w)dw \right| \right) d\xi \\
&\quad + |\xi|^\lambda 1_{|\xi| \leq \tilde{h}^{-1}} \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w)s(x, w)dw \right| \right) \right] \\
&\leq \frac{1}{|\theta_1(\zeta)|} 1_{|\xi| \leq \tilde{h}^{-1}} \left[ |D_\zeta \ln \phi_1(\zeta)| \int_0^{h^{-1}} |\xi|^\lambda \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w)s(x, w)dw \right| \right) d\xi \\
&\quad + |\xi|^\lambda \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w)s(x, w)dw \right| \right) \right].
\end{align*}
\]

Second, 

\[
\Psi^+_V,\lambda,\chi, s(\zeta, h) = \sup_{x \in \text{supp}(X)} \left| \int \Psi_V,\lambda,\chi_1(\zeta, x, w, h)s(x, w)dw \right|
\]

\[
= \sup_{x \in \text{supp}(X)} \left| \int \left( \frac{1}{2\pi |\theta_1(\zeta)|} i^{-\frac{1}{2}} \right) \left( -i\xi \right)^\lambda \kappa(h\xi) \exp(-i\xi \phi_V(\zeta, w)d\xi) s(x, w)dw \right|
\]

63
so that

\[
\int \Psi_{V,\lambda, X_1,s}(\zeta, h) d\zeta \leq \int_0^{h^{-1}} (1 + |\zeta|)^{-\gamma_0} \exp (-\alpha_\phi |\zeta|^{\beta_0}) \left( \int_0^{h^{-1}} |\zeta|^{\lambda (1 + |\zeta|)^{\gamma_0}} \exp \left( \frac{1}{\alpha_\phi s} \xi \frac{|\zeta|^{\beta_0}}{\chi_n(\zeta, w)} \right) d\zeta \right) d\zeta
\]

\[
\leq (1 + h^{-1})^{1 - \gamma_0} \exp (-\alpha_\phi (h^{-1})^{\beta_0}) (1 + h^{-1})^{\lambda + \gamma_0 + 1} \exp \left( \frac{1}{\alpha_\phi s} (h^{-1})^{\beta_0} \right)
\]

Third,

\[
\Psi_{V,\lambda, X_1,s}(\zeta, h) = \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda, X_1}(\zeta, x, h) s(x, w) d\zeta \right|
\]

\[
\leq \left| \zeta \right|^{1 \{ |\zeta| \leq h^{-1} \}} \left( \sup_{x \in \text{supp}(X)} \left| \int \frac{\phi_V(\zeta, w)}{\chi_n(\zeta, w)} s(x, w) d\zeta \right| \right)
\]

so that

\[
h^{-1} \int \Psi_{V,\lambda, X_1,s}(\zeta, h) d\zeta \leq h^{-1} \int_0^{h^{-1}} \left| \zeta \right|^{1 \{ |\zeta| \leq h^{-1} \}} \left( \sup_{x \in \text{supp}(X)} \left| \int \frac{\phi_V(\zeta, w)}{\chi_n(\zeta, w)} s(x, w) d\zeta \right| \right) \exp \left( \frac{1}{\alpha_\phi s} \xi \frac{|\zeta|^{\beta_0}}{\chi_n(\zeta, w)} \right) d\zeta
\]

\[
\leq (1 + h^{-1})^{\gamma_0 + \lambda + 1} \exp \left( \frac{1}{\alpha_\phi s} (h^{-1})^{\beta_0} \right)
\]

Finally,

\[
\Psi_{V,\lambda, f,w,s}(\zeta, h) = \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda, f,w}(\zeta, x, h) s(x, w) d\zeta \right|
\]

\[
\leq \left| \zeta \right|^{1 \{ |\zeta| \leq h^{-1} \}} \left( \sup_{x \in \text{supp}(X)} \left| \int \frac{\phi_V(\zeta, w)}{f_w(w)} s(x, w) d\zeta \right| \right)
\]
Thus the remainder term in eqn.(9) is
\[
\| \sum_{j=1}^{J} \left( \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right) s_{j}(x, w) \| d\zeta \leq (1 + h^{-1})^{\gamma_{o} + \lambda + 2} \exp(\alpha_{e_{x}}(h^{-1})^{\beta_{o}}) .
\]

Putting four terms together gives the desired result. \( \Box \)

\textbf{Proof of Theorem 4.7}  
(i) By the assumption 4.10(i), we have
\[
\sup_{j=1, \ldots, J} \sup_{(x, w) \in \text{supp}(X, W)} \left| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right| \leq o(\alpha_{1n}) + o_{p}(\alpha_{1n})^{1/2} + o_{p}(\alpha_{1n}) \leq o_{p}(\alpha_{1n}^{1/2}).
\]
Thus the remainder term in eqn.(9) is \( o_{p}\left(\left(\alpha_{1n}^{1/2}\right)^{2}\right) = o_{p}(\alpha_{1n}) \) by letting \( \hat{g}_{V_{j}, \lambda_{j}}(x, w) = \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) \). We also have
\[
\left| \sum_{j=1}^{J} \left( \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right) s_{j}(x, w) \right| \leq \sum_{j=1}^{J} \sup_{(x, w) \in \text{supp}(X, W)} \left| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right| \sup_{x \in \text{supp}(X)} \int \left| s_{j}(x, w) \right| d\zeta
\]
\[
\leq \sum_{j=1}^{J} \sup_{(x, w) \in \text{supp}(X, W)} \left| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right| \sup_{x \in \text{supp}(X)} \int \left| s_{j}(x, w) \right| d\zeta \leq \left\| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right\|_{\infty} \sup_{x \in \text{supp}(X)} \int \left| s_{j}(x, w) \right| d\zeta \leq \left\| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right\|_{\infty} .
\]

since \( \sup_{x \in \text{supp}(X)} \int \left| s_{j}(x, w) \right| d\zeta < \infty \). Then the result immediately follows.

(ii) By the assumption 4.10(ii), we have
\[
\sup_{j=1, \ldots, J} \sup_{(x, w) \in \text{supp}(X, W)} \left| \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) - g_{V_{j}, \lambda_{j}}(x, w, h_{n}) \right| \leq o(\alpha_{2n}) + o_{p}(\alpha_{2n})^{1/2} + o_{p}(\alpha_{2n}) \leq o_{p}(\alpha_{2n}^{1/2}).
\]
Thus the remainder term in eqn.(9) is \( o_{p}\left(\left(\alpha_{2n}^{1/2}\right)^{2}\right) = o_{p}(\alpha_{2n}) \) by letting \( \hat{g}_{V_{j}, \lambda_{j}}(x, w) = \hat{g}_{V_{j}, \lambda_{j}}(x, w, h_{n}) \). We also have

65
\[
\sum_{j=1}^{J} \int \left( g_{V,\lambda_j}(x, w, h_n) - g_{V,\lambda_j}(x, w) \right) s_j(x, w) dw
\]

\[
= \sum_{j=1}^{J} \int B_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw + \sum_{j=1}^{J} \int L_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw + \sum_{j=1}^{J} \int R_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw.
\]

For the first term,

\[
\sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int B_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw \right| \leq \sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int B_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw \right|.
\]

Note that

\[
\sup_{x \in \text{supp}(X)} \left| \int B_{V,\lambda}(x, w, h_n) s(x, w) dw \right|
\]

\[
= \sup_{x \in \text{supp}(X)} \left| \int (g_{V,\lambda}(x, w, h) - g_{V,\lambda}(x, w, 0)) s(x, w) dw \right|
\]

\[
= \sup_{x \in \text{supp}(X)} \left| \int \left( \frac{1}{2\pi} \int \kappa(h) (\exp(-i\zeta x) - 1) s(x, w) dw \right) d\zeta \right|
\]

\[
\leq \frac{1}{\pi} \int_{\xi/h}^{\infty} \left| (\kappa(h) - 1) \right| \left| \zeta \right|^\lambda \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) dw \right| \right) d\zeta
\]

\[
\leq \int_{\xi/h}^{\infty} \left| \zeta \right|^\lambda \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) dw \right| \right) d\zeta
\]

\[
\leq \int_{\xi/h}^{\infty} \left| \zeta \right|^\lambda \left( 1 + \left| \zeta \right| \right)^{\gamma_{\phi_s}} \exp(\alpha_{\phi_s} \left| \zeta \right|^\beta_{\phi_s}) d\zeta
\]

\[
= O \left( \left( \frac{\xi}{h} \right)^{\gamma_{\phi_s} + \lambda + 1} \exp \left( \alpha_{\phi_s} \left( \frac{\xi}{h} \right)^{\beta_{\phi_s}} \right) \right)
\]

\[
= O \left( (h^{-1})^{\gamma_{\lambda, B, s}} \exp \left( \alpha_{B, s} (h^{-1})^{\beta_{B, s}} \right) \right).
\]

Thus we have

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

For the second term,

\[
\sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int L_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw \right| \leq \sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int L_{V,\lambda_j}(x, w, h_n) s_j(x, w) dw \right|.
\]
Note that
\[
\sup_{x \in \text{supp}(X)} \left| \int L_{V,\lambda}(x, w, h_n)s(x, w)dw \right| \\
= \sup_{x \in \text{supp}(X)} \left| \int \int \Psi_{V,\lambda,1}(\zeta, x, w, h) \left( \hat{E}[e^{\iota \zeta X_2}] - E[e^{\iota \zeta X_2}] \right) \\
+ \Psi_{V,\lambda,X_1}(\zeta, x, w, h) \left( \hat{E}[X_1e^{\iota \zeta X_2}] - E[X_1e^{\iota \zeta X_2}] \right) \\
+ \Psi_{V,\lambda,X_1}(\zeta, x, w, h) \left( \hat{E}[Ve^{\iota \zeta X_2}k_h(W - w)] - E[Ve^{\iota \zeta X_2}k_h(W - w)] \right) \\
+ \Psi_{V,\lambda,f_w}(\zeta, x, w, h) \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) \right| d\zeta(x, w)dw \\
\leq \int \left[ \left( \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda,1}(\zeta, x, w, h)s(x, w)dw \right| \left| \hat{E}[e^{\iota \zeta X_2}] - E[e^{\iota \zeta X_2}] \right| \right) \\
+ \left( \sup_{x \in \text{supp}(X)} \left| \int \Psi_{V,\lambda,X_1}(\zeta, x, w, h)s(x, w)dw \right| \left| \hat{E}[X_1e^{\iota \zeta X_2}] - E[X_1e^{\iota \zeta X_2}] \right| \right) \\
\times \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[Ve^{\iota \zeta X_2}k_h(W - w)] - E[Ve^{\iota \zeta X_2}k_h(W - w)] \right| \right) \right] d\zeta \\
= \int \left[ \Psi_{V,\lambda,1,s}(\zeta, h) \left| \hat{E}[e^{\iota \zeta X_2}] - E[e^{\iota \zeta X_2}] \right| + \Psi_{V,\lambda,X_1,s}(\zeta, h) \left| \hat{E}[X_1e^{\iota \zeta X_2}] - E[X_1e^{\iota \zeta X_2}] \right| \\
+ h^{-1} \Psi_{V,\lambda,X_1,s}(\zeta, h) \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[Ve^{\iota \zeta X_2}k_h(W - w)] - E[Ve^{\iota \zeta X_2}k_h(W - w)] \right| \right) \\
+ h^{-1} \Psi_{V,\lambda,f_w,s}(\zeta, h) \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right| \right) \right] d\zeta.
\]

Then we have
\[
E \left[ \sup_{x \in \text{supp}(X)} \left| \int L_{V,\lambda}(x, w, h_n)s(x, w)dw \right| \right] \\
\leq \int \left[ \Psi_{V,\lambda,1,s}(\zeta, h)E \left\{ \left| \hat{E}[e^{\iota \zeta X_2}] - E[e^{\iota \zeta X_2}] \right|^2 \right\}^{1/2} \right] \\
+ \Psi_{V,\lambda,X_1,s}(\zeta, h)E \left\{ \left| \hat{E}[X_1e^{\iota \zeta X_2}] - E[X_1e^{\iota \zeta X_2}] \right|^2 \right\}^{1/2} \right] \\
+ h^{-1} \Psi_{V,\lambda,X_1,s}(\zeta, h)E \left\{ \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[Ve^{\iota \zeta X_2}k_h(W - w)] - E[Ve^{\iota \zeta X_2}k_h(W - w)] \right| \right)^2 \right\}^{1/2} \right] \\
+ h^{-1} \Psi_{V,\lambda,f_w,s}(\zeta, h)E \left\{ \left( \sup_{w \in \text{supp}(W)} \left| \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right| \right)^2 \right\}^{1/2} \right] d\zeta.
\]
\[
\begin{aligned}
\leq & \int \left[ \Psi_{V,\lambda,s}^+ (\zeta, h) \left\{ E \left( \left( \hat{E}[e^{i\zeta X_2} - E[e^{i\zeta X_2}]]^2 \right) \right) \right\}^{1/2} \\
+ & \Psi_{V,\lambda,s}^+ (\zeta, h) \left\{ E \left( \left( \hat{E}[X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}]]^2 \right) \right) \right\}^{1/2} \\
+ & h^{-1} \Psi_{V,\lambda,v,s}^+ (\zeta, h) \left\{ E \left( \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[V e^{i\zeta X_2} h_k h(W - w)] - E[V e^{i\zeta X_2} h_k h(W - w)] \right) \right)^2 \right) \right\}^{1/2} \\
+ & h^{-1} \Psi_{V,\lambda,f_w,s}^+ (\zeta, h) \left\{ E \left( \left( \sup_{w \in \text{supp}(W)} \left( \hat{E}[h_k (W - w)] - E[h_k (W - w)] \right) \right)^2 \right) \right\}^{1/2} d\zeta \\
= & n^{-1/2} \int \left[ \Psi_{V,\lambda,s}^+ (\zeta, h) \left\{ E \left( \left( e^{i\zeta X_2} - E[e^{i\zeta X_2}] \right)^2 \right) \right\}^{1/2} \\
+ & \Psi_{V,\lambda,s}^+ (\zeta, h) \left\{ E \left( \left( X_1 e^{i\zeta X_2} - E[X_1 e^{i\zeta X_2}] \right)^2 \right) \right\}^{1/2} \\
+ & h^{-1} \Psi_{V,\lambda,v,s}^+ (\zeta, h) \left\{ E \left( \left( \sup_{w \in \text{supp}(W)} \left( V e^{i\zeta X_2} h_k h(W - w) - E[V e^{i\zeta X_2} h_k h(W - w)] \right) \right)^2 \right) \right\}^{1/2} \\
+ & h^{-1} \Psi_{V,\lambda,f_w,s}^+ (\zeta, h) \left\{ E \left( \left( \sup_{w \in \text{supp}(W)} \left( h_k (W - w) - E[h_k (W - w)] \right) \right)^2 \right) \right\}^{1/2} d\zeta \\
\leq & n^{-1/2} \left( \sum_{A=1}^{X_1} \int \Psi_{V,\lambda,s}^+ (\zeta, h) d\zeta + h^{-1} \sum_{B=\lambda,v,f_w} \int \Psi_{V,\lambda,B,s}^+ (\zeta, h) d\zeta \right) \\
= & n^{-1/2} \Psi_{V,\lambda,s}^+ (h),
\end{aligned}
\]

where \( \Psi_{V,\lambda,s}^+ (h) = O \left( (1 + h^{-1})^{\gamma_{\alpha\phi} + \gamma_1 - \gamma_0 + 2} \exp \left( (\alpha_{\phi} s_1 \{ \beta_{\phi} \geq \beta_0 \} - \alpha_1 s_0 \{ \beta_{\phi} \leq \beta_0 \}) (h^{-1})^{\max(\beta_0, \beta_{\phi,1})} \right) \right. \) It follows by Markov’s inequality that

\[
\sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int_{L_{V,j,\lambda}} s_j(x, w) dw \right| = O_p \left( n^{-1/2} (1 + h^{-1})^{\gamma_{\alpha\phi} + \gamma_1 - \gamma_0 + 2} \exp \left( (\alpha_{\phi} s_1 \{ \beta_{\phi} \geq \beta_0 \} - \alpha_1 s_0 \{ \beta_{\phi} \leq \beta_0 \}) (h^{-1})^{\max(\beta_0, \beta_{\phi,1})} \right) \right).
\]

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

\[
\sup_{x \in \text{supp}(X)} \left| \int R_1 s(x, w) \, dw \right| \\
\leq \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) \, dw \right| \right) \exp \left( \int_0^\infty |\delta \tilde{Q}_1(\zeta)| \, d\zeta \right) \theta_1(\zeta) d\zeta \\
\leq \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) \, dw \right| \right) \exp \left( \int_0^\infty |\delta \tilde{q}_1(\zeta)| \, d\zeta \right) \left( \int_0^\infty |\delta \tilde{q}_1(\zeta)| \, d\zeta \right) d\zeta \\
\leq \exp(\gamma(p)) \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) \, dw \right| \right) \left( \int_0^\infty |\delta \tilde{q}_1(\zeta)| \, d\zeta \right) \left( \int_0^\infty |\delta \tilde{q}_1(\zeta)| \, d\zeta \right) d\zeta \\
\leq \exp(\gamma(p)) \int_0^\infty |\zeta|^\lambda |\kappa(h\zeta)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) \, dw \right| \right) \\
\times \int_0^\infty \left| \frac{\delta \tilde{q}_1(\zeta)}{\theta_1(\zeta)} - \frac{\theta_1(\zeta) \delta \tilde{q}_1(\zeta)}{(\theta_1(\zeta))^2} \right| \left( 1 + \frac{\delta \tilde{q}_1(\zeta)}{\theta_1(\zeta)} \right)^{-1} \left( 1 + \frac{\theta_1(\zeta)}{\theta_1(\zeta)} \right)^{-1} \, d\zeta \\
\leq \exp(\gamma(p)) \Upsilon(h) \tilde{\Phi}_n^2 |\zeta| |\kappa(h\zeta)| \left( \sup_{x \in \text{supp}(X)} \left| \int \phi_V(\zeta, w) s(x, w) \, dw \right| \right) \\
\times \int_0^\infty \left( \frac{1}{\theta_1(\zeta)} + \frac{\theta_1(\zeta)}{\theta_1(\zeta)} \right) d\xi d\zeta \\
= \exp(\gamma(p)) \Upsilon(h) \tilde{\Phi}_n^2 \Psi^{+}_{V,\lambda,s}(h) \\
\leq \Upsilon(h) \tilde{\Phi}_n^2 \Psi^{+}_{V,\lambda,s}(h).
\]

we note that

\[
\Upsilon(h) \tilde{\Phi}_n^2 \Psi^{+}_{V,\lambda,s}(h) = \left( \Upsilon(h) \tilde{\Phi}_n^2 n^{1/2} \right) n^{-1/2} \Psi^{+}_{V,\lambda,s}(h),
\]

\[
= o_p(1) O_p \left( n^{-1/2} (h^{-1})^{\gamma_{L,s}} \exp \left( (\alpha_{L,s}(h^{-1})^{\beta_{L,s}}) \right) \right)
\]

\[
= o_p \left( n^{-1/2} (h^{-1})^{\gamma_{L,s}} \exp \left( (\alpha_{L,s}(h^{-1})^{\beta_{L,s}}) \right) \right).
\]
Because all other terms are also bounded by the upper bound for \( \sup_{x \in \text{supp}(X)} | \int R_{11} s(x, w) dw | \) as shown in the proof of Theorem 4.4, we have

\[
\sup_{x \in \text{supp}(X)} \left| \sum_{j=1}^{J} \int R_{V_j, \lambda_j}(x, w, h_n) s_j(x, w) dw \right| = o_p \left( n^{-1/2} \right).
\]

Thus putting all together gives the desired result. \( \square \)

**Proof of Theorem 4.8**

By the assumption 4.11, we have

\[
\max_{j=1, \ldots, J} \sup_{(x, w) \in \text{supp}(X, W)} \left| \hat{g}_{V_j, \lambda_j}(x, w, h_n) - g_{V_j, \lambda_j}(x, w, h_n) \right|
\]

\[
= \max_{j=1, \ldots, J} \sup_{(x, w) \in \text{supp}(X, W)} \left| B_{V_j, \lambda_j}(x, w, h_n) + L_{V_j, \lambda_j}(x, w, h_n) + R_{V_j, \lambda_j}(x, w, h_n) \right|
\]

\[
= o(n^{-1/2}) + o_p(n^{-1/4}) + o_p(n^{-1/2})
\]

\[
= o_p(n^{-1/4}).
\]

Thus the remainder term in eqn.(10) is \( o_p \left( n^{-1/4} \right) \) when we let \( \hat{g}_{V_j, \lambda_j}(x, w) = \hat{g}_{V_j, \lambda_j}(x, w, h_n) \) and \( \hat{f}_W(w) = \hat{f}_W(w) \). We also have

\[
\sum_{j=1}^{J} \int \left( \hat{g}_{V_j, \lambda_j}(x, w, h_n) - g_{V_j, \lambda_j}(x, w, h_n) \right) s_j(x, w) dw dx
\]

\[
= \sum_{j=1}^{J} \int L_{V_j, \lambda_j}(x, w, h_n) s_j(x, w) dw dx + \sum_{j=1}^{J} \int \left( B_{V_j, \lambda_j}(x, w, h_n) + R_{V_j, \lambda_j}(x, w, h_n) \right) s_j(x, w) dw dx.
\]

Note that

\[
\left| \sum_{j=1}^{J} \int \left( B_{V_j, \lambda_j}(x, w, h_n) + R_{V_j, \lambda_j}(x, w, h_n) \right) s_j(x, w) dw dx \right|
\]

\[
\leq \left( \max_{j=1, \ldots, J} \sup_{(x, w) \in \text{supp}(X, W)} \left| B_{V_j, \lambda_j}(x, w, h_n) + R_{V_j, \lambda_j}(x, w, h_n) \right| \right) \sum_{j=1}^{J} \int |s_j(x, w)| dw dx
\]

\[
= o_p(n^{-1/2}),
\]

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

\[
b(\hat{g}(h_n), \hat{f}(h_n)) - b(g, f) = \sum_{j=1}^{J} \int L_{V_j, \lambda_j}(x, w, h_n) s_j(x, w) dw dx
\]

\[
+ \int \left( \hat{f}_W(w) - f_W(w) \right) s_{J+1}(x, w) dw dx + o_p(n^{-1/2}).
\]

We also note that

\[
\sum_{j=1}^{J} \int L_{V_j, \lambda_j}(x, w, h_n) s_j(x, w) dw dx + \int \left( \hat{f}_W(w) - f_W(w) \right) s_{J+1}(x, w) dw dx
\]
For the first term in the integrand of eqn. (23), we have

\[ \lim_{h \to 0} \sum_{j=1}^{J} \int \int L_{V_j, \lambda_j} (x, w, \tilde{h}) s_j(x, w) dw dx \]

\[ + \lim_{h \to 0} \int \int \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx \]

\[ + \left\{ \lim_{h \to 0} \int \int \left( L_{V_j, \lambda_j} (x, w, h_n) - L_{V_j, \lambda_j} (x, w, \tilde{h}) \right) s_j(x, w) dw dx \right\} s_{J+1}(x, w) dw dx \]

\[ = \left\{ \lim_{h \to 0} \sum_{j=1}^{J} \int \int \left\{ \int \left[ \Psi_{V_j, \lambda_j, 1} (\zeta, x, w, \tilde{h}) \left( \hat{E}[e^{i \zeta X_2}] - E[e^{i \zeta X_2}] \right) \right. \right. \]

\[ + \left. \Psi_{V_j, \lambda_j, x_{V_j}} (\zeta, x, w, \tilde{h}) \left( \hat{E}[V_j e^{i \zeta X_2} k_h(W - w)] - E[V_j e^{i \zeta X_2} k_h(W - w)] \right) \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) \right\} s_j(x, w) dw dx \]

\[ + \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx \].

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

\[ \lim_{h \to 0} \sum_{j=1}^{J} \int \int L_{V_j, \lambda_j} (x, w, \tilde{h}) s_j(x, w) dw dx + \lim_{h \to 0} \int \int \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx \]

\[ = \sum_{j=1}^{J} \int \left\{ \lim_{h \to 0} \int \int \Psi_{V_j, \lambda_j, 1} (\zeta, x, w, \tilde{h}) s_j(x, w) dw dx \left( \hat{E}[e^{i \zeta X_2}] - E[e^{i \zeta X_2}] \right) \right. \]

\[ + \left. \lim_{h \to 0} \int \int \Psi_{V_j, \lambda_j, x_{V_j}} (\zeta, x, w, \tilde{h}) s_j(x, w) \left( \hat{E}[V_j e^{i \zeta X_2} k_h(W - w)] - E[V_j e^{i \zeta X_2} k_h(W - w)] \right) \right. \]

\[ \left. d \zeta \right\} s_{J+1}(x, w) dw dx \]

\[ + \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx. \]

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

\[ \lim_{h \to 0} \sum_{j=1}^{J} \int \int L_{V_j, \lambda_j} (x, w, \tilde{h}) s_j(x, w) dw dx + \lim_{h \to 0} \int \int \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx \]

\[ = \sum_{j=1}^{J} \int \left\{ \lim_{h \to 0} \int \int \Psi_{V_j, \lambda_j, 1} (\zeta, x, w, \tilde{h}) s_j(x, w) dw dx \left( \hat{E}[e^{i \zeta X_2}] - E[e^{i \zeta X_2}] \right) \right. \]

\[ + \left. \lim_{h \to 0} \int \int \Psi_{V_j, \lambda_j, x_{V_j}} (\zeta, x, w, \tilde{h}) s_j(x, w) \left( \hat{E}[V_j e^{i \zeta X_2} k_h(W - w)] - E[V_j e^{i \zeta X_2} k_h(W - w)] \right) \right. \]

\[ \left. d \zeta \right\} s_{J+1}(x, w) dw dx \]

\[ + \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{J+1}(x, w) dw dx. \]

For the first term in the integrand of eqn. (23), we have
\[
\frac{1}{\theta_1(\zeta)} \int_{\zeta}^{\pm \infty} (-i\xi) \kappa(h\xi) \exp(-i\xi x) \phi_V(x, w) d\xi
\]

Similarly, for the second term, we have

\[
\frac{1}{\theta_1(\zeta)} \int_{\zeta}^{\pm \infty} \left( \int \exp(-i\xi x) \int s(x, w) \phi_V(x, w) d\xi \right) (-i\xi) \kappa(h\xi) d\xi
\]

We also note that for the third term, the

\[
\frac{1}{\theta_1(\zeta)} \int_{\zeta}^{\pm \infty} \left( \int \exp(-i\xi x) \int s(x, w) \phi_V(x, w) d\xi \right) (-i\xi) \kappa(h\xi) d\xi
\]
and for the fourth term,

\[
\lim_{h \to 0} \int \int \Psi_{V,\lambda,fw}(\zeta, x, w, \tilde{h})s(x, w) \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) dw dx
\]

\[
= \lim_{h \to 0} \int \int \left\{ \frac{-1}{2\pi} (-i\zeta)^{\lambda} \kappa(\tilde{h}\zeta) \exp(-i\zeta x) \frac{\phi_V(\zeta, w)}{f_W(w)} \right\} s(x, w) \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) dw dx
\]

\[
= \frac{-1}{2\pi} (-i\zeta)^{\lambda} \int \exp(-i\zeta x) \lim_{h \to 0} \int \frac{1}{f_W(w)} s(x, w) \phi_V(\zeta, w) \times \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) dw dx \lim_{h \to 0} \kappa(\tilde{h}\zeta)
\]

\[
= \frac{-1}{2\pi} (-i\zeta)^{\lambda} \int \exp(-i\zeta x) \lim_{h \to 0} \int \frac{1}{f_W(w)} s(x, w) \phi_V(\zeta, w) \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) dw dx
\]

\[
\equiv \hat{E}[Z_{V,\lambda,fw}(s, \zeta; W)] - E[Z_{V,\lambda,fw}(s, \zeta; W)],
\]

where \( Z_{V,\lambda,XV}(s, \zeta; V, X_2, W) \) and \( Z_{V,\lambda,fw}(s, \zeta; W) \) are defined in the statement of the theorem.

Thus it follows that

\[
\lim_{h \to 0} \sum_{j=1}^J \int \int L_{V,\lambda_j}(x, w, \tilde{h})s_j(x, w) dw dx + \lim_{h \to 0} \int \left( \hat{E}[k_h(W - w)] - E[k_h(W - w)] \right) s_{j+1}(x, w) dw dx
\]

\[
= \sum_{j=1}^J \left\{ \Psi_{V,\lambda_j,1,s_j}(\zeta) \left( \hat{E}[e^{i\zeta X_2}] - E[e^{i\zeta X_2}] \right) + \Psi_{V,\lambda_j,2,s_j}(\zeta) \left( \hat{E}[X_1 e^{i\zeta X_2}] - E[X_1 e^{i\zeta X_2}] \right)
\]

\[
+ \left( \hat{E}[Z_{V,\lambda_j,XV}(s_j, \zeta; V_j, X_2, W)] - E[Z_{V,\lambda_j,XV}(s_j, \zeta; V_j, X_2, W)] \right)
\]

\[
+ \left( \hat{E}[Z_{V,\lambda_j,fw}(s_j, \zeta; W)] - E[Z_{V,\lambda_j,fw}(s_j, \zeta; W)] \right) \right\} d\zeta
\]

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

\[
= \hat{E} \left[ \tilde{\psi}_s(V, X_1, X_2, W) \right],
\]

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

\[
|\tilde{\psi}_s(v, x_1, x_2, w)| \leq C \max\{1, |x_1|\} \tilde{\Psi}_{V,\lambda,s}.
\]

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

The second term of eqn.(22) can be shown to be \( o_p(n^{-1/2}) \) because it can be written as an \( h_n \)-dependent sample average \( \hat{E} \left[ \tilde{\psi}_s(V, X_1, X_2, W, h_n) \right] \) where \( \tilde{\psi}_s(V, X_1, X_2, W, h_n) \) is such that \( \lim_{h_n \to 0} E \left[ \tilde{\psi}_s(V, X_1, X_2, W, h_n) \right] = 0 \). The similar procedure to the case of \( \hat{E} \left[ \tilde{\psi}_s(V, X_1, X_2, W) \right] \) is used just by replacing \( \kappa \left( \tilde{h}\zeta \right) \) by \( \kappa \left( h_n\xi \right) - \kappa \left( \tilde{h}\zeta \right) \) and \( k_h(\cdot) \) by \( (k_{h_n}(\cdot) - k_h(\cdot)) \), and taking the limit as \( h_n \to 0 \) and \( \tilde{h} \to 0 \).  \( \square \)

**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
where \( R_{V,\lambda}(\tilde{g}_{V,\lambda}(x, w, h_n), (\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w))) \) is a remainder term in which \( \tilde{g}_{V,\lambda}(x, w, h_n) \) 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,

\[
\max_{V = 1, Y, \lambda = 0, 1} \sup_{(x, w) \in (X, W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)| = O_p(\varepsilon_n),
\]

\[
\varepsilon_n = (h_n^{-1})\gamma_1 \beta_1 \exp\left(\alpha_1 (h_n^{-1})\beta_1\right) + n^{-1/2} (h_n^{-1})\gamma_2 \exp\left(\alpha_2 (h_n^{-1})\beta_2\right) \to 0.
\]

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 \((x, w) \in \Gamma_x\). 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 \( \tau_n \) uniformly for \((x, w) \in \Gamma_x\), we have that

\[
\max_{V = 1, Y, \lambda = 0, 1} \sup_{(x, w) \in (X, W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)| = O(1)O_p(\tau_n^{-3})O_p(\varepsilon_n) = O_p(\varepsilon_n/\tau_n^3).
\]

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 \((x, w) \in \Gamma_x\). These terms involve a finite sum of \( (i) \) finite products of the functions \( \tilde{g}_{V,\lambda}(x, w, h_n) \) for \( V = 1, Y \) and \( \lambda = 0, 1 \); \( (ii) \) division by a product of at most 4 functions of the form \( g_{1,0}(x, w, h_n) \); and \( (iii) \) pairwise products of functions of the form \( \tilde{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w) \) for \( V = 1, Y \) and \( \lambda = 0, 1 \). First, the contribution of \( (i) \) is bounded in probability uniformly for \((x, w) \in \Gamma_x\) because

\[
|\hat{g}_{V,\lambda}(x, w, h_n)| \leq |g_{V,\lambda}(x, w)| + |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)|
\]

\[
\leq |g_{V,\lambda}(x, w)| + |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)|
\]

\[
\leq |g_{V,\lambda}(x, w)| + \max_{V = 1, Y, \lambda = 0, 1} \sup_{(x, w) \in (X, W)} |\hat{g}_{V,\lambda}(x, w, h_n) - g_{V,\lambda}(x, w)|
\]

\[
= O_p(1) + o_p(1)
\]

\[
= O_p(1).
\]

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

\[
\hat{g}_{1,0}(x, w, h_n) = g_{1,0}(x, w) \left( 1 + \frac{\tilde{g}_{1,0}(x, w, h_n) - g_{1,0}(x, w)}{g_{1,0}(x, w)} \right)
\]

\[
= f_{X|W}(x \mid w) \left( 1 + \frac{\tilde{g}_{1,0}(x, w, h_n) - g_{1,0}(x, w)}{f_{X|W}(x \mid w)} \right)
\]

\[
= f_{X|W}(x \mid w) \left( 1 + O_p\left( \frac{\varepsilon_n}{\tau_n^3} \right) \right).
\]

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 \((x, w) \in \Gamma_x\)

\[
\hat{g}_{1,0}(x, w, h_n) = f_{X|W}(x \mid w) (1 + o_p(1)) 
\]

\[
\geq \tau_n/2
\]

74
with probability approaching one since \( f_{X|W}(x \mid w) \geq \tau_n \) for \((x, w) \in \Gamma\), by construction. Therefore we have that the contribution of (ii) is \( \bar{g}_{1, \lambda}(x, w, h_n) = O_p(\tau_n^{-4}) \). Finally, the contribution of (iii) is \( O_p(\epsilon_n^2) \). Putting all together, we have

\[
R_{V, \lambda}(\bar{g}_{V, \lambda}(x, w, h_n), (\hat{g}_{V, \lambda}(x, w, h_n) - g_{V, \lambda}(x, w))
= O_p(1)O_p(\tau_n^{-4})O_p(\epsilon_n^2) = O_p\left(\frac{\epsilon_n}{\tau_n}\right) = o_p\left(\frac{\epsilon_n}{\tau_n}\right)
\]

so that

\[
\sup_{(x, w) \in \Gamma} \left| \hat{\beta}(x, w, h_n) - \beta(x, w) \right| = O_p\left(\frac{\epsilon_n}{\tau_n}\right) + o_p\left(\frac{\epsilon_n}{\tau_n}\right) = o_p(1). \]

\[\□\]

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.(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 Taylor expansions.

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

\[
\sup_{x \in \mathbb{M}} \left| \hat{\beta}_m(x) - \beta_m(x) \right|
= \sup_{x \in \mathbb{M}} \left| \int_{S_W} \left( \hat{\beta}(x, w, h_n) - \beta(x, w) \right) m(w) dw \right|
= \sup_{x \in \mathbb{M}} \left| \int_{S_W} \sum_{V=1, Y \lambda=0, 1} m(w) s_{V, \lambda}(x, w) (\hat{g}_{V, \lambda}(x, w, h_n) - g_{V, \lambda}(x, w)) dw \right| + o_p(1)
= \sup_{x \in \mathbb{M}} \left| \sum_{V=1, Y \lambda=0, 1} \sum_{x, w, h_n} m(w) s_{V, \lambda}(x, w) (\hat{g}_{V, \lambda}(x, w, h_n) - g_{V, \lambda}(x, w)) dw \right| + o_p(1)
= O \left( \tau^{-3} (h_n^{-1})^{\gamma_{1, \beta}} \exp(\alpha_{B, s}(h_n^{-1})^{\beta_{u, s}}) \right) + O_p \left( \tau^{-3} n^{-1/2} (h_n^{-1})^{\gamma_{1, \beta}} \exp(\alpha_{L, s}(h_n^{-1})^{\beta_{u, s}}) \right),
\]

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

\[
\sup_{x \in \mathbb{M}} \left| \hat{\beta}_{f_{W}}(x) - \beta_{f_{W}}(x) \right|
= \sup_{x \in \mathbb{M}} \left| \int_{S_W} \left( \hat{\beta}(x, w, h_n)m(w)f_{W}(w) - \beta(x, w)m(w)f_{W}(w) \right) dw \right|
= \sup_{x \in \mathbb{M}} \left| \int_{S_W} m(w)f_{W}(w) \left( \hat{\beta}(x, w, h_n) - \beta(x, w) \right) dw \right|
+ \left( \int_{S_W} \beta(x, w)m(w) \left( f_{W}(w) - f_{W}(w) \right) dw \right) + o_p(1)
\]
where the last equality is attained by Theorem 4.7.

(iii) From the definitions of $\hat{\beta}_{m_{fW|x}}(x)$ and $\hat{\beta}_{m_{fW|x}}(x)$, we have

$$\sup_{x \in M} \left| \hat{\beta}_{m_{fW|x}}(x) - \hat{\beta}_{m_{fW|x}}(x) \right|$$

$$= \sup_{x \in M} \left| \int_{S^w_{\beta(h_n)}} \left( \hat{\beta}(x, w_{n}, w_{h_n})m(w) \frac{g_{1,0}(x, w_{n}, w_{h_n})f_W(w)}{g_{1,0}(x, w_{n}, w_{h_n})f_W(w)} - \beta(x, w_{n})m(w) \frac{g_{1,0}(x, w, dw)}{g_{1,0}(x, w)dw} \right) \right| + o_p(1)$$

$$= \sup_{x \in M} \left| \int_{S^w_{\beta(h_n)}} \left( m(w) \frac{g_{1,0}(x, w, dw)}{g_{1,0}(x, w)dw} \left( \hat{\beta}(x, w_{n}, w_{h_n}) - \beta(x, w_{n}) \right) \right) + o_p(1)$$

$$= \sup_{x \in M} \left| \sum_{V = 1, Y = 0, 1} \int_{S^w_{\beta(h_n)}} m(w) f_{W|x}(w | x) s_{V,x}(w) (\hat{g}_{V,x}(x, w, h_n) - g_{V,x}(x, w)) dw \right| + o_p(1)$$

where the last equality is attained by Theorem 4.7.

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

\[
\hat{\beta}_{\tilde{m}} - \beta_{\tilde{m}} = \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \left( \hat{\beta}(x, w, h_n) - \beta(x, w) \right) \tilde{m}(x, w) dw dx
\]

\[
= \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \sum_{V=1}^{Y \lambda=0,1} \tilde{m}(x, w) s_{V,\lambda}(x, w) \left( \hat{g}(x, w, h_n) - g_{V,\lambda}(x, w) \right) dw dx + o_p(n^{-1/2})
\]

\[
= \sum_{V=1}^{Y \lambda=0,1} \int_{S^w_{\beta(h_n)}} \tilde{m}(x, w) s_{V,\lambda}(x, w) \left( \hat{g}(x, w, h_n) - g_{V,\lambda}(x, w) \right) dw dx + o_p(n^{-1/2})
\]

\[
= \sum_{V=1, Y \lambda=0,1} \tilde{E} \left[ \psi_{V,\lambda}(m_{V,\lambda}; V, X_1, X_2, W) \right] + o_p(n^{-1/2})
\]

\[
= \tilde{E} \left[ \sum_{V=1, Y \lambda=0,1} \psi_{V,\lambda}(m_{V,\lambda}; V, X_1, X_2, W) \right] + o_p(n^{-1/2}).
\]

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

(iii) From the definitions of \( \hat{\beta}_{\tilde{m}_{fW|x}} \) and \( \beta_{\tilde{m}_{fW|x}} \), we have

\[
\hat{\beta}_{\tilde{m}_{fW|x}} - \beta_{\tilde{m}_{fW|x}} = \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \left[ \hat{\beta}(x, w, h_n) \tilde{m}(x, w) \frac{g_{1,0}(x, w, h_n)}{\tilde{g}(x, w, h_n)} \right] dx dw
\]

\[
- \beta(x, w) \tilde{m}(x, w) \frac{g_{1,0}(x, w)}{\tilde{g}(x, w)} \left( \hat{f}(w) - f(w) \right) + \left( \frac{g_{1,0}(x, w)}{\tilde{g}(x, w)} \right)^2 \left( \int_{S^w_{\beta(h_n)}} \tilde{g}(x, w, h_n) dw - \int_{S^w_{\beta(h_n)}} g_{1,0}(x, w, h_n) dw \right) \right] dw dx + o_p(n^{-1/2})
\]

\[
= \sum_{V=1, Y \lambda=0,1} \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \tilde{m}(x, w) f_{W|x}(w | x) s_{V,\lambda}(x, w) \left( \hat{g}(x, w, h_n) - g_{V,\lambda}(x, w) \right) dw dx
\]

\[
+ \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \beta(x, w) \tilde{m}(x, w) \frac{f_{W|x}(w | x)}{f(x)} \left( \hat{g}(x, w, h_n) - g_{1,0}(x, w) \right) dw dx
\]

\[
+ \int_{S^w_{\beta(h_n)}} \int_{S^w_{\beta(h_n)}} \beta(x, w) \tilde{m}(x, w) \frac{f_{W|x}(w | x)}{f(x)} \left( \hat{f}(w) - f(w) \right) dw dx
\]
\[- \int_{S^\beta_{(\cdot, h_n)}} \beta(x, w) \tilde{m}(x, w) \frac{f_{W|X}(w | x)}{f_X(x)} \left( \int_{S^\beta_{(\cdot, h_n)}} \hat{g}_{1,0}(x, w, h_n)dw - \int_{S^\beta_{(\cdot, h_n)}} g_{1,0}(x, w)dw \right) dwx + o_p(n^{-1/2}) \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} E \left[ \psi_{V,\lambda} \left( \hat{m}_{W|X s_{V,\lambda}; V, X_1, X_2, W} \right) \right] + E \left[ \psi_{1,0} \left( P_1; 1, X_1, X_2, W \right) \right] \]

\[- E \left[ \psi_{1,0} \left( P_2; 1, X_1, X_2, W \right) \right] + E \left[ \psi_{f} \left( P_3; W \right) \right] + o_p(n^{-1/2}) \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} E \left[ \psi_{V,\lambda} \left( \hat{m}_{W|X s_{V,\lambda}; V, X_1, X_2, W} \right) \right] + \psi_{1,0} \left( P_1; 1, X_1, X_2, W \right) \]

\[- \psi_{1,0} \left( P_2; 1, X_1, X_2, W \right) + \psi_{f} \left( P_3; W \right) \]

Let \( \psi_{\tilde{m}_{W|X}} (v, x_1, x_2, \tilde{w}) \equiv \sum_{V=1,Y} \sum_{\lambda=0,1} \psi_{V,\lambda} \left( \tilde{m}_{W|X s_{V,\lambda}; v, x_1, x_2, \tilde{w}} \right) + \psi_{1,0} \left( P_1; 1, x_1, x_2, \tilde{w} \right) \)

\[- \psi_{1,0} \left( P_2; 1, x_1, x_2, \tilde{w} \right) + \psi_{f} \left( P_3; \tilde{w} \right) \]

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 \( \hat{\tilde{m}}_{W|X} \) and \( \beta \hat{\tilde{m}}_{W|X} \), we have

\[\hat{\tilde{m}}_{W|X} - \beta \hat{\tilde{m}}_{W|X} \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^\beta_{(\cdot, h_n)}} \int_{S^\beta_{(\cdot, h_n)}} \beta(x, w, h_n) \tilde{m}(x, w) \hat{g}_{1,0}(x, w, h_n) f_W(w) - \beta(x, w) \tilde{m}(x, w) g_{1,0}(x, w) f_W(w) \] \[\left( \beta(x, w, h_n) - \beta(x, w) \right) dwx + o_p(n^{-1/2}) \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^\beta_{(\cdot, h_n)}} \int_{S^\beta_{(\cdot, h_n)}} \tilde{m}(x, w) g_{1,0}(x, w) f_W(w) \left( \beta(x, w, h_n) - \beta(x, w) \right) dwx \]

\[+ \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^\beta_{(\cdot, h_n)}} \beta(x, w) \tilde{m}(x, w) f_W(w) \left( \hat{g}_{1,0}(x, w, h_n) - g_{1,0}(x, w) \right) dwx \]

\[+ \sum_{V=1,Y} \sum_{\lambda=0,1} \int_{S^\beta_{(\cdot, h_n)}} \beta(x, w) \tilde{m}(x, w) f_W(w) \left( \hat{f}_W(w) - f_W(w) \right) dwx + o_p(n^{-1/2}) \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} E \left[ \psi_{V,\lambda} \left( \tilde{m}_{W|X s_{V,\lambda}; V, X_1, X_2, W} \right) \right] + E \left[ \psi_{1,0} \left( \beta \tilde{m}_{W|X}; 1, X_1, X_2, W \right) \right] \]

\[+ E \left[ \psi_{f} \left( \beta \tilde{m}_{W|X}; W \right) \right] + o_p(n^{-1/2}) \]

\[= \sum_{V=1,Y} \sum_{\lambda=0,1} \psi_{V,\lambda} \left( \tilde{m}_{W|X s_{V,\lambda}; V, X_1, X_2, W} \right) + \psi_{1,0} \left( \beta \tilde{m}_{W|X}; 1, X_1, X_2, W \right) \]

\[+ \psi_f \left( \beta \tilde{m}_{W|X}; \tilde{w} \right) \]

The result is immediate from the application of Theorem 4.8. \( \square \)
References


<table>
<thead>
<tr>
<th>Example</th>
<th>$U_x$</th>
<th>$W$</th>
<th>$U_1, U_2$</th>
<th>$U_u$</th>
<th>$f_1(X)$</th>
<th>$f_2(W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N(0, 0.5)</td>
<td>N(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>2</td>
<td>N(0, 0.5)</td>
<td>N(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>3</td>
<td>N(0, 0.5)</td>
<td>N(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>4</td>
<td>N(0, 0.5)</td>
<td>N(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>5</td>
<td>N(0, 0.5)</td>
<td>L(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>6</td>
<td>N(0, 0.5)</td>
<td>L(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>7</td>
<td>N(0, 0.5)</td>
<td>L(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>8</td>
<td>L(0, 0.5)</td>
<td>N(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>9</td>
<td>L(0, 0.5)</td>
<td>N(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>10</td>
<td>L(0, 0.5)</td>
<td>N(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>11</td>
<td>L(0, 0.5)</td>
<td>N(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>12</td>
<td>L(0, 0.5)</td>
<td>N(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>13</td>
<td>L(0, 0.5)</td>
<td>L(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>14</td>
<td>L(0, 0.5)</td>
<td>L(0, 1)</td>
<td>N(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
<tr>
<td>15</td>
<td>L(0, 0.5)</td>
<td>L(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>erf(w)</td>
</tr>
<tr>
<td>16</td>
<td>L(0, 0.5)</td>
<td>L(0, 1)</td>
<td>L(0, 0.25)</td>
<td>N(0, 0.09)</td>
<td>erf(x)</td>
<td>S(w)</td>
</tr>
</tbody>
</table>
**Table 2. Monte Carlo simulation results for the examples**

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5</td>
<td>6.75</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02278</td>
<td>0.00063</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05150</td>
<td>0.04910</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.07628</td>
<td>0.04972</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 2</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5</td>
<td>6.75</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.05632</td>
<td>0.01278</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05751</td>
<td>0.05261</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.11383</td>
<td>0.06540</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 3</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5</td>
<td>6.75</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.05670</td>
<td>0.01295</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05563</td>
<td>0.05142</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.11233</td>
<td>0.06437</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 4</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5</td>
<td>6.75</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02510</td>
<td>0.00105</td>
</tr>
<tr>
<td>Variance</td>
<td>0.06109</td>
<td>0.05606</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.08619</td>
<td>0.05711</td>
</tr>
</tbody>
</table>

**Example 1**

- **Monte Carlo simulation results for the examples**
- **Fourier (optimal)**
- **Bandwidth**: 6.5, 6.75, 7, 7.25, 7.5, 7.75, 8
- **Bias squared**: 0.02278, 0.00063, 0.00789, 0.03124, 0.05596, 0.08108, 0.11129
- **Variance**: 0.05150, 0.04910, 0.04500, 0.04168, 0.03921, 0.03708, 0.03490
- **Mean square error**: 0.07628, 0.04972, 0.05289, 0.07293, 0.09517, 0.11815, 0.14619

**Example 2**

- **Monte Carlo simulation results for the examples**
- **Fourier (optimal)**
- **Bandwidth**: 2, 2.25, 2.5, 2.75, 3
- **Bias squared**: 0.11903, 0.02781, 0.00889, 0.01441, 0.03896
- **Variance**: 0.02763, 0.01001, 0.00570, 0.00395, 0.00253
- **Mean square error**: 0.14665, 0.03782, 0.01458, 0.01063, 0.00351

**Example 3**

- **Monte Carlo simulation results for the examples**
- **Fourier (optimal)**
- **Bandwidth**: 2, 2.25, 2.5, 2.75, 3
- **Bias squared**: 0.07261, 0.00642, 0.00028, 0.00406, 0.01462
- **Variance**: 0.02636, 0.00976, 0.00351, 0.00253, 0.00277
- **Mean square error**: 0.10143, 0.01653, 0.00698, 0.00227, 0.00342

**Example 4**

- **Monte Carlo simulation results for the examples**
- **Fourier (optimal)**
- **Bandwidth**: 2, 2.25, 2.5, 2.75, 3
- **Bias squared**: 0.02510, 0.00105, 0.00665, 0.02879, 0.05272
- **Variance**: 0.06109, 0.05606, 0.05139, 0.04761, 0.04480
- **Mean square error**: 0.08619, 0.05711, 0.05804, 0.07640, 0.09752

83
<table>
<thead>
<tr>
<th>Example 5</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5 6.75 7 7.25 7.5 7.75 8</td>
<td>6.9</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02570 0.00243 0.00271 0.01728 0.03476 0.05357 0.07686 0.00018</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05002 0.04580 0.04192 0.03879 0.03649 0.03451 0.03250 0.04340</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.07572 0.04823 0.04463 0.05607 0.07125 0.08808 0.10936 0.00458</td>
<td></td>
</tr>
<tr>
<td>Local linear</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.65</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.51750 0.16099 0.11275 0.11591 0.13800 0.16885 0.20353 0.11153</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.33088 0.03015 0.01083 0.00571 0.00367 0.00337 0.00308 0.00626</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.84833 0.19114 0.12358 0.12161 0.14167 0.17155 0.20570 0.11868</td>
<td></td>
</tr>
<tr>
<td>No measurement error</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.55</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.14816 0.01653 0.00574 0.01004 0.02346 0.04582 0.07519 0.00588</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05036 0.01326 0.00686 0.00476 0.00384 0.00337 0.00308 0.00626</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.19851 0.02980 0.01480 0.02730 0.04919 0.07827 0.01214 0.00124</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 6</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5 6.75 7 7.25 7.5 7.75 8</td>
<td>6.9</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.00124 0.00581 0.03165 0.06624 0.09755 0.12761 0.16230 9.44E-05</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05245 0.05287 0.07490 0.10641 0.13543 0.16352 0.19620 0.04963</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.19851 0.02980 0.01480 0.02730 0.04919 0.07827 0.01214 0.00124</td>
<td></td>
</tr>
<tr>
<td>Local linear</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.65</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.62539 0.26934 0.20344 0.20593 0.23413 0.27374 0.31792 0.20067</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.53643 0.03186 0.01225 0.00649 0.00418 0.00337 0.00308 0.00715</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>1.16182 0.30120 0.21569 0.21241 0.23831 0.27685 0.32043 0.20881</td>
<td></td>
</tr>
<tr>
<td>No measurement error</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.55</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.24395 0.05439 0.03219 0.04192 0.06719 0.09755 0.12761 0.01034</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05451 0.01522 0.00814 0.00562 0.00442 0.00378 0.00308 0.00745</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.29846 0.06961 0.04033 0.04754 0.07161 0.10725 0.15032 0.03999</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 7</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5 6.75 7 7.25 7.5 7.75 8</td>
<td>6.9</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02668 0.00271 0.03165 0.06624 0.09755 0.12761 0.16230 9.44E-05</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05121 0.04706 0.04325 0.04017 0.03788 0.03591 0.03390 0.04954</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.07810 0.04991 0.04754 0.07161 0.09755 0.12761 0.16230 0.04963</td>
<td></td>
</tr>
<tr>
<td>Local linear</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.65</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.62539 0.26934 0.20344 0.20593 0.23413 0.27374 0.31792 0.20067</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.53643 0.03186 0.01225 0.00649 0.00418 0.00337 0.00308 0.00715</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>1.16182 0.30120 0.21569 0.21241 0.23831 0.27685 0.32043 0.20881</td>
<td></td>
</tr>
<tr>
<td>No measurement error</td>
<td>2 2.25 2.5 2.75 3</td>
<td>3.25 3.5 2.55</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.24395 0.05439 0.03219 0.04192 0.06719 0.09755 0.12761 0.01034</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05451 0.01522 0.00814 0.00562 0.00442 0.00378 0.00308 0.00745</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.29846 0.06961 0.04033 0.04754 0.07161 0.10725 0.15032 0.03999</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 8</th>
<th>Fourier</th>
<th>(optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5 6.75 7 7.25 7.5 7.75 8</td>
<td>6.5</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.01244 0.00581 0.03165 0.06624 0.09755 0.12761 0.16230 9.44E-05</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.05245 0.05287 0.07490 0.10641 0.13543 0.16352 0.19620 0.04963</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.19851 0.02980 0.01480 0.02730 0.04919 0.07827 0.01214 0.00124</td>
<td></td>
</tr>
<tr>
<td>Example 9</td>
<td>Fourier (optimal)</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>6.5  6.75  7  7.25  7.5  7.75  8  6.9</td>
<td></td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02500  0.00164  0.00436  0.02206  0.04202  0.06287  0.08826  0.00066</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.08595  0.05236  0.04810  0.04464  0.04205  0.03981  0.03752  0.03973</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.08195  0.05400  0.05246  0.06670  0.08408  0.10268  0.12579  0.05039</td>
<td></td>
</tr>
<tr>
<td>Local linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  3</td>
<td></td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.39873  0.27315  0.20482  0.17235  0.16437  0.17336  0.19303  0.16437</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.07042  0.01974  0.00908  0.00707  0.00584  0.00511  0.00462  0.00742</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.46915  0.29289  0.21390  0.17798  0.16861  0.17687  0.19607  0.18661</td>
<td></td>
</tr>
<tr>
<td>No measurement error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  2.7</td>
<td></td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.01103  0.00443  0.00230  0.00310  0.00769  0.01868  0.03771  0.00272</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.02563  0.01411  0.00935  0.00584  0.00511  0.00462  0.00462  0.00742</td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.03666  0.01854  0.01165  0.01017  0.01353  0.02379  0.04233  0.01014</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 10</th>
<th>Fourier (optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5  6.75  7  7.25  7.5  7.75  8  6.65</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.00201  0.00531  0.03214  0.06847  0.10110  0.13215  0.16783  0.00070</td>
</tr>
<tr>
<td>Variance</td>
<td>0.04967  0.04580  0.04221  0.03927  0.03706  0.03513  0.03316  0.04734</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.05168  0.05111  0.07435  0.10774  0.13815  0.16727  0.20097  0.05039</td>
</tr>
<tr>
<td>Local linear</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  2.95</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.51643  0.37207  0.29260  0.25596  0.24892  0.26250  0.28901  0.24844</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05180  0.01917  0.01012  0.00647  0.00472  0.00380  0.00494  0.00888</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.56823  0.39124  0.30272  0.26243  0.25364  0.26631  0.29225  0.25343</td>
</tr>
<tr>
<td>No measurement error</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  2.6</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.03756  0.02328  0.01766  0.02009  0.03101  0.05186  0.08283  0.01771</td>
</tr>
<tr>
<td>Variance</td>
<td>0.02854  0.01557  0.01012  0.00757  0.00626  0.00548  0.00494  0.00888</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.06609  0.03885  0.02778  0.02766  0.03727  0.05735  0.08777  0.02658</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 11</th>
<th>Fourier (optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5  6.75  7  7.25  7.5  7.75  8  6.9</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.02952  0.00531  0.03214  0.06847  0.10110  0.13215  0.16783  0.00070</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05658  0.05210  0.04792  0.04451  0.04196  0.03974  0.03748  0.04952</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.08611  0.05495  0.05081  0.06325  0.07949  0.09725  0.11957  0.04909</td>
</tr>
<tr>
<td>Local linear</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  2.9</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.31110  0.21349  0.15940  0.13635  0.13471  0.14804  0.17107  0.13328</td>
</tr>
<tr>
<td>Variance</td>
<td>0.04883  0.01880  0.01007  0.00652  0.00472  0.00371  0.00324  0.00498</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.35993  0.39124  0.30272  0.26243  0.25364  0.26631  0.29225  0.25343</td>
</tr>
<tr>
<td>No measurement error</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>2  2.25  2.5  2.75  3  3.25  3.5  2.75</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.01215  0.00488  0.00246  0.00313  0.00753  0.01814  0.03653  0.00313</td>
</tr>
<tr>
<td>Variance</td>
<td>0.02565  0.01455  0.01012  0.00757  0.00626  0.00548  0.00494  0.00734</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.03780  0.03885  0.02778  0.02766  0.03727  0.05735  0.08777  0.02658</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 12</th>
<th>Fourier (optimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>6.5  6.75  7  7.25  7.5  7.75  8  6.65</td>
</tr>
<tr>
<td>Bias squared</td>
<td>0.00405  0.00302  0.02630  0.06015  0.09126  0.12122  0.15589  6.57E-05</td>
</tr>
<tr>
<td>Variance</td>
<td>0.05835  0.05318  0.07259  0.10325  0.13196  0.15982  0.19234  0.05182</td>
</tr>
<tr>
<td>Mean square error</td>
<td>0.06751  0.03891  0.02774  0.02744  0.03167  0.05644  0.08620  0.02643</td>
</tr>
</tbody>
</table>
Table 2. Monte Carlo simulation results for the examples (continued)

<table>
<thead>
<tr>
<th>Example 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>No measurement error</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>No measurement error</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>No measurement error</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>No measurement error</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Bias squared</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Mean square error</td>
</tr>
<tr>
<td>Example</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Example 1</td>
</tr>
<tr>
<td>Example 2</td>
</tr>
<tr>
<td>Example 3</td>
</tr>
<tr>
<td>Example 4</td>
</tr>
<tr>
<td>Example 5</td>
</tr>
<tr>
<td>Example 6</td>
</tr>
<tr>
<td>Example 7</td>
</tr>
<tr>
<td>Example 8</td>
</tr>
<tr>
<td>Example 9</td>
</tr>
<tr>
<td>Example 10</td>
</tr>
<tr>
<td>Example 11</td>
</tr>
<tr>
<td>Example 12</td>
</tr>
<tr>
<td>Example 13</td>
</tr>
<tr>
<td>Example 14</td>
</tr>
<tr>
<td>Example 15</td>
</tr>
<tr>
<td>Example 16</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>at x=1 and w=1</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05834)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02785)</td>
</tr>
<tr>
<td>at x=-0.75 and w=-0.75</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05844)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02797)</td>
</tr>
<tr>
<td>at x=-0.5 and w=-0.5</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05702)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02876)</td>
</tr>
<tr>
<td>at x=0.25 and w=0.25</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05729)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02701)</td>
</tr>
<tr>
<td>at x=0 and w=0</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05833)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02901)</td>
</tr>
<tr>
<td>at x=0.5 and w=0.5</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05762)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02841)</td>
</tr>
<tr>
<td>at x=0.75 and w=0.75</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05799)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.02926)</td>
</tr>
<tr>
<td>at x=1 and w=1</td>
</tr>
<tr>
<td>Fourier</td>
</tr>
<tr>
<td>(0.05766)</td>
</tr>
<tr>
<td>Local linear</td>
</tr>
<tr>
<td>(0.03038)</td>
</tr>
</tbody>
</table>

Observations: 1544

Notes: Standard errors obtained by bootstrap methods are in parentheses. Data are from the Children of the NLSY linked to their mothers in the main NLSY79. All variables are standardized having means of zeros and standard deviations of one. Mother’s AFQT is used as conditioning instrument. Income is after-tax and after-transfer. 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. Year refers to the NLSY survey year; income refers to the previous year’s income.
Table 4.2. Impact of permanent family income on children’s math achievement

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=-1</td>
<td>0.70890</td>
<td>0.65565</td>
<td>0.65465</td>
<td>0.65465</td>
<td>0.6543</td>
<td>0.65107</td>
<td>0.63969</td>
<td>0.61812</td>
</tr>
<tr>
<td></td>
<td>(0.08261)</td>
<td>(0.08316)</td>
<td>(0.07899)</td>
<td>(0.07860)</td>
<td>(0.07981)</td>
<td>(0.07545)</td>
<td>(0.07851)</td>
<td>(0.07282)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=-0.75</td>
<td>0.70200</td>
<td>0.64984</td>
<td>0.64886</td>
<td>0.64886</td>
<td>0.64865</td>
<td>0.64537</td>
<td>0.63425</td>
<td>0.60954</td>
</tr>
<tr>
<td></td>
<td>(0.08449)</td>
<td>(0.07872)</td>
<td>(0.07340)</td>
<td>(0.07874)</td>
<td>(0.07808)</td>
<td>(0.08142)</td>
<td>(0.07867)</td>
<td>(0.07520)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=-0.5</td>
<td>0.69689</td>
<td>0.64559</td>
<td>0.64463</td>
<td>0.64463</td>
<td>0.64442</td>
<td>0.64120</td>
<td>0.63027</td>
<td>0.60954</td>
</tr>
<tr>
<td></td>
<td>(0.08168)</td>
<td>(0.07777)</td>
<td>(0.07784)</td>
<td>(0.07530)</td>
<td>(0.08061)</td>
<td>(0.07571)</td>
<td>(0.07334)</td>
<td>(0.07471)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=0</td>
<td>0.69182</td>
<td>0.64160</td>
<td>0.64065</td>
<td>0.64065</td>
<td>0.64044</td>
<td>0.63727</td>
<td>0.62649</td>
<td>0.60604</td>
</tr>
<tr>
<td></td>
<td>(0.07989)</td>
<td>(0.07735)</td>
<td>(0.07732)</td>
<td>(0.07686)</td>
<td>(0.07497)</td>
<td>(0.07572)</td>
<td>(0.07316)</td>
<td>(0.07256)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=0.25</td>
<td>0.69350</td>
<td>0.64285</td>
<td>0.64190</td>
<td>0.64190</td>
<td>0.64169</td>
<td>0.63850</td>
<td>0.62768</td>
<td>0.60716</td>
</tr>
<tr>
<td></td>
<td>(0.08012)</td>
<td>(0.07841)</td>
<td>(0.07542)</td>
<td>(0.07452)</td>
<td>(0.07801)</td>
<td>(0.07653)</td>
<td>(0.07317)</td>
<td>(0.07447)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=0.5</td>
<td>0.69348</td>
<td>0.64349</td>
<td>0.64255</td>
<td>0.64255</td>
<td>0.64234</td>
<td>0.63912</td>
<td>0.62820</td>
<td>0.60750</td>
</tr>
<tr>
<td></td>
<td>(0.08297)</td>
<td>(0.07555)</td>
<td>(0.07553)</td>
<td>(0.07531)</td>
<td>(0.07630)</td>
<td>(0.07486)</td>
<td>(0.07464)</td>
<td>(0.07127)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=1</td>
<td>0.69686</td>
<td>0.64666</td>
<td>0.64572</td>
<td>0.64572</td>
<td>0.64551</td>
<td>0.64224</td>
<td>0.63114</td>
<td>0.61010</td>
</tr>
<tr>
<td></td>
<td>(0.07704)</td>
<td>(0.07737)</td>
<td>(0.07770)</td>
<td>(0.07770)</td>
<td>(0.07326)</td>
<td>(0.07570)</td>
<td>(0.07160)</td>
<td>(0.07291)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=0.75</td>
<td>0.69284</td>
<td>0.64259</td>
<td>0.64258</td>
<td>0.64258</td>
<td>0.64258</td>
<td>0.64258</td>
<td>0.64258</td>
<td>0.64258</td>
</tr>
<tr>
<td></td>
<td>(0.04215)</td>
<td>(0.04393)</td>
<td>(0.04512)</td>
<td>(0.04270)</td>
<td>(0.04645)</td>
<td>(0.04848)</td>
<td>(0.04337)</td>
<td>(0.04469)</td>
</tr>
<tr>
<td></td>
<td>Fourier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at x=1</td>
<td>0.70197</td>
<td>0.65136</td>
<td>0.65041</td>
<td>0.65041</td>
<td>0.65041</td>
<td>0.65041</td>
<td>0.65041</td>
<td>0.65041</td>
</tr>
<tr>
<td></td>
<td>(0.08101)</td>
<td>(0.07410)</td>
<td>(0.07496)</td>
<td>(0.07221)</td>
<td>(0.07787)</td>
<td>(0.07595)</td>
<td>(0.07508)</td>
<td>(0.07212)</td>
</tr>
<tr>
<td>Observations</td>
<td>658</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors obtained by bootstrap methods are in parentheses. Data are from the Children of the NLSY linked to their mothers in the main NLSY79. All variables are standardized having means of zeros and standard deviations of one. Mother’s AFQT is used as conditioning instrument. Income is after-tax and after-transfer. Error-laden measurement of permanent family income is an average of family incomes in 1994, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income. Year refers to the NLSY survey year; income refers to the previous year’s income.
Table 5.1. Impact of current family income on children's reading achievement

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>at x=-1</td>
<td>0.74166</td>
<td>0.68550</td>
<td>0.68445</td>
<td>0.68445</td>
<td>0.68422</td>
<td>0.68071</td>
<td>0.66881</td>
<td>0.64626</td>
</tr>
<tr>
<td></td>
<td>(0.07067)</td>
<td>(0.06712)</td>
<td>(0.06707)</td>
<td>(0.06352)</td>
<td>(0.06541)</td>
<td>(0.06659)</td>
<td>(0.06139)</td>
<td>(0.06496)</td>
</tr>
<tr>
<td></td>
<td>0.10350</td>
<td>0.10029</td>
<td>0.10023</td>
<td>0.10023</td>
<td>0.10022</td>
<td>0.10008</td>
<td>0.09961</td>
<td>0.09872</td>
</tr>
<tr>
<td></td>
<td>(0.03201)</td>
<td>(0.03287)</td>
<td>(0.03207)</td>
<td>(0.03124)</td>
<td>(0.03212)</td>
<td>(0.03224)</td>
<td>(0.03135)</td>
<td>(0.03387)</td>
</tr>
<tr>
<td>at x=-0.75 and w=-1</td>
<td>0.73213</td>
<td>0.67781</td>
<td>0.67679</td>
<td>0.67679</td>
<td>0.67679</td>
<td>0.67657</td>
<td>0.67315</td>
<td>0.66583</td>
</tr>
<tr>
<td></td>
<td>(0.06972)</td>
<td>(0.06654)</td>
<td>(0.06597)</td>
<td>(0.06607)</td>
<td>(0.06680)</td>
<td>(0.06320)</td>
<td>(0.06223)</td>
<td>(0.06176)</td>
</tr>
<tr>
<td></td>
<td>0.10801</td>
<td>0.10439</td>
<td>0.10432</td>
<td>0.10432</td>
<td>0.10431</td>
<td>0.10414</td>
<td>0.10356</td>
<td>0.10294</td>
</tr>
<tr>
<td></td>
<td>(0.03316)</td>
<td>(0.03416)</td>
<td>(0.03381)</td>
<td>(0.03188)</td>
<td>(0.03226)</td>
<td>(0.03159)</td>
<td>(0.03128)</td>
<td>(0.03213)</td>
</tr>
<tr>
<td>at x=-0.5 and w=-0.5</td>
<td>0.72451</td>
<td>0.67177</td>
<td>0.67077</td>
<td>0.67077</td>
<td>0.67055</td>
<td>0.66720</td>
<td>0.65583</td>
<td>0.64242</td>
</tr>
<tr>
<td></td>
<td>(0.06668)</td>
<td>(0.06625)</td>
<td>(0.06713)</td>
<td>(0.06500)</td>
<td>(0.06310)</td>
<td>(0.06250)</td>
<td>(0.06270)</td>
<td>(0.06215)</td>
</tr>
<tr>
<td></td>
<td>0.11033</td>
<td>0.10652</td>
<td>0.10645</td>
<td>0.10645</td>
<td>0.10643</td>
<td>0.10625</td>
<td>0.10562</td>
<td>0.10443</td>
</tr>
<tr>
<td></td>
<td>(0.03273)</td>
<td>(0.03282)</td>
<td>(0.03320)</td>
<td>(0.03250)</td>
<td>(0.03209)</td>
<td>(0.03125)</td>
<td>(0.03142)</td>
<td>(0.03202)</td>
</tr>
<tr>
<td>at x=0 and w=0</td>
<td>0.71472</td>
<td>0.66441</td>
<td>0.66346</td>
<td>0.66346</td>
<td>0.66325</td>
<td>0.65996</td>
<td>0.64879</td>
<td>0.62759</td>
</tr>
<tr>
<td></td>
<td>(0.06752)</td>
<td>(0.06219)</td>
<td>(0.06327)</td>
<td>(0.06402)</td>
<td>(0.06253)</td>
<td>(0.06562)</td>
<td>(0.06062)</td>
<td>(0.06135)</td>
</tr>
<tr>
<td></td>
<td>0.11269</td>
<td>0.10871</td>
<td>0.10864</td>
<td>0.10864</td>
<td>0.10863</td>
<td>0.10842</td>
<td>0.10773</td>
<td>0.10644</td>
</tr>
<tr>
<td></td>
<td>(0.03206)</td>
<td>(0.03230)</td>
<td>(0.03252)</td>
<td>(0.03180)</td>
<td>(0.03304)</td>
<td>(0.03413)</td>
<td>(0.03230)</td>
<td>(0.03418)</td>
</tr>
<tr>
<td>at x=0.25 and w=0.25</td>
<td>0.71246</td>
<td>0.66303</td>
<td>0.66210</td>
<td>0.66210</td>
<td>0.66188</td>
<td>0.65859</td>
<td>0.64741</td>
<td>0.62620</td>
</tr>
<tr>
<td></td>
<td>(0.06785)</td>
<td>(0.06642)</td>
<td>(0.06310)</td>
<td>(0.06204)</td>
<td>(0.06397)</td>
<td>(0.06343)</td>
<td>(0.06186)</td>
<td>(0.05920)</td>
</tr>
<tr>
<td></td>
<td>0.11508</td>
<td>0.11095</td>
<td>0.11088</td>
<td>0.11088</td>
<td>0.11086</td>
<td>0.11064</td>
<td>0.10998</td>
<td>0.10846</td>
</tr>
<tr>
<td></td>
<td>(0.03167)</td>
<td>(0.03240)</td>
<td>(0.03225)</td>
<td>(0.03268)</td>
<td>(0.03240)</td>
<td>(0.03181)</td>
<td>(0.03152)</td>
<td>(0.03233)</td>
</tr>
<tr>
<td>at x=0.5 and w=0.5</td>
<td>0.71194</td>
<td>0.66317</td>
<td>0.66225</td>
<td>0.66225</td>
<td>0.66203</td>
<td>0.65872</td>
<td>0.64745</td>
<td>0.62609</td>
</tr>
<tr>
<td></td>
<td>(0.06777)</td>
<td>(0.06439)</td>
<td>(0.06427)</td>
<td>(0.06349)</td>
<td>(0.06644)</td>
<td>(0.06164)</td>
<td>(0.06194)</td>
<td>(0.06078)</td>
</tr>
<tr>
<td></td>
<td>0.11756</td>
<td>0.11329</td>
<td>0.11321</td>
<td>0.11321</td>
<td>0.11320</td>
<td>0.11295</td>
<td>0.11212</td>
<td>0.11057</td>
</tr>
<tr>
<td></td>
<td>(0.03282)</td>
<td>(0.03259)</td>
<td>(0.03292)</td>
<td>(0.03273)</td>
<td>(0.03250)</td>
<td>(0.03166)</td>
<td>(0.03153)</td>
<td>(0.03149)</td>
</tr>
<tr>
<td>at x=0.75 and w=0.75</td>
<td>0.71315</td>
<td>0.66483</td>
<td>0.66392</td>
<td>0.66392</td>
<td>0.66370</td>
<td>0.66033</td>
<td>0.64891</td>
<td>0.62726</td>
</tr>
<tr>
<td></td>
<td>(0.06837)</td>
<td>(0.06485)</td>
<td>(0.06385)</td>
<td>(0.06616)</td>
<td>(0.06334)</td>
<td>(0.06356)</td>
<td>(0.06207)</td>
<td>(0.06006)</td>
</tr>
<tr>
<td></td>
<td>0.11998</td>
<td>0.11558</td>
<td>0.11550</td>
<td>0.11550</td>
<td>0.11548</td>
<td>0.11521</td>
<td>0.11431</td>
<td>0.11262</td>
</tr>
<tr>
<td></td>
<td>(0.03212)</td>
<td>(0.03410)</td>
<td>(0.03292)</td>
<td>(0.03242)</td>
<td>(0.03293)</td>
<td>(0.03344)</td>
<td>(0.03193)</td>
<td>(0.03168)</td>
</tr>
<tr>
<td>at x=1 and w=1</td>
<td>0.71610</td>
<td>0.66802</td>
<td>0.66712</td>
<td>0.66712</td>
<td>0.66690</td>
<td>0.66346</td>
<td>0.65180</td>
<td>0.62971</td>
</tr>
<tr>
<td></td>
<td>(0.06455)</td>
<td>(0.06513)</td>
<td>(0.06262)</td>
<td>(0.06241)</td>
<td>(0.06342)</td>
<td>(0.06242)</td>
<td>(0.06206)</td>
<td>(0.06945)</td>
</tr>
<tr>
<td></td>
<td>0.12259</td>
<td>0.11809</td>
<td>0.11800</td>
<td>0.11800</td>
<td>0.11799</td>
<td>0.11769</td>
<td>0.11670</td>
<td>0.11485</td>
</tr>
<tr>
<td></td>
<td>(0.03241)</td>
<td>(0.03543)</td>
<td>(0.03272)</td>
<td>(0.03200)</td>
<td>(0.03295)</td>
<td>(0.03285)</td>
<td>(0.03261)</td>
<td>(0.03165)</td>
</tr>
<tr>
<td>Observations</td>
<td>1274</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors obtained by bootstrap methods are in parentheses. Data are from the Children of the NLSY linked to their mothers in the main NLSY79. All variables are standardized having means of zeros and standard deviations of one. Mother’s AFQT is used as conditioning instrument. Income is after-tax and after-transfer. 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. Reading score is obtained by taking a simple average of the reading recognition and reading comprehension scores. Year refers to the NLSY survey year; income refers to the previous year’s income.
Table 5.2. Impact of permanent family income on children’s reading achievement

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fourier</td>
<td>Local linear</td>
<td>Fourier</td>
<td>Local linear</td>
<td>Fourier</td>
<td>Local linear</td>
<td>Fourier</td>
<td>Local linear</td>
</tr>
<tr>
<td>at x=-1 and w=-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourier</td>
<td>0.71174</td>
<td>(0.09026)</td>
<td>0.71174</td>
<td>(0.09026)</td>
<td>0.71174</td>
<td>(0.09026)</td>
<td>0.71174</td>
<td>(0.09026)</td>
</tr>
<tr>
<td>Local linear</td>
<td>0.11540</td>
<td>(0.04548)</td>
<td>0.11540</td>
<td>(0.04548)</td>
<td>0.11540</td>
<td>(0.04548)</td>
<td>0.11540</td>
<td>(0.04548)</td>
</tr>
</tbody>
</table>

Observations 651

Notes: Standard errors obtained by bootstrap methods are in parentheses. Data are from the Children of the NLSY linked to their mothers in the main NLSY79. All variables are standardized having means of zeros and standard deviations of one. Mother’s AFQT is used as conditioning instrument. Income is after-tax and after-transfer. Error-laden measurement of permanent family income is an average of family incomes in 1994, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income. Reading score is obtained by taking a simple average of the reading recognition and reading comprehension scores. Year refers to the NLSY survey year; income refers to the previous year’s income.
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: 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 2: Causal effects - conditioning instrument and measurement error
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: Causal effects - impact of family income on child achievement
Notes: Our estimator is used for current family income (top) and permanent family income (bottom). Error-laden measurement of current family income is family income in 1998. Error-laden measurement of permanent family income is an average of family incomes in 1994, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 4: Impact of family income on children’s math scores (Fourier)
Notes: Local linear estimator is used for current family income (top) and permanent family income (bottom). Error-laden measurement of current family income is family income in 1998. Error-laden measurement of permanent family income is an average of family incomes in 1994, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 5: Impact of family income on children’s math scores (Local linear)
Notes: Our estimator is used for current family income (top) and permanent family income (bottom). Error-laden measurement of current family income is family income in 1998. Error-laden measurement of permanent family income is an average of family incomes in 1994, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 6: Impact of family income on children’s reading scores (Fourier)
Notes: Local linear estimator is used for current family income (top) and permanent family income (bottom). Error-laden measurement of current family income is family income in 1998. Error-laden measurement of permanent family income is an average of family incomes in 1996, 1996 and 1998. Family income in 2000 is used as additional error-laden measurement of family income.

Figure 7: Impact of family income on children’s reading scores (Local linear)