Estimating average marginal effects in nonseparable structural systems

Susanne Schennach
Halbert White
Karim Chalak

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP31/07

Estimating Average Marginal Effects in Nonseparable Structural Systems

Susanne Schennach
University of Chicago

Halbert White
UC San Diego

Karim Chalak
Boston College

December 3, 2007

Abstract

We provide nonparametric estimators of derivative ratio-based average marginal effects of an endogenous cause, X, on a response of interest, Y, for a system of recursive structural equations. The system need not exhibit linearity, separability, or monotonicity. Our estimators are local indirect least squares estimators analogous to those of Heckman and Vytlacil (1999, 2001) who treat a latent index model involving a binary X. We treat the traditional case of an observed exogenous instrument (OXI) and the case where one observes error-laden proxies for an unobserved exogenous instrument (PXI). For PXI, we develop and apply new results for estimating densities and expectations conditional on mismeasured variables. For both OXI and PXI, we use infinite order flat-top kernels to obtain uniformly convergent and asymptotically normal nonparametric estimators of instrument-conditioned effects, as well as root-n consistent and asymptotically normal estimators of average effects.

Acknowledgement 0.1 *We thank Stephan Hoderlein, Andres Santos, and Suyong Song for their comments. Any errors are the authors' responsibility. S. M. Schennach acknowledges support from the National Science Foundation via grant SES-0452089.*

JEL Classification Numbers: C13, C14, C31.
Keywords: derivative ratio effect, endogeneity, indirect least squares, instrumental variables, measurement error, nonparametric estimator, nonseparable structural equation.

1 Introduction

This paper studies identification and estimation of measures of the marginal effect of an endogenous cause in a system of structural equations with exogenous instruments. As in Altonji and Matzkin (2005), Hoderlein (2005[1]), and Hoderlein and Mammen (2007),

[1] A more recent version of this working paper is available as Hoderlein (2007).

our structural equations involve general measurable functions: we do not impose linearity, monotonicity, or separability. Our estimators are correspondingly nonparametric. Our results complement the work of these authors, and they complement and extend prior work on nonparametric instrumental variables (IV) methods imposing separability or monotonicity, such as that of Angrist and Imbens (1994), Angrist, Imbens, and Rubin (1996), Heckman (1997), Heckman and Vytlacil (1999, 2001, 2005), Blundell and Powell (2000), Chesher (2003), Darolles, Florens, and Renault (2003), Imbens and Newey (2003), Matzkin (2003, 2004), Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2006), Heckman, Urzua, and Vytlacil (2006), Santos (2006), and Hahn and Ridder (2007) among others.

As Darolles, Florens, and Renault (2003, p.1-2) note, several different notions of instrumental variables appear in the nonlinear IV literature. Chalak and White (2007a) (CW) propose a taxonomy of instruments based on their role in identifying structural effects of interest. Although CW's taxonomy is developed in the linear parametric context, it applies generally, as they explicitly note. Among the various possibilities, we focus here on the use of classical exogenous instruments to study effect measures constructed as ratios of certain derivatives, derivative ratio (DR) effect measures, for short. The motivations for considering DR effects are several: First, in classical linear structural systems with exogenous instruments, these effects motivate and underlie Haavelmo's (1943) classical method of indirect least squares (ILS). In more general systems, Heckman (1997) and Heckman and Vytlacil (1999, 2001, 2005) show that DR effects correspond to a variety of structurally meaningful weighted averages of effects of interest; the corresponding estimators are "local IV," or, more aptly, local ILS (LILS) estimators (as suggested by Heckman and Vytlacil, 2007). The intuitive appeal of DR effect measures in these cases and the computational ease of the associated LILS estimators make DR effect measures a natural candidate for application to the general case.

Alternatives to the use of exogenous instruments include the use of conditioning instruments, which deliver a conditional independence relation central to structural identification. This gives a different class of effect measures, precisely those considered by Altonji and Matzkin (2005), Hoderlein (2005), White and Chalak (2006), Hoderlein and Mammen (2007), and Chalak and White (2007b), among others, for general nonseparable systems.

We pay particular attention to the structural content and interpretation of DR effect measures. As we discuss, their relative ease of interpretation hinges crucially on whether or not the structural equation determining the endogenous cause of interest, say X, is separable, regardless of the separability of the structural equation relating the response of interest, say Y, to X. When X is separably determined, the structural content of the derivative ratio as a measure of average marginal effect is easily appreciated, even for nonseparably determined Y.

In the fully nonseparable case, the DR effect is still a measure of a well-defined weighted average marginal effect. Nonseparability for X leads only to changes in the weighting functions employed in constructing the average derivatives of interest. Interestingly, these weight changes typically do not preclude the use of DR effects to test the hypothesis that an endogenous cause X has an effect on the response Y. Knowing how DR effect measures behave in the general case also provides the necessary foundation for formal tests of the properties of the underlying structure, such as whether X is separably determined or not.

We study two cases elucidated by CW: the traditional *observed exogenous instrument* (OXI) case, where the exogenous instrument is observed without error; and the *proxies for unobserved exogenous instrument* (PXI) case, where the exogenous instrument is not directly observable, but error-contaminated measurements are available to serve as proxy instruments, as in Butcher and Case (1994). In the linear parametric case treated by CW, estimation methods for the OXI and PXI cases are identical, despite the interesting fact that in the PXI case, the (error-laden) proxy instruments are correlated with the reduced form errors, yielding inconsistent reduced form estimators. As CW explain, ILS and standard

IV methods generally yield consistent estimators of the effects of interest nevertheless.

Once one goes beyond the linear parametric case, however, the OXI and PXI cases require fundamentally different estimation methods. We can treat the OXI case using any of a variety of familiar nonparametric methods, such as kernel or sieve methods. The PXI case demands an innovative approach, however. In fact, our PXI results are the first to cover nonparametric generalizations of the linear parametric case using instrument proxies.

For the OXI case, we apply infinite order ("flat-top") kernels (Politis and Romano, 1999) to estimate functionals of the distributions of the observable variables that we then combine to obtain new estimators of the average marginal effect represented by the DR effect measure. We obtain uniform convergence rates and asymptotic normality for estimators of instrument-conditioned average marginal effects as well as root-n consistency and asymptotic normality for estimators of their weighted averages.

For the PXI case, we build on recent results of Schennach (2004a, 2004b) to obtain a variety of new results. Specifically, we show that two error-contaminated measurements of the unobserved exogenous instrument are sufficient to identify objects of interest and to deliver consistent estimators. Our general estimation theory covers densities of mismeasured variables and expectations conditional on mismeasured variables, as well as their derivatives with respect to the mismeasured variable. We provide uniform convergence rates over expanding intervals (and, in some cases, over the whole real line) as well as asymptotic normality results in fully nonparametric settings. We also consider nonlinear functionals of such nonparametric quantities and establish their root-n consistency and asymptotic normality. This analysis thus provides numerous general-purpose asymptotic results of independent interest, beyond the PXI case.

The plan of the paper is as follows. In Section 2 we specify a recursive structural system that generates the data and define the DR effect measures of interest. We provide formal conditions ensuring the identification of the DR effect measures, that is, the equality of the counterfactually based effects of interest with well-defined standard stochastic objects. We devote particular attention to the interpretation of these DR effect measures in a range of special cases. As mentioned above, DR effect measures are naturally estimated by nonparametric local ILS methods. Section 3 treats the OXI case. We provide results establishing consistency and asymptotic normality for our nonparametric estimators. Section 4 develops new general results for estimation of densities and functionals of densities of mismeasured variables. As an application, we treat the PXI case, ensuring the identification of the objects of interest and providing estimation results analogous to those of Section 3. Section 5 contains a discussion of the results, and Section 6 provides a summary and discussion of directions for future research. All proofs are gathered into the Mathematical Appendix.

2 Data Generation and Structural Identification

2.1 Data Generation

We begin by specifying a recursive structural system that generates the data. In such systems, there is an inherent ordering of the system variables: "predecessor" variables may determine "successor" variables, but not vice versa. For example, when X determines Y, then Y cannot determine X. In such cases, we say for convenience that Y *succeeds* X, and we write $Y \Leftarrow X$ as a shorthand notation.

Assumption 2.1 *Let a recursive structural system generate random variables* $\{U, X, Y, Z\}$ *such that* $Y \Leftarrow (U, X, Z)$, $X \Leftarrow (U, Z)$, *and* $Z \Leftarrow U$. *In addition:* (i) *Let* υ_x, υ_y, *and* υ_z *be measurable functions such that* $U_x \equiv \upsilon_x(U), U_y \equiv \upsilon_y(U), U_z \equiv \upsilon_z(U)$ *are random vectors of countable dimensions* ℓ_x, ℓ_y, *and* ℓ_z *respectively;* (ii) (X, Y, Z) *is generated as*

$$Z \stackrel{c}{=} p(U_z)$$

$$X \stackrel{c}{=} q(Z, U_x)$$

$$Y \stackrel{c}{=} r(X, U_y),$$

where $p, q,$ *and* r *are unknown measurable scalar-valued functions;* (iii) $E(X)$ *and* $E(Y)$ *are finite;* (iv) *The realizations of* X *and* Y *are observed; those of* U *are not.*

We consider scalar $X, Y,$ and Z for simplicity; extensions are straightforward. We explicitly assume observability of X and Y and unobservability of U. We separately treat cases in which Z is observable (Section 3) or unobservable (Section 4). An important feature here is that the unobserved causes $U_x, U_y,$ and U_z may be multi-dimensional. Indeed, the unobserved causes need not even be finite dimensional.

The response functions $p, q,$ and r embody the structural (causal) relations between the system variables. (In what follows we use "structural" and "causal" synonymously.) We use the $\overset{c}{=}$ notation to emphasize the causal structure of these relations, as in CW. Assuming only measurability for $p, q,$ and r permits but does not require linearity, monotonicity in variables, or separability between observables and unobservables.

The structure of Assumption 2.1 can arise in numerous economic applications. For example, this structural system can correspond to a nonparametric demand system with a heterogeneous population, as in Hoderlein (2005).

Our interest attaches to the effect of X on Y. Specifically, when the derivative exists, consider the marginal effect of continuously distributed X on Y, $D_x r(X, U_y)$, where $D_x \equiv (\partial/\partial x)$. If r were linear and separable, say,

$$r(X, U_y) = X\beta_0 + U_y'\alpha_y,$$

then $D_x r(X, U_y) = \beta_0$. Generally we will not require linearity or separability, so $D_x r(X, U_y)$ is no longer constant but generally depends on both X and U_y. To handle dependence on the unobservable U_y, we consider certain average marginal effects, defined below.

Generally, X and U_y may be correlated or otherwise dependent, in which case X is "endogenous." Just as in the linear separable case, when X is endogenous, the availability of suitable instrumental variables permits identification and estimation of effects of interest.

The structure above permits Z to play this instrumental role, given a suitable exogeneity condition. To specify this, we follow Dawid (1979) and write $X \perp Y$ when random variables X and Y are independent and $X \not\perp Y$ otherwise.

Assumption 2.2 $U_z \perp (U_x, U_y)$.

As we make no assumption regarding the relation of U_x and U_y, we may have $U_x \not\perp U_y$, which, given Assumption 2.1, implies that X is endogenous: $X \not\perp U_y$. On the other hand, Assumptions 2.1 and 2.2 imply $Z \perp (U_x, U_y)$, so that Z is exogenous with respect to both U_x and U_y in the classical sense.

As discussed in CW, a variety of different conditional independence relations and exclusion restrictions can be employed to identify effects of interest. For example, with the structure of Assumption 2.1, certain structural effects can be identified using Hoderlein's (2005) assumption 2.3, which states that $Z \perp U_y \mid U_x$, and where U_x is assumed to be identified due to further structure, "e.g. monotonicity of [q] in [U_x] and [$Z \perp U_x$]" (Hoderlein, 2005, p. 5). But this implies that $X \perp U_y \mid U_x$, a conditional independence assumption similar to that imposed in Altonji and Matzkin (2005), White and Chalak (2006), and Hoderlein and Mammen (2007). Such conditional independence conditions are neither necessary nor sufficient for Assumption 2.2, and, as is apparent by inspection, the structural effects identified under the various exogeneity conditions can easily differ. Which exogeneity condition is appropriate in any particular instance depends on the specifics of the economic structure, as extensively discussed by CW.

2.2 Identification

In classical linear separable structural systems with exogenous instruments, the effect of X on Y can be recovered from the reduced form as the ratio of the effect of Z on Y to that of Z on X; the effect of interest can then be estimated using Haavelmo's (1943) ILS method. In more general cases, information about the marginal effect of X on Y can similarly be extracted, based on the ratio of the marginal effect of Z on Y to that of Z on X, that is, as a derivative ratio.

To see how this works, consider first the effect of Z on X. The starting point for a study of this effect is the conditional expectation of X given $Z = z$,

$$\begin{aligned} \mu_X(z) &\equiv E(X \mid Z = z) && (1) \\ &= \int q(z, u_x) dF(u_x|z), && (2) \end{aligned}$$

where $dF(u_x|z)$ denotes the conditional density of U_x given $Z = z$. The existence of μ_X in eq.(1) is guaranteed whenever $E(X) < \infty$, regardless of any underlying structure. Thus, μ_X is stochastically meaningful whenever it exists, as it is simply an aspect of the joint distribution of X and Z.

If the structure provided by Assumptions 2.1(*i-iii*) holds and the conditional distribution of U_x given Z is regular (e.g., Dudley, 2002, ch.10.2), then the integral representation of eq.(2) also holds. (In what follows, we implicitly assume the regularity of all referenced conditional distributions.) Eq.(2) provides μ_X with some structural content; specifically it is an average response. As we discuss shortly, there is nevertheless not yet sufficient content to use μ_X to identify effects of interest.

When Z does not determine U (recall Assumption 2.1 ensures $Z \Leftarrow U$), the structurally meaningful *average counterfactual response* of X to Z is given by

$$\rho_X(z) \equiv \int q(z, u_x) \; dF(u_x), \qquad (3)$$

where $dF(u_x)$ denotes the unconditional density of U_x. Given differentiability of q and an interchange of integral and derivative (see, e.g., White and Chalak (2006, theorem 3.3(ii)),

$$D_z\rho_X(z) = \int D_z q(z, u_x) \; dF(u_x), \qquad (4)$$

ensuring that $D_z\rho_X$ represents the average marginal effect of Z on X.

Our assumptions ensure that Z is exogenous with respect to U_x ($Z \perp U_x$), so that

$$\int q(z, u_x)\ dF(u_x|z) = \int q(z, u_x)\ dF(u_x),$$

as $Z \perp U_x$ implies $dF(u_x|z) = dF(u_x)$. That is, $\mu_X = \rho_X$. Moreover, $D_z\mu_X = D_z\rho_X$, so μ_X is now fully informative about the structurally meaningful $D_z\rho_X$. When, as is true here, stochastic objects like μ_X are identified with a structurally meaningful object, we say that they are *structurally identified*. Similarly, when structurally meaningful objects like ρ_X are identified with stochastic objects, we say they are *stochastically identified*. If stochastic identification holds uniquely with a representation solely in terms of observable random variables, then we say that both the stochastic object and its structural counterpart are *fully identified*. Thus, with μ_X and $D_z\mu_X$ fully identified, both ρ_X and $D_z\rho_X$ can be estimated from data under mild conditions.

Similarly, we can write

$$\mu_Y(z) \equiv E(Y \mid Z = z) \tag{5}$$

$$= \int r(q(z, u_x), u_y)\ dF(u_x, u_y|z), \tag{6}$$

where $dF(u_x, u_y|z)$ denotes the conditional density of (U_x, U_y) given $Z = z$. The finiteness of $E(Y)$ ensures that μ_Y exists. In the absence of further assumptions, μ_Y is also purely a stochastic object. The integral representation of eq.(6) holds under Assumptions 2.1(*i-iii*). The requirement that Z succeeds U and the exogeneity of Z with respect to (U_x, U_y), jointly ensured by Assumptions 2.1 and 2.2, structurally identify μ_Y as the average counterfactual response of Y to Z. That is, $\mu_Y = \rho_Y$, where

$$\rho_Y(z) \equiv \int r(q(z, u_x), u_y)\ dF(u_x, u_y), \tag{7}$$

and $dF(u_x, u_y)$ denotes the unconditional density of (U_x, U_y).

Further, given differentiability, the derivative $D_z\mu_Y$ is structurally identified as $D_z\rho_Y$. We can interpret this as an average marginal effect of Z on Y. Specifically, given differentiability of q and r and the interchange of derivative and integral, we have

$$D_z \rho_Y(z) = \int D_z[r(q(z,u_x),u_y)]\ dF(u_x,u_y). \tag{8}$$

This involves the marginal effect of X on Y as a consequence of the chain rule:

$$\begin{aligned} D_z \rho_Y(z) &= \int D_x r(q(z,u_x),u_y)\ D_z q(z,u_x)\ dF(u_x,u_y) \\ &= \int \left[\int D_x r(q(z,u_x),u_y) dF(u_y|\ u_x)\right]\ D_z q(z,u_x)\ dF(u_x), \end{aligned}$$

where $dF(u_y|u_x)$ denotes the conditional density of U_y given $U_x = u_x$.

The analog of the ratio of reduced form coefficients exploited by Haavelmo's (1943) ILS estimator is the derivative ratio

$$\beta(z) \equiv D_z \mu_Y(z)\ /\ D_z \mu_X(z). \tag{9}$$

This ratio is a population analog of the local ILS estimator, introduced by Heckman and Vytlacil (1999, 2001) as a "local instrumental variable" for a case with X binary and $q(z,u_x) = 1\{q_1(z) - u_x \geq 0\}$. As defined, $\beta(z)$ is purely a stochastic object.

Observe that $\beta(z)$ is well defined only when the numerator and denominator are well defined and the denominator does not vanish. The latter condition is the analog of the classical requirement that the instrumental variable Z must be "relevant." We thus define the *support* of β to be the set on which $\beta(z)$ is well defined, $S_\beta \equiv \{z : f_Z(z) > 0, |D_z \mu_X(z)| > 0\}$, where $f_Z(\cdot)$ is the density of Z. The requirement that $f_Z(z) > 0$ ensures that both $D_z \mu_Y(z)$ and $D_z \mu_X(z)$ are well defined. When $X, Y,$ and Z are observable, we may consistently estimate β on its support under mild conditions; this is the subject of Section 3. We show in Section 4 that we can consistently estimate β even when Z is not observable.

2.3 Interpreting DR Effects

If both the numerator and denominator of $\beta(z)$ are structurally identified, then so is $\beta(z)$. In particular, when $\beta(z)$ is structurally identified, it represents a specific weighted average of the marginal effect of interest, $D_x r(X, U_y)$, as the expressions above imply $\beta = \beta^*$, where

$$\beta^*(z) \equiv D_z \rho_Y(z) \; / \; D_z \rho_X(z) \tag{10}$$

$$= \int \; [\int D_x r(q(z, u_x), u_y) \; dF(u_y | \; u_x)] \; \varsigma(z, u_x) \; dF(u_x), \tag{11}$$

for $z \in S_{\beta^*} \equiv \{z : f_Z(z) > 0, |D_z \rho_X(z)| > 0\}$. The weights $\varsigma(z, u_x)$ are given by

$$\varsigma(z, u_x) \equiv D_z q(z, u_x) \; / \; \int D_z q(z, u_x) \; dF(u_x),$$

and for each $z \in S_{\beta^*}$,

$$\int \varsigma(z, u_x) \; dF(u_x) = 1.$$

We can also represent $\beta^*(z)$ and $\varsigma(z, U_x)$ in terms of certain conditional expectations. Specifically, under our assumptions, we have

$$\beta^*(z) = E[\; E(D_x r(X, U_y) \mid Z = z, U_x) \; \varsigma(z, U_x) \;]$$

$$\varsigma(z, U_x) = D_z q(z, U_x) \; / \; E(D_z q(Z, U_x) \mid Z = z).$$

Thus, $\beta^*(z)$ provides a measure of average marginal effect that emphasizes $E(D_x r(X, U_y) \mid Z = z, U_x)$ for values of $D_z q(z, U_x)$ that are large relative to $E(D_z q(Z, U_x) \mid Z = z)$.

Note that, while the weights $\varsigma(z, U_x)$ can be negative, they are necessarily positive when $q(z, u_x)$ is strictly monotone in z for almost all u_x, with common sign for $D_z q(z, u_x)$. This is often plausible given that Z is an instrument for X. In this case, an estimator of $\beta^*(z)$ can clearly be used to test the null hypothesis that X has no effect on Y, since then $\beta^*(z) = 0$ if and only if $\int D_x r(q(z, u_x), u_y) \; dF(u_y | \; u_x) = 0$ for almost every u_x in the support of U_x.

To gain further insight, we consider the form taken by β^* in some important special cases. First, when r is linear, we have $r(x, u_y) = x\beta_0 + u_y$. It is immediate that regardless of the form of q, $\beta^*(z) = \beta_0$ for all $z \in S_{\beta^*}$.

Next, consider the case where X is separably determined: $q(z, u_x) = q_1(z) + u_x$. (There is no loss of generality in specifying scalar u_x under separability.) In this case we have

$\varsigma(z, u_x) \equiv 1$ for $z \in S_{\beta^*}$. When r is also separable, so that $r(x, u_y) = r_1(x) + u_y$ (see, e.g., Newey and Powell, 2003; Darolles, Florens, and Renault, 2003), we have $\beta^*(z) = \beta^*_{ss}(z)$ for $z \in S_{\beta^*}$, where

$$\begin{aligned}\beta^*_{ss}(z) &\equiv \int D_x r_1(q_1(z) + u_x)\ dF(u_x) \\ &= E(D_x r_1(X) \mid Z = z).\end{aligned}$$

In fact, separability for r does not play a critical role; when r is nonseparable we have $\beta^*(z) = \beta^*_{ns}(z)$ for $z \in S_{\beta^*}$, where

$$\begin{aligned}\beta^*_{ns}(z) &\equiv \int D_x r(q(z, u_x), u_y)\ dF(u_y, u_x) \\ &= E(D_x r(X, U_y) \mid Z = z).\end{aligned}$$

Both β^*_{ss} and β^*_{ns} are easily interpretable quantities.

It remains to consider nonseparable q. First, when r is separable, we have $\beta^*(z) = \beta^*_{sn}(z)$ for $z \in S_{\beta^*}$, where

$$\begin{aligned}\beta^*_{sn}(z) &\equiv \int D_x r_1(q(z, u_x))\ \varsigma(z, u_x)\ dF(u_x) \\ &= E[\ E(D_x r_1(X) \mid Z = z)\ \varsigma(z, U_x)\ ].\end{aligned}$$

This is still a weighted average of an expected marginal effect, namely $E(D_x r_1(X) \mid Z = z)$, but now the nonseparability of q necessitates the presence of the weights $\varsigma(z, U_x)$. When r is nonseparable, we are back to the general case, with $\beta^*_{nn}(z) \equiv \beta^*(z)$ for $z \in S_{\beta^*}$.

To gain additional insight for nonseparable q, we note that the independence imposed in Assumption 2.2 ensures $E[\ E(D_x r(X, U_y) \mid Z = z, U_x)\ ] = E(D_x r(X, U_y) \mid Z = z)$. Adding and subtracting this in the expression for $\beta^*_{nn}(z)$, we get

$$\beta^*_{nn}(z) = E(D_x r(X, U_y) \mid Z = z) - E[\ E(D_x r(X, U_y) \mid Z = z, U_x)\ (1 - \varsigma(z, U_x))\ ].$$

From the fact that $E[\varsigma(z, U_x)] = 1$ and (e.g.) Cauchy-Schwarz, it follows that when suitable second moments exist,

$$|\beta^*_{nn}(z) - E(D_x r(X, U_y) \mid Z = z)|\ \leq \delta(z)\ \sigma_{\varsigma}(z),$$

where

$$\delta^2(z) \equiv E[\ \{\ E(D_x r(X, U_y) \mid Z = z, U_x) - E(D_x r(X, U_y)|Z = z)\ \}^2\ ]$$

is a measure of the conditional variation of $D_x r(X, U_y)$, and

$$\sigma_\varsigma^2(z) \equiv E[\ (1 - \varsigma(z, U_x))^2\ ]$$

is a measure of the departure of q from separability. Thus, we see that the smaller are either $\delta(z)$ or $\sigma_\varsigma(z)$, the closer $\beta^*_{nn}(z)$ is to the simple average derivative

$$\beta^*_{ns}(z) = E(D_x r(X, U_y) \mid Z = z).$$

From these results, we see that DR effects generally deliver a measure of average marginal effect. This is perfectly straightforward to interpret when X is separably determined. The measure is more nuanced otherwise, due to the presence of the weights $\varsigma(z, U_x)$. For all the reasons discussed in the introduction, however, the less obvious interpretation of DR effects in the general case by no means renders them uninteresting.

Another interesting case arises when q is nonseparable but an "index monotonicity" relation holds. Let $X \stackrel{c}{=} q(Z, U_x)$, for vector-valued U_x. There always exist measurable functions V_x and q_2, *scalar*- and vector-valued respectively, such that $U_x = q_2(V_x)$ and q_2 is one-to-one, so that $V_x = q_2^{-1}(U_x)$. Further, the independence of Z and U_x ensures independence of Z and V_x (and vice versa). Let $q_1(Z, V_x) \equiv q(Z, q_2(V_x))$. Then we also have $q(Z, U_x) = q_1(Z, q_2^{-1}(U_x))$. In this representation, the unobservables enter as a scalar index, $V_x = q_2^{-1}(U_x)$. Thus, such a scalar index representation $X \stackrel{c}{=} q_1(Z, V_x)$ always exists.

If in addition q_1 is such that $q_1(z, v_x)$ is monotone in v_x for each z, we say that *index monotonicity* holds for q. This special case parallels the assumption of "monotonicity of the endogenous regressor in the unobserved component" in Imbens and Newey (2003) (see also Chesher (2003) and Matzkin (2003), for example). In this case, an explicit expression for q_1 can be given along the lines of Imbens and Newey (2003) or Hoderlein (2005). Specifically, let V_x have the uniform distribution. (This can always be ensured. If $\tilde{V}_x$ is non-uniform with distribution $\tilde{F}$, then $V_x = \tilde{F}(\tilde{V}_x)$ is uniform.) Let $F(x|\ z)$ denote the conditional CDF of X given $Z = z$. As $V_x = F(X|Z)$ is uniform and $F(\cdot \mid z)$ is invertible, we have $X = F^{-1}(V_x|\ Z)$, where $F^{-1}(\cdot \mid z)$ is the inverse of $F(\cdot \mid z)$ with respect to its first argument. Further, $F^{-1}(v_x|\ z)$ is monotone in v_x for each z. As q_1 is monotone in v_x for each z, it must be that

$$q_1(z, v_x) = F^{-1}(v_x|\ z).$$

Further, when X and Z are observable, $V_x = F(X \mid Z)$ can be consistently estimated. The same is true for q_1 and $D_z q_1$.

To examine the identification of effects of interest with index monotonicity, define

$$\tilde{\mu}_Y(z, v_x) \equiv E(Y \mid Z = z, V_x = v_x) = \int r(q_1(z, v_x), u_y)\ dF(u_y|\ z, v_x) \qquad \text{and}$$

$$\tilde{\rho}_Y(z, v_x) \equiv \int r(q_1(z, v_x), u_y)\ dF(u_y|\ v_x).$$

Under exogeneity, we have structural identification: $\tilde{\mu}_Y = \tilde{\rho}_Y$. This suggests an alternative average effect measure when r and q are nonseparable and $D_z q_1(z, v_x) \neq 0$, namely

$$\beta^*_m(z, v_x) \equiv D_z \tilde{\rho}_Y(z, v_x)\ /\ D_z q_1(z, v_x) = \int D_x r(q_1(z, v_x), u_y)\ dF(u_y|v_x).$$

Averaging this over V_x (equivalently U_x) gives

$$\begin{aligned}\bar{\beta}^*_m(z) &\equiv \int \beta^*_m(z, v_x)\ dF(v_x) \\ &= \int D_x r(q_1(z, v_x), u_y)\ dF(u_y, v_x) \equiv \beta^*_{ns}(z).\end{aligned}$$

Now let $\beta_m(z, v_x) \equiv D_z \tilde{\mu}_Y(z, v_x)\ /\ D_z q_1(z, v_x)$, and define

$$\bar{\beta}_m(z) \equiv \int \beta_m(z, v_x)\ dF(v_x|z).$$

Under mild conditions, exogeneity ensures structural identification: $\bar{\beta}_m = \bar{\beta}^*_m$. Thus, full identification of $D_z \mu_Y$ and index monotonicity for q ensure that we can fully identify and estimate $\bar{\beta}^*_m = \beta^*_{ns}$, even when q is nonseparable. (When q is separable, index monotonicity necessarily holds.) As β^*_{ns} and β^*_{nn} generally differ in the absence of index monotonicity for q, it is possible to test this property by comparing estimators of $\bar{\beta}^*_m$ and β^*_{nn}. Here we leave aside formal analysis of $\bar{\beta}_m$, as index monotonicity is a strong assumption, as emphasized in Hoderlein and Mammen (2007). This is especially so when the unobservables are vector-valued; further, the estimation theory is much more involved than that for β. We also leave formal treatment of tests for separability or index monotonicity to other work.

2.4 Formal Identification Results

We now record our identification results as formal statements. These succinctly summarize our discussion above and serve as a later reference. For these results, we let supp$(\cdot)$ denote the support of the indicated random variable, that is, the smallest Borel set that contains the indicated random variable with probability one. Proposition 2.1 formalizes existence of the relevant stochastic objects, Proposition 2.2 formalizes structural identification, and Proposition 2.3 formalizes possible forms for β^*.

Proposition 2.1 *Suppose that* (X, Y, Z) *are random variables such that* $E(X)$ *and* $E(Y)$ *are finite.* (i) *Then there exist measurable real-valued functions* μ_X *and* μ_Y *defined on* supp(Z) *by eqs.*(1) *and* (5). (ii) *Suppose also that* μ_X *and* μ_Y *are differentiable on* supp(Z). *Then there exists a measurable real-valued function* β *defined on* S_β *by eq.*(9).

Proposition 2.2 *Suppose Assumptions 2.1(i)-(iii) and Assumption 2.2 hold.* (i) *Then there exist measurable real-valued functions* ρ_X *and* ρ_Y *defined on* supp(Z) *by eqs.*(3) *and* (7) *respectively. Further, eqs.*(2) *and* (6) *hold, so that* μ_X *and* μ_Y *are structurally identified on* supp(Z) *as* $\mu_X = \rho_X$ *and* $\mu_Y = \rho_Y$. (ii) *Suppose also that* μ_X *and* μ_Y *are differentiable on* supp(Z). *Then* ρ_X *and* ρ_Y *are differentiable on* supp(Z), *and* $D_z\mu_X$ *and* $D_z\mu_Y$ *are structurally identified on* supp(Z) *as* $D_z\mu_X = D_z\rho_X$ *and* $D_z\mu_Y = D_z\rho_Y$. *In addition, there exists a measurable real-valued function* β^* *defined on* S_{β^*} *by eq.*(10), *and* β *is structurally identified on* $S_\beta = S_{\beta^*}$ *as* $\beta = \beta^*$. (iii) *If Assumption 2.1(iv) also holds and* μ_X *and* μ_Y

have representations in terms of observable random variables, then $\rho_X, \rho_Y, D_z\rho_X$, *and* $D_z\rho_Y$ *are fully identified on* $\text{supp}(Z)$, *and* β *and* β^* *are fully identified on* $S_\beta = S_{\beta^*}$.

Proposition 2.3 *Suppose the conditions of Proposition 2.2 hold and that* $z \to q(z, u_x)$ *is differentiable on* $\text{supp}(Z)$ *for each* $u_x \in \text{supp}(U_x)$ *and* $x \to r(x, u_y)$ *is differentiable on* $\text{supp}(X)$ *for each* $u_y \in \text{supp}(U_y)$. (i) *If eqs.*(4) *and* (8) *hold for each* $z \in \text{supp}(Z)$, *then eq.*(11) *holds, so* $\beta^*(z) = \beta^*_{nn}(z)$ *for all* $z \in S_{\beta^*}$. (ii) *Further, for all* $z \in S_{\beta^*}$: (a) *if* r *is linear, then* $\beta^*(z) = \beta_0$; (b) *if* r *and* q *are separable, then* $\beta^*(z) = \beta^*_{ss}(z)$; (c) *if* q *is separable and* r *is nonseparable, then* $\beta^*(z) = \beta^*_{ns}(z)$; (d) *if* q *is nonseparable and* r *is separable, then* $\beta^*(z) = \beta^*_{sn}(z)$; *and* (e) *if* q *and* r *are nonseparable and an index monotonicity condition holds for* q, *then* $\bar{\beta}^*_m(z) = \beta^*_{ns}(z)$.

Several remarks are in order. First, Proposition 2.1 makes no reference at all to any underlying structure: it applies to any random variables. Next, note that the identification results of Propositions 2.1 and 2.2 do not require that X is continuously distributed or that q or r are differentiable, as these conditions are not necessary for the existence of $D_z\mu_X$ or $D_z\mu_Y$. In such cases, the specific representations of Proposition 2.3 do not necessarily hold, as differentiability for q and r is explicitly required there. Nevertheless, β^* can still have a useful interpretation as a generalized average marginal effect, similar to that analyzed by Heckman and Vytlacil (2007). For brevity and conciseness, we leave aside a more detailed examination of these possibilities here. Finally, we need not require that Z is everywhere continuously distributed; local versions of these results hold on open neighborhoods where Z is continuously distributed.

2.5 Estimation Framework

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

$$\beta^*_w \equiv \int_{S_{\beta^*}} \beta^*(z)\, w(z)\; dz \quad \text{or} \quad \beta^*_{wf_Z} \equiv \int_{S_{\beta^*}} \beta^*(z)\, w(z)\, f_Z(z)\, dz,$$

where $w(\cdot)$ is a user-supplied weight function. Tables 1A and 1B in Heckman and Vytlacil (2005) summarize the appropriate weights needed to generate policy parameters of interest, such as the average treatment effect or the effect of treatment on the treated, in the context of a latent index model. Under structural identification, we have $\beta_w^* = \beta_w$ and $\beta_{wf_Z}^* = \beta_{wf_Z}$, where

$$\beta_w \equiv \int_{S_\beta} \beta(z)\, w(z)\ dz \quad \text{and} \quad \beta_{wf_Z} \equiv \int_{S_\beta} \beta(z)\, w(z)\, f_Z(z)\, dz. \tag{12}$$

We thus focus on estimating stochastically identified β, β_w, and β_{wf_Z}.

To encompass these objects, we focus on estimating quantities of the general form

$$g_{V,\lambda}(z) \equiv D_z^\lambda \left(E\left[V \mid Z = z\right]\ f_Z(z)\right), \tag{13}$$

where $D_z^\lambda \equiv (\partial^\lambda / \partial z^\lambda)$ denotes the derivative operator of degree λ, and V is a generic random variable that will stand for X, Y, or the constant ($V \equiv 1$).

Note that special cases of eq.(13) include densities

$$f_Z(z) = g_{1,0}(z),$$

conditional expectations

$$\mu_Y(z) = g_{Y,0}(z)\ /\ g_{1,0}(z),$$

and, when they exist, their derivatives

$$D_z \mu_Y(z) = \frac{g_{Y,1}(z)}{g_{1,0}(z)} - \frac{g_{Y,0}(z)}{g_{1,0}(z)} \frac{g_{1,1}(z)}{g_{1,0}(z)}.$$

Once we know the asymptotic properties of estimators of $g_{V,\lambda}(z)$, we easily obtain the asymptotic properties of estimators of $\beta(z)$, β_w, or β_{wf_Z}.

As discussed above, we treat two distinct cases. In the first case (OXI), we observe Z, ensuring that X, Y, and Z permit estimation of β and related objects of interest. In the second case (PXI), we do not observe Z but instead observe a proxy $Z_1 \stackrel{c}{=} Z + U_1$ (with $U_1 \perp Z$). In the absence of further information, β is no longer empirically accessible.

The difficulty can be seen as follows. Under our assumptions, Z_1 is a "valid" and "relevant" standard instrument; thus, for linear r and q, we can structurally identify $D_z\mu_{Y,1}(z)$ / $D_z\mu_{X,1}(z_1) = cov(Y, Z_1)/cov(X, Z_1) = cov(Y, Z)/cov(X, Z) = D_z\mu_Y(z)$ / $D_z\mu_X(z)$ as $D_z\rho_Y(z)$ / $D_z\rho_X(z) = \beta_0$, where $\mu_{Y,1}(z) \equiv E(Y \mid Z_1 = z)$ and $\mu_{X,1}(z) \equiv E(X \mid Z_1 = z)$. This fails without linearity, as $D_z\mu_{Y,1}(z)$ / $D_z\mu_{X,1}(z)$ generally differs from $D_z\mu_Y(z)$ / $D_z\mu_X(z)$. Thus, even with structural identification of $D_z\mu_Y(z)$ / $D_z\mu_X(z)$, $D_z\mu_{Y,1}(z)$ / $D_z\mu_{X,1}(z)$ does not generally have a structural meaning – it remains a stochastic object. In other words, substituting a proxy for an instrument, while harmless in fully linear settings, generally leads to inconsistent estimates of structural effects in nonlinear settings.

As we show, however, β can be estimated if we can observe two error-contaminated proxies for Z,

$$Z_1 \stackrel{c}{=} Z + U_1 \qquad Z_2 \stackrel{c}{=} Z + U_2,$$

where U_1 and U_2 are random variables satisfying assumptions given below.

3 Estimation with Observed Exogenous Instruments

3.1 Asymptotics: General Theory

We first state results for generic Z and V, with $g_{V,\lambda}$ as defined above. Our first conditions specify some relevant properties of Z and V. For notational convenience in what follows, we may write "$\sup_{z\in\mathbb{R}}$" or "$\inf_{z\in\mathbb{R}}$" in place of "$\sup_{z\in\text{supp}(Z)}$" or "$\inf_{z\in\text{supp}(Z)}$". By convention, we also take the value of any referenced function to be zero except when $z \in \text{supp}(Z)$.

Assumption 3.1 *Z is a random variable with continuous density f_Z such that* $\sup_{z\in\mathbb{R}}$ $f_Z(z) < \infty$.

Among other things, this ensures that $f_Z(z) > 0$ for all $z \in \text{supp}(Z)$.

Assumption 3.2 *V is a random variable such that* (i) $E(|V|) < \infty$; (ii) $E(V^2) < \infty$ *and* $\sup_{z\in\mathbb{R}} E\left[V^2|Z=z\right] < \infty$; (iii) $\inf_{z\in\mathbb{R}} E\left[V^2|Z=z\right] > 0$; (iv) *for some* $\delta > 0$, $E\left[|V|^{2+\delta}\right] < \infty$ *and* $\sup_{z\in\mathbb{R}} E\left[|V|^{2+\delta}|Z=z\right] < \infty$.

Assumptions 3.1(i) and 3.2(i) ensure that $g_{V,0}(z)$ is well defined. Next, we impose smoothness on $g_{V,0}$. Let $\mathbb{N} \equiv \{0, 1, ...\}$ and $\overline{\mathbb{N}} \equiv \mathbb{N} \cup \{\infty\}$.

Assumption 3.3 *$g_{V,0}$ is continuously differentiable of order $\Lambda \in \overline{\mathbb{N}}$ on $\mathbb{R}$.*

Given a sample of n independent and identically distributed (IID) observations $\{V_i, Z_i\}$, a natural kernel estimator for $g_{V,\lambda}(z)$ is

$$\hat{g}_{V,\lambda}(z,h) = D_z^\lambda \hat{E}\left[\frac{V}{h}\, k\left(\frac{Z-z}{h}\right)\right] = (-1)^\lambda h^{-1-\lambda} \hat{E}\left[V\, k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right],$$

where $k(\cdot)$ is a user-specified kernel, $k^{(\lambda)}(z) \equiv D_z^\lambda k(z)$, $h > 0$ is the kernel bandwidth, and the operator $\hat{E}[\cdot]$ denotes a sample average: for any random variable W, $\hat{E}[W] \equiv n^{-1}\sum_{i=1}^n W_i$, where $W_1, \dots, W_n$ is a sample of random variables, distributed identically as W. We specify our choice of kernel as follows

Assumption 3.4 *The real-valued kernel $z \to k(z)$ is measurable and symmetric, $\int k(z)dz = 1$, and its Fourier transform $\xi \to \kappa(\xi)$ is such that:* (i) *κ has two bounded derivatives;* (ii) *κ is compactly supported (without loss of generality, we take the support to be $[-1,1]$); and* (iii) *there exists $\bar{\xi} > 0$ such that $\kappa(\xi) = 1$ for $|\xi| < \bar{\xi}$.*

Requiring that the kernel's Fourier transform is compactly supported implies that the kernel is continuously differentiable to any order. Politis and Romano (1999) call a kernel whose Fourier transform is constant in the neighborhood of the origin, as in (iii), a "flat-top" kernel. When the derivatives of the Fourier transform vanish at the origin, all moments of the kernel vanish, by the well-known Moment Theorem. Such kernels are thus also called "infinite order" kernels. These have the property that, if the function to be estimated is infinitely many times differentiable, the bias of the kernel estimator shrinks faster than any positive power of h. The use of infinite order kernels is not essential for the OXI case, but is especially advantageous in the PXI case, where fast convergence rates are more difficult to achieve. We use infinite order kernels in both cases to maintain a fully comparable analysis.

Our first result decomposes the kernel estimation error.

Lemma 3.1 *Suppose that $\{V_i, Z_i\}$ is a sequence of identically distributed random variables satisfying Assumptions 3.1, 3.2(i) and 3.3, and that Assumption 3.4 holds. Then for each $\lambda = 0, ..., \Lambda$, $z \in \mathrm{supp}(Z)$, and $h > 0$*

$$\hat{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z) = B_{V,\lambda}(z,h) + L_{V,\lambda}(z,h), \tag{14}$$

where $B_{V,\lambda}(z,h)$ is a nonrandom "bias term" defined as

$$B_{V,\lambda}(z,h) \equiv g_{V,\lambda}(z,h) - g_{V,\lambda}(z),$$

with

$$g_{V,\lambda}(z,h) \equiv D_z^{\lambda} E\left[\frac{V}{h} k\left(\frac{Z-z}{h}\right)\right] = (-1)^{\lambda} E\left[V h^{-\lambda-1} k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right];$$

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

$$L_{V,\lambda}(z,h) = \hat{E}\left[\ell_{V,\lambda}(z,h;V,Z)\right],$$

with

$$\ell_{V,\lambda}(z,h;v,\tilde{z}) \equiv (-1)^{\lambda} h^{-\lambda-1} v k^{(\lambda)}\left(\frac{\tilde{z}-z}{h}\right) - E\left[(-1)^{\lambda} h^{-\lambda-1} V k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right].$$

Proofs can be found in the Mathematical Appendix.

To obtain rate of convergence results for our kernel estimators, we impose further smoothness conditions on $g_{V,0}$ and specify convergence rates for the bandwidth.

Assumption 3.5 *For $\zeta \in \mathbb{R}$, let $\phi_V(\zeta) \equiv E\left[V e^{\mathbf{i}\zeta Z}\right] = \int g_{V,0}(z) e^{i\zeta z} dz$. There exist constants $C_\phi > 0$, $\alpha_\phi \leq 0$, $\beta_\phi \geq 0$, and $\gamma_\phi \in \mathbb{R}$, such that $\beta_\phi \, \gamma_\phi \geq 0$ and*

$$|\phi_V(\zeta)| \leq C_\phi (1 + |\zeta|)^{\gamma_\phi} \exp\left(\alpha_\phi |\zeta|^{\beta_\phi}\right). \tag{15}$$

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

This Fourier transform bound directly relates to conditions on the derivatives of $g_{V,0}$. If for some $\gamma_\phi < 0$, $g_{V,0}$ admits $\Lambda = -\gamma_\phi$ derivatives that are absolutely integrable over $\mathbb{R}$, then Assumption 3.5 is satisfied with $\alpha_\phi = 0$. The situation where $\alpha_\phi < 0$ corresponds to the case where $g_{V,0}$ is infinitely many times differentiable ($\Lambda = \infty$). This Fourier bound is particularly advantageous when combined with an infinite order kernel, because the order of magnitude of the estimation bias is then directly related to the constants α_ϕ and β_ϕ. A further advantage is that Assumption 3.5 exactly parallels the assumptions needed for the PXI case, thus facilitating comparisons.

We choose the kernel bandwidth h according to the next condition.

Assumption 3.6 $\{h_n\}$ *is a sequence of positive numbers such that as* $n \to \infty$, $h_n \to 0$, *and for given* $\lambda \in \{0, ..., \Lambda\}$, $nh_n^{2\lambda+1} \to \infty$.

Taken together, our moment and bandwidth conditions are standard in the kernel estimation literature (e.g. Haerdle and Linton, 1994; Andrews, 1995; Pagan and Ullah, 1999).

The decomposition of Lemma 3.1 and the assumptions just given enable us to state our first main result. We give this in a form that somewhat departs from the usual asymptotics for kernel estimators, but that facilitates the analysis for the various quantities of interest and eases comparisons with the PXI case.

Theorem 3.2 *Let the conditions of Lemma 3.1 hold with* $\{V_i, Z_i\}$ *IID.*

(*i*) *Suppose in addition that Assumption 3.5 holds for given* $\lambda \in \{0, ..., \Lambda\}$. *Then for* $h > 0$,

$$\sup_{z\in\mathbb{R}} |B_{V,\lambda}(z,h)| = O\left(\left(h^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h^{-1}\right)^{\beta_B}\right)\right),$$

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

(ii) For each $z \in \text{supp}(Z)$ *and* $h > 0$, $E[L_{V,\lambda}(z,h)] = 0$, *and if Assumption 3.2(ii) also holds then*

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

where

$$\Omega_{V,\lambda}(z,h) \equiv E\left[\left(\ell_{V,\lambda}(z,h;V,Z)\right)^{2}\right]$$

is finite and satisfies

$$\sqrt{\sup_{z\in\mathbb{R}} \Omega_{V,\lambda}(z,h)} = O\left(h^{-\lambda-1/2}\right). \tag{16}$$

Further,

$$\sup_{z\in\mathbb{R}} |L_{V,\lambda}(z,h)| = O_p\left(n^{-1/2}h^{-\lambda-1}\right). \tag{17}$$

If in addition $h_n \to 0$ *as* $n \to \infty$, *then for each* $z \in \text{supp}(Z)$

$$h_n^{2\lambda+1}\Omega_{V,\lambda}(z,h_n) \to E\left[V^2|Z=z\right] f_Z(z) \int \left(k^{(\lambda)}(z)\right)^2 dz \tag{18}$$

and if Assumption 3.2(iii) also holds, then $\Omega_{V,\lambda}(z,h_n) > 0$ *for all* n *sufficiently large.*

(iii) If in addition to the conditions of (ii), Assumptions 3.2(iv) and 3.6 for given $\lambda \in \{0,...,\Lambda\}$ *also hold, then for each* $z \in \text{supp}(Z)$

$$n^{1/2}\left(\Omega_{V,\lambda}(z,h_n)\right)^{-1/2} L_{V,\lambda}(z,h_n) \xrightarrow{d} N(0,1). \tag{19}$$

As we use nonparametric estimators $\hat{g}_{V,\lambda}$ as building blocks for more complex quantities of interest such as β_w and β_{wf_Z}, we now consider a functional b of a k-vector $g \equiv (g_{V_1,\lambda_1},\ldots,g_{V_k,\lambda_k})$. Specifically, we establish the asymptotic properties of $b(\hat{g}(\cdot,h)) - b(g) \equiv b(\hat{g}_{V_1,\lambda_1}(\cdot,h),\ldots,\hat{g}_{V_k,\lambda_k}(\cdot,h)) - b(g_{V_1,\lambda_1},\ldots,g_{V_k,\lambda_k})$. We first impose minimum convergence rates. For conciseness, we state these in a high-level form; primitive conditions obtain via Theorem 3.2.

Assumption 3.7 *For given* $\lambda \in \{0,...,\Lambda\}$, $\sup_{z\in\mathbb{R}} |B_{V,\lambda}(z,h_n)| = o\left(n^{-1/2}\right)$ *and* $\sup_{z\in\mathbb{R}} |L_{V,\lambda}(z,h_n)| = o_p\left(n^{-1/4}\right)$.

The following theorem consists of two parts. The first part provides an asymptotically linear representation, useful for analyzing a scalar estimator constructed as a functional of a vector of estimators. The second part gives a convenient asymptotic normality and root-n consistency result useful for analyzing β_w and β_{wf_Z}. In this result we explicitly consider a finite family of random variables $\{V_1, ..., V_J\}$ satisfying Assumptions 3.2, 3.3, and 3.5. We require that these conditions hold uniformly, with the same constants $\delta, \Lambda, C_\phi, \alpha_\phi, \beta_\phi, \gamma_\phi$ for all V in the family. As the family is finite, this can always be ensured by taking the constants $\delta, \Lambda, C_\phi, \alpha_\phi, \beta_\phi, \gamma_\phi$ to be the worst-case values among all V in the family.

Theorem 3.3 *For $\Lambda, J \in \mathbb{N}$, let $\lambda_1, \ldots, \lambda_J$ belong to $\{0, ..., \Lambda\}$, and suppose that $\{V_{1i}, ..., V_{Ji}, Z_i\}$ is an IID sequence of random vectors such that $\{V_{ji}, Z_i\}$ satisfies the conditions of Theorem 3.2 and Assumption 3.7 for $j = 1, ..., J$ with identical choices of k and h_n.*

Let the real-valued functional b be such that for any $\tilde{g} \equiv (\tilde{g}_{V_1,\lambda_1}, \ldots, \tilde{g}_{V_J,\lambda_J})$ in an L_∞ neighborhood of the J-vector $g \equiv (g_{V_1,\lambda_1}, ..., g_{V_J,\lambda_J})$,

$$b(\tilde{g}) - b(g) = \sum_{j=1}^{J} \int \left(\tilde{g}_{V_j,\lambda_j}(z) - g_{V_j,\lambda_j}(z)\right) s_j(z)\, dz + \sum_{j=1}^{J} O\left(\left\|\tilde{g}_{V_j,\lambda_j} - g_{V_j,\lambda_j}\right\|_\infty^2\right) \quad (20)$$

for some real-valued functions s_j, $j = 1, ..., J$. If s_j is such that $\sup_{z\in\mathbb{R}} |s_j(z)| < \infty$, $\int |s_j(z)|\, dz < \infty$, and $E\left[\left(V_j s_j^{(\lambda_j)}(Z)\right)^2\right] < \infty$ (with $s_j^{(\lambda_j)}(z) \equiv D_z^{\lambda_j} s_j(z)$) for each $j = 1, ..., J$, then

$$b(\hat{g}(\cdot, h_n)) - b(g) = \sum_{j=1}^{J} \hat{E}\left[\psi_{V_j,\lambda_j}(s_j; V_j, Z)\right] + o_p\left(n^{-1/2}\right),$$

where

$$\psi_{V_j,\lambda_j}(s_j; v_j, z) \equiv \left(v_j s_j^{(\lambda_j)}(z) - E\left[V_j s_j^{(\lambda_j)}(Z)\right]\right), \quad j = 1, ..., J.$$

Moreover,

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

where

$$\Omega_b \equiv E\left[\left(\sum_{j=1}^{J} \psi_{V_j,\lambda_j}(s_j; V_j, Z)\right)^2\right] < \infty.$$

Interestingly, this result provides "nonparametric first step correction terms", $\psi_{V_j,\lambda_j}(s_j; v_j, z)$, similar to the correction terms $\alpha(z)$ introduced in Newey (1994). Whereas Newey (1994) provides correction terms for conditional expectations and densities (and derivatives thereof), we provide correction terms for quantities of the form $g_{V,\lambda}(z)$. Naturally, our correction term for $g_{1,0}(z)$ reduces to Newey's correction term for densities. Also, applying Theorem 3.3 to a nonlinear functional of the ratio $g_{V,0}(z)/g_{1,0}(z)$ recovers Newey's correction term for conditional expectations.

3.2 Asymptotics: OXI Case

We now apply our general asymptotic results to our main quantities of interest, eqs.(9) and (12). First we treat the following nonparametric estimator of $\beta(z)$:

$$\hat{\beta}(z, h_n) \equiv D_z\hat{\mu}_Y(z, h_n) \;/\; D_z\hat{\mu}_X(z, h_n) \tag{21}$$

for $z \in \text{supp}(Z)$, where

$$\begin{aligned} D_z\hat{\mu}_Y(z,h) &\equiv \frac{\hat{g}_{Y,1}(z,h)}{\hat{g}_{1,0}(z,h)} - \frac{\hat{g}_{Y,0}(z,h)}{\hat{g}_{1,0}(z,h)}\frac{\hat{g}_{1,1}(z,h)}{\hat{g}_{1,0}(z,h)} \qquad \text{and} \\ D_z\hat{\mu}_X(z,h) &\equiv \frac{\hat{g}_{X,1}(z,h)}{\hat{g}_{1,0}(z,h)} - \frac{\hat{g}_{X,0}(z,h)}{\hat{g}_{1,0}(z,h)}\frac{\hat{g}_{1,1}(z,h)}{\hat{g}_{1,0}(z,h)}. \end{aligned}$$

Applying Theorem 3.2 and a straightforward Taylor expansion, we obtain

Theorem 3.4 *Suppose that* $\{X_i, Y_i, Z_i\}$ *is an IID sequence of random variables satisfying the conditions of Theorem 3.2 for* $V = 1, X, Y$*, with* $\Lambda \geq 1$ *and* $\lambda = 0, 1$*, and with identical choices of* k *and* h_n*. Further, suppose* $\max_{V=1,X,Y} \max_{\lambda=0,1} \sup_{z\in\mathbb{R}} |g_{V,\lambda}(z)| < \infty$*, and for* $\tau > 0$*, define*

$$\mathbf{Z}_\tau \equiv \{z \in \mathbb{R} : f_Z(z) \geq \tau \text{ and } |D_z\mu_X(z)| \geq \tau\}.$$

Then

$$\sup_{z\in\mathbf{Z}_\tau} \left|\hat{\beta}(z,h_n) - \beta(z)\right| = O\left(\tau^{-4}\left(h_n^{-1}\right)^{\gamma_{1,B}} \exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)\right) + O_p\left(\tau^{-4} n^{-1/2}\left(h_n^{-1}\right)^2\right),$$

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

$$\sup_{z \in \mathbf{Z}_{\tau_n}} \left| \hat{\beta}(z, h_n) - \beta(z) \right| = o_p(1).$$

The delta method secures the next result.

Theorem 3.5 *Suppose that* $\{X_i, Y_i, Z_i\}$ *is an IID sequence satisfying the conditions of Theorem 3.2 for* $V = 1, X, Y$, *with* $\Lambda \geq 1$ *and* $\lambda = 0, 1$, *and with identical choices for* k *and* $\{h_n\}$. *Further, suppose* $\max_{V=1,X,Y} \max_{\lambda=0,1} |g_{V,\lambda}(z)| < \infty$. *Then for all* $z \in \text{supp}(Z)$ *such that* $|D_z \mu_X(z)| > 0$,

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

provided that $\left(\max_{V=1,X,Y} \max_{\lambda=0,1} \left(n^{1/2} h_n^{\lambda+1/2} \right) |B_{V,\lambda}(z, h_n)| \right) \xrightarrow{p} 0$ *and that*

$$\Omega_\beta(z, h_n) \equiv E\left[(\ell_\beta(z, h_n; X, Y, Z))^2 \right]$$

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

$$\begin{aligned} \ell_\beta(z, h; x, y, z) \equiv & \sum_{\lambda=0,1} (s_{X,1,\lambda}(z) \ell_{1,\lambda}(z, h; 1, z) + s_{X,X,\lambda}(z) \ell_{X,\lambda}(z, h; x, z) \\ & + s_{Y,1,\lambda}(z) \ell_{1,\lambda}(z, h; 1, z) + s_{Y,Y,\lambda}(z) \ell_{Y,\lambda}(z, h; y, z)) \end{aligned} \tag{22}$$

$$\begin{aligned} s_{Y,Y,1}(z) &\equiv \frac{1}{D_z \mu_X(z)} \frac{1}{g_{1,0}(z)} \\ s_{Y,Y,0}(z) &\equiv -\frac{1}{D_z \mu_X(z)} \frac{g_{1,1}(z)}{g_{1,0}(z)} \frac{1}{g_{1,0}(z)} \\ s_{Y,1,1}(z) &\equiv -\frac{1}{D_z \mu_X(z)} \frac{g_{Y,0}(z)}{g_{1,0}(z)} \frac{1}{g_{1,0}(z)} \\ s_{Y,1,0}(z) &\equiv \frac{1}{D_z \mu_X(z)} \left(2 \frac{g_{Y,0}(z)}{g_{1,0}(z)} \frac{g_{1,1}(z)}{g_{1,0}(z)} - \frac{g_{Y,1}(z)}{g_{1,0}(z)} \right) \frac{1}{g_{1,0}(z)} \end{aligned}$$

$$\begin{aligned} s_{X,Y,1}(z) &\equiv \frac{\beta(z)}{D_z \mu_X(z)} \frac{1}{g_{1,0}(z)} \\ s_{X,Y,0}(z) &\equiv -\frac{\beta(z)}{D_z \mu_X(z)} \frac{g_{1,1}(z)}{g_{1,0}(z)} \frac{1}{g_{1,0}(z)} \\ s_{X,1,1}(z) &\equiv -\frac{\beta(z)}{D_z \mu_X(z)} \frac{g_{X,0}(z)}{g_{1,0}(z)} \frac{1}{g_{1,0}(z)} \\ s_{X,1,0}(z) &\equiv \frac{\beta(z)}{D_z \mu_X(z)} \left(2 \frac{g_{X,0}(z)}{g_{1,0}(z)} \frac{g_{1,1}(z)}{g_{1,0}(z)} - \frac{g_{X,1}(z)}{g_{1,0}(z)} \right) \frac{1}{g_{1,0}(z)}. \end{aligned}$$

As described in Section 2, weighted functions of β, β_w and β_{wf_Z}, defined in eq.(12) are also of interest. We now propose the following estimators for these:

$$\hat{\beta}_w \equiv \int_{S_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(z,h_n)\, w(z)\, dz$$
$$\hat{\beta}_{wf_Z} \equiv \int_{S_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(z,h_n)\, w(z)\, \hat{g}_{1,0}(z,h_n)\, dz,$$

where $S_{\hat{\beta}(\cdot,h_n)} \equiv \{z : \hat{g}_{1,0}(z,h_n) > 0, |D_z\hat{\mu}_X(z,h_n)| > 0\}$. We next restrict the weights.

Assumption 3.8 *Let $\mathcal{W}$ be a bounded measurable subset of* $\mathbb{R}$. (i) *The weighting function $w : \mathbb{R} \to \mathbb{R}$ is measurable and supported on $\mathcal{W}$*; (ii) $\inf_{z\in\mathcal{W}} f_Z(z) > 0$ *and* $\inf_{z\in\mathcal{W}} |D_z\mu_X(z)| > 0$; (iii) $\max_{V=1,X,Y} \max_{\lambda=0,1} \sup_{z\in\mathcal{W}} |g_{V,\lambda}(z)| < \infty$.

The asymptotic distributions of these estimators follow by straightforward application of Theorem 3.3, noting that, with probability approaching one, the integrals over the random set $S_{\hat{\beta}(\cdot,h_n)}$ equal the same integral over the set $\mathcal{W}$, because under our assumptions the denominators in the expression for $\hat{\beta}(z,h_n)$ converge uniformly to functions that are bounded away from zero over $\mathcal{W}$. Due to the weighted estimators' semiparametric nature, root-n consistency and asymptotic normality hold.

Theorem 3.6 *Suppose the conditions of Theorem 3.3 hold for $V = 1, X, Y,$ and $\lambda = 0, 1$, and that Assumption 3.8 also holds. Then*

$$n^{1/2}\Omega_w^{-1/2}\left(\hat{\beta}_w - \beta_w\right) \xrightarrow{d} N(0,1),$$

provided that

$$\Omega_w \equiv E\left[\left(\psi_{\beta_w}(X,Y,Z)\right)^2\right]$$

is finite and positive for all n sufficiently large, where

$$\psi_{\beta_w}(x,y,z) \equiv \sum_{\lambda=0,1} (\psi_{1,\lambda}(ws_{X,1,\lambda};1,z) + \psi_{X,\lambda}(ws_{X,X,\lambda};x,z)$$
$$+\psi_{1,\lambda}(ws_{Y,1,\lambda};1,z) + \psi_{Y,\lambda}(ws_{Y,Y,\lambda};y,z)),$$

$ws_{A,V,\lambda}$ *denotes the function mapping* z *to* $w(z)\, s_{A,V,\lambda}(z)$, *and where* $\psi_{V,\lambda}(s; v, z)$ *is defined in Theorem 3.3.*

Theorem 3.7 *Suppose the conditions of Theorem 3.3 hold for* $V = 1, X, Y$, *and* $\lambda = 0, 1$, *and that Assumption 3.8 also holds. Then*

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

provided that

$$\Omega_{wf_Z} \equiv E\left[\left(\psi_{\beta_{wf_z}}(X,Y,Z)\right)^2\right]$$

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

$$\begin{aligned}\psi_{\beta_{wf_Z}}(x,y,z) \equiv\ & \{\sum_{\lambda=0,1}(\psi_{1,\lambda}(wf_Z s_{X,1,\lambda}; 1, z) + \psi_{X,\lambda}(wf_Z s_{X,X,\lambda}; x, z)\\ & +\psi_{1,\lambda}(wf_Z s_{Y,1,\lambda}; 1, z) + \psi_{Y,\lambda}(wf_Z s_{Y,Y,\lambda}; y, z))\}\\ & +\psi_{1,0}(w\beta; 1, z),\end{aligned}$$

$ws_{A,V,\lambda}$ *denotes the function mapping* z *to* $w(z)\, f_Z(z) s_{A,V,\lambda}(z)$, $w\beta$ *denotes the function mapping* z *to* $w(z)\,\beta(z)$, *and where* $\psi_{V,\lambda}(s; v, z)$ *is defined in Theorem 3.3.*

It is straightforward to show that the asymptotic variances in Theorems 3.2, 3.3, 3.5, 3.6, and 3.7 can be consistently estimated, although we do not provide explicit theorems due to space limitations. In the cases of Theorems 3.2 or 3.5, this estimation can be accomplished, respectively, by substituting conventional kernel nonparametric estimates into eq.(18), or by calculating the variance of eq.(22) through a similar technique. In the case of Theorems 3.3, 3.6, and 3.7, we directly provide an expression for the influence function, from which the asymptotic variance is easy to calculate.

4 Estimation with Proxies for Unobserved Exogenous Instruments

When Z cannot be observed, the estimators of Section 3 are not feasible. In this section we consider estimators based on error-laden measurements of Z. This delivers nonparametric and semi-parametric analogs of the PXI estimators introduced by CW.

4.1 A General Representation Result

We begin by obtaining a representation in terms of observables for $g_{V,\lambda}$ with generic V when Z is unobserved, using two error-contaminated measurements of Z:

$$Z_1 = Z + U_1 \qquad Z_2 = Z + U_2.$$

We impose the following conditions on Z, V, U_1, and U_2. For succinctness, some conditions may overlap those previously given.

Assumption 4.1 $E[|Z|] < \infty$, $E[|U_1|] < \infty$, *and* $E[|V|] < \infty$.

Assumption 4.2 $E[U_1|Z, U_2] = 0$, $U_2 \perp Z$, *and* $E[V|Z, U_2] = E[V|Z]$.

The next assumption formalizes the measurement of Z.

Assumption 4.3 $Z_1 = Z + U_1$ *and* $Z_2 = Z + U_2$.

We now show that $g_{V,\lambda}$ can be defined solely in terms of the joint distribution of V, Z_1, and Z_2. Thus, if these are observable, then $g_{V,\lambda}$ is empirically accessible. This result generalizes Schennach (2004b), which focused on the $\lambda = 0$ case.

Lemma 4.1 *Suppose Assumptions 3.1, 4.1 - 4.3, and 3.3 hold. Then for each* $\lambda \in \{0, ..., \Lambda\}$ *and* $z \in \text{supp}(Z)$

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

where for each real ζ,

$$\phi_V(\zeta) \equiv E\left[V e^{\mathbf{i}\zeta Z}\right] = \frac{E\left[V e^{\mathbf{i}\zeta Z_2}\right]}{E\left[e^{\mathbf{i}\zeta Z_2}\right]} \exp\left(\int_0^{\zeta} \frac{\mathbf{i} E\left[Z_1 e^{\mathbf{i}\xi Z_2}\right]}{E\left[e^{\mathbf{i}\xi Z_2}\right]} d\xi \right).$$

4.2 Estimation

Our estimator is motivated by a smoothed version of $g_{V,\lambda}(z)$.

Lemma 4.2 *Suppose Assumptions 3.1, 4.1, and 3.3 hold, and let k satisfy Assumption 3.4. For $h > 0$ and for each $\lambda \in \{0, ..., \Lambda\}$ and $z \in \text{supp}(Z)$ now let*

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

Then

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

By lemma 1 of the appendix of Pagan and Ullah (1999, p.362), we have $\lim_{h\to 0} g_{V,\lambda}(z,h) = g_{V,\lambda}(z)$, so we also define $g_{V,\lambda}(z,0) \equiv g_{V,\lambda}(z)$. Motivated by Lemma 4.2, we now propose the estimator

$$\hat{g}_{V,\lambda}(z,h_n) \equiv \frac{1}{2\pi} \int (-\mathbf{i}\zeta)^{\lambda} \kappa(h_n\zeta)\, \hat{\phi}_V(\zeta) \exp(-\mathbf{i}\zeta z)\, d\zeta, \tag{23}$$

with $h_n \to 0$ as $n \to \infty$, where, motivated by Lemma 4.1,

$$\hat{\phi}_V(\zeta) \equiv \frac{\hat{E}\left[Ve^{\mathbf{i}\zeta Z_2}\right]}{\hat{E}\left[e^{\mathbf{i}\zeta Z_2}\right]} \exp\left(\int_0^{\zeta} \frac{\mathbf{i}\hat{E}\left[Z_1 e^{\mathbf{i}\xi Z_2}\right]}{\hat{E}\left[e^{\mathbf{i}\xi Z_2}\right]} d\xi\right), \tag{24}$$

and $\hat{E}[\cdot]$ denotes a sample average, as above.

4.3 Asymptotics: General Theory

The results of this section extensively generalize those of Schennach (2004a, 2004b), to include (i) the $\lambda \neq 0$ case (ii) uniform convergence results and (iii) general semiparametric functionals of $g_{V,\lambda}$, and hence will be applicable beyond our PXI case. Parallel to Lemma 3.1, we first decompose the estimation error into components that will be further characterized in subsequent results.

Lemma 4.3 *Suppose that $\{V_i, Z_i, U_{1i}, U_{2i}\}$ is a sequence of identically distributed random variables satisfying Assumptions 3.1, 4.1 - 4.3, and 3.3, and that Assumption 3.4 holds. Then for each $\lambda = 0, ..., \Lambda$, $z \in \text{supp}(Z)$, and $h > 0$,*

$$\hat{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z) = B_{V,\lambda}(z,h) + L_{V,\lambda}(z,h) + R_{V,\lambda}(z,h), \tag{25}$$

where $B_{V,\lambda}(z,h)$ is a nonrandom "bias term" defined as

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

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

$$L_{V,\lambda}(z,h) = \hat{E}\left[\ell_{V,\lambda}(z,h;V,Z_1,Z_2)\right],$$

with

$$\begin{aligned} \ell_{V,\lambda}(z,h;v,z_1,z_2) \equiv & \int \Psi_{V,\lambda,1}(\xi,z,h)\left(e^{\mathbf{i}\xi z_2} - E\left[e^{\mathbf{i}\xi Z_2}\right]\right)d\xi \\ & + \int \Psi_{V,\lambda,Z_1}(\xi,z,h)\left(z_1 e^{\mathbf{i}\xi z_2} - E\left[Z_1 e^{\mathbf{i}\xi Z_2}\right]\right)d\xi \\ & + \int \Psi_{V,\lambda,V}(\xi,z,h)\left(v e^{\mathbf{i}\xi z_2} - E\left[V e^{\mathbf{i}\xi Z_2}\right]\right)d\xi, \end{aligned}$$

where, for $A = 1, Z_1$, and V, we let $\theta_A(\zeta) \equiv E\left[Ae^{\mathbf{i}\zeta Z_2}\right]$ and define

$$\begin{aligned} \Psi_{V,\lambda,1}(\xi,z,h) \equiv & -\frac{1}{2\pi}\frac{\phi_V(\xi)}{\theta_1(\xi)}\exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^\lambda \kappa(h\xi) \\ & -\frac{1}{2\pi}\frac{\mathbf{i}\theta_{Z_1}(\xi)}{(\theta_1(\xi))^2}\int_\xi^{\pm\infty}\exp(-\mathbf{i}\zeta z)(-\mathbf{i}\zeta)^\lambda\kappa(h\zeta)\phi_V(\zeta)\,d\zeta \\ \Psi_{V,\lambda,Z_1}(\xi,z,h) \equiv & \frac{1}{2\pi}\frac{\mathbf{i}}{\theta_1(\xi)}\int_\xi^{\pm\infty}\exp(-\mathbf{i}\zeta z)(-\mathbf{i}\zeta)^\lambda\kappa(h\zeta)\phi_V(\zeta)\,d\zeta \\ \Psi_{V,\lambda,V}(\xi,z,h) \equiv & \frac{1}{2\pi}\frac{\phi_1(\xi)}{\theta_1(\xi)}\exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^\lambda\kappa(h\xi), \end{aligned}$$

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

We already have conditions sufficient to describe the asymptotic properties of the bias term defined in Lemma 4.3.

Theorem 4.4 *Let the conditions of Lemma 4.3 hold with $\{V_i, Z_i, U_{1i}, U_{2i}\}$ IID, and suppose in addition that Assumption 3.5 holds for given $\lambda \in \{0, ..., \Lambda\}$. Then for $h > 0$,*

$$\sup_{z\in\mathbb{R}} |B_{V,\lambda}(z,h)| = O\left(\left(h^{-1}\right)^{\gamma_{\lambda,B}} \exp\left(\alpha_B \left(h^{-1}\right)^{\beta_B}\right)\right),$$

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

This result is closely parallel to Theorem 3.2(i). Our next result parallels Theorem 3.2(ii) and (iii). For this, we first ensure that $L_{V,\lambda}(z,h)$ has finite variance.

Assumption 4.4 $E\left[Z_1^2\right] < \infty, E\left[V^2\right] < \infty$.

To obtain the rate for $\Omega_{V,\lambda}(z,h) = \text{var}(n^{1/2} L_{V,\lambda}(z,h))$, we impose bounds on the tail behavior of the Fourier transforms involved, as is common in the deconvolution literature (e.g. Fan, 1991; Fan and Truong, 1993). These rates are analogous to Assumption 3.5.

Assumption 4.5 (i) *For each $\zeta \in \mathbb{R}$, let $\phi_1(\zeta) \equiv E\left[e^{\mathbf{i}\zeta Z}\right]$ satisfy*

$$\left|\frac{D_\zeta \phi_1(\zeta)}{\phi_1(\zeta)}\right| \leq C_1 (1 + |\zeta|)^{\gamma_1} \tag{26}$$

for some $C_1 > 0$ and $\gamma_1 \geq 0$; and for C_ϕ, α_ϕ, β_ϕ, and γ_ϕ, as in Assumption 3.5,

$$|\phi_1(\zeta)| \leq C_\phi (1 + |\zeta|)^{\gamma_\phi} \exp\left(\alpha_\phi |\zeta|^{\beta_\phi}\right);$$

(ii) *For each $\zeta \in \mathbb{R}$, let $\theta_1(\zeta) \equiv E\left[e^{\mathbf{i}\zeta Z_2}\right]$ satisfy*

$$|\theta_1(\zeta)| \geq C_\theta (1 + |\zeta|)^{\gamma_\theta} \exp\left(\alpha_\theta |\zeta|^{\beta_\theta}\right) \tag{27}$$

for some $C_\theta > 0$ and $\alpha_\theta \leq 0$, $\beta_\theta \geq \beta_\phi \geq 0$, and $\gamma_\theta \in \mathbb{R}$, such that $\gamma_\theta \beta_\theta \geq 0$.

For conciseness, we express our bounds in the form $(1 + |\zeta|)^\gamma \exp\left(\alpha |\zeta|^\beta\right)$, thereby simultaneously covering the ordinarily smooth ($\alpha = 0$, $\beta = 0$) and supersmooth ($\alpha \neq 0$, $\beta \neq 0$) cases. Note that the lack of a term $\exp\left(\alpha_1 |\zeta|^{\beta_1}\right)$ in eq.(26) results in a negligible loss of generality, as $D_\zeta \phi_1(\zeta)/\phi_1(\zeta) = D_\zeta \ln \phi_1(\zeta)$, and $\ln \phi_1(\zeta)$ is typically a power of ζ for large ζ, even if $\phi_1(\zeta)$ is associated with a supersmooth distribution. The tail behaviors of $\phi_1(\zeta)$ and $\phi_V(\zeta)$ have the same effect on the convergence rate; we may thus impose the same bound without loss of generality. The lower bound on $\theta_1(\zeta)$ is implied by separate lower bounds on $E\left[e^{\mathbf{i}\zeta Z}\right]$ and $E\left[e^{\mathbf{i}\zeta U_2}\right]$, as independence ensures $E\left[e^{\mathbf{i}\zeta Z_2}\right] = E\left[e^{\mathbf{i}\zeta Z}\right] E\left[e^{\mathbf{i}\zeta U_2}\right]$.

By using the infinite order kernels of Assumption 3.4, we ensure that the rate of convergence of the estimator is never limited by the order of the kernel but only by the smoothness of the data generating process. This can be especially helpful when the densities of Z_2 and Z are both supersmooth, in which case an infinite order kernel can often deliver a convergence rate n^{-r} for some $r > 0$. In contrast a traditional finite-order kernel only achieves a $(\ln n)^{-r}$ rate. Although our theory can easily be adapted to cover finite-order kernels, as in (Schennach, 2004b), we focus on infinite order kernels to exploit their better rates.

The next bounds parallel Assumption 3.2(iv) and help to establish asymptotic normality of the kernel regression estimators.

Assumption 4.6 *For some* $\delta > 0$, $E\left[|Z_1|^{2+\delta}\right] < \infty, \sup_{z\in\mathbb{R}} E\left[Z_1^{2+\delta}|Z_2 = z\right] < \infty, E\left[|V|^{2+\delta}\right] < \infty$, *and* $\sup_{z\in\mathbb{R}} E\left[V^{2+\delta}|Z_2 = z\right] < \infty$.

The next assumption imposes a lower bound on the bandwidth that will be used when establishing asymptotic normality.

Assumption 4.7 *If* $\beta_\theta = 0$ *in Assumption 4.5, then for given* $\lambda \in \{0, ..., \Lambda\}$, $h_n^{-1} = O\left(n^{-\eta} n^{(3/2)/\left(3-\gamma_\theta+\gamma_\phi+\gamma_1+\lambda\right)}\right)$ *for some* $\eta > 0$*; otherwise* $h_n^{-1} = O\left((\ln n)^{\beta_\theta^{-1}-\eta}\right)$ *for some* $\eta > 0$.

Theorem 4.5 *Let the conditions of Lemma 4.3 hold with* $\{V_i, Z_i, U_{1i}, U_{2i}\}$ *IID.* (i) *Then for each* $z \in \operatorname{supp}(Z)$ *and* $h > 0, E\left[L_{V,\lambda}(z,h)\right] = 0$, *and if Assumption 4.4 also holds, then*

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

where

$$\Omega_{V,\lambda}(z,h) \equiv E\left[(\ell_{V,\lambda}(z,h;V,Z_1,Z_2))^2\right] < \infty.$$

Further, if Assumption 4.5 holds then

$$\sqrt{\sup_{z\in\mathbb{R}} \Omega_{V,\lambda}(z,h)} = O\left(\left(h^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h^{-1}\right)^{\beta_L}\right)\right), \tag{28}$$

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*

$$\sup_{z\in\mathbb{R}} |L_{V,\lambda}(z,h)| = O_p\left(n^{-1/2}\left(h^{-1}\right)^{\gamma_{\lambda,L}} \exp\left(\alpha_L \left(h^{-1}\right)^{\beta_L}\right)\right);$$

(ii) If Assumptions 4.6 and 4.7 also hold, and if for each $z\in\mathbb{R}$, $\Omega_{V,\lambda}(z,h_n)>0$ *for all* n *sufficiently large, then for each* $z\in \mathrm{supp}(Z)$

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

Finally, we establish a bound on the remainder $R_{V,\lambda}(z,h_n)$. For this, we introduce restrictions on the moments of Z_2.

Assumption 4.8 $E[|Z_2|]<\infty, E[|Z_1 Z_2|]<\infty$, *and* $E[|VZ_2|]<\infty$.

We provide two bounds for $R_{V,\lambda}(z,h_n)$. The first is relevant when one requires a limiting distribution. When instead we only need a convergence rate, a lower bandwidth bound slightly different than that of Assumption 4.7 applies.

Assumption 4.9 *If* $\beta_\theta = 0$ *in Assumption 4.5, then* $h_n^{-1} = O\left(n^{-\eta} n^{(2+2\gamma_1-2\gamma_\theta)^{-1}}\right)$ *for some* $\eta>0$*; otherwise* $h_n^{-1} = O\left((\ln n)^{\beta_\theta^{-1}-\eta}\right)$ *for some* $\eta>0$.

Note that neither of Assumption 4.7 or 4.9 is necessarily stronger than the other.

Theorem 4.6 *(i) Suppose the conditions of Theorem 4.5 hold, together with Assumption 4.8. Then*

$$\begin{aligned}\sup_{z\in\mathbb{R}} |R_{V,\lambda}(z,h_n)| &= O_p\left(n^{(-1/2)+\epsilon}\left(1+h_n^{-1}\right)^{1+\gamma_1-\gamma_\theta}\exp\left(-\alpha_\theta\left(h_n^{-1}\right)^{\beta_\theta}\right)\right)\\ &\quad\times O_p\left(n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{\lambda,L}}\exp\left(\alpha_L\left(h_n^{-1}\right)^{\beta_L}\right)\right)\end{aligned}$$

for some $\varepsilon > 0$. *(ii) If Assumption 4.9 holds in place of Assumption 4.7, then*

$$\sup_{z\in\mathbb{R}} |R_{V,\lambda}(z,h_n)| = o_p\left(n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{\lambda,L}}\exp\left(\alpha_L\left(h_n^{-1}\right)^{\beta_L}\right)\right).$$

We can now collect Theorems 4.4-4.6 into two straightforward corollaries, one establishing a convergence rate and one establishing asymptotic normality.

Corollary 4.7 *If the conditions of Theorem 4.6(ii) hold, then*

$$\begin{aligned}\sup_{z\in\mathbb{R}} |\hat{g}_{V,\lambda}(z,h_n) - g_{V,\lambda}(z,0)| &= O\left(\left(h_n^{-1}\right)^{\gamma_{\lambda,B}}\exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)\right) + \\ &\quad + O_p\left(n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{\lambda,L}}\exp\left(\alpha_L\left(h_n^{-1}\right)^{\beta_L}\right)\right).\end{aligned}$$

The following assumption ensures that the bias and higher-order terms will never dominate the asymptotically linear terms.

Assumption 4.10 *For given* $\lambda \in \{0,...,\Lambda\}, h_n \to 0$ *at a rate such that for each* $z \in \text{supp}(Z)$ *such that* $\Omega_{V,\lambda}(z,h_n) > 0$ *for all* n *sufficiently large, we have* $n^{1/2}\left(\Omega_{V,\lambda}(z,h_n)\right)^{-1/2} |B_{V,\lambda}(z,h_n)| \xrightarrow{p} 0$ *and* $n^{1/2}\left(\Omega_{V,\lambda}(z,h_n)\right)^{-1/2}|R_{V,\lambda}(z,h_n)| \xrightarrow{p} 0$.

For our next result, it is not sufficient to require that $B_{V,\lambda}(z,h)$ and $R_{V,\lambda}(z,h)$ are small relative to the bound given in eq.(28), because the latter is an upper bound. Instead, Assumption 4.10 ensures a lower bound on $\Omega_{V,\lambda}(z,h_n)$. While we give this assumption in a fairly high-level form for clarity, one can state more primitive (but also more cumbersome) sufficient conditions using techniques given in Schennach (2004b).

Corollary 4.8 *If the conditions of Theorem 4.6(i) and Assumption 4.10 hold, then for each* $z \in \text{supp}(Z)$ *such that* $\Omega_{V,\lambda}(z,h_n) > 0$ *for all* n *sufficiently large, we have*

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

Just as in the OXI case, we now consider the case of a functional b of a finite vector $g \equiv (g_{V_1,\lambda_1}, \ldots, g_{V_J,\lambda_J})$ of quantities of the general form of eq.(13) and seek the asymptotic properties of $b(\hat{g}(\cdot, h)) - b(g) \equiv b((\hat{g}_{V_1,\lambda_1}(\cdot, h), \ldots, \hat{g}_{V_J,\lambda_J}(\cdot, h))) - b((g_{V_1,\lambda_1}, \ldots, g_{V_J,\lambda_J}))$.

We first require minimum convergence rates, which we state here in a high-level form for conciseness — primitive conditions can be obtained via Theorems 4.4-4.6.

Assumption 4.11 *For given* $\lambda \in \{0, ..., \Lambda\}$, $\sup_{z\in\mathbb{R}} |B_{V,\lambda}(z, h_n)| = o\left(n^{-1/2}\right)$, $\sup_{z\in\mathbb{R}} |L_{V,\lambda}(z, h_n)| = o_p\left(n^{-1/4}\right)$, *and* $\sup_{z\in\mathbb{R}} |R_{V,\lambda}(z, h_n)| = o_p\left(n^{-1/2}\right)$.

The following theorem consists of two parts, one establishing the validity of an asymptotically linear representation, useful for analyzing a scalar estimator constructed as a functional of a vector of estimators. The second part gives a convenient asymptotic normality and root-n consistency result useful for analyzing β_w and β_{wf_Z}.

Theorem 4.9 *For given* Λ, $J \in \mathbb{N}$, *let* $\lambda_1, \ldots, \lambda_J$ *belong to* $\{0, ..., \Lambda\}$, *and suppose that* $\{V_{1i}, ..., V_{Ji}, Z_i, U_{1i}, U_{2i}\}$ *is an IID sequence of random vectors such that* $\{V_{ji}, Z_i, U_{1i}, U_{2i}\}$ *satisfies the conditions of Corollary 4.8 and Assumption 4.11 for* $j = 1, ..., J$, *with identical choices of* k *and* h_n.

Let the real-valued functional b *satisfy, for any* $\tilde{g} \equiv (\tilde{g}_{V_1,\lambda_1}, \ldots, \tilde{g}_{V_J,\lambda_J})$ *in an* L_∞ *neighborhood of the* $k-$*vector* $g \equiv (g_{V_1,\lambda_1}, ..., g_{V_J,\lambda_J})$,

$$b(\tilde{g}) - b(g) = \sum_{j=1}^{J} \int \left(\tilde{g}_{V_j,\lambda_j}(z) - g_{V_j,\lambda_j}(z)\right) s_j(z)\, dz + \sum_{j=1}^{J} O\left(\left\|\tilde{g}_{V_j,\lambda_j} - g_{V_j,\lambda_j}\right\|_\infty^2\right) \quad (29)$$

for some real-valued functions $s_j, j = 1, ..., J$. *If* s_j *is such that* $\int |s_j(z)|\, dz < \infty$ *and* $\int \bar{\Psi}_{s_j,V_j,\lambda_j}(\xi)\, d\xi < \infty$, *where*

$$\begin{aligned} \bar{\Psi}_{s,V,\lambda}(\xi) &\equiv \frac{1}{|\theta_1(\xi)|}\left(1 + \frac{|\theta_{Z_1}(\xi)|}{|\theta_1(\xi)|}\right)\left(\{\int_{|\xi|}^{\infty} |\sigma_s(\zeta)|\, |\zeta|^\lambda\, |\phi_V(\zeta)|\, d\zeta\} + |\sigma_s(\xi)|\, |\xi|^\lambda\, |\phi_1(\xi)|\right) \\ \theta_{Z_1}(\xi) &\equiv E\left[Z_1 e^{i\xi Z_2}\right] \\ \sigma_s(\xi) &\equiv \int s(z)\, e^{\mathbf{i}\xi z} dz, \end{aligned}$$

for each $j=1,...,J$, then

$$b\left(\hat{g}\left(\cdot,h_n\right)\right)-b\left(g\right)=\sum_{j=1}^{J}\hat{E}\left[\psi_{V_j,\lambda_j}\left(s_j;V_j,Z_1,Z_2\right)\right]+o_p\left(n^{-1/2}\right),$$

where

$$\begin{aligned}\psi_{V,\lambda}\left(s;v,z_1,z_2\right) &\equiv \int\Psi_{s,V,\lambda,1}\left(\xi\right)\left(e^{\mathbf{i}\xi z_2}-E\left[e^{\mathbf{i}\xi Z_2}\right]\right)d\xi\\ &+\int\Psi_{s,V,\lambda,Z_1}\left(\xi\right)\left(z_1e^{\mathbf{i}\xi z_2}-E\left[Z_1e^{\mathbf{i}\xi Z_2}\right]\right)d\xi\\ &+\int\Psi_{s,V,\lambda,V}\left(\xi\right)\left(ve^{\mathbf{i}\xi z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right)d\xi,\end{aligned}$$

with

$$\begin{aligned}\Psi_{s,V,\lambda,1}\left(\xi\right) &\equiv -\frac{1}{2\pi}\frac{\phi_V\left(\xi\right)}{\theta_1\left(\xi\right)}\sigma_s^{\dagger}\left(\xi\right)\left(-\mathbf{i}\xi\right)^{\lambda}-\frac{1}{2\pi}\frac{\mathbf{i}\theta_{Z_1}\left(\xi\right)}{\left(\theta_1\left(\xi\right)\right)^2}\int_{\xi}^{\pm\infty}\sigma_s^{\dagger}\left(\zeta\right)\left(-\mathbf{i}\zeta\right)^{\lambda}\phi_V\left(\zeta\right)d\zeta\\ \Psi_{s,V,\lambda,Z_1}\left(\xi\right) &\equiv \frac{1}{2\pi}\frac{\mathbf{i}}{\theta_1\left(\xi\right)}\int_{\xi}^{\pm\infty}\sigma_s^{\dagger}\left(\zeta\right)\left(-\mathbf{i}\zeta\right)^{\lambda}\phi_V\left(\zeta\right)d\zeta\\ \Psi_{s,V,\lambda,V}\left(\xi\right) &\equiv \frac{1}{2\pi}\frac{\phi_1\left(\xi\right)}{\theta_1\left(\xi\right)}\sigma_s^{\dagger}\left(\xi\right)\left(-\mathbf{i}\xi\right)^{\lambda},\end{aligned}$$

where $\dagger$ denotes the complex conjugate. Moreover,

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

where

$$\Omega_b=E\left[\left(\sum_{j=1}^{k}\psi_{V_j,\lambda_j}\left(s_j;V_j,Z_1,Z_2\right)\right)^2\right]<\infty.$$

4.4 Asymptotics: PXI Case

Having derived general asymptotic results, we now apply them to the main quantities of interest (eqs.(9) and (12)). Consider the following nonparametric estimator of $\beta\left(z\right)$:

$$\hat{\beta}\left(z,h\right)\equiv D_z\hat{\mu}_Y(z,h)\ /\ D_z\hat{\mu}_X(z,h) \tag{30}$$

for $z\in\text{supp}(Z)$, where, using the kernel estimators $\hat{g}$ of the preceding section, we have

$$\begin{aligned}D_z\hat{\mu}_Y(z,h) &\equiv \frac{\hat{g}_{Y,1}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)}-\frac{\hat{g}_{Y,0}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)}\frac{\hat{g}_{1,1}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)} \qquad \text{and}\\ D_z\hat{\mu}_X(z,h) &\equiv \frac{\hat{g}_{X,1}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)}-\frac{\hat{g}_{X,0}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)}\frac{\hat{g}_{1,1}\left(z,h\right)}{\hat{g}_{1,0}\left(z,h\right)}.\end{aligned}$$

Combining the results from the previous section with a straightforward Taylor expansion yields the following result.

Theorem 4.10 *Suppose that* $\{X_i, Y_i, Z_i, U_{1i}, U_{2i}\}$ *is an IID sequence satisfying the conditions of Corollary 4.7 for* $V = 1, X, Y,$ *with* $\Lambda \geq 1$ *and* $\lambda = 0, 1,$ *and with identical choices of* k *and* h_n*. Further, suppose* $\max_{V=1,X,Y} \max_{\lambda=0,1} \sup_{z\in\mathbb{R}} |g_{V,\lambda}(z)| < \infty,$ *and for* $\tau > 0,$ *define*

$$\mathbf{Z}_\tau \equiv \{z \in \mathbb{R} : f_Z(z) \geq \tau \text{ and } |D_z \mu_X(z)| \geq \tau\}.$$

Then

$$\begin{aligned}\sup_{z\in\mathbf{Z}_\tau} \left|\hat{\beta}(z, h_n) - \beta(z)\right| &= O\left(\tau^{-4}\left(h_n^{-1}\right)^{\gamma_{1,B}} \exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)\right) + \\ &\quad + O_p\left(\tau^{-4} n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{1,L}} \exp\left(\alpha_L\left(h^{-1}\right)^{\beta_L}\right)\right),\end{aligned}$$

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

$$\sup_{z\in\mathbf{Z}_{\tau_n}} \left|\hat{\beta}(z, h_n) - \beta(z)\right| = o_p(1).$$

The delta method secures the next result.

Theorem 4.11 *Suppose that* $\{X_i, Y_i, Z_i, U_{1i}, U_{2i}\}$ *is an IID sequence satisfying the conditions of Corollary 4.8 for* $V = 1, X, Y,$ *with* $\Lambda \geq 1$ *and* $\lambda = 0, 1,$ *and with identical choices of* k *and* h_n*. Further, suppose* $\max_{V=1,X,Y} \max_{\lambda=0,1} \sup_{z\in\mathbb{R}} |g_{V,\lambda}(z)| < \infty$*. Then for all* $z \in \text{supp}(Z)$ *such that* $|D_z \mu_X(z)| > 0,$

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

provided that

$$\Omega_\beta(z, h) = E\left[\left(\ell_\beta(z, h; X, Y, Z_1, Z_2)\right)^2\right]$$

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

$$\begin{aligned}\ell_\beta(z, h; x, y, z_1, z_2) &= \sum_{\lambda=0,1} (s_{X,1,\lambda}(z)\,\ell_{1,\lambda}(z, h; 1, z_1, z_2) + s_{X,X,\lambda}(z)\,\ell_{X,\lambda}(z, h; x, z_1, z_2) \\ &\quad + s_{Y,1,\lambda}(z)\,\ell_{1,\lambda}(z, h; 1, z_1, z_2) + s_{Y,Y,\lambda}(z)\,\ell_{Y,\lambda}(z, h; y, z_1, z_2)),\end{aligned}$$

and where $s_{X,1,\lambda}(z)$, $s_{X,X,\lambda}(z)$, $s_{Y,1,\lambda}(z)$, *and* $s_{Y,Y,\lambda}(z)$ *for* $\lambda = 0, 1$ *are as defined in Theorem 3.5.*

We now consider semiparametric functionals taking the forms of eq.(12) and analyze the estimators

$$\begin{aligned} \hat{\beta}_w &= \int_{S_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(z, h_n) w(z) dz \\ \hat{\beta}_{wf_Z} &= \int_{S_{\hat{\beta}(\cdot,h_n)}} \hat{\beta}(z, h_n) w(z) \hat{g}_{1,0}(z, h_n) dz, \end{aligned}$$

where $S_{\hat{\beta}(\cdot,h_n)} \equiv \{z : \hat{g}_{1,0}(z, h_n) > 0, |D_z\hat{\mu}_X(z,h)| > 0\}$.

The asymptotic distributions of these estimators follow by straightforward application of Theorem 4.9, analogously to the OXI case. Thanks to their semiparametric nature, root$-n$ consistency and asymptotic normality is possible.

Theorem 4.12 *Suppose the conditions of Theorem 4.9 hold for* $V = 1, X, Y$ *and* $\lambda = 0, 1$, *and that Assumption 3.8 holds. Then*

$$n^{1/2}\Omega_w^{-1/2}\left(\hat{\beta}_w - \beta_w\right) \xrightarrow{d} N(0, 1),$$

provided that

$$\Omega_w \equiv E\left[\left(\psi_{\beta_w}(X, Y, Z_1, Z_2)\right)^2\right]$$

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

$$\begin{aligned} \psi_{\beta_w}(x, y, z_1, z_2) \equiv & \sum_{\lambda=0,1} (\psi_{1,\lambda}(ws_{X,1,\lambda}; 1, z_1, z_2) + \psi_{X,\lambda}(ws_{X,X,\lambda}; x, z_1, z_2) \\ & +\psi_{1,\lambda}(ws_{Y,1,\lambda}; 1, z_1, z_2) + \psi_{Y,\lambda}(ws_{Y,Y,\lambda}; y, z_1, z_2)), \end{aligned}$$

$ws_{A,V,\lambda}$ *denotes the function mapping* z *to* $w(z) s_{A,V,\lambda}(z)$, *and where* $\psi_{V,\lambda}$ *is defined in Theorem 4.9.*

Theorem 4.13 *Suppose the conditions of Theorem 4.9 hold for* $V = 1, X, Y$ *and* $\lambda = 0, 1$, *and that Assumption 3.8 holds. Then*

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

provided that

$$\Omega_{wf_Z} \equiv E\left[\left(\psi_{\beta_{wf_z}}(X, Y, Z_1, Z_2)\right)^2\right]$$

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

$$\begin{aligned}\psi_{\beta_{wf_Z}}(x, y, z_1, z_2) \;\; &\equiv \;\; \{\sum_{\lambda=0,1}(\psi_{1,\lambda}(wf_Z s_{X,1,\lambda}; 1, z_1, z_2) + \psi_{X,\lambda}(wf_Z s_{X,X,\lambda}; x, z_1, z_2) \\ &\quad +\psi_{1,\lambda}(wf_Z s_{Y,1,\lambda}; 1, z_1, z_2) + \psi_{Y,\lambda}(wf_Z s_{Y,Y,\lambda}; y, z_1, z_2))\} \\ &\quad +\psi_{1,0}(w\beta; 1, z_1, z_2),\end{aligned}$$

$wf_Z s_{A,V,\lambda}$ *denotes the function mapping* z *to* $w(z)\, f_Z(z) s_{A,V,\lambda}(z)$, $w\beta$ *denotes the function mapping* z *to* $w(z)\,\beta(z)$, *and where* $\psi_{V,\lambda}$ *is defined in Theorem 4.9.*

Although we do not provide explicit theorems due to space limitations, it is straightforward to show that the asymptotic variances in Theorems 4.9, 4.12, 4.13 can be consistently estimated, since we provide an explicit expression for the appropriate influence functions. In the cases of Theorems 4.5, 4.8, and 4.11, the bandwidth-dependence of the variance is nontrivial, and it is not guaranteed that the same bandwidth sequence used for the point estimators provides suitably consistent estimators of the asymptotic variance. Consequently, it may be more convenient to rely on subsampling methods for purposes of inference. Fortunately, powerful subsampling methods designed to handle generic convergence rates (such as ours) are available from Bertail, Politis, Haefke, and White (2004). These require nothing more than the existence of a limiting distribution for a suitably normalized estimator, precisely as we have already established in our results above.

While the above treatment covers proxies for instruments whose measurement errors satisfy conditional mean or independence assumptions, more general forms of proxies contaminated by either "nonclassical" or "Berkson-type"[2] measurement errors could be considered by adapting the techniques developed in Hu and Schennach (2007) or Schennach (2007), respectively.

5 Discussion

The estimation results of Sections 3 and 4 apply to *any* random variables satisfying the given regularity conditions, and these do not involve structural relations. Thus, in the absence of further conditions, these estimators have no necessary structural meaning. To interpret estimators of $\beta(z)$ as measuring an average marginal effect, Assumptions 2.1 and 2.2 suffice, as Proposition 2.2 ensures. When Assumption 2.2 fails, analysis analogous to that of White and Chalak (2006, section 5.1) shows that $\beta(z) = \gamma(z)\beta^*(z) + \delta(z)$, where $\gamma(z)$ and $\delta(z)$ are not stochastically identified, but generally satisfy $\gamma(z) \neq 1$ and $\delta(z) \neq 0$. When Assumption 2.1 fails, then $\beta^*(z)$ is no longer even defined. Thus, Assumptions 2.1 and 2.2 are crucial to giving a structural interpretation to an estimator of $\beta(z)$.

In sharp contrast to the linear case, here we rely on $U_z \perp U_x$ to structurally identify $\beta(z)$. In the linear case, this assumption is not necessary, and in the PXI case, a convenient simplification renders even $D_z\mu_{Y,1}(z)$ / $D_z\mu_{X,1}(z)$ structurally identified (see CW). The simplicity of the linear case masks the fundamental differences between OXI and PXI.

Inspecting the measurement assumptions of Section 4 (Assumptions 4.1, 4.2, 4.4, 4.6, and 4.8) reveals an asymmetry in the properties assumed of Z_1 and Z_2 and/or U_1 and U_2. Although this asymmetry may be present in some applications, in others symmetry may be more plausible. In the latter situations, one can construct two estimators of $\beta(z)$, say $\hat{\beta}_1(z, h_n)$ and $\hat{\beta}_2(z, h_n)$, by interchanging the roles of Z_1 and Z_2. Using these, one can construct a weighted estimator with superior asymptotic efficiency, having the GLS form

$$[\iota'\hat{\Sigma}(z, h_n)^{-1}\iota]^{-1}\iota'\hat{\Sigma}(z, h_n)^{-1}\hat{\boldsymbol{\beta}}(z, h_n),$$

[2] A instrument proxy contaminated by a Berkson-type error can be directly used as an instrument, unless we wish to identify effects conditional on the true instrument instead of its proxy.

where $\iota \equiv (1,1)'$, and $\hat{\Sigma}(z, h_n)$ represents an estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(z, h_n) \equiv (\hat{\beta}_1(z, h_n), \hat{\beta}_2(z, h_n))'$ (suitably scaled). The estimator $\hat{\Sigma}(z, h_n)$ can be constructed using subsampling, as in Section 4. The same approach applies to estimating functionals of β.

More generally, one may have multiple error-laden measurements of an unobserved exogenous instrument Z, say $(Z_1, ..., Z_k), k > 2$. Depending on the measurement properties plausible for these, one can construct a vector of consistent asymptotically normal estimators $\hat{\boldsymbol{\beta}}(z, h_n) \equiv (\hat{\beta}_1(z, h_n), ..., \hat{\beta}_\ell(z, h_n))'$, where $\ell \geq k$. From these, one can construct a relatively efficient estimator as a GLS weighted combination of the elements of $\hat{\boldsymbol{\beta}}(z, h_n)$, analogous to the case with $\ell = 2$ given above.

6 Summary and Concluding Remarks

In this paper we provide consistent and asymptotically normal nonparametric estimators of average marginal effects of an endogenous cause, X, on a response of interest, Y, for a general system of structural equations. The system is general in that we do not assume linearity, separability, or monotonicity for the structural relations. Our estimators are local indirect least squares (LILS) estimators analogous to those introduced by Heckman and Vytlacil (1999, 2001) for an index model involving a binary X. We treat two cases, the traditional OXI case and the PXI case, where the exogenous instrument cannot be observed, but where as few as two error-laden proxies are available.

For the OXI case, we use the infinite order ("flat-top") kernels of Politis and Romano (1999), obtaining uniform convergence rates as well as asymptotic normality for estimators of identified instrument-conditioned average marginal effects and root$-n$ consistency of their weighted averages. For the PXI case, we develop new results for estimating densities and expectations conditional on mismeasured variables, as well as their derivatives with respect to the mismeasured variable. We provide uniform convergence rates, as well as asymptotic normality results in fully nonparametric settings. We also consider nonlinear functionals of such nonparametric quantities and use these to establish root-n consistency and asymptotic normality for estimators applicable to the PXI case. Previously, only results for the quite special linear PXI case were available (CW); by covering the general nonseparable case, the present results necessarily also cover the widely applicable PXI cases in which one or the other of q or r is separable: $\beta^*_{ss}, \beta^*_{sn}$, and β^*_{ns}.

There are a variety of interesting directions for further research. In particular, it is of interest to develop the proposed tests of the separability of q based on our estimators. It also appears relatively straightforward to develop estimators analogous to those given here for average marginal effects of endogenous causes in non-recursive ("simultaneous") nonseparable systems. Finally, it appears feasible and is of considerable interest to extend the methods developed here to provide nonparametric analogs of the various extended instrumental variables estimators analyzed by CW.

A Appendix

The proofs of Lemma 3.1, Theorem 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 are fairly standard and can be found in the supplementary material.

Proof of Lemma 4.1. Assumption 4.1 ensures that all expectations below exist and are finite. Given Assumptions 4.2 and 4.3, we have

$$\begin{aligned}
\frac{\mathbf{i}E\left[Z_1 e^{\mathbf{i}\xi Z_2}\right]}{E\left[e^{\mathbf{i}\xi Z_2}\right]} &= \frac{\mathbf{i}E\left[Ze^{\mathbf{i}\xi(Z+U_2)}\right]+\mathbf{i}E\left[E\left[U_1|Z,U_2\right]e^{\mathbf{i}\xi(Z+U_2)}\right]}{E\left[e^{\mathbf{i}\xi(Z+U_2)}\right]} \\
&= \frac{\mathbf{i}E\left[Ze^{\mathbf{i}\xi(Z+U_2)}\right]}{E\left[e^{\mathbf{i}\xi(Z+U_2)}\right]} = \frac{\mathbf{i}E\left[Ze^{\mathbf{i}\xi Z}\right]}{E\left[e^{\mathbf{i}\xi Z}\right]}\frac{E\left[e^{\mathbf{i}\xi U_2}\right]}{E\left[e^{\mathbf{i}\xi U_2}\right]} \\
&= \frac{\mathbf{i}E\left[Ze^{\mathbf{i}\xi Z}\right]}{E\left[e^{\mathbf{i}\xi Z}\right]} = D_\xi \ln\left(E\left[e^{\mathbf{i}\xi Z}\right]\right).
\end{aligned}$$

It follows that for each real ζ,

$$\begin{aligned}
\phi_V(\zeta) &\equiv E\left[Ve^{\mathbf{i}\zeta Z}\right] = \frac{E\left[Ve^{\mathbf{i}\zeta Z}\right]E\left[e^{\mathbf{i}\zeta U_2}\right]}{E\left[e^{\mathbf{i}\zeta Z}\right]E\left[e^{\mathbf{i}\zeta U_2}\right]}E\left[e^{\mathbf{i}\zeta Z}\right] \\
&= \frac{E\left[Ve^{\mathbf{i}\zeta Z_2}\right]}{E\left[e^{\mathbf{i}\zeta Z_2}\right]}E\left[e^{\mathbf{i}\zeta Z}\right] \\
&= \frac{E\left[Ve^{\mathbf{i}\zeta Z_2}\right]}{E\left[e^{\mathbf{i}\zeta Z_2}\right]}\exp\left(\ln\left(E\left[e^{\mathbf{i}\zeta Z}\right]\right)-\ln 1\right) \\
&= \frac{E\left[Ve^{\mathbf{i}\zeta Z_2}\right]}{E\left[e^{\mathbf{i}\zeta Z_2}\right]}\exp\left(\int_0^\zeta D_\xi \ln\left(E\left[e^{\mathbf{i}\xi Z}\right]\right)d\xi\right) \\
&= \frac{E\left[Ve^{\mathbf{i}\zeta Z_2}\right]}{E\left[e^{\mathbf{i}\zeta Z_2}\right]}\exp\left(\int_0^\zeta \frac{\mathbf{i}E\left[Z_1e^{\mathbf{i}\xi Z_2}\right]}{E\left[e^{\mathbf{i}\xi Z_2}\right]}d\xi\right).
\end{aligned}$$

For each $\lambda \in \{0, ..., \Lambda\}$ and $z \in \text{supp}(Z)$, we have

$$\frac{1}{2\pi}\int(-\mathbf{i}\zeta)^\lambda \phi_V(\zeta)\exp(-\mathbf{i}\zeta z)\,d\zeta = \frac{1}{2\pi}\int(-\mathbf{i}\zeta)^\lambda E\left[Ve^{\mathbf{i}\zeta Z}\right]\exp(-\mathbf{i}\zeta z)\,d\zeta.$$

The expression on the right is the inverse Fourier transform of $(-\mathbf{i}\zeta)^\lambda E\left[Ve^{\mathbf{i}\zeta Z}\right]$. Integration by parts, valid under Assumptions 3.1 and 3.3, gives

$$\begin{aligned}(-\mathbf{i}\zeta)^{\lambda} E\left[Ve^{\mathbf{i}\zeta Z}\right] &= (-\mathbf{i}\zeta)^{\lambda} \int E\left[V|Z=z\right] f_Z(z)\, e^{\mathbf{i}\zeta z} dz \\ &= (-1)^{\lambda} \int E\left[V|Z=z\right] f_Z(z)\, D_z^{\lambda} e^{\mathbf{i}\zeta z} dz \\ &= \int \left(D_z^{\lambda}\left(E\left[V|Z=z\right] f_Z(z)\right)\right) e^{\mathbf{i}\zeta z} dz \\ &= \int g_{V,\lambda}(z)\;\; e^{\mathbf{i}\zeta z} dz.\end{aligned}$$

As the final expression is the Fourier transform of $g_{V,\lambda}(z)$, the conclusion follows. ■

Proof of Lemma 4.2. Assumptions 3.1, 4.1, 3.3, and 3.4 ensure the existence of

$$\begin{aligned}g_{V,\lambda}(z,h) &\equiv \int \frac{1}{h} k\left(\frac{\tilde{z}-z}{h}\right) g_{V,\lambda}(\tilde{z})\, d\tilde{z} \\ &= \int \frac{1}{h} k\left(\frac{\tilde{z}-z}{h}\right) D_{\tilde{z}}^{\lambda}\left(E\left[V|Z=\tilde{z}\right] f_Z(\tilde{z})\right) d\tilde{z}.\end{aligned}$$

By the Convolution Theorem, the inverse Fourier Transform (FT) of the product of $\kappa(h\zeta)$ and $(-\mathbf{i}\zeta)^{\lambda} E\left[Ve^{\mathbf{i}\zeta Z}\right]$ is the convolution between the inverse FT of $\kappa(h\zeta)$ and the inverse FT of $(-\mathbf{i}\zeta)^{\lambda} E\left[Ve^{\mathbf{i}\zeta Z}\right]$. The inverse FT of $\kappa(h\zeta)$ is $h^{-1}k(z/h)$, and the inverse FT of $(-\mathbf{i}\zeta)^{\lambda} E\left[Ve^{\mathbf{i}\zeta Z}\right]$ is $D_z^{\lambda}\left(E\left[V|Z=z\right] f_Z(z)\right)$. It follows that

$$\begin{aligned}g_{V,\lambda}(z,h) &= \frac{1}{2\pi} \int \kappa(h\zeta)\left((-\mathbf{i}\zeta)^{\lambda} E\left[Ve^{\mathbf{i}\zeta Z}\right]\right) \exp(-\mathbf{i}\zeta z)\, d\zeta \\ &= \frac{1}{2\pi} \int \kappa(h\zeta)(-\mathbf{i}\zeta)^{\lambda} \phi_V(\zeta) \exp(-\mathbf{i}\zeta z)\, d\zeta.\end{aligned}$$

■

Proof of Lemma 4.3. Assumptions 3.1, 4.1, 3.3, and 3.4 ensure the existence of $g_{V,\lambda}(z)$ and $g_{V,\lambda}(z,h)$. Adding and subtracting appropriately gives eq.(25), where for any $\bar{g}_{V,\lambda}(z,h)$

$$\begin{aligned}B_{V,\lambda}(z,h) &\equiv g_{V,\lambda}(z,h) - g_{V,\lambda}(z) \\ L_{V,\lambda}(z,h) &\equiv \bar{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h) \\ R_{V,\lambda}(z,h) &\equiv \hat{g}_{V,\lambda}(z,h) - \bar{g}_{V,\lambda}(z,h).\end{aligned}$$

We now derive the form that $\bar{g}_{V,\lambda}(z,h)$ must have in order for $L_{V,\lambda}(z,h)$ to be a linearization of $\hat{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h)$.

Recall that for $A = 1, Z_1$, and V, we let $\theta_A(\zeta) \equiv E\left[Ae^{\mathbf{i}\zeta Z_2}\right]$. Also, write $\hat{\theta}_A(\zeta) \equiv \hat{E}\left[Ae^{\mathbf{i}\zeta Z_2}\right]$ and $\delta\hat{\theta}_A(\zeta) \equiv \hat{\theta}_A(\zeta) - \theta_A(\zeta)$. We first state a useful representation for $\hat{\theta}_V(\zeta)/\hat{\theta}_1(\zeta)$:

$$\frac{\hat{\theta}_V(\zeta)}{\hat{\theta}_1(\zeta)} = \frac{\theta_V(\zeta) + \delta\hat{\theta}_V(\zeta)}{\theta_1(\zeta) + \delta\hat{\theta}_1(\zeta)} = q_V(\zeta) + \delta\hat{q}_V(\zeta), \tag{31}$$

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

$$\delta\hat{q}_V(\zeta) = \left(\frac{\delta\hat{\theta}_V(\zeta)}{\theta_1(\zeta)} - \frac{\theta_V(\zeta)\,\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1} \tag{32}$$

or

$$\begin{aligned}
\delta\hat{q}_V(\zeta) &= \delta_1\hat{q}_V(\zeta) + \delta_2\hat{q}_V(\zeta), \qquad \text{with} \qquad (33)\\
\delta_1\hat{q}_V(\zeta) &\equiv \frac{\delta\hat{\theta}_V(\zeta)}{\theta_1(\zeta)} - \frac{\theta_V(\zeta)\,\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\\
\delta_2\hat{q}_V(\zeta) &\equiv \frac{\theta_V(\zeta)}{\theta_1(\zeta)}\left(\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^2\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1} - \frac{\delta\hat{\theta}_V(\zeta)}{\theta_1(\zeta)}\frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\left(1 + \frac{\delta\hat{\theta}_1(\zeta)}{\theta_1(\zeta)}\right)^{-1}.
\end{aligned}$$

Similarly, for $Q_z(\xi) \equiv \int_0^\xi (\mathbf{i}\theta_z(\zeta)/\theta_1(\zeta))d\zeta$, $\delta\hat{Q}_z(\xi) \equiv \int_0^\xi (\mathbf{i}\hat{\theta}_z(\zeta)/\hat{\theta}_1(\zeta))d\zeta - Q_z(\xi)$, and some random function $\delta\bar{Q}_z(\xi)$ such that $\left|\delta\bar{Q}_z(\xi)\right| \leq \left|\delta\hat{Q}_z(\xi)\right|$ for all ξ,

$$\exp\left(Q_z(\xi) + \delta\hat{Q}_z(\xi)\right) = \exp\left(Q_z(\xi)\right)\left(1 + \delta\hat{Q}_z(\xi) + \frac{1}{2}[\exp\left(\delta\bar{Q}_z(\xi)\right)]\left(\delta\hat{Q}_z(\xi)\right)^2\right). \tag{34}$$

Substituting eqs.(31) and (34) into

$$\begin{aligned}
&\hat{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h)\\
=\;& \frac{1}{2\pi}\int \kappa(h\xi)\left[\frac{\hat{\theta}_V(\xi)}{\hat{\theta}_1(\xi)}\exp\left(\int_0^\xi \frac{\mathbf{i}\hat{\theta}_{Z_1}(\zeta)}{\hat{\theta}_1(\zeta)}d\zeta\right) - \frac{\theta_V(\xi)}{\theta_1(\xi)}\exp\left(\int_0^\xi \frac{\mathbf{i}\theta_{Z_1}(\zeta)}{\theta_1(\zeta)}d\zeta\right)\right]d\xi
\end{aligned}$$

and keeping the terms linear in $\delta\hat{\theta}_1(\zeta)$ or $\delta\hat{\theta}_{Z_1}(\zeta)$ gives the linearization of $\hat{g}_{V,\lambda}(z,h)$, denoted $\bar{g}_{V,\lambda}(z,h)$:

$$
\begin{aligned}
&\bar{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h) \\
= \;& \frac{1}{2\pi}\int \exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^{\lambda}\kappa(h\xi)\phi_V(\xi)[\int_0^{\xi}\left(\frac{\mathbf{i}\delta\hat{\theta}_{Z_1}(\zeta)}{\theta_1(\zeta)} - \frac{\mathbf{i}\theta_{Z_1}(\zeta)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)d\zeta]d\xi \\
&+\frac{1}{2\pi}\int \exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^{\lambda}\kappa(h\xi)\left(\frac{\delta\hat{\theta}_V(\xi)}{\theta_1(\xi)}\phi_1(\xi) - \frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\phi_V(\xi)\right)d\xi.
\end{aligned}
$$

Using the identity

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

for any absolutely integrable function f, we obtain

$$
\begin{aligned}
L_{V,\lambda}(z,h) \;\equiv\;& \bar{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h) \\
=\;& \frac{1}{2\pi}\int\int_{\zeta}^{\pm\infty} \exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^{\lambda}\kappa(h\xi)\phi_V(\xi)\,d\xi\left(\frac{\mathbf{i}\delta\hat{\theta}_{Z_1}(\zeta)}{\theta_1(\zeta)} - \frac{\mathbf{i}\theta_{Z_1}(\zeta)\delta\hat{\theta}_1(\zeta)}{(\theta_1(\zeta))^2}\right)d\zeta \\
&+ \frac{1}{2\pi}\int \exp(-\mathbf{i}\xi z)(-\mathbf{i}\xi)^{\lambda}\kappa(h\xi)\left(\frac{\delta\hat{\theta}_V(\xi)}{\theta_1(\xi)}\phi_1(\xi) - \frac{\delta\hat{\theta}_1(\xi)}{\theta_1(\xi)}\phi_V(\xi)\right)d\xi \\
=\;& \sum_{A=1,Z_1,V}\int \Psi_{V,\lambda,A}(\xi,z,h)\,\delta\hat{\theta}_A(\xi)\,d\xi \\
=\;& \sum_{A=1,Z_1,V}\int \Psi_{V,\lambda,A}(\xi,z,h)\left(\hat{E}\left[Ae^{\mathbf{i}\xi Z_2}\right] - E\left[Ae^{\mathbf{i}\xi Z_2}\right]\right)d\xi \qquad (35)\\
=\;& \hat{E}\left[\sum_{A=1,Z_1,V}\int \Psi_{V,\lambda,A}(\xi,z,h)\left(Ae^{\mathbf{i}\xi Z_2} - E\left[Ae^{\mathbf{i}\xi Z_2}\right]\right)d\xi\right] \\
=\;& \hat{E}\left[\ell_{V,\lambda}(z,h;V,Z_1,Z_2)\right],
\end{aligned}
$$

where $\Psi_{V,\lambda,A}(\xi,z,h)$ and $\ell_{V,\lambda}(z,h;V,Z_1,Z_2)$ are defined in the statement of the Lemma. ■

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

Proof of Theorem 4.4. By Parseval's identity, we have

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

where we use Assumption 3.4 to ensure $\kappa(\zeta) = 1$ for $|\zeta| \leq \bar{\xi}$ and $\sup_{\zeta}|\kappa(\zeta)| < \infty$. Thus, Assumption 3.5 (eq.(15)) yields

$$\begin{aligned}
|B(z,h)| &\preceq \int_{\bar{\xi}/h}^{\infty}\zeta^{\lambda}(1+\zeta)^{\gamma_\phi}\exp\left(\alpha_\phi\zeta^{\beta_\phi}\right)d\zeta \\
&= O\left(\left(\bar{\xi}/h\right)^{\gamma_\phi+\lambda+1}\exp\left(\alpha_\phi\left(\bar{\xi}/h\right)^{\beta_\phi}\right)\right) \\
&= O\left(\left(h^{-1}\right)^{\gamma_{\lambda,B}}\exp\left(\alpha_B\left(h^{-1}\right)^{\beta_B}\right)\right).
\end{aligned}$$

■

Lemma A.1 *Suppose the conditions of Lemma 4.3 hold. For each* ξ *and* h*, and for* $A = 1, Z_1, V$*, let* $\Psi^{+}_{V,\lambda,A}(\xi,h) \equiv \sup_{z\in\mathbb{R}}|\Psi_{V,\lambda,A}(\xi,z,h)|$*, and define*

$$\Psi^{+}_{V,\lambda}(h) = \sum_{A=1,Z_1,V}\int\Psi^{+}_{V,\lambda,A}(\zeta,h)\,d\zeta.$$

If Assumption 4.5 also holds, then for $h > 0$

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

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

$$\begin{aligned}
\Psi^{+}_{V,\lambda,1}(\xi,h) &\equiv \sup_{z\in\mathbb{R}} \left|\Psi_{V,\lambda,1}(\xi,z,h)\right| \\
&\leq \sup_{z\in\mathbb{R}} \frac{|\phi_V(\xi)|}{|\theta_1(\xi)|} \left|\exp(-\mathbf{i}\xi z)\right| |\xi|^{\lambda} |\kappa(h\xi)| \\
&\quad + \sup_{z\in\mathbb{R}} \frac{|\theta_{Z_1}(\xi)|}{|\theta_1(\xi)|^2} \int_{\xi}^{\pm\infty} \left|\exp(-\mathbf{i}\zeta z)\right| |\zeta|^{\lambda} |\kappa(h\zeta)| |\phi_V(\zeta)| \, d\zeta \\
&\leq \frac{|\phi_V(\xi)|}{|\theta_1(\xi)|} |\xi|^{\lambda} |\kappa(h\xi)| + \frac{|\theta_{Z_1}(\xi)|}{|\theta_1(\xi)|^2} \int_{\xi}^{\pm\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| |\phi_V(\zeta)| \, d\zeta \\
&= \frac{1}{|\theta_1(\xi)|} \left(|\phi_V(\xi)| |\xi|^{\lambda} |\kappa(h\xi)| + \frac{|\theta_{Z_1}(\xi)|}{|\theta_1(\xi)|} \int_{\xi}^{\pm\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| |\phi_V(\zeta)| \, d\zeta \right) \\
&= \frac{1}{|\theta_1(\xi)|} \left(|\phi_V(\xi)| |\xi|^{\lambda} |\kappa(h\xi)| + \frac{|\phi_1'(\xi)|}{|\phi_1(\xi)|} \int_{\xi}^{\pm\infty} |\zeta|^{\lambda} |\kappa(h\zeta)| |\phi_V(\zeta)| \, d\zeta \right)
\end{aligned}$$

where we use the fact that

$$\begin{aligned}
\frac{\theta_{Z_1}(\xi)}{\theta_1(\xi)} &= \frac{E\left[Z_1 e^{\mathbf{i}\xi Z_2}\right]}{E\left[e^{\mathbf{i}\xi Z_2}\right]} = \frac{E\left[(Z+U_1)\, e^{\mathbf{i}\xi(Z+U_2)}\right]}{E\left[e^{\mathbf{i}\xi(Z+U_2)}\right]} = \frac{E\left[Z e^{\mathbf{i}\xi(Z+U_2)}\right] + E\left[E\left[U_1|Z,U_2\right] e^{\mathbf{i}\xi(Z+U_2)}\right]}{E\left[e^{\mathbf{i}\xi(Z+U_2)}\right]} \\
&= \frac{E\left[Z e^{\mathbf{i}\xi(Z+U_2)}\right]}{E\left[e^{\mathbf{i}\xi(Z+U_2)}\right]} = \frac{E\left[Z e^{\mathbf{i}\xi Z}\right]}{E\left[e^{\mathbf{i}\xi Z}\right]} \frac{E\left[e^{\mathbf{i}\xi U_2}\right]}{E\left[e^{\mathbf{i}\xi U_2}\right]} = \frac{-\mathbf{i}(d/d\xi) E\left[e^{\mathbf{i}\xi Z}\right]}{E\left[e^{\mathbf{i}\xi Z}\right]} = -\mathbf{i}\frac{(d/d\xi)\phi_1(\xi)}{\phi_1(\xi)}
\end{aligned}$$

Integrating $\Psi^{+}_{V,\lambda,1}(\xi,h)$ with respect to ξ and using Assumption 4.5 gives

$$\begin{aligned}
&\int \Psi^{+}_{V,\lambda,1}(\xi,h)\, d\xi \\
&\preceq \int \frac{1}{|\theta_1(\xi)|} \left(|\phi_V(\xi)| |\xi|^{\lambda} 1\left(|\xi| \leq h^{-1}\right) + \frac{|\phi_1'(\xi)|}{|\phi_1(\xi)|} 1\left(|\xi| \leq h^{-1}\right) \int_{|\xi|}^{h^{-1}} |\zeta|^{\lambda} |\phi_V(\zeta)| \, d\zeta \right) d\xi \\
&\preceq \int (1+|\xi|)^{-\gamma_\theta} \exp\left(-\alpha_\theta |\xi|^{\beta_\theta}\right) 1\left(|\xi| \leq h^{-1}\right) \times \\
&\quad \times \left((1+|\xi|)^{\gamma_\phi} \exp\left(\alpha_\phi |\xi|^{\beta_\phi}\right) |\xi|^{\lambda} + (1+|\xi|)^{\gamma_1} \int_0^{h^{-1}} |\zeta|^{\lambda} (1+|\zeta|)^{\gamma_\phi} \exp\left(\alpha_\phi |\zeta|^{\beta_\phi}\right) d\zeta \right) d\xi \\
&\preceq \int_0^{h^{-1}} (1+|\xi|)^{-\gamma_\theta} \exp\left(-\alpha_\theta |\xi|^{\beta_\theta}\right) \times \\
&\quad \times \left((1+|\xi|)^{\gamma_\phi+\lambda} \exp\left(\alpha_\phi |\xi|^{\beta_\phi}\right) + (1+|\xi|)^{\gamma_1} \int_0^{h^{-1}} |\zeta|^{\lambda} (1+|\zeta|)^{\gamma_\phi} \exp\left(\alpha_\phi |\zeta|^{\beta_\phi}\right) d\zeta \right) d\xi
\end{aligned}$$

$$
\begin{aligned}
&\preceq \left(1+h^{-1}\right)^{1-\gamma_\theta} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) \times \\
&\quad \times \left(\left(1+h^{-1}\right)^{\gamma_\phi+\lambda} \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right) + \left(1+h^{-1}\right)^{\gamma_1} \left(1+h^{-1}\right)^{\lambda+\gamma_\phi+1} \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right)\right) \\
&\preceq \left(1+h^{-1}\right)^{1-\gamma_\theta} \left(1+h^{-1}\right)^{\gamma_\phi+\lambda} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right) \left(1+\left(1+h^{-1}\right)^{1+\gamma_1}\right) \\
&\preceq \left(1+h^{-1}\right)^{2-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right).
\end{aligned}
$$

Next,

$$
\begin{aligned}
\Psi^+_{V,\lambda,Z_1}(\xi,h) &\equiv \sup_{z\in\mathbb{R}} \left|\Psi_{V,\lambda,Z_1}(\xi,z,h)\right| \\
&\leq \sup_{z\in\mathbb{R}} \frac{1}{|\theta_1(\xi)|} \int_\xi^{\pm\infty} \left|\exp\left(-\mathbf{i}\zeta z\right)\right| |\zeta|^\lambda \left|\kappa(h\zeta)\right| \left|\phi_V(\zeta)\right| d\zeta \\
&\leq \frac{1}{|\theta_1(\xi)|} \int_\xi^{\pm\infty} |\zeta|^\lambda \left|\kappa(h\zeta)\right| \left|\phi_V(\zeta)\right| d\zeta,
\end{aligned}
$$

so that

$$
\begin{aligned}
\int \Psi^+_{V,\lambda,Z_1}(\xi,h)\, d\xi &\preceq \int [\frac{1}{|\theta_1(\xi)|} 1\left(|\xi| \leq h^{-1}\right) \int_\xi^{h^{-1}} |\zeta|^\lambda \left|\phi_V(\zeta)\right| d\zeta]\; d\xi \\
&\preceq h^{-1}\left(1+h^{-1}\right)^{-\gamma_\theta} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) \left(1+h^{-1}\right)^{\gamma_\phi+\lambda+1} \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right) \\
&\preceq \left(1+h^{-1}\right)^{2-\gamma_\theta+\gamma_\phi+\lambda} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) \exp\left(\alpha_\phi \left(h^{-1}\right)^{\beta_\phi}\right).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\Psi^+_{V,\lambda,V}(\xi,h) &\equiv \sup_{z\in\mathbb{R}} \left|\Psi_{V,\lambda,V}(\xi,z,h)\right| \\
&\leq \sup_{z\in\mathbb{R}} \frac{|\phi_1(\xi)|}{|\theta_1(\xi)|} \left|\exp\left(-\mathbf{i}\xi z\right)\right| |\xi|^\lambda \left|\kappa(h\xi)\right| \\
&= \sup_{z\in\mathbb{R}} \frac{|\phi_1(\xi)|}{|\theta_1(\xi)|} |\xi|^\lambda \left|\kappa(h\xi)\right| \\
&= \frac{|\phi_1(\xi)|}{|\theta_1(\xi)|} |\xi|^\lambda \left|\kappa(h\xi)\right|,
\end{aligned}
$$

so that

$$\begin{aligned}
\int \Psi^{+}_{V,\lambda,V}(\xi,h)\,d\xi &\preceq \int_0^{h^{-1}} \frac{|\phi_1(\xi)|}{|\theta_1(\xi)|}|\xi|^{\lambda}\,d\xi \\
&\preceq h^{-1}\left(1+h^{-1}\right)^{-\gamma_\theta}\exp\left(-\alpha_\theta\left(h^{-1}\right)^{\beta_\theta}\right)\left(1+h^{-1}\right)^{\gamma_\phi+\lambda}\exp\left(\alpha_\phi\left(h^{-1}\right)^{\beta_\phi}\right) \\
&\preceq \left(1+h^{-1}\right)^{1-\gamma_\theta+\gamma_\phi+\lambda}\exp\left(-\alpha_\theta\left(h^{-1}\right)^{\beta_\theta}\right)\exp\left(\alpha_\phi\left(h^{-1}\right)^{\beta_\phi}\right).
\end{aligned}$$

Collecting together these rates delivers the desired result. ■

Lemma A.2 *For a finite integer J, let $\{P_{n,j}(z_2)\}$ define a sequence of nonrandom real-valued continuously differentiable functions of a real variable z_2, $j=1,...,J$. Let A_j and Z_2 be random variables satisfying $E\left[A_j^{2+\delta}|Z_2=z_2\right]\leq C$ for some $C,\delta>0$ for all $z_2\in$* supp(Z), *$j=1,...,J$, such that $\sup_{n\geq N}\sigma_n<\infty$ and $\inf_{n\geq N}\sigma_n>0$ for some $N\in\mathbb{N}^+$, where*

$$\sigma_n \equiv \left(var[\sum_{j=1}^{J} A_j P_{n,j}(Z_2)]\right)^{1/2}.$$

If $\sup_{z_2\in\mathbb{R}}|D_{z_2}P_{n,j}(z_2)| = O\left(n^{(3/2)-\eta}\right)$ for some $\eta>0$, $j=1,...,J$, then

$$\sigma_n^{-1}n^{1/2}\left(\hat{E}\left[\sum_{j=1}^{J}A_jP_{n,j}(Z_2)\right]-E\left[\sum_{j=1}^{J}A_jP_{n,j}(Z_2)\right]\right)\xrightarrow{d}N(0,1).$$

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

Proof of Theorem 4.5. (i) The fact that $E\left[L_{V,\lambda}(z,h)\right]=0$ follows directly from eq.(35). Next, Assumption 4.4(i) ensures the existence and finiteness of

$$\begin{aligned}
E\left[\left(L_{V,\lambda}(z,h)\right)^2\right] &= E\left[\left(\hat{E}\left[\ell_{V,\lambda}(z,h;V,Z_1,Z_2)\right]\right)^2\right] \\
&= n^{-1}E\left[\left(\ell_{V,\lambda}(z,h;V,Z_1,Z_2)\right)^2\right] = n^{-1}\Omega_{V,\lambda}(z,h).
\end{aligned}$$

Specifically, from eq.(35), we have

$$\begin{aligned}
\Omega_{V,\lambda}(z,h) &\equiv E\left[n\left(\bar{g}_{V,\lambda}(z,h)-g_{V,\lambda}(z,h)\right)^2\right] = E\left[\left(\sum_{A=1,Z_1,V}\int \Psi_{V,A}(\xi,z,h)\,n^{1/2}\delta\hat{\theta}_A(\xi)\,d\xi\right)^2\right] \\
&= \sum_{A_1=1,Z_1,V}\sum_{A_2=1,Z_1,V}\int\int \Psi_{V,\lambda,A_1}(\zeta,z,h)\,E\left[n\delta\hat{\theta}_{A_1}(\zeta)\,\delta\hat{\theta}^{\dagger}_{A_2}(\xi)\right]\left(\Psi_{V,\lambda,A_2}(\xi,z,h)\right)^{\dagger}d\zeta d\xi \\
&= \sum_{A_1=1,Z_1,V}\sum_{A_2=1,Z_1,V}\int\int \Psi_{V,\lambda,A_1}(\zeta,z,h)\,V_{A_1A_2}(\zeta,\xi)\left(\Psi_{V,\lambda,A_2}(\xi,z,h)\right)^{\dagger}d\zeta d\xi,
\end{aligned}$$

where

$$\begin{aligned}
V_{A_1A_2}(\zeta,\xi) &\equiv E\left[n\delta\hat{\theta}_{A_1}(\zeta)\,\delta\hat{\theta}^{\dagger}_{A_2}(\xi)\right] = E\left[n\left(\hat{\theta}_{A_1}(\zeta)-\theta_{A_1}(\zeta)\right)\left(\hat{\theta}^{\dagger}_{A_2}(\xi)-\theta^{\dagger}_{A_2}(\xi)\right)\right] \\
&= E\left[\left(A_1e^{\mathbf{i}\zeta Z_2}-\theta_{A_1}(\zeta)\right)\left(A_2e^{-\mathbf{i}\xi Z_2}-\theta^{\dagger}_{A_2}(\xi)\right)\right] \\
&= E\left[A_1e^{\mathbf{i}\zeta Z_2}A_2e^{-\mathbf{i}\xi Z_2}\right]-\theta_{A_1}(\zeta)\,E\left[A_2e^{-\mathbf{i}\xi Z_2}\right]-E\left[A_1e^{\mathbf{i}\zeta Z_2}\right]\theta^{\dagger}_{A_2}(\xi)+\theta_{A_1}(\zeta)\,\theta^{\dagger}_{A_2}(\xi) \\
&= E\left[A_1e^{\mathbf{i}\zeta Z_2}A_2e^{-\mathbf{i}\xi Z_2}\right]-\theta_{A_1}(\zeta)\,\theta^{\dagger}_{A_2}(\xi)-\theta_{A_1}(\zeta)\,\theta^{\dagger}_{A_2}(\xi)+\theta_{A_1}(\zeta)\,\theta^{\dagger}_{A_2}(\xi) \\
&= E\left[A_1A_2e^{\mathbf{i}(\zeta-\xi)Z_2}\right]-\theta_{A_1}(\zeta)\,\theta^{\dagger}_{A_2}(\xi) \\
&= \theta_{(A_1A_2)}(\zeta-\xi)-\theta_{A_1}(\zeta)\,\theta_{A_2}(-\xi).
\end{aligned}$$

By Assumption 4.4(i),

$$\begin{aligned}
\left|V_{A_1A_2}(\zeta,\xi)\right| &= \left|\theta_{(A_1A_2)}(\zeta-\xi)-\theta_{A_1}(\zeta)\,\theta_{A_2}(-\xi)\right| \\
&\leq E\left[\left|A_1A_2\right|\left|e^{\mathbf{i}(\zeta-\xi)Z_2}\right|\right]+E\left[\left|A_1\right|\left|e^{\mathbf{i}\zeta z}\right|\right]E\left[\left|A_2\right|\left|e^{-\mathbf{i}\xi Z_2}\right|\right] \\
&\leq E\left[\left|A_1A_2\right|\right]+E\left[\left|A_1\right|\right]E\left[\left|A_2\right|\right]\preceq 1.
\end{aligned}$$

It follows that

$$\begin{aligned}
\Omega_{V,\lambda}(z,h) &\leq \sum_{A_1=1,Z_1,V}\sum_{A_2=1,Z_1,V}\int\int\left|\Psi_{V,\lambda,A_1}(\zeta,z,h)\right|\left|V_{A_1A_2}(\zeta,\xi)\right|\left|\left(\Psi_{V,\lambda,A_2}(\xi,z,h)\right)^{\dagger}\right|d\zeta d\xi \\
&\preceq \sum_{A_1=1,Z_1,V}\sum_{A_2=1,Z_1,V}\int\int\left|\Psi_{V,\lambda,A_1}(\zeta,z,h)\right|\left|\left(\Psi_{V,\lambda,A_2}(\xi,z,h)\right)\right|d\zeta d\xi \\
&= \left(\sum_{A=1,Z_1,V}\int\left|\Psi_{V,\lambda,A}(\zeta,z,h)\right|d\zeta\right)^2 \leq \left(\sum_{A=1,Z_1,V}\int \Psi^{+}_{V,\lambda,A}(\zeta,h)\,d\zeta\right)^2 = \left(\Psi^{+}_{V,\lambda}(h)\right)^2,
\end{aligned}$$

where

$$\Psi^{+}_{V,\lambda,A}(\zeta,h) = \sup_{z\in\mathbb{R}} \left|\Psi_{V,\lambda,A}(\zeta,z,h)\right| \tag{36}$$

$$\begin{aligned} \Psi^{+}_{V,\lambda}(h) &= \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\zeta,h)\,d\zeta \qquad (37) \\ &= O\left(\left(1+h^{-1}\right)^{2-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)\left(h^{-1}\right)^{\beta_\theta}\right)\right). \end{aligned}$$

The last order of magnitude is shown in Lemma A.1. Hence, we have shown eq.(28).

Next, we turn to uniform convergence. From eq.(35), we have

$$\begin{aligned} \sup_{z\in\mathbb{R}} \left|\bar{g}_{V,\lambda}(z,h)-g_{V,\lambda}(z,h)\right| &= \sup_{z\in\mathbb{R}} \left|\sum_{A=1,Z_1,V} \int \Psi_{V,\lambda,A}(\xi,z,h)\left(\hat{E}\left[Ve^{\mathbf{i}\xi Z_2}\right]-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right)d\xi\right| \\ &\le \sum_{A=1,Z_1,V} \int \left(\sup_{z\in\mathbb{R}} \left|\Psi_{V,\lambda,A}(\xi,z,h)\right|\right)\left|\hat{E}\left[Ve^{\mathbf{i}\xi Z_2}\right]-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right| d\xi \\ &= \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\left|\hat{E}\left[Ve^{\mathbf{i}\xi Z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right]\right| d\xi \end{aligned}$$

where $\Psi^{+}_{V,\lambda,A}(\xi,h)$ is as defined above and where the integrals are finite since $|\hat{E}\{Ve^{\mathbf{i}\xi Z_2}- E\left[Ve^{\mathbf{i}\xi Z_2}\right]\}| \le 1$ and since Lemma A.1 implies that $\sum_{A=1,Z_1,V}\int \Psi^{+}_{V,\lambda,A}(\xi,h)\,d\xi < \infty$.

We then have:

$$\begin{aligned} & E\left[\sup_{z\in\mathbb{R}} \left|\bar{g}_{V,\lambda}(z,h)-g_{V,\lambda}(z,h)\right|\right] \\ \le\ & \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\, E\left[\left(\left|\hat{E}\left[Ve^{\mathbf{i}\xi Z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right]\right|^2\right)^{1/2}\right] d\xi \\ \le\ & \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\left(E\left[\left|\hat{E}\left[Ve^{\mathbf{i}\xi Z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right]\right|^2\right]\right)^{1/2} d\xi \\ =\ & \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\left(n^{-1}E\left[\left|Ve^{\mathbf{i}\xi Z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right|^2\right]\right)^{1/2} d\xi \\ =\ & n^{-1/2} \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\left(E\left[\left|Ve^{\mathbf{i}\xi Z_2}-E\left[Ve^{\mathbf{i}\xi Z_2}\right]\right|^2\right]\right)^{1/2} d\xi \\ \preceq\ & n^{-1/2} \sum_{A=1,Z_1,V} \int \Psi^{+}_{V,\lambda,A}(\xi,h)\,d\xi \\ =\ & n^{-1/2}\Psi^{+}_{V,\lambda}(h), \end{aligned}$$

where $\Psi^{+}_{V,\lambda}(h) = O\left((1+h^{-1})^{2-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)\left(h^{-1}\right)^{\beta_\theta}\right)\right)$, as shown in Lemma A.1. By Markov's inequality it follows that

$$\begin{aligned}
\sup_{z\in\mathbb{R}} |L_{V,\lambda}(z,h)| &= \sup_{z\in\mathbb{R}} |\bar{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h)| \\
&= O_p\left(n^{-1/2}\left(1+h^{-1}\right)^{2-\gamma_\theta+\gamma_\phi+\lambda+\gamma_r} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)\left(h^{-1}\right)^{\beta_\theta}\right)\right).
\end{aligned}$$

(ii) To show asymptotic normality, we apply Lemma A.2 to

$$\ell_{V,\lambda}(z,h_n;V,Z_1,Z_2) = \sum_{A=1,Z_1,V} \int \Psi_{V,\lambda,A}(\xi,z,h_n) \;\; A e^{\mathbf{i}\xi Z_2} \; d\xi$$

with

$$P_{n,A}(z_2) = \int \Psi_{V,\lambda,A}(\xi,z,h_n)\, e^{\mathbf{i}\xi z_2} d\xi,$$

for $A = 1, Z_1, V$, where z is fixed.

Our previous conditions ensure that for some finite $N, \sup_{n>N} \Omega_{V,\lambda}(z,h_n) = \sup_{n>N} \mathrm{var}[\ell_{V,\lambda}(z,h_n;V,Z_1,Z_2)] < \infty$, and we assume $\inf_{n>N} \Omega_{V,\lambda}(z,h_n) > 0$. It remains to verify $\sup_{z\in\mathbb{R}} |D_{z_2}P_{n,A}(z_2)| = O\left(n^{(3/2)-\eta}\right)$. For this, we use Lemma A.1. Specifically,

$$\begin{aligned}
\sup_{z_2\in\mathbb{R}} |D_{z_2}P_{n,A}(z_2)| &= \sup_{z_2\in\mathbb{R}} \left|\int \mathbf{i}\xi \Psi_{V,\lambda,A}(\xi,z,h_n)\, e^{\mathbf{i}\xi z_2} d\xi\right| \\
&\leq \sup_{z_2\in\mathbb{R}} \int |\xi| \left|\Psi_{V,\lambda,A}(\xi,z,h_n)\right| d\xi \\
&= 2\int_0^{h^{-1}} |\xi| \left|\Psi_{V,\lambda,A}(\xi,z,h_n)\right| d\xi \\
&\leq 2\int_0^{h^{-1}} |\xi|\, \Psi^{+}_{V,\lambda}(\xi,h_n)\, d\xi \\
&\preceq \int_0^{h^{-1}} |\xi| \left(1+h_n^{-1}\right)^{2-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)\left(h_n^{-1}\right)^{\beta_\theta}\right) d\xi \\
&\preceq \left(1+h_n^{-1}\right)^{3-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)\left(h_n^{-1}\right)^{\beta_\theta}\right)
\end{aligned}$$

Assumption 4.7 requires that if $\beta_\theta \neq 0$, we have $h_n^{-1} = O\left((\ln n)^{1/\beta_\theta-\eta}\right)$ for some $\eta > 0$, so that

$$\sup_{z_2\in\mathbb{R}} |D_{z_2}P_{n,A}(z_2)| \preceq \left(1+(\ln n)^{1/\beta_\theta-\eta}\right)^{3-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \exp\left(\left(\alpha_\phi 1\left(\beta_\phi=\beta_\theta\right)-\alpha_\theta\right)(\ln n)^{1-\eta\beta_\theta}\right).$$

The right-hand side grows more slowly than any power of n so we certainly have $\sup_{z_2\in\mathbb{R}} |D_{z_2}P_{n,A}(z_2)| = O\left(n^{(3/2)-\eta}\right)$.

If $\beta_\theta = 0$, Assumption 4.7 requires that $h_n^{-1} = O\left(n^{-\eta}n^{(3/2)/\left(3-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1\right)}\right)$ so that

$$\begin{aligned}\sup_{z_2\in\mathbb{R}} |D_{z_2}P_{n,A}(z_2)| &\preceq \left(1+n^{-\eta}n^{(3/2)/\left(3-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1\right)}\right)^{3-\gamma_\theta+\gamma_\phi+\lambda+\gamma_1} \\ &\preceq \left(1+n^{-\eta}n^{3/2}\right) \\ &= O_p\left(n^{(3/2)-\eta}\right).\end{aligned}$$

■

Lemma A.3 *Let A and Z_2 be random variables satisfying $E\left[|A|^2\right]<\infty$ and $E[|A|\,|Z_2|]<\infty$ and let $(A_i,Z_{2,i})_{i=1,\ldots,n}$ be a corresponding IID sample. Then, for any $u,U\geq 0$ and $\epsilon>0$,*

$$\sup_{\zeta\in[-Un^u,Un^u]}\left|\hat{E}\left[A\exp\left(\mathbf{i}\zeta Z_2\right)\right]-E\left[A\exp\left(\mathbf{i}\zeta Z_2\right)\right]\right| = O_p\left(n^{-1/2+\epsilon}\right). \tag{38}$$

Proof. See Lemma 6 in Schennach (2004a). ■

Proof of Theorem 4.6. We substitute expansions (31) and (34) into

$$\begin{aligned}\hat{g}_{V,\lambda}(z,h)-g_{V,\lambda}(z,h) &= \int \exp\left(-\mathbf{i}\xi z\right)\left(-\mathbf{i}\xi\right)^\lambda \kappa(h\xi) \\ &\quad\times\left(\frac{\hat{\theta}_V(\xi)}{\hat{\theta}_1(\xi)}\exp\left(\int_0^\xi \frac{\mathbf{i}\hat{\theta}_{Z_1}(\zeta)}{\hat{\theta}_1(\zeta)}d\zeta\right)-\phi_V(\xi)\right)d\xi\end{aligned}$$

and remove the terms linear in $\delta\hat{\theta}_A(\zeta)$ for $A=1,Z_1,V$. For notational simplicity, we write h instead of h_n here. We then find that $|\hat{g}_{V,\lambda}(z,h)-\bar{g}_{V,\lambda}(z,h)| \preceq \sum_{j=1}^7 R_j$, where

$$\begin{aligned}R_1 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)|\,|\delta_1\hat{q}_V(\xi)|\,|\phi_1(\xi)|\left(\int_0^\xi |\delta_1\hat{q}_{Z_1}(\zeta)|\,d\zeta\right)d\xi \\ R_2 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)|\,|\delta_2\hat{q}_V(\xi)|\,|\phi_1(\xi)|\,d\xi \\ R_3 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)|\,|\delta_2\hat{q}_V(\xi)|\,|\phi_1(\xi)|\left(\int_0^\xi |\delta_1\hat{q}_{Z_1}(\zeta)|\,d\zeta\right)d\xi\end{aligned}$$

$$
\begin{aligned}
R_4 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)| |q_V(\xi)| |\phi_1(\xi)| \left(\int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)| \, d\zeta \right) d\xi \\
R_5 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)| |\delta \hat{q}_V(\xi)| |\phi_1(\xi)| \left(\int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)| \, d\zeta \right) d\xi \\
R_6 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)| |q_V(\xi)| |\phi_1(\xi)| \frac{1}{2} \exp\left(\left| \delta \bar{Q}_{Z_1}(\xi) \right| \right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta \right)^2 d\xi \\
R_7 &= \int_0^\infty |\xi|^\lambda |\kappa(h\xi)| |\delta \hat{q}_V(\xi)| |\phi_1(\xi)| \frac{1}{2} \exp\left(\left| \delta \bar{Q}_{Z_1}(\xi) \right| \right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta \right)^2 d\xi.
\end{aligned}
$$

These terms can then be bounded in terms of $\Psi^+_{V,\lambda}(h)$, defined in eq.(37), and

$$
\begin{aligned}
\Upsilon(h) &\equiv \left(1 + h^{-1}\right) \left(\sup_{\xi \in [-h^{-1}, h^{-1}]} \frac{|\phi_1'(\xi)|}{|\phi_1(\xi)|} \right) \left(\sup_{\xi \in [-h^{-1}, h^{-1}]} |\theta_1(\xi)|^{-1} \right) \\
&= O\left(\left(1 + h^{-1}\right)^{1+\gamma_1-\gamma_\theta} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta} \right) \right) \\
\hat{\Phi}_n &\equiv \max_{A=1,Z_1,V} \sup_{\zeta \in \left[-h_n^{-1}, h_n^{-1}\right]} \left| \hat{\theta}_A(\zeta) - \theta_A(\zeta) \right| = O_p\left(n^{-1/2+\epsilon}\right) \text{ for any } \epsilon > 0.
\end{aligned}
$$

The latter order of magnitude follows from Lemma A.3, given Assumptions 4.7 and 4.8. Also, we note that

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

Now, we have

$$
\begin{aligned}
R_1 &\leq \int_0^\infty |\kappa(h\xi)| \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|^2} \right) \hat{\Phi}_n |\phi_0(\xi)| \left(\int_0^\xi |\delta_1 \hat{q}_{Z_1}(\zeta)| \, d\zeta \right) d\xi \\
&\preceq \Upsilon(h) \hat{\Phi}_n \int_0^\infty |\kappa(h\xi)| \left(1 + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|} \right) |\phi_0(\xi)| \left(\int_0^\xi |\delta_1 \hat{q}_{Z_1}(\zeta)| \, d\zeta \right) d\xi \\
&= \Upsilon(h) \hat{\Phi}_n \int_0^\infty \{ \int_\zeta^\infty |\kappa(h\xi)| \left(1 + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|} \right) |\phi_0(\xi)| \, d\xi \} |\delta_1 \hat{q}_{Z_1}(\zeta)| \, d\zeta \\
&= \Upsilon(h) \hat{\Phi}_n \int_0^\infty \{ \int_\zeta^\infty |\kappa(h\xi)| \left(|\phi_0(\xi)| + |\phi_V(\xi)| \right) d\xi \} |\delta_1 \hat{q}_{Z_1}(\zeta)| \, d\zeta
\end{aligned}
$$

$$
\begin{aligned}
&\leq \Upsilon(h)\,\hat{\Phi}_n^2 \int_0^\infty \{\int_\zeta^\infty |\kappa(h\xi)|\,(|\phi_0(\xi)| + |\phi_V(\xi)|)\,d\xi\} \left(1 + \frac{|\theta_{Z_1}(\zeta)|}{|\theta_1(\zeta)|}\right) \frac{1}{|\theta_1(\zeta)|} d\zeta \\
&\preceq \Upsilon(h)\,\hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h) \\
&= O_p\left(\left(1+h^{-1}\right)^{1+\gamma_1-\gamma_\theta} \exp\left(-\alpha_\theta \left(h^{-1}\right)^{\beta_\theta}\right) n^{-1+2\epsilon} \left(h^{-1}\right)^{\gamma_L} \exp\left(\alpha_L \left(h^{-1}\right)^{\beta_L}\right)\right),
\end{aligned}
$$

as required for part (i). Below, we show that the remaining terms are similarly behaved.

For part (ii), we note that

$$
\Upsilon(h)\,\hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h) = \left(\Upsilon(h)\,\hat{\Phi}_n^2 n^{1/2}\right) n^{-1/2} \Psi_{V,\lambda}^+(h).
$$

As Lemma A.1 implies that $n^{-1/2}\Psi_{V,\lambda}^+(h)$ is $O_p\left(n^{-1/2}\left(h^{-1}\right)^{\gamma_L} \exp\left(\alpha_L \left(h^{-1}\right)^{\beta_L}\right)\right)$, we only need to show that $\left(\Upsilon(h_n)\,\hat{\Phi}_n^2 n^{1/2}\right) = o_p(1)$.

If $\beta_\theta \neq 0$, the assumptions of the Theorem ensure that $h_n^{-1} \preceq (\ln n)^{(1/\beta_\theta)-\eta}$, so that

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

where the last line follows since $\ln n$ dominates $(\ln n)^{1-\eta\beta}$ and $\ln\ln n$ and since $-1/2+2\epsilon < 0$.

If $\beta_\theta = 0$, the assumptions of the Theorem ensure that $h_n^{-1} \preceq n^{(1+\gamma_1-\gamma_\theta)^{-1}/2-\eta}$, and

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

selecting $\epsilon < \eta/2$.

The remaining terms can be similarly bounded, as they all have the same $\Upsilon(h)\,\hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h)$ leading term:

$$
\begin{aligned}
R_2 &\leq \int_0^\infty |\kappa(h\xi)| \left| \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|^2} \frac{1}{|\theta_1(\xi)|} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} + \frac{1}{|\theta_1(\xi)|^2} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \right| |\phi_0(\xi)|\, d\xi \\
&\preceq \Upsilon(h)\,\hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \int_0^\infty |\kappa(h\xi)| \frac{1}{|\theta_1(\xi)|} \left| \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|} + 1 \right| |\phi_0(\xi)|\, d\xi \\
&\preceq \Upsilon(h)\,\hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \left(\int_0^\infty \frac{|\kappa(h\xi)|\,|\phi_V(\xi)|}{|\theta_1(\xi)|} d\xi + \int_0^\infty \frac{|\kappa(h\xi)|\,|\phi_0(\xi)|}{|\theta_1(\xi)|} d\xi \right) \\
&= \Upsilon(h)\,\hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h)\,(1 + o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
R_3 &\preceq \Upsilon(h)\,\hat{\Phi}_n \int_0^\infty |\kappa(h\xi)|\,|\delta_2 \hat{q}_V(\xi)|\,|\phi_0(\xi)|\, d\xi \\
&= \Upsilon(h)\,\hat{\Phi}_n R_2 = o_p(1)\, R_2;
\end{aligned}
$$

$$
\begin{aligned}
R_4 &= \int_0^\infty |\kappa(h\xi)|\,|\phi_V(\xi)|\, \{ \int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)|\, d\zeta \} d\xi \\
&\preceq \Upsilon(h)\,\hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \int_0^\infty \frac{\int_\zeta^\infty |\kappa(h\xi)|\,|\phi_V(\xi)|\, d\xi}{|\theta_1(\zeta)|} d\zeta \\
&= \Upsilon(h)\,\hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h)\,(1 + o_p(1));
\end{aligned}
$$

$$
\begin{aligned}
R_5 &\leq \int_0^\infty |\kappa(h\xi)| \left(\frac{1}{|\theta_1(\xi)|} + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|^2} \right) \hat{\Phi}_n |1 + o_p(1)|^{-1} |\phi_0(\xi)|\, \{ \int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)|\, d\zeta \} d\xi \\
&= \Upsilon(h)\,\hat{\Phi}_n |1 + o_p(1)|^{-1} \int_0^{h^{-1}} |\kappa(h\xi)| \left(1 + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|} \right) |\phi_0(\xi)|\, \{ \int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)|\, d\zeta \} d\xi \\
&= \Upsilon(h)\,\hat{\Phi}_n |1 + o_p(1)|^{-1} \\
&\quad \times [\int_0^\infty |\kappa(h\xi)|\,|\phi_0(\xi)|\, \{ \int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)|\, d\zeta \} d\xi \\
&\quad + \int_0^\infty |\kappa(h\xi)|\,|\phi_V(\xi)|\, \{ \int_0^\xi |\delta_2 \hat{q}_{Z_1}(\zeta)|\, d\zeta \} d\xi] \\
&= \Upsilon(h)\,\hat{\Phi}_n (1 + o_p(1))\, R_4 = o_p(1)\, R_4;
\end{aligned}
$$

$$
\begin{aligned}
R_6 &\leq \int_0^\infty |\kappa(h\xi)| \, |\phi_V(\xi)| \frac{1}{2} \exp\left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right)^2 d\xi \\
&\leq \frac{1}{2} \exp(o_p(1)) \int_0^\infty |\kappa(h\xi)| \, |\phi_V(\xi)| \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) d\xi \\
&\preceq \frac{1}{2} \exp(o_p(1)) \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \\
&\quad \times \int_0^\infty |\kappa(h\xi)| \, |\phi_V(\xi)| \left(\int_0^\xi \left(\frac{1}{|\theta_1(\zeta)|} + \frac{|\theta_{Z_1}(\zeta)|}{|\theta_1(\zeta)|^2}\right) d\zeta\right) d\xi \\
&= \frac{1}{2} \exp(o_p(1)) \Upsilon(h) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \\
&\quad \times \int_0^\infty \{\int_\zeta^\infty |\kappa(h\xi)| \, |\phi_V(\xi)| \, d\xi\} \left(\frac{1}{|\theta_1(\zeta)|} + \frac{|\theta_{Z_1}(\zeta)|}{|\theta_1(\zeta)|^2}\right) d\zeta \\
&= O_p(1) \Upsilon(h) \hat{\Phi}_n^2 \Psi_{V,\lambda}^+(h);
\end{aligned}
$$

$$
\begin{aligned}
R_7 &\leq \int_0^\infty |\kappa(h\xi)| \left(1 + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|}\right) \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} |\phi_0(\xi)| \\
&\quad \times \frac{1}{2} \exp\left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right)^2 d\xi \\
&\preceq \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \\
&\quad \times \int_0^\infty |\kappa(h\xi)| \left(1 + \frac{|\theta_V(\xi)|}{|\theta_1(\xi)|}\right) |\phi_0(\xi)| \\
&\quad \times \exp\left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right)^2 d\xi \\
&\preceq \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \\
&\quad \times \int_0^\infty |\kappa(h\xi)| \, |\phi_0(\xi)| \exp\left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right)^2 d\xi \\
&\quad + A \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} \\
&\quad \times \int_0^\infty |\kappa(h\xi)| \, |\phi_V(\xi)| \exp\left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right) \left(\int_0^\xi |\delta \hat{q}_{Z_1}(\zeta)| \, d\zeta\right)^2 d\xi \\
&\preceq \Upsilon(h) \hat{\Phi}_n |1 + o_p(1)|^{-1} R_6 \\
&= o_p(1) R_6.
\end{aligned}
$$

■

Proof of Theorem 4.9. The given assumptions clearly ensure that

$$\begin{aligned}
& \max_{j=1,\ldots J}\sup_{z\in\mathbb{R}}\left|\hat{g}_{V_j,\lambda_j}(z,h_n)-g_{V_j,\lambda_j}(z)\right| \\
= & \max_{j=1,\ldots J}\sup_{z\in\mathbb{R}}\left|B_{V_j,\lambda_j}(z,h_n)+L_{V_j,\lambda_j}(z,h_n)+R_{V_j,\lambda_j}(z,h_n)\right| \\
= & o_p\left(n^{-1/2}\right)+o_p\left(n^{-1/4}\right)=o_p\left(n^{-1/4}\right),
\end{aligned}$$

so that when $\tilde{g}_{V_j,\lambda_j}(z)=\hat{g}_{V_j,\lambda_j}(z,h_n)$, the remainder term in eq.(29) is $o_p\left(\left(n^{-1/4}\right)^2\right)=o_p\left(n^{-1/2}\right)$. Moreover, we have

$$\begin{aligned}
& \sum_{j=1}^{J}\int\left(\hat{g}_{V_j,\lambda_j}(z,h)-g_{V_j,\lambda_j}(z)\right)s_j(z)\,dz \\
= & \sum_{j=1}^{J}\int L_{V_j,\lambda_j}(z,h)\,s_j(z)\,dz+\sum_{j=1}^{J}\int\left(B_{V_j,\lambda_j}(z,h)+R_{V_j,\lambda_j}(z,h)\right)s_j(z)\,dz,
\end{aligned}$$

where

$$\begin{aligned}
& \left|\sum_{j=1}^{J}\int\left(B_{V_j,\lambda_j}(z,h_n)+R_{V_j,\lambda_j}(z,h_n)\right)s_j(z)\,dz\right| \\
\leq & \left(\max_{j=1,\ldots,J}\sup_{z\in\mathbb{R}}\left|B_{V_j,\lambda_j}(z,h_n)+R_{V_j,\lambda_j}(z,h_n)\right|\right)\sum_{j=1}^{J}\int\left|s_j(z)\right|dz \\
= & o_p\left(n^{-1/2}\right),
\end{aligned}$$

since $\int\left|s_j(z)\right|dz<\infty$ and $\max_{j=1,\ldots J}\sup_{z\in\mathbb{R}}\max\left\{\left|B_{V_j,\lambda_j}(z,h_n)\right|,\left|R_{V_j,\lambda_j}(z,h_n)\right|\right\}=o_p\left(n^{-1/2}\right)$ by assumption. It follows that

$$b\left(\hat{g}(\cdot,h_n)\right)-b(g)=\sum_{j=1}^{J}\int L_{V_j,\lambda_j}(z,h_n)\,s_j(z)\,dz+o_p\left(n^{-1/2}\right).$$

Next, we note that

$$\begin{aligned}
& \int L_{V_j,\lambda_j}(z,h_n)\,s_j(z)\,dz \\
= & \lim_{\tilde{h}\to 0}\int L_{V_j,\lambda_j}\left(z,\tilde{h}\right)s_j(z)\,dz+\lim_{\tilde{h}\to 0}\int\left(L_{V_j,\lambda_j}(z,h_n)-L_{V_j,\lambda_j}\left(z,\tilde{h}\right)\right)s_j(z)\,dz,
\end{aligned} \tag{39}$$

where the first term will be shown to be a standard sample average while the second will shown to be asymptotically negligible.

By the definition of $L_{V_j,\lambda_j}\left(z,\tilde{h}\right)$ (see Lemma 4.3), we have

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\int L_{V_j,\lambda_j}\left(z,\tilde{h}\right)s_j\left(z\right)dz \\
=\;&\lim_{\tilde{h}\to 0}\sum_{A=1,Z_1,V_j}\int\{\int \Psi_{V_j,\lambda_j,A}\left(\xi,z,\tilde{h}\right)\left(\hat{E}\left[Ae^{\mathbf{i}\xi Z_2}\right]-E\left[Ae^{\mathbf{i}\xi Z_2}\right]\right)d\xi\}s_j\left(z\right)dz.
\end{aligned}$$

Under the assumption that $\int\bar{\Psi}_{s_j,V_j,\lambda_j}\left(\xi\right)d\xi<\infty$, the integrand is absolutely integrable (for any given sample), thus enabling us to interchange integrals as well as limits in the sequel:

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\int L_{V_j,\lambda_j}\left(z,\tilde{h}\right)s_j\left(z\right)dz \\
=\;&\lim_{\tilde{h}\to 0}\sum_{A=1,Z_1,V_j}\int\left(\int \Psi_{V_j,\lambda_j,A}\left(\xi,z,\tilde{h}\right)s_j\left(z\right)dz\right)\left(\hat{E}\left[Ae^{\mathbf{i}\xi Z_2}\right]-E\left[Ae^{\mathbf{i}\xi Z_2}\right]\right)d\xi.
\end{aligned}$$

The innermost integrals can be calculated explicitly:

$$\begin{aligned}
&\lim_{\tilde{h}\to 0}\int \Psi_{V,\lambda,1}\left(\xi,z,\tilde{h}\right)s\left(z\right)dz \\
=\;&-\frac{1}{2\pi}\frac{\phi_V\left(\xi\right)}{\theta_1\left(\xi\right)}\left(\int\exp\left(-\mathbf{i}\xi z\right)s\left(z\right)dz\right)\left(-\mathbf{i}\xi\right)^{\lambda}\lim_{\tilde{h}\to 0}\kappa\left(\tilde{h}\xi\right) \\
&-\frac{1}{2\pi}\frac{\mathbf{i}\theta_{Z_1}\left(\xi\right)}{\left(\theta_1\left(\xi\right)\right)^2}\int_{\xi}^{\pm\infty}\left(\int\exp\left(-\mathbf{i}\xi z\right)s\left(z\right)dz\right)\left(-\mathbf{i}\zeta\right)^{\lambda}\lim_{\tilde{h}\to 0}\kappa\left(\tilde{h}\zeta\right)\phi_V\left(\zeta\right)d\zeta \\
=\;&-\frac{1}{2\pi}\frac{\phi_V\left(\xi\right)}{\theta_1\left(\xi\right)}\sigma_s^{\dagger}\left(\xi\right)\left(-\mathbf{i}\xi\right)^{\lambda} \\
&-\frac{1}{2\pi}\frac{\mathbf{i}\theta_{Z_1}\left(\xi\right)}{\left(\theta_1\left(\xi\right)\right)^2}\int_{\xi}^{\pm\infty}\sigma_s^{\dagger}\left(\zeta\right)\left(-\mathbf{i}\zeta\right)^{\lambda}\phi_V\left(\zeta\right)d\zeta \\
\equiv\;&\Psi_{s,V,\lambda,1}\left(\xi\right),
\end{aligned}$$

where $\Psi_{s,V,\lambda,1}\left(\xi\right)$ is defined in the statement of the theorem. Similarly,

$$\begin{aligned}
\lim_{\tilde{h}\to 0}\int \Psi_{V,\lambda,Z_1}\left(\xi,z,\tilde{h}\right)s\left(z\right)dz &= \frac{1}{2\pi}\frac{\mathbf{i}}{\theta_1\left(\xi\right)}\int_{\xi}^{\pm\infty}\sigma_s^{\dagger}\left(\zeta\right)\left(-\mathbf{i}\zeta\right)^{\lambda}\phi_V\left(\zeta\right)d\zeta\equiv\Psi_{s,V,\lambda,Z_1}\left(\xi\right) \\
\lim_{\tilde{h}\to 0}\int \Psi_{V,\lambda,V}\left(\xi,z,\tilde{h}\right)s\left(z\right)dz &= \frac{1}{2\pi}\frac{\phi_1\left(\xi\right)}{\theta_1\left(\xi\right)}\sigma_s^{\dagger}\left(\xi\right)\left(-\mathbf{i}\xi\right)^{\lambda}\equiv\Psi_{s,V,\lambda,V}\left(\xi\right).
\end{aligned}$$

It follows that

$$\lim_{\tilde{h}\to 0}\int L_{V_j,\lambda_j}\left(z,\tilde{h}\right)s_j\left(z\right)dz = \sum_{A=1,Z_1,V_j}\int \Psi_{s_j,V_j,\lambda_j,A}\left(\xi\right)\left(\hat{E}\left[Ae^{\mathbf{i}\xi Z_2}\right]-E\left[Ae^{\mathbf{i}\xi Z_2}\right]\right)d\xi$$
$$= \hat{E}\left[\psi_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2\right)\right],$$

as defined in the statement of the theorem. The assumption that $\int \bar{\Psi}_{s_j,V_j,\lambda_j}\left(\xi\right)d\xi < \infty$ also implies that

$$\left|\psi_{s_j,V_j,\lambda_j}\left(v,z_1,z_2\right)\right| \leq C\max\left\{1,\left|v\right|,\left|z_1\right|\right\}\int \bar{\Psi}_{s_j,V_j,\lambda_j}\left(\xi\right)d\xi$$

for some $C<\infty$. Since $E\left[V_j^2\right]<\infty$ and $E\left[Z_1^2\right]<\infty$ by assumption, $E[|\psi_{s_j,V_j,\lambda_j}(V_j, Z_1,Z_2)|^2]<\infty$, and it follows by the Lindeberg-Levy central limit theorem that $\hat{E}\left[\psi_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2\right)\right]$ is root$-n$ consistent and asymptotically normal.

The second term of eq.(39) can be shown to be $o_p\left(n^{-1/2}\right)$ by noting that it can be written as an h_n-dependent sample average $\hat{E}\left[\tilde{\psi}_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2,h_n\right)\right]$, where $\tilde{\psi}_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2,h\right)$ is such that $\lim_{h\to 0}E\left[\left|\tilde{\psi}_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2,h\right)\right|^2\right]=0$. The manipulations are similar to the treatment of $\hat{E}\left[\psi_{s_j,V_j,\lambda_j}\left(V_j,Z_1,Z_2\right)\right]$ above, replacing $\kappa\left(\tilde{h}\zeta\right)$ by $\left(\kappa\left(h_n\zeta\right)-\kappa\left(\tilde{h}\zeta\right)\right)$ and taking the limit as $\tilde{h}\to 0$ and $h_n\to 0$. ■

Proof of Theorem 4.10. Consider a Taylor expansion of $\hat{\beta}\left(z,h\right)-\beta\left(z\right)$ in $\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)$ to first order:

$$\hat{\beta}\left(z,h\right)-\beta\left(z\right)$$
$$=\sum_{A=X,Y}\sum_{V=1,A}\sum_{\lambda=0,1}s_{A,V,\lambda}\left(z\right)\left(\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)\right)+R_{A,V,\lambda}\left(\bar{g}_{V,\lambda}\left(z,h\right),\left(\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)\right)\right), \tag{40}$$

where the $s_{A,V,\lambda}\left(z\right)$ are given in the statement of Theorem 3.5 and where $R_{A,V,\lambda}[\bar{g}_{V,\lambda}\left(z,h\right), (\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right))]$ is a remainder term in which for every (z,h), $\bar{g}_{V,\lambda}\left(z,h\right)$ lies between $\hat{g}_{V,\lambda}\left(z,h\right)$ and $g_{V,\lambda}\left(z\right)$. (We similarly use an overbar $^-$ to denote any function of $g_{V,\lambda}\left(z\right)$ in which $g_{V,\lambda}\left(z\right)$ has been replaced by $\bar{g}_{V,\lambda}\left(z,h\right)$.)

We first note that, by Corollary 4.7,

$$\max_{V=1,X,Y}\max_{\lambda=0,1}\sup_{z\in\mathbb{R}}\left|\hat{g}_{V,\lambda}\left(z,h_n\right)-g_{V,\lambda}\left(z\right)\right|=O_p\left(\varepsilon_n\right),$$

where $\varepsilon_n\equiv\left(h_n^{-1}\right)^{\gamma_{1,B}}\exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)+n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{1,L}}\exp\left(\alpha_L\left(h_n^{-1}\right)^{\beta_L}\right)\rightarrow 0$.

The first terms in the summation in eq.(40) can be shown to be $O_p\left(\varepsilon_n/\tau^4\right)$ uniformly for $z\in\mathbf{Z}_\tau$ as follows. Each $s_{A,V,\lambda}\left(z\right)$ term consists of products of functions of the form $g_{V,\lambda}\left(z\right)$ (which are uniformly bounded over $\mathbb{R}$ by assumption) divided by products of at most 4 functions of the form $g_{1,0}\left(z\right)$ or $D_z\mu_X(z)$, which are by construction bounded below by τ uniformly for $z\in\mathbf{Z}_\tau$. It follows that $\sup_{z\in\mathbf{Z}_\tau}\left|s_{A,V,\lambda}\left(z\right)\left(\hat{g}_{V,\lambda}\left(z,h_n\right)-g_{V,\lambda}\left(z\right)\right)\right|=O\left(1\right)O_p\left(\tau^{-4}\right)O_p\left(\varepsilon_n\right)=O_p\left(\varepsilon_n/\tau^4\right)$.

The remainder terms in eq.(40) can be shown to be $o_p\left(\varepsilon_n/\tau^4\right)$ uniformly for $z\in\mathbf{Z}_\tau$ as follows. Without deriving their explicit form, it is clear that these involve a finite sum of (i) finite products of the functions $\bar{g}_{V,\lambda}\left(z,h\right)$ for $V=1,X,Y$ and $\lambda=0,1$; (ii) division by a product of at most 5 functions of the form $\bar{g}_{1,0}\left(z,h\right)$ or $D_z\bar{\mu}_X(z)$; and (iii) pairwise products of functions of the form $\left(\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)\right)$. The contribution of (i) is bounded in probability uniformly for $z\in\mathbb{R}$ since

$$\begin{aligned}\left|\bar{g}_{V,\lambda}\left(z,h\right)\right| &\leq \left|g_{V,\lambda}\left(z\right)\right|+\left|\bar{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)\right|\\ &\leq \left|g_{V,\lambda}\left(z\right)\right|+\left|\hat{g}_{V,\lambda}\left(z,h\right)-g_{V,\lambda}\left(z\right)\right|\end{aligned}$$

where $\left|g_{V,\lambda}\left(z\right)\right|$ is uniformly bounded over $\mathbb{R}$ by assumption and $\sup_{z\in\mathbb{R}}\left|\hat{g}_{V,\lambda}\left(z,h_n\right)-g_{V,\lambda}\left(z\right)\right|\leq O_p\left(\varepsilon_n\right)=o_p(1)$. The contribution of (ii) is bounded by noting that for $z\in\mathbf{Z}_\tau$

$$\begin{aligned}\bar{g}_{1,0}\left(z,h_n\right) &= g_{1,0}\left(z\right)\left(1+\frac{\bar{g}_{1,0}\left(z,h_n\right)-g_{1,0}\left(z\right)}{g_{1,0}\left(z\right)}\right)\\ &= f_Z\left(z\right)\left(1+\frac{\bar{g}_{1,0}\left(z,h_n\right)-g_{1,0}\left(z\right)}{f_Z\left(z\right)}\right)\\ &= f_Z\left(z\right)\left(1+O_p\left(\frac{\varepsilon_n}{\tau}\right)\right).\end{aligned}$$

Now choose $\{\tau_n\}$ such that $\tau_n > 0$, $\tau_n \to 0$ as $n \to \infty$, and $\varepsilon_n/\tau_n^4 \to 0$. It follows that $\varepsilon_n/\tau_n \to 0$ as well. Hence for $z \in \mathbf{Z}_{\tau_n}$ we have

$$\bar{g}_{1,0}\left(z, h_n\right) = f_Z\left(z\right)\left(1 + o_p\left(1\right)\right).$$

Since $f_Z\left(z\right) \geq \tau_n$ for $z \in \mathbf{Z}_{\tau_n}$ by construction, we also have $f_Z\left(z\right)\left(1 + o_p\left(1\right)\right) \geq \tau_n/2$ with probability approaching one (w.p.a. 1). Similar reasoning holds for $D_z\bar{\mu}_X(z)$. Hence, the denominator is bounded below by $\left(\tau_n/2\right)^5$ w.p.a. 1, where the power 5 arises from the presence of up to 5 of such terms. Finally, the contribution of (iii) is simply $O_p\left(\varepsilon_n^2\right)$. Collecting all three orders of magnitudes, we obtain

$$O_p\left(1\right) O_p\left(\tau_n^{-5}\right) O_p\left(\varepsilon_n^2\right) = O_p\left(\frac{\varepsilon_n^2}{\tau_n^5}\right) = O_p\left(\frac{\varepsilon_n}{\tau_n^4}\right) O_p\left(\frac{\varepsilon_n}{\tau_n}\right) = O_p\left(\frac{\varepsilon_n}{\tau_n^4}\right) o_p\left(1\right) = o_p\left(\frac{\varepsilon_n}{\tau_n^4}\right),$$

so that

$$\sup_{z \in \mathbf{Z}_{\tau_n}} \left|\hat{\beta}\left(z, h_n\right) - \beta\left(z\right)\right| = o_p\left(\frac{\varepsilon_n}{\tau_n^4}\right) = o_p(1).$$

■

Proof of Theorem 4.11. The delta method applies directly to show that the asymptotic normality of $\hat{g}_{V,\lambda}\left(z, h_n\right) - g_{V,\lambda}\left(z\right)$ provided by Corollary 4.8 carries over to $\hat{\beta}\left(z, h_n\right) - \beta\left(z\right)$, as a first-order Taylor expansion of $\hat{\beta}\left(z, h_n\right) - \beta\left(z\right)$ in $\hat{g}_{V,\lambda}\left(z, h_n\right) - g_{V,\lambda}\left(z\right)$ yields

$$\hat{\beta}\left(z, h_n\right) - \beta\left(z\right) = \sum_{A=X,Y} \sum_{V=1,A} \sum_{\lambda=0,1} s_{A,V,\lambda}\left(z\right)\left(\hat{g}_{V,\lambda}\left(z, h_n\right) - g_{V,\lambda}\left(z\right)\right) + R_n,$$

where the $s_{A,V,\lambda}\left(z\right)$ terms are as defined in Theorem 3.5 and where the remainder term R_n is necessarily negligible since, under the assumptions that $\max_{V=1,X,Y} \max_{\lambda=0,1} \left|g_{V,\lambda}\left(z\right)\right| < \infty$, $f_Z\left(z\right) > 0$ and $\left|D_z\mu_X(z)\right| > 0$, the first derivative terms $s_{A,V,\lambda}\left(z\right)$ are continuous. ■

B Supplementary material

Proof of Lemma 3.1. This result holds by construction. ■

Lemma B.1 *Suppose Assumption 3.4 holds. Then* $\sup_{z\in\mathbb{R}}\left|k^{(\lambda)}(z)\right|<\infty$, $\int\left|k^{(\lambda)}(z)\right|dz<\infty$, $0<\int\left|k^{(\lambda)}(z)\right|^2dz<\infty$, $\int\left|k^{(\lambda)}(z)\right|^{2+\delta}dz<\infty$, *and* $|z|\left|k^{(\lambda)}(z)\right|\to 0$ *as* $|z|\to\infty$.

Proof. The Fourier transform of $k^{(\lambda)}(z)$ is $(-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)$, which is bounded by assumption and therefore absolutely integrable, given the assumed compact support of $\kappa(\zeta)$. Hence $k^{(\lambda)}(z)$ is bounded, since $\left|k^{(\lambda)}(z)\right|=\left|\int(-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)e^{-i\zeta z}d\zeta\right|\leq\int|\zeta|^{\lambda}|\kappa(\zeta)|d\zeta<\infty$. Note that $\int\left|k^{(\lambda)}(z)\right|^2dz>0$ unless $k^{(\lambda)}(z)=0$ for all $z\in\mathbb{R}$, which would imply that $k(z)$ is a polynomial, making it impossible to satisfy $\int k(z)dz=1$. Hence, $\int\left|k^{(\lambda)}(z)\right|^2dz>0$.

The Fourier transform of $z^2k^{(\lambda)}(z)$ is $-(d^2/d\zeta^2)\left((-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)\right)$. By the compact support of $\kappa(\zeta)$, if $\kappa(\zeta)$ has two bounded derivatives then so does $(-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)$, and it follows that $-(d^2/d\zeta^2)\left((-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)\right)$ is absolutely integrable. By the Riemann-Lebesque Lemma, the inverse Fourier transform of $\mathbf{i}(d^2/d\zeta^2)\left((-\mathbf{i}\zeta)^{\lambda}\kappa(\zeta)\right)$ is such that $z^2k^{(\lambda)}(z)\to 0$ as $|z|\to\infty$. Hence, we know that there exists C such that

$$\left|k^{(\lambda)}(z)\right|\leq\frac{C}{1+z^2},$$

and the function on the right-hand side satisfies all the remaining properties stated in the lemma. ■

Proof of Theorem 3.2. (i) The order of magnitude of the bias is derived in the proof of Theorem 4.4 in the foregoing appendix. The convergence rate of $B_{V,\lambda}(z,h)$ is also derived in Theorem 4.4.

(ii) The facts that $E\left[L_{V,\lambda}(z,h)\right]=0$ and $E\left[L_{V,\lambda}^2(z,h)\right]=n^{-1}\Omega_{V,\lambda}(z,h)$ hold by construction. Next, Assumptions 3.2(ii) and 3.4 ensure that

$$\begin{aligned}\Omega_{V,\lambda}(z,h) &= E\left[\left((-1)^{\lambda}h^{-\lambda-1}Vk^{(\lambda)}\left(\frac{Z-z}{h}\right)\right)^2\right]-\left(E\left[(-1)^{\lambda}h^{-\lambda-1}Vk^{(\lambda)}\left(\frac{Z-z}{h}\right)\right]\right)^2\\ &\leq E\left[\left((-1)^{\lambda}h^{-\lambda-1}Vk^{(\lambda)}\left(\frac{Z-z}{h}\right)\right)^2\right]\end{aligned}$$

$$
\begin{aligned}
&= h^{-2\lambda-1} E\left[E\left[V^2|Z\right] h^{-1}\left(k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right)^2\right] \\
&\preceq h^{-2\lambda-1} E\left[h^{-1}\left(k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right)^2\right] \\
&\qquad \text{(by Assumption 3.2(}ii\text{) and Jensen's inequality)} \\
&= h^{-2\lambda-1} \int h^{-1}\left(k^{(\lambda)}\left(\frac{\tilde{z}-z}{h}\right)\right)^2 f_Z(\tilde{z})\, d\tilde{z} \\
&= h^{-2\lambda-1} \int \left(k^{(\lambda)}(u)\right)^2 f_Z(z+hu)\, du \\
&\qquad \text{(after a change of variable from } \tilde{z} \text{ to } z+hu\text{)} \\
&\preceq h^{-2\lambda-1} \int \left(k^{(\lambda)}(u)\right)^2 du \qquad \text{(by Assumption 3.1(}i\text{)} \\
&\preceq h^{-2\lambda-1} \qquad \text{(by Lemma B.1)}
\end{aligned}
$$

and hence

$$
\sqrt{\sup_{z\in\mathbb{R}} \Omega_{V,\lambda}(z,h)} = O\left(h^{-\lambda-1/2}\right).
$$

We now establish the uniform convergence rate. Using Parseval's identity, we have

$$
\begin{aligned}
L_{V,\lambda}(z,h) &= \hat{E}\left[(-1)^\lambda h^{-\lambda-1} V k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right] - E\left[(-1)^\lambda h^{-\lambda-1} V k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right] \\
&= \frac{1}{2\pi}\int \left(\hat{E}\left[Ve^{\mathbf{i}\zeta Z}\right] - E\left[Ve^{\mathbf{i}\zeta Z}\right]\right)(-\mathbf{i}\zeta)^\lambda \kappa(h\zeta)\, e^{-\mathbf{i}\zeta z} d\zeta,
\end{aligned}
$$

so it follows that

$$
\left|L_{V,\lambda}(z,h)\right| \leq \frac{1}{2\pi}\int \left|\hat{E}\left[Ve^{\mathbf{i}\zeta Z}\right] - E\left[Ve^{\mathbf{i}\zeta Z}\right]\right| |\zeta|^\lambda \left|\kappa(h\zeta)\right| d\zeta,
$$

and that

$$
\begin{aligned}
E\left[\left|L_{V,\lambda}(z,h)\right|\right] &= \frac{1}{2\pi}\int E\left[\left|\hat{E}\left[Ve^{\mathbf{i}\zeta Z}\right] - E\left[Ve^{\mathbf{i}\zeta Z}\right]\right|\right] |\zeta|^\lambda \left|\kappa(h\zeta)\right| d\zeta \\
&\leq \frac{1}{2\pi}\int (E[\left(\hat{E}\left[Ve^{\mathbf{i}\zeta Z}\right] - E\left[Ve^{\mathbf{i}\zeta Z}\right]\right) \\
&\qquad \times \left(\hat{E}\left[Ve^{\mathbf{i}\zeta Z}\right] - E\left[Ve^{\mathbf{i}\zeta Z}\right]\right)^\dagger])^{1/2} |\zeta|^\lambda \left|\kappa(h\zeta)\right| d\zeta \\
&\leq \frac{1}{2\pi}\int \left(n^{-1} E\left[Ve^{\mathbf{i}\zeta Z} V e^{-\mathbf{i}\zeta Z}\right]\right)^{1/2} |\zeta|^\lambda \left|\kappa(h\zeta)\right| d\zeta \\
&= n^{-1/2}\frac{1}{2\pi}\int \left(E\left[V^2\right]\right)^{1/2} |\zeta|^\lambda \left|\kappa(h\zeta)\right| d\zeta
\end{aligned}
$$

$$\begin{aligned}
&\preceq n^{-1/2} \int |\zeta|^{\lambda} |\kappa(h\zeta)| \, d\zeta \\
&= n^{-1/2} h^{-1-\lambda} \int |\xi|^{\lambda} |\kappa(\xi)| \, d\xi \\
&\preceq n^{-1/2} h^{-\lambda-1}.
\end{aligned}$$

Hence, by the Markov inequality,

$$\sup_{z \in \mathbb{R}} |L_{V,\lambda}(z,h)| = O_p\left(n^{-1/2} h^{-\lambda-1}\right).$$

When $h_n \to 0$, lemma 1 in the appendix of Pagan and Ullah (1999, p.362) applies to yield:

$$\begin{aligned}
h_n^{2\lambda+1} \Omega_{V,\lambda}(z,h_n) &= E\left[h_n^{-1}\left((-1)^{\lambda} V k^{(\lambda)}\left(\frac{Z-z}{h_n}\right)\right)^2\right] \\
&\quad - h_n\left(E\left[(-1)^{\lambda} h_n^{-1} V k^{(\lambda)}\left(\frac{Z-z}{h_n}\right)\right]\right)^2 \\
&= E\left[E\left[V^2|Z\right] h_n^{-1}\left(k^{(\lambda)}\left(\frac{Z-z}{h_n}\right)\right)^2\right] \\
&\quad - h_n\left(E\left[E\left[V|Z\right] h^{-1} k^{(\lambda)}\left(\frac{Z-z}{h_n}\right)\right]\right)^2 \\
&\to E\left[V^2|Z=z\right] f_Z(z) \int \left(k^{(\lambda)}(z)\right)^2 dz.
\end{aligned}$$

By Assumptions 3.1 and 3.2(iii), $E\left[V^2|Z=z\right] f_Z(z) > 0$ for $z \in \text{supp}(Z)$ and 3.4 ensures $\int \left(k^{(\lambda)}(z)\right)^2 dz > 0$ by Lemma B.1, so that $h_n^{2\lambda+1}\Omega_{V,\lambda}(z,h_n) > 0$ for all n sufficiently large.

(iii) To show asymptotic normality, we verify that $\ell_{V,\lambda}(z,h_n;V,Z)$ satisfies the hypotheses of the Lindeberg-Feller Central Limit Theorem for IID triangular arrays (indexed by n). The Lindeberg condition is: For all $\varepsilon > 0$,

$$\lim_{n\to\infty} Q_{n,h_n}(z,\varepsilon) \to 0,$$

where

$$Q_{n,h}(z,\varepsilon) \equiv \left(\Omega_{V,\lambda}(z,h)\right)^{-1} E\left[1\left(|\ell_{V,\lambda}(z,h;V,Z)| \geq \varepsilon \left(\Omega_{V,\lambda}(z,h)\right)^{1/2} n^{1/2}\right) |\ell_{V,\lambda}(z,h;V,Z)|^2\right].$$

Using the inequality $E\left[1\left[W \geq \eta\right] W^2\right] \leq \eta^{-\delta} E\left[W^{2+\delta}\right]$ for any $\delta > 0$, we have

$$Q_{n,h}(z,\varepsilon) \leq \left(\Omega_{V,\lambda}(z,h)\right)^{-1}\left(\varepsilon\left(\Omega_{V,\lambda}(z,h)\right)^{1/2} n^{1/2}\right)^{-\delta} E\left[|\ell_{V,\lambda}(z,h;V,Z)|^{2+\delta}\right],$$

where Assumption 3.2(iii) ensures that

$$
\begin{aligned}
E\left[\left|\ell_{V,\lambda}(z,h;V,Z)\right|^{2+\delta}\right] &= h^{-\lambda(2+\delta)}h^{-1-\delta}E\left[h^{-1}\left|V\right|^{2+\delta}\left|k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right|^{2+\delta}\right] \\
&= h^{-\lambda(2+\delta)}h^{-1-\delta}E\left[h^{-1}E\left[\left|V\right|^{2+\delta}|Z\right]\left|k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right|^{2+\delta}\right] \\
&\preceq h^{-\lambda(2+\delta)}h^{-1-\delta}E\left[h^{-1}\left|k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right|^{2+\delta}\right] \\
&\preceq h^{-\lambda(2+\delta)}h^{-1-\delta}.
\end{aligned}
$$

The results above and Assumption 3.2(iv) ensure that for any given z there exist $0 < A_{1,z}, A_{2,z} < \infty$ such that $A_{1,z}h_n^{-2\lambda-1} < \Omega_{V,\lambda}(z,h_n) < A_{2,z}h_n^{-2\lambda-1}$ for all h_n sufficiently small. Hence, we have

$$
\begin{aligned}
Q_{n,h_n}(z,\varepsilon) &\preceq \left(\varepsilon h_n^{-\lambda-1/2}n^{1/2}\right)^{-\delta}\frac{h_n^{-\lambda(2+\delta)}h_n^{-1-\delta}}{h_n^{-2\lambda-1}} \\
&= \left(\varepsilon h_n^{-\lambda-1/2}n^{1/2}h_n^{\lambda}h_n\right)^{-\delta} \\
&= \varepsilon^{-\delta}\left(nh_n\right)^{-\delta/2} \to 0
\end{aligned}
$$

provided $nh_n \to \infty$, which is implied by Assumption 3.6: $h_n \to 0, nh_n^{2\lambda+1} \to \infty$. ■

Proof of Theorem 3.3. The $O\left(\left\|\tilde{g}_{V_j,\lambda_j} - g_{V_j,\lambda_j}\right\|_\infty^2\right)$ remainder in eq.(20) can be dealt with as in the proof above of Theorem 4.9. Next, we note that

$$
\int s(z)\left(\hat{g}_{V,\lambda}(z,h) - g_V(z)\right)dz = L + B_h + R_h,
$$

where

$$
\begin{aligned}
L &= \hat{E}\left[Vs^{(\lambda)}(Z)\right] - E\left[Vs^{(\lambda)}(Z)\right] = \hat{E}\left[\psi_{V,\lambda}(s;V,Z)\right] \\
B_h &= \int s(z)\left(g_{V,\lambda}(z,h) - g_{V,\lambda}(z)\right)dz \\
R_h &= \int s(z)\left(\hat{g}_{V,\lambda}(z,h) - g_{V,\lambda}(z,h)\right)dz - \left(\hat{E}\left[Vs^{(\lambda)}(Z)\right] - E\left[Vs^{(\lambda)}(Z)\right]\right).
\end{aligned}
$$

We then have, by Assumption 3.7,

$$\begin{aligned}|B_{h_n}| &\equiv \left|\int s(z)\left(g_{V,\lambda}(z,h_n)-g_{V,\lambda}(z)\right)dz\right| \le \int |s(z)|\left|g_{V,\lambda}(z,h_n)-g_{V,\lambda}(z)\right|dz \\ &= \int |s(z)|\left|B_{V,\lambda}(z,h_n)\right|dz = o_p\left(n^{-1/2}\right)\int |s(z)|\,dz = o_p\left(n^{-1/2}\right).\end{aligned}$$

Next,

$$\begin{aligned}R_h &= \int s(z)\left(\hat{g}_{V,\lambda}(z,h)-g_{V,\lambda}(z,h)\right)dz - \left(\hat{E}\left[s^{(\lambda)}(Z)V\right]-E\left[s^{(\lambda)}(Z)V\right]\right) \\ &= (-1)^{\lambda}\int s(z)\left(\hat{E}\left[V\frac{1}{h^{1+\lambda}}k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right]-E\left[V\frac{1}{h^{1+\lambda}}k^{(\lambda)}\left(\frac{Z-z}{h}\right)\right]\right)dz \\ &\quad -\left(\hat{E}\left[Vs^{(\lambda)}(Z)\right]-E\left[Vs^{(\lambda)}(Z)\right]\right) \\ &= \int s^{(\lambda)}(z)\left(\hat{E}\left[V\frac{1}{h}k\left(\frac{Z-z}{h}\right)\right]-E\left[V\frac{1}{h}k\left(\frac{Z-z}{h}\right)\right]\right)dz \\ &\quad -\left(\hat{E}\left[Vs^{(\lambda)}(Z)\right]-E\left[Vs^{(\lambda)}(Z)\right]\right) \\ &= \int (\hat{E}\left[Vs^{(\lambda)}(z)\frac{1}{h}k\left(\frac{Z-z}{h}\right)-Vs^{(\lambda)}(Z)\right] \\ &\quad -E\left[Vs^{(\lambda)}(z)\frac{1}{h}k\left(\frac{Z-z}{h}\right)-Vs^{(\lambda)}(Z)\right])dz \\ &= \hat{E}\left[V\left(s^{(\lambda)}(z,h)-s^{(\lambda)}(Z)\right)-E\left[V\left(s^{(\lambda)}(z,h)-s^{(\lambda)}(Z)\right)\right]\right]\end{aligned}$$

where

$$s^{(\lambda)}(\tilde{z},h)=\int s^{(\lambda)}(z,h)\frac{1}{h}k\left(\frac{\tilde{z}-z}{h}\right)dz.$$

Hence, R_{h_n} is a zero-mean sample average where the variance of each individual IID term goes to zero, implying that $R_{h_n}=o_p\left(n^{-1/2}\right)$. ■

Proof of Theorem 3.4. This proof is virtually identical to the proof of Theorem 4.10 in the foregoing appendix, with $\varepsilon_n=\left(h_n^{-1}\right)^{\gamma_{1,B}}\exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)+n^{-1/2}\left(h_n^{-1}\right)^2$ instead of $\varepsilon_n=\left(h_n^{-1}\right)^{\gamma_{1,B}}\exp\left(\alpha_B\left(h_n^{-1}\right)^{\beta_B}\right)+n^{-1/2}\left(h_n^{-1}\right)^{\gamma_{1,L}}\exp\left(\alpha_L\left(h_n^{-1}\right)^{\beta_L}\right)$. ■

Proof of Theorem 3.5. This proof is virtually identical to the proof of Theorem 4.11, invoking Theorem 3.2 instead of Corollary 4.8. ■

References

Altonji, J. and R. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053-1102.

Andrews, D.W.K. (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560-596.

Angrist, J. and G. Imbens (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables" (with Discussion), *Journal of the American Statistical Association*, 91, 444-455.

Bertail, P., D. Politis, C. Haefke, and H. White (2004), "Subsampling the Distribution of Diverging Statistics with Applications to Finance," *Journal of Econometrics*, 120, 295-326.

Blundell R. and J. Powell (2000), "Endogeneity in Nonparametric and Semiparametric Regression Models," University of California, Berkeley, Department of Economics Discussion Paper.

Butcher, K. and A. Case (1994), "The Effects of Sibling Sex Composition on Women's Education and Earnings," *The Quarterly Journal of Economics*, 109, 531-563.

Chalak, K. and H. White (2007a), "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," UCSD Department of Economics Discussion Paper.

Chalak, K. and H. White (2007b), "Identification with Conditioning Instruments in Causal Systems," UCSD Department of Economics Discussion Paper.

Chernozhukov, V. and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., G. Imbens, and W. Newey (2007), "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4-14.

Chesher, A., (2003), "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.

Darolles, S., J. Florens, and E. Renault (2003), "Nonparametric Instrumental Regression," University of Tolouse GREMAQ Working Paper.

Dawid, P. (1979), "Conditional Independence and Statistical Theory," *Journal of the Royal Statistical Society*, Series B, 41, 1-31.

Dudley, R. (2002). *Real Analysis and Probability.* New York: Cambridge University Press.

Fan, J. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of Statistics*, 19, 1257-1272.

Fan, J. and Y.K. Truong (1993), "Nonparametric Regression with Errors in Variables," *Annals of Statistics*, 21, 1900-1925.

Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1-12.

Haerdle, W. and O. Linton (1994), "Applied Nonparametric Methods," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. IV. Amsterdam: Elsevier, ch.38.

Hahn, J. and G. Ridder (2007), "Conditional Moment Restrictions and Triangular Simultaneous Equations," IEPR Working Paper No. 07.3

Heckman, J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441-462.

Heckman J. and E. Vytlacil (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences* 96, 4730-4734.

Heckman, J. and E. Vytlacil (2001), ""Local Instrumental Variables," in C. Hsiao, K. Morimune, and J. Powell (eds.) in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya.* Cambridge: Cambridge University Press, pp. 1-46.

Heckman, J. and E. Vytlacil (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

Heckman, J. and E. Vytlacil (2007), "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," University of Chicago, Department of Economics Discussion Paper.

Heckman, J., S. Urzua, and E. Vytlacil (2006), "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389-432.

Hoderlein, S. (2005), "Nonparametric Demand Systems, Instrumental Variables and a Heterogeneous Population," Mannheim University, Department of Economics Working Paper.

Hoderlein, S. (2007) "How Many Consumers are Rational?," Mannheim University, Department of Economics Working Paper.

Hoderlein, S. and E. Mammen (2007), "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513-1518.

Hu, Y. and S. Schennach (2007), "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, forthcoming.

Imbens, G. and W. Newey (2003), "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," manuscript.

Matzkin, R. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71 1339-1375.

Matzkin, R. (2004), "Unobservable Instruments," Northwestern University Department of Economics Working Paper.

Newey, W. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

Newey, W. and J. Powell (2003), "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.

Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics.* Cambridge: Cambridge University Press.

Politis, D.N. and J.P. Romano (1999), "Multivariate Density Estimation with General Flat-Top Kernels of Infinite Order," *Journal of Multivariate Analysis*, 68, 1-25.

Santos, A. (2006), "Instrumental Variables Methods for Recovering Continuous Linear Functionals," Stanford University Department of Economics Working Paper.

Schennach, S.M. (2004a), "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33-75.

Schennach, S.M. (2004b), "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046-1093.

Schennach, S.M. (2007), "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75, 201-239.

White, H. and K. Chalak (2006), "A Unified Framework for Defining and Identifying Causal Effects," UCSD Department of Economics Discussion Paper.