Identification, Estimation and Efficiency of Nonparametric and Semiparametric Models in Microeconometrics

David Tomás Jacho–Chávez

A Dissertation submitted to the Department of Economics in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

London School of Economics and Political Science

London, July 2006

Declaration

I hereby declare:

- No part of this doctoral dissertation has been presented to any University for any degree.
- Chapter 2 was undertaken as joint work with Professor Oliver B. Linton and Professor Arthur Lewbel.

David Tomás Jacho–Chávez

Abstract

The focal point of this thesis is on identification and estimation of nonparametric models, as well as the efficiency and higher order properties of a class of semiparametric estimators in Microeconometrics.

We present a new identification result for a particular nonparametric model that nests many popular parametric/nonparametric Econometric models as special cases. Estimators are proposed and their asymptotic properties derived; in particular, they are shown to be consistent and asymptotically pointwise normally distributed. We implement these estimators for the nonparametric estimation and testing of production functions in 4 industries within the Chinese economy in the years 1995 and 2001.

The statistical properties of an entire family of semiparametric estimators for Limited Dependent Variables models are also analyzed. The derived theoretical results have direct applicability to a wide range of estimation problems. In particular, we derive the semiparametric efficiency bounds and show that some of the already–proposed estimators achieve these bounds. A connection with the Programme Evaluation literature is established as well.

Finally, we derive an asymptotic approximation to the Mean Square Error of this class of semiparametric estimators to aid the choice of smoothing parameter. It is demonstrated that this choice can be made on the basis of bias alone. Possible extensions in this framework are also discussed.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to extend my most sincere thanks and appreciation to my supervisor, Oliver Linton, for his support, priceless advice and patience. Oliver's optimism and good humour were always very helpful during the completion of this dissertation. His valuable insights were crucial for this work. I will always consider myself incredibly fortunate to have worked under his supervision.

Jezamin Lim, my partner, deserves special mention. She has been my inspiration and source of strength during good and bad times. Her love, understanding and support have made this dissertation possible. I will be forever in debt with her.

Throughout my higher education, I have received financial support from the Department of Economics at the London School of Economics and Political Science, and the Escuela Superior Politécnica del Litoral (Ecuador). Many thanks goes to everyone at these places.

Finally, I would like to thank my parents, family and friends for their unconditional encouragement and caring in all these years.

Chapter 1

Introduction

Nonparametric and semiparametric specifications are common in Econometric models. In Microeconomics, they are informative about consumer or firm behavior while imposing a minimum set of restrictions in order to achieve identification of the main features of interest. Furthermore, non/semiparametric econometric estimators can be constructed (in some cases efficiently) to remain consistent in situations where parametric models are not. This robustness property remains the most important aspect in this literature.

The calculation of fully nonparametric estimators, while modeling economic relationships, imposes enormous data requirements when a large number of variables is involved. This problem, known as the 'curse of dimensionality', can be alleviated by the use of certain functional structures that may be imposed by Economic theory. These nonparametric functional forms are more restrictive than a fully nonparametric specification, but their assumptions are weaker than those with a finite–dimensional parametric specification. These particular forms (as we will show in this dissertation) can be used to identify the model itself, and we build estimators that achieve faster rates of convergence relative to their nonparametric counterparts. On the other hand, semiparametric models are also very attractive alternatives to reduce this 'curse of dimensionality'. They impose some parametric restrictions on the economic relationship that we try to model, while allowing functional form of many of its other components to remain unknown. A large number of the derived semiparametric estimators achieve parametric rates of convergence, and attain the semiparametric efficiency bounds induced by their underlying assembly.

This thesis makes contributions to these two related efforts to reduce the necessity of functional form restrictions, which are needed to identify and estimate efficiently Econometric models. The outline of the thesis is as follows:

Chapter 2: Identification and Nonparametric Estimation of a Transformed Additively Separable Model. We first introduce a flexible structure for a function of random variables that nests many features of Econometric models as special cases. This particular form involves a smooth monotonic transformation of another smooth function, which is assumed to be separable, either additively or multiplicatively, with respect to one of its argument. For example: This particular functional form could represent the conditional mean or quantile function of the observed outcome in Limited Dependent Variable Models. It could also represent homothetic functions widely used in Economic modeling. In a regression model with unknown transformation of the dependent variable, the conditional distribution of the dependent variable given the observed regressors, also shares this functional form.

Unlike other authors in this literature, we make full use of the implied strictly monotonic link function in the examples above to achieve nonparametric identification of all their unknown components. Furthermore, we propose a computationally simple nonparametric estimation algorithm that does not require any optimization or matching. The resulting estimators have pointwise asymptotic normal distributions under regularity conditions. Their rates of convergence are also faster than those of a fully nonparametric alternative. We also find that they perform fairly well in comparison with other nonparametric estimators previously proposed in the literature, for a variety of Monte Carlo designs dealing with small sample sizes.

Finally, the idea of Generalized Homothetic functions is introduced in order to estimate production functions for various industries within the Chinese economy during the years 1995 and 2001. We also estimate and test a range of parametric specifications for comparison purposes. In certain industries, we find that their implied measures of input substitution and scale are very different to those implied by our more flexible specification.

Chapter 3: Efficiency Bounds in Semiparametric Models defined by Moment Restrictions using an Estimated Conditional Probability Density. In this chapter, we calculate the semiparametric efficiency bounds for an entire class of semiparametric estimators proposed in the literature of Limited Dependent Variable models. Regardless of the inherent nonlinearities of this type of model, these estimators are computationally easy to calculate as they have 'closed' formulae. In some cases, they resemble Least Squares or Linear Instrumental Variable estimators in a simple linear regression model. They are also robust to measurement errors, endogeneity and heteroskedasticity of unknown form.

It is shown that these bounds are sharp. That is, two previously proposed estimators in the literature achieve these bounds. The bounds are also applied to estimators of treatment effects. We find that they are also efficient among a particular family of propensity–score–weighted estimators. We also find that using an estimate rather than the real conditional probability density in the construction of these estimators is more efficient. Furthermore, the general result presented in this chapter can be directly applied to any new estimator that belongs to this class. All these results seem to be new in the literature.

Our theoretical findings are confirmed in a simulation study involving a variety of designs, frameworks and kernel–based estimators. We find marginal improvements in terms of fitting criteria when Local Linear instead of Local Constant regression is used for the nonparametric component of the estimator.

Chapter 4: Optimal Bandwidth Choice for Estimation of Inverse Conditional-Density-Weighted Expectations. The kernel–based implementation of the class of semiparametric estimators discussed in Chapter 3 requires the choice of a smoothing parameter. This chapter characterizes its optimal value. We obtain a 'closed' formula for the optimal bandwidth by minimizing the leading terms of a second–order mean squared error expansion of the estimator with respect to this smoothing parameter.

It turns out that we can choose the optimal value for the bandwidth based on bias alone. In particular, we show that there are two sources of biases: a 'smoothing' bias, and a 'degrees–of–freedom' bias. The optimal bandwidth makes these biases' contributions to the asymptotic mean square error have the same order of magnitude. Based on the derived formula for the optimal smoothing parameter, a simple 'plug–in' estimator for the optimal bandwidth is proposed. We prove its consistency under regularity conditions.

We examine the quality of the asymptotic approximation in finite samples for simple Monte Carlo designs. The proposed 'plug–in' estimator for the optimal bandwidth is also shown to perform fairly well under various circumstances and sample sizes. Finally, we examine how our results adapt in the presence of discrete regressors. The potential use of a Bootstrap bandwidth selection mechanism is also presented.

Each chapter can be read independently from each other. This means that there is variation in the notation used within each chapter.

Chapter 2

Identification and Nonparametric Estimation of a Transformed Additively Separable Model

2.1 Introduction

For vector x and scalar z, let $r(x,z)$ be a function that, along with its derivatives, can be consistently estimated nonparametrically. Unconstrained nonparametric estimation of r is usually unattractive when $x \in \Re^d$ is multidimensional, because the rate of convergence decreases rapidly as d increases, yielding very imprecise estimates with samples of practical size, see Stone (1980). We may overcome this curse of dimensionality by making assumptions about the functional form of r that are stronger than those of a fully nonparametric estimator, but weaker than those of a finite–dimensional parametric model, see Stone (1986). In the fully nonparametric framework, one such dimension–reduction method is to assume there exist functions H, G and F such that

$$r(x,z) = H[M(x,z)] = H[G(x) + F(z)] \tag{2.1.1}$$

where $M(x,z) \equiv G(x) + F(z)$ and H is strictly monotonic. This chapter provides an identification result that allows us to recover H, M, G and F in the above specification. An estimation algorithm is then proposed when $r(x,z)$ represents a conditional mean function for a given sample $\{Y_i, X_i, Z_i\}_{i=1}^n$. We also provide limiting distributions for the resulting nonparametric estimators of each component of (2.1.1), as well as present evidence of their small sample performance in a limited Monte Carlo experiment.

This framework encompasses a large class of economic models. For example, the function $r(x,z)$ could be utility or consumer cost functions recovered from estimated consumer demand functions via revealed preference theory, or it could also be a production or producer cost function that can be recovered directly from a data set. When $H[m] = m$, the identity function, Chiang (1984), Simon and Blume (1994), Bairam (1994), and Chung (1994) reviewed popular parametric functional forms used in economics. In demand analysis, Goldman and Uzawa (1964) provideed an overview of the variety of separability concepts implicit in such specifications.

Many methods have been developed for the identification and estimation of strongly or additively separable models, where $r(x,z) = \sum_{k=1}^{d} G_k(x_k) + F(z)$ or its generalized version $r(x,z) = H[\sum_{k=1}^{d} G_k(x_k) + F(z)]$. Friedman and Stutzle (1981), Breiman and Friedman (1985), Andrews (1991), Tjøstheim and Auestad (1994) and Linton and Nielsen (1995) are examples of the prior and Linton and Härdle (1996), and Horowitz and Mammen (2004) proposed estimators of the latter for known H. Horowitz (2001) used this assumed strong separability in order to identify the components of the model when H is entirely unknown, and proposed a kernel–based consistent and asymptotically normal estimator. When $d = 1$, specification (2.1.1) is nested in the class of models Horowitz considers. However, many econometric models imply link functions, H, that are strictly monotonic but otherwise unknown (see examples below). By making use of this extra information, our identification result does not requires G to be additive in its argument, in order to achieve full identification.

As strong separability may be too restrictive in the context of an empirical application, models satisfying equation (2.1.1) are called weakly separable. They offer a more flexible specification that allows for some interaction among regressors, as well as a faster rate of convergence compared to fully unrestricted nonparametric estimation. Pinkse (2001) provides a general nonparametric estimator for this class of models under weaker conditions on M, i.e. no separability, and on H, which is assumed to be increasing only. However, in this partly separable specification, Pinkse's estimator will compute M up to an arbitrary monotonic transformation; while ours, by making use of the assumed strict monotonicity of H, provides the unique (up to sign–scale and location normalization) M, and by virtue of marginal integration, the unique G and F.

These transformed partly additive models could also arise as ordinary partly additive regression models in which the dependent variable is censored, truncated or binary. These would be models in which $Y^* = G(X) + F(Z) + \varepsilon$ for some unobserved Y^* and ε, with ε independent of (X,Z) with an absolutely continuous distribution function, and what is observed is (Y,X,Z), where $Y = Y^* 1(Y^* \geq 0)$, or $Y = Y^* | Y^* \geq 0$, or $Y = 1(Y^* \geq 0)$, in which case $r(x,z) = E[Y|X = x, Z = z]$ or $r(x,z) = \text{med}[Y|X = x, Z = z]$. The function H would then be the distribution or a quantile function of ε. Threshold or selection equations in particular are commonly of this form, having $Y = 1[G(X) + \varepsilon \geq -z]$, where $-z$ is some threshold, e.g. price or bid, with $G(X) + \varepsilon$ being willingness to pay. In this sense, our identification result is similar to Lewbel and Linton (2002) for the censored or truncated regression, though it is applicable to a wider range of Limited Dependent Variable models, and makes use of the extra assumed separability.

Model (2.1.1) may also become evident in a regression model with unknown transformation of the dependent variable, $F(z) = G(x) + \varepsilon$, where ε has absolutely continuous distribution function H which is independent of x, F is an unknown monotonic transformation and G, an unknown regression function. It follows that the conditional distribution z given x, $F_{Z|X}$, is given by $H(F(z) - G(x)) \equiv r(z,x)$, where $F_{Z|X} \equiv r(z,x)$. For this model, Ekeland, Heckman, and Nesheim (2004) provided an identification result that also exploits separability between x and z, but not the monotonicity of H as we do here. More generally, Matzkin (2003) considered identification of models of the form $Y = m(X, Z, \varepsilon)$ with ε independent of (X, Z). In this framework, our model makes no assumption about the role of unobservables, as well as provides no estimates of these other than a limiting distribution theory for estimates of r.

Moreover, the proposed identification result may also be extended to the transformed multiplicative sub–models of the form $H[M(x,z)] = H[G(x)F(z)]$, which are very common in production literature. Particularly, a function $r(\widetilde{x}, z)$ is said to be homothetic if and only if $r(\widetilde{x}, z) = H[M^*(\widetilde{x}, z)]$ where H is strictly monotonic and M^* is linearly homogeneous, i.e. $M^*(\lambda\widetilde{x}, \lambda z) = \lambda M^*(\widetilde{x}, z)$ or equivalently $M^*(\widetilde{x}, z) = \lambda^{-1} M^*(\lambda\widetilde{x}, \lambda z)$. If $\lambda = z^{-1}$ and $x = \widetilde{x}/z$, it follows that $M(x,z) \equiv G(x)F(z)$, where $G(x) = M^*(x, 1)$ and $F(z) = z$. Our estimator can readily be used in order to identify this homothetic model, as well as a more general class of functions where $F(z)$ is not a simple power function of z. We implement our methodology for the estimation of generalized homothetic production functions for four industries in the People's Republic of China. For this, we have built an `R` package (see Ihaka and Gentleman (1996)), `JLLprod`, incorporating functions that implement the techniques proposed here.

Although the functions H, G and F may not be of direct interest in some applications, our proposed estimators might still be useful for testing whether or not functions have the proposed separability, by comparing $\widehat{r}(x,z)$ with $\widehat{H}[\widehat{G}(x) + \widehat{F}(z)]$, or in the production theory context, to test whether production functions are generalized homothetic, by comparing $F(z) = z$ with $\widehat{F}(z)$. In addition, the more general model $r(x,z,w) = H[M(x,z), w]$ can also be identified when $M(x,z)$ is additive or multiplicative and H is strictly monotonic with respect to its first argument.

Section 2.2 sets out the main identification results. Our proposed estimation algorithm is presented in Section 2.3. Section 2.4 analyzes the asymptotic properties of the estimators. A Monte Carlo experiment is presented in Section 2.5 comparing our estimators to those proposed by Linton and Nielsen (1995), and Linton and Härdle (1996), both of which use knowledge of H, and with Horowitz (2001). This section also provides an empirical illustration of our methodology for the estimation of generalized production functions in four industries within the Chinese economy for the years 1995 and 2001. Finally, Section 4.6 concludes and briefly outlines possible extensions.

2.2 Identification

The main identification idea is presented in this section. Firstly, observe that (2.1.1) is unchanged if G and F are replaced by $G+c_G$ and $F+c_F$, respectively, and $H(m)$ is replaced by $\widetilde{H}(m) = H(m - c_G - c_F)$. Similarly, (2.1.1) remains unchanged if G and F are replaced by cG and cF respectively, for some $c \neq 0$ and $H(m)$ is replaced by $\widetilde{H}(m) = H(m/c)$. Therefore, as it is commonly the case in the nonparametric literature, location and scale normalizations are needed to make identification possible. We describe and discuss these normalizations below, but first, we state the following conditions which are assumed to hold throughout our exposition.

Assumption I:

(I1) Let $W \equiv (X, Z)$ be a $(d+1)$–dimensional random vector with support $\Psi_x \times \Psi_z$, where $\Psi_x \subseteq \Re^d$, and $\Psi_z \subseteq \Re$, for some $d \geq 1$. The distribution of W is absolutely continuous with respect to Lebesgue measure with probability density $f_W(w) > 0$ for all $w = (x, z) \in \Psi_x \times \Psi_z$. There exists functions r, H, G and F such that $r(x, z) = H[G(x) + F(z)]$ for all $w \equiv (x, z) \in \Psi_x \times \Psi_z$.

(I2) (i) The function H is strictly monotonic and H, G and F are continuous and differentiable with respect to any mixture of their arguments. (ii) F has finite first derivative, $f(z)$, over its entire support, and $f(z_0) = 1$ for some $z_0 \in int(\Psi_z)$. (iii) Let $H(0) = r_0$, where r_0 is a constant. In addition, (iv) Let $r(x, z) \in \Psi_{r(x,z_0)}$ for all $w \equiv (x, z) \in \Psi_x \times \Psi_z$, where $\Psi_{r(x,z)}$ is the image of the function $r(x, z)$.

Assumption (I1) specifies the model. The functions M, thus G and F are not nonparametrically identified if (X, Z) has discrete elements, a restriction which is common in nonparametric models with unknown link function (see Horowitz (2001)). Assumption (I2) defines the required location and scale normalizations that makes identification possible. It also requires that the image of $r(x, z)$ over its entire support is replicated once r is evaluated at z_0 for all x. This assumption also implies that $s(x, z) \equiv \partial r(x, z)/\partial z$ is a well defined function for all $w \in \Psi_x \times \Psi_z$. Then, for the random variables $r(X,Z)$ and $s(X,Z)$, let us define the function $q(t,z)$ by

$$q(t,z) = E\left[s(X,Z) \mid r(X,Z) = t, Z = z\right]. \tag{2.2.1}$$

The assumed strict monotonicity of H ensures that H^{-1}, the inverse function of H, is well defined over its entire support; also, let $h(M) = H^{(1)}(M)$ be the first derivative of H.

Theorem 2.2.1 *Let Assumption I hold. Then,*

$$M(x,z) \equiv G(x) + F(z) = \int_{r_0}^{r(x,z)} \frac{dt}{q(t,z_0)}. \tag{2.2.2}$$

Proof. It follows from Assumption (I1) that $s(x,z) = h[M(x,z)]\, f(z)$, so

$$\begin{aligned} E\left[s(X,Z) \mid r(X,Z) = t, Z = z_0\right] &= E\left[h[M(X,Z)]\, f(Z) \mid r(X,Z) = t, Z = z_0\right] \\ &= E\left[h\left[H^{-1}(r(X,Z))\right] f(Z) \mid r(X,Z) = t, Z = z_0\right] \\ &= h\left[H^{-1}(t)\right] f(z_0), \text{ and} \end{aligned}$$

$q(t,z_0) = h\left[H^{-1}(t)\right] f(z_0)$. Then using the change of variables $m = H^{-1}(t)$, and noticing that $h\left[H^{-1}(t)\right] = h(m)$ and $dt = h(m)\, dm$, we obtain

$$\begin{aligned} \int_{r_0}^{r(x,z)} \frac{dt}{q(t,z_0)} &= \int_{r_0}^{r(x,z)} \frac{dt}{h\left[H^{-1}(t)\right] f(z_0)} \\ &= \int_{H^{-1}[r_0]}^{H^{-1}[r(x,z)]} \frac{h(m)\, dm}{h(m)\, f(z_0)} \\ &= \left(H^{-1}[r(x,z)] - H^{-1}[r_0]\right)(1/f(z_0)) = M(x,z) \equiv G(x) + F(z), \end{aligned}$$

as required. ■

In the special case of an identity link function, i.e. $H(m) = m$, q has a simple form $q(t,z_0) = f(z_0) \equiv q(z_0)$ which is constant over all t and equals 1 by Assumption (I2). It is clear from the proof of this theorem, that without knowledge of z_0 and r_0, Assumptions (I2)(ii) and (I2)(iii), the function $M(x,z)$ can only be identified up to a sign–scale factor $1/f(z_0)$, and location constant $H^{-1}[r_0](1/f(z_0))$ provided $|f(z_0)| > 0$ and $\left|H^{-1}[r_0]\right| < \infty$. In addition, (I2)(iv) depends on a range of (X,Z) that is large enough to obtain the function $r(X,Z)$ everywhere in the interval r_0 to $r(x,z)$. That is, it ensures that q exists everywhere on $\Psi_{r(x,z)} \times \Psi_z$, making $M(x,z)$ identifiable for all x and z.

Lewbel and Linton (2002) also used a similar result (2.2.2) in the nonparametric censored regression setup, $Y = \max[0, M(W) - \varepsilon]$. Their estimator assumes independence between W and ε with $E(\varepsilon) = 0$. For the proposed partly separable case, Theorem 2.2.1 above replicates their Theorem 3 (page 769), but with additional normalizations. In particular, $q(t, z_0) = F_\varepsilon\left[\mathfrak{F}^{-1}(t)\right] f(z_0)$, where F_ε is the cumulative distribution function of ε and $\mathfrak{F}(m) = \int_{-\infty}^{m} F_\varepsilon(e)\, de$. As is in our case, their location constant must be known a priori. The assumed additive separability with respect to z also adds an extra normalization on Ψ_z.

For the multiplicative model, $M(x,z) = G(x) F(z)$, the following assumption and corollary provides the necessary identification.

Assumption I*:

(I*1) Let $W \equiv (X, Z)$ be a $(d+1)$–dimensional random vector with support $\Psi_x \times \Psi_z$, where $\Psi_x \subseteq \Re^d$, and $\Psi_z \subseteq \Re$, for some $d \geq 1$. The distribution of W is absolutely continuous with respect to Lebesgue measure with probability density $f_W(w) > 0$ for all $w = (x,z) \in \Psi_x \times \Psi_z$. There exists functions r, H, G and F such that $r(x,z) = H[G(x) F(z)]$ for all $w \equiv (x,z) \in \Psi_x \times \Psi_z$.

(I*2) (i) The function H is strictly monotonic and H, G and F are continuous and differentiable with respect to any mixture of their arguments. (ii) F has finite first derivative, $f(z)$, such that $F(z_0)/f(z_0) = 1$ for some $z_0 \in int(\Psi_z)$. (iii) Let $H(1) = r_1$, where r_1 is a constant. In addition, (iv) Let $r(x,z) \in \Psi_{r(x,z_0)}$ for all $w = (x,z) \in \Psi_x \times \Psi_z$, where $\Psi_{r(x,z)}$ is the image of the function $r(x,z)$.

Corollary 2.2.2 *Let Assumption I* hold. Then,*

$$M(x,z) = G(x) F(z) = \exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{q(t, z_0)} \right). \tag{2.2.3}$$

Proof. See Appendix. ■

If r_l is greater than $r(x,z)$, for any nonnegative constant, r_l, then the integrals of the form $\int_{r_l}^{r(x,z)}$ above are to be interpreted as $-\int_{r_l}^{r(x,z)}$, for $l = 0, 1$. Once $M(x,z)$ has been pulled out of the unknown (but strictly monotonic) function H in (2.2.2) or (2.2.3), we may recover G and F by standard marginal integration, see Linton and Nielsen (1995). Let P_1 and P_2 be deterministic discrete or continuous weighting functions with $\int_{\Psi_z} dP_1(z) = 1$ and $\int_{\Psi_x} dP_2(x) = 1$. These integrals should be interpreted in the Stieltjes sense. Let p_1 and p_2

be the densities of P_1 and P_2 with respect to Lebesgue measure in $\Re$ and $\Re^d$ respectively. Then

$$\alpha_{P_1}(x) = \int_{\Psi_z} M(x,z)\, dP_1(z), \text{ and } \alpha_{P_2}(z) = \int_{\Psi_x} M(x,z)\, dP_2(x).$$

In the additive model, $\alpha_{P_1}(x) = G(x) + c_1$ and $\alpha_{P_2}(z) = F(z) + c_2$, where $c_1 = \int_{\Psi_z} F(z)\, dP_1(z)$ and $c_2 = \int_{\Psi_x} G(x)\, dP_2(x)$. While in the multiplicative case, $\alpha_{P_1}(x) = c_1 G(x)$ and $\alpha_{P_2}(z) = c_2 F(z)$. Hence, $\alpha_{P_1}(x)$ and $\alpha_{P_2}(z)$ are, up to identifiability, the components of M in both additive ($c = c_1 + c_2$) and multiplicative structures ($c = c_1 \times c_2$).

Given the definition of $r(x,z)$, it follows that

$$H(M(x,z)) = E[r(X,Z)|\, M(X,Z) = M(x,z)],$$

thus the function H may also be identified. If $r(x,z) \equiv E[Y|\, X = x, Z = z]$ for some random Y, then the equality $H(M(x,z)) = E[Y|\, M(X,Z) = M(x,z)]$ may also be used to identify H.

We could replace the sign–scale normalization in Assumption (I2)(ii), by another that assumes there is a bounded, non–negative function, ω, such that

$$\int \frac{\omega(z_0)}{f(z_0)} dz_0 = 1,$$

with ω integrating to one over its compact support. For the applied researcher, a normalization restriction such as (I2) is empirically appealing because it entails the selection of a single value rather than a whole function. From a practical point of view, it will also ease computational time. Besides, these restrictions may well be imposed by economic theory. For example, the neoclassical production function (positive but decreasing marginal products with respect to each factor) of two inputs, with constant returns to scale, implies that its two production factors, K, capital and, L, labor are essential in the sense that positive inputs of both factors are needed for a positive output. If $r(K,L)$ represents such a function, $r_1 = r(0,L) = r(K,0) \equiv \min_{K,L} r(K,L)$ is a natural choice. Furthermore, such a production function has a multiplicative structure (see Section 2.5) with $F(L) = L$, in which case $f(L) = 1$ and any $L_0 > 0$ may be chosen and full identification can be achieved.

Strict monotonicity of the link function plays an important role in these results. Because of it, the conditional mean of $s(x,z)$ given r and z is a well–defined function, with a known structure which is separable in z. This contrasts with Horowitz (2001) and Ekeland, Heckman, and Nesheim (2004), where strict monotonicity is neither assumed nor is it exploited for identification, rather it is the separability of the partial derivatives of $r(x,z)$ that is used instead. It is also worth noticing that our identification result does not need the existence of stochastic variation in $s(x,z)$ – it could be known or take on random values

– once conditioned on r and z.

2.3 Estimation

In this section, for the case $r(x,z) \equiv E[Y|\, X = x, Z = z]$, we describe estimators of M, G, F and H based on replacing the unknown functions $r(x,z)$, $s(x,z)$ and $q(t,z)$ in (2.2.2) by multidimensional smoothers. Since an estimator of the partial derivative of the regression surface, $r(x,z)$, with respect to z is needed, a natural choice of smoother will be a Local Polynomial estimator, which produces estimators for r and s simultaneously. These nonparametric estimators also have better boundary behavior and the ability to adapt to non–uniform designs, among other desirable properties (see Fan and Gijbels (1996)).

For a given random sample $\{Y_i, X_i, Z_i\}_{i=1}^{n}$, estimators of M, G, F and H in the additive case, can be constructed by following these steps:

1) Obtain a consistent estimator of $\widehat{r}_i = \widehat{r}(X_i, Z_i)$ and $\widehat{s}_i = \widehat{s}(X_i, Z_i)$ by local p_1-th order polynomial regression of Y_i on X_i and Z_i with corresponding kernel K_1, and bandwidth sequence $h_1 = h_1(n)$ for $i = 1, \ldots, n$.

2) Obtain a consistent estimator of $q(t,z)$, given z_0 for all t, by local p_2-th order polynomial regression of $\widehat{s}_i$ on $\widehat{r}_i$ and Z_i with corresponding kernel K_2 and bandwidth sequence $h_2 = h_2(n)$ for $i = 1, \ldots, n$. Denote this estimate as $\widehat{q}(t, z_0) = \widehat{E}[\,\widehat{s}|\widehat{r}(X,Z) = t, Z = z_0]$.

3) For a constant r_0, define an estimate of $M(x,z) \equiv G(x) + F(z)$ by

$$\widehat{M}(x,z) = \int_{r_0}^{\widehat{r}(x,z)} \frac{dt}{\widehat{q}(t,z_0)}. \tag{2.3.1}$$

4) Estimate $G(x)$ and $F(z)$ consistently up to an additive constant by marginal integration,

$$\widehat{\alpha}_{P_1}(x) = \int_{\Psi_z} \widehat{M}(x,z)\, dP_1(z), \tag{2.3.2}$$

$$\widehat{\alpha}_{P_2}(z) = \int_{\Psi_x} \widehat{M}(x,z)\, dP_2(x). \tag{2.3.3}$$

5) Now for $\widetilde{c} = (1/2)\left[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x)\, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z)\, dP_1(z)\right]$, define $\widetilde{G}(x) = \widehat{\alpha}_{P_1}(x) - \widetilde{c}$, $\widetilde{F}(z) = \widehat{\alpha}_{P_2}(z) - \widetilde{c}$ and $\widetilde{M}(X_i, Z_i) \equiv \widetilde{G}(X_i) + \widetilde{F}(Z_i) + \widetilde{c}$, then we can obtain a consistent estimator of $H(m)$ by local p_*-th polynomial regression of Y_i or $\widehat{r}(X_i, Z_i)$

on $\widetilde{M}(X_i, Z_i)$ with corresponding kernel k_* and bandwidth sequence $h_* = h_*(n)$ for $i = 1, \ldots, n$. Denote this estimate as $\widehat{H}(m)$.

If we are interested in estimating a partly multiplicative model instead, we can replace steps 3–5 by:

3*) For a constant r_1, define an estimate of $M(x, z) \equiv G(x) F(z)$ by

$$\widehat{M}(x, z) = \exp\left(\int_{r_1}^{\widehat{r}(x,z)} \frac{dt}{\widehat{q}(t, z_0)}\right).$$

4*) Estimate $G(x)$ and $F(z)$ consistently up to a scale factor by marginal integration,

$$\widehat{\alpha}_{P_1}(x) = \int_{\Psi_z} \widehat{M}(x, z)\, dP_1(z),$$
$$\widehat{\alpha}_{P_2}(z) = \int_{\Psi_x} \widehat{M}(x, z)\, dP_2(x).$$

5*) Now for $\widetilde{c} = (1/2)\left[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x)\, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z)\, dP_1(z)\right]$, define $\widetilde{G}(x) = \widehat{\alpha}_{P_1}(x)/\widetilde{c}$, $\widetilde{F}(z) = \widehat{\alpha}_{P_2}(z)/\widetilde{c}$, and $\widetilde{M}(X_i, Z_i) \equiv \widetilde{G}(X_i)\widetilde{F}(Z_i)\widetilde{c}$, then we can obtain a consistent estimator of $H(m)$ by local p_*-th polynomial regression of Y_i or $\widehat{r}(X_i, Z_i)$ on $\widetilde{M}(X_i, Z_i)$ with corresponding kernel k_* and bandwidth sequence $h_* = h_*(n)$ for $i = 1, \ldots, n$. Denote this estimate as $\widehat{H}(m)$.

We can immediately observe how important the joint–unconstrained nonparametric estimation of r and s is in step 1. They will not only be used in estimating q in step 2, but r along with the preset r_0 (r_1) will also define the limits of the integral in (2.3.1) in step 3 (3*). Operationally, because of estimation error in step 1, the function $\widehat{q}(t, z_0)$ is only observed for $t \in range(\widehat{r}(X_i, z_0))$, but we continue it beyond this support by linear extrapolation (with slope equal to the derivative of $\widehat{q}$ at the corresponding end of the support) elsewhere in step 3 (3*). Moreover, (2.3.1) can be easily evaluated using numerical integration. A convenient choice of $P_1(z)$ and $P_2(x)$, in (2.3.2) and (2.3.3), are $F_z(z)$ and $F_x(x)$, the distribution functions of Z and X respectively. Given that they are in general unknown in practice, we can replace them by their empirical analog, $\widehat{F}_z(z)$ and $\widehat{F}_x(x)$, so we have $\widehat{\alpha}_1(x) \equiv n^{-1}\sum_{i=1}^{n} \widehat{M}(x, Z_i)$ and $\widehat{\alpha}_2(z) = n^{-1}\sum_{i=1}^{n} \widehat{M}(X_i, z)$. Finally, notice that $\widehat{H}$ in step 5 (5*) involves a simple univariate nonparametric regression.

Known Link Function

In many practical situations, especially with binary and survival time data, the conditional distribution of Y given (X, Z) belongs to a known family with known link function, H, for example the logit and probit link functions are common for binary data, and the logarithm transform for Poisson count data, see McCullagh and Nelder (1989). More generally, if H is twice continuously differentiable such that $h(M) = H^{(1)}(M) \equiv \partial H(m)/\partial m|_{m=M} \neq 0$ over its entire support, the function $q(t, z_0)$ in Theorem 2.2.1 and Corollary 2.2.2 can be replaced by $q_{add}(t) \equiv h\left[H^{-1}(t)\right]$ in the additive case, or by $q_{mult}(t) \equiv h\left[H^{-1}(t)\right] H^{-1}(t)$ in the multiplicative one, so scale normalization is not needed. Specifically,

$$\begin{aligned} \int_{r_0}^{r(x,z)} \frac{dt}{q_{add}(t)} + H^{-1}[r_0] &= \int_{r_0}^{r(x,z)} \frac{dt}{h[H^{-1}(t)]} + H^{-1}[r_0] \\ &= H^{-1}[r(x,z)] = M(x,z) \equiv G(x) + F(z), \end{aligned} \tag{2.3.4}$$

and similarly

$$\begin{aligned} \exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{q_{mult}(t)} + \ln\left(H^{-1}[r_1]\right) \right) &= \exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{h[H^{-1}(t)]\, H^{-1}(t)} + H^{-1}[r_1] \right) \\ &= H^{-1}[r(x,z)] = M(x,z) \equiv G(x)\, F(z), \end{aligned} \tag{2.3.5}$$

by a change of variables $m = H^{-1}(t)$, such that $dt = h(m)\, dm$. The above equalities hold for any r_l such that $H^{-1}[r_l] < \infty$ for $l = 0, 1$, so it does not require a location normalization as well. Notice that $q(t, z_0) = q_{add}(t)\,(1/f(z_0))$ and $q(t, z_0) = q_{mult}(t)\,(F(z_0)/f(z_0))$, in the additive and multiplicative case respectively.

After replacing the unknown conditional mean function $r(x, z)$, in (2.3.4) and (2.3.5), by local p_1-th order polynomial regression of Y on X and Z with kernel K_1, and bandwidth sequence $h_1 = h_1(n)$, we obtain $\widehat{M}(x, z) = H^{-1}[\widehat{r}(x, z)]$, which corresponds to the estimator proposed by Linton and Härdle (1996) and to that proposed by Linton and Nielsen (1995) for the identity link. In the fully additive case, $G(x) = \sum_{k=1}^{d} G(x_k)$, they also derive the asymptotic properties of $\widehat{M}$, $\widehat{G}_k$ and $\widehat{F}$. In Section 2.5, we compare the performance of our procedure to that of these two estimators in the special case when H is known and $d = 1$.

2.4 Asymptotic Properties

This section gives assumptions under which we present theorems providing the pointwise distribution of our estimators of M, G, F and H for some $z = z_0$ and $r = r_0$. This is done for the additive case in conditional mean function estimation as described in the previous section. The technical issues involving the distribution of M and H are those of generated regressors, see Ahn (1995), Ahn (1997), Su and Ullah (2004), Su and Ullah (2006), and Lewbel and Linton (2006). Once the asymptotic normal distribution of M is established, the asymptotic properties of G and F will follow from ordinary marginal integration results.

ASSUMPTION E:

(E1) The kernels K_l, $l = 1, 2$, satisfy $K_1 = \Pi_{j=1}^{d+1} k_1(w_j)$, $K_2 = \Pi_{j=1}^{2} k_2(v_j)$, and k_l, $l = 1, 2$, are bounded, symmetric about zero, with compact support $[-c_l, c_l]$ and satisfy the property that $\int_{\Re} k_l(u)\, du = 1$. For $l = 1$ and 2, the functions $H_{l\mathbf{j}} = u^{\mathbf{j}} K_l(u)$ for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p_l + 1$ are Lipschitz continuous. The matrices $\mathbf{M}_r$ and $\mathbf{M}_q$, multivariate moments of the kernels K_1 and K_2 respectively (defined in the Appendix) are nonsingular.

(E2) The densities f_W of W_i, and f_V of V_i for $W_i^\top \equiv (X_i^\top, Z_i)$ and $V_i \equiv (r_i, Z_i)$ respectively are uniformly bounded and they are also bounded away from zero on their compact support.

(E3) For some $\xi > 2$, $E[|\varepsilon_{r,i}|^\xi] < \infty$, $E[|\varepsilon_{q,i}|^\xi] < \infty$, and $E[|\varepsilon_{r,i}\varepsilon_{q,i}|^\xi] < \infty$ where $\varepsilon_{r,i} = Y_i - r(X_i, Z_i)$ and $\varepsilon_{q,i} = S_i - q(r_i, Z_i)$. Also, $E\left[\varepsilon_{r,i}^2 \middle| X_i = x, Z_i = z\right] \equiv \sigma_r^2(x, z)$, be such that $\nu_{P_1}(x) \equiv \int p_1^2(z)\, \sigma_r^2(x, z)\, f_W^{-1}(x, z)\, q^{-2}(r, z_0)\, dz < \infty$ and $\nu_{P_2}(z) \equiv \int p_2^2(x)\, \sigma_r^2(x, z)\, f_W^{-1}(x, z)\, q^{-2}(r, z_0)\, dx < \infty$.

(E4) The function $r(\cdot)$ is $(p_1 + 1)$ times partially continuously differentiable and the function $q(\cdot)$ is $(p_2 + 1)$ times partially continuously differentiable. The corresponding $(p_1 + 1)$th or $(p_2 + 1)$th order partial derivatives are Lipschitz continuous on their compact support.

(E5) The bandwidth sequences h_1, and h_2 go to zero as $n \to \infty$, and satisfy the following conditions:

(i) $nh_1^{d+1} h_2^{2(p_2+1)} \to c \in [0, \infty)$,

(ii) $n^{1/2} h_1^{d+1} h_2^2 / \ln n \to \infty$, $n^{1/2} h_1^{2(p_1+1)} h_2^{-2} \to 0$,

(iii) $nh_1^{d+1} h_1^{2(p_1+1)} \to c \in [0, \infty)$, and $nh_1^{d+1} h_1^{2p_1} h_2^2 \to c \in [0, \infty)$.

Assumptions (E1)–(E4) provide the regularity conditions needed for the existence of an asymptotic distribution. The estimation error $\varepsilon_{q,i}$, in Assumption (E3), is such that $E[\varepsilon_{q,i}|\, r\,(X_i, z) = r, Z_i = z] = 0$. However, $E\,[\varepsilon_{q,i}|\, X_i = x, Z_i = z] \neq 0$, so we write $\varepsilon_{q,i} = g_q\,(x, z) + \eta_i$, where $E\,[\eta_i|\, X_i = x, Z_i = z] = 0$ by construction. Assumption (E4) ensures Taylor–series expansions to appropriate orders.

Let $\nu_{1n} = n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n} + h_1^{p_1+1}$ and $\nu_{2n} = n^{-1/2} h_2^{-1} \sqrt{\ln n} + h_2^{p_2+1}$, then by Theorem 6 (page 593) in Masry (1996a), $\max\limits_{1\leq j\leq n} \parallel \widehat{r}\,(W_j) - r\,(W_j) \parallel = O_p\,(\nu_{1n})$, $\max\limits_{1\leq j\leq n} \parallel \widehat{s}\,(W_j) - s\,(W_j) \parallel = O_p\left(h_1^{-1}\nu_{1n}\right)$ and $\sup\limits_{v} \parallel \widehat{q}\,(v) - q\,(v) \parallel = O_p\,(\nu_{2n})$ if the unobserved $\{V_1, \ldots, V_n\}$ were used in constructing $\widehat{q}$. Because $\{\widehat{V}_1, \ldots, \widehat{V}_n\}$ were used instead, the approximation error is accounted for in Assumption (E5)(ii), which implies that $(h_2^{-1}\nu_{1n})^2 = o(n^{-1/2}h_2^{-1})$ and so $h_2^{-1}\nu_{1n} = o\,(1)$, where the appearance of h_2^{-1} is because of the use of Taylor–series expansions in our proofs. Assumption (E5) permits various choices of bandwidths for given polynomial orders. For example, if $p_1 = p_2 = 3$, we could set $h_1 \propto n^{-1/9}$, and $h_2 = bb \times h_1$ when $d = 1$, for a nonzero scalar bb, as in our Monte Carlo experiment in Section 2.5. More generally, in view of Assumption (E5)(iii), $h_1 \propto n^{-1/[2(p_1+1)+d]}$ and $h_2 \propto n^{-1/[2p_2+3]}$ will work for a variety of combinations of d, p_1, and p_2.

Theorem 2.4.1 *Suppose that Assumption I holds. Then, under Assumption E, there exists a bounded continuous function $\mathcal{B}\,(x, z)$ such that*

$$\sqrt{nh_1^{d+1}}\left(\widehat{M}\,(x, z) - M\,(x, z) - \mathcal{B}\,(x, z)\right) \xrightarrow{d} N\left[0, \frac{\sigma_r^2\,(x, z)}{q^2\,(r, z_0)\, f_W\,(x, z)} \left[\mathbf{M}_r^{-1}\Gamma_r\mathbf{M}_r^{-1}\right]_{0,0}\right],$$

where $[\mathbf{A}]_{0,0}$ means the upper-left element of matrix $\mathbf{A}$.

Proof. The proof of this theorem, along with definitions of each component, is given in the Appendix. ■

We should mention that there are four sources of biases, defined in the Appendix, i.e. $\mathcal{B}\,(x, z) = h_1^{p_1+1}\mathcal{B}_1\,(x, z) + h_1^{p_1}h_2\mathcal{B}_2\,(x, z) + h_2^{p_2+1}\mathcal{B}_3\,(x, z) + h_1^{p_1+1}\mathcal{B}_4\,(x, z)$, where $\mathcal{B}_3$ corresponds to the ordinary nonparametric bias of $\widehat{q}$ if the unobserved r and s were used instead in step 2, and $\mathcal{B}_4$ corresponds to the standard nonparametric bias while calculating $\widehat{r}$ in step 1 weighted by $q^{-1}\,(r, z_0)$. $\mathcal{B}_1$ and $\mathcal{B}_2$ are because of the use of generated regressor $\widehat{r}$, and generated response $\widehat{s}$ in constructing $\widehat{q}$ respectively in step 2.

Given this result,

$$\begin{aligned} E\{\widehat{M}\,(x, z)\} - M\,(x, z) &= O(h_1^{p_1+1}) + O(h_1^{p_1}h_2) + O(h_2^{p_2+1}), \text{ and} \\ Var\{\widehat{M}\,(x, z)\} &= O(n^{-1}h_1^{-(d+1)}), \end{aligned}$$

and these orders of magnitude also hold at boundary points by virtue of using Local Polynomial regression in each step. By employing generic marginal integration of this preliminary smoother, as described in step 4, we obtain by straightforward calculation the following result:

Corollary 2.4.2 *Suppose that Assumption I holds. Then, under Assumption E*

$$\sqrt{nh_1^d}\left(\widehat{\alpha}_{P_1}(x)-\alpha_{P_1}(x)-\int\mathcal{B}(x,z)\,dP_1(z)\right)\xrightarrow{d}N\left[0,\nu_{P_1}(x)\left[\mathbf{M}_r^{-1}\Gamma_r^1\mathbf{M}_r^{-1}\right]_{0,0}\right], \quad (2.4.1)$$

$$\sqrt{nh_1}\left(\widehat{\alpha}_{P_2}(z)-\alpha_{P_2}(z)-\int\mathcal{B}(x,z)\,dP_2(x)\right)\xrightarrow{d}N\left[0,\nu_{P_2}(z)\left[\mathbf{M}_r^{-1}\Gamma_r^2\mathbf{M}_r^{-1}\right]_{0,0}\right]. \quad (2.4.2)$$

where $[\mathbf{A}]_{0,0}$ *means the upper-left element of matrix* $\mathbf{A}$.

Proof. The proof follows from results in Linton and Nielsen (1995) and Linton and Härdle (1996), and therefore is omitted. ■

Our procedure is similar to many other kernel–based multi–stage nonparametric procedures in that the first estimation step does not contribute to the asymptotic variance of the final stage estimators, see Linton (2000), Xiao, Linton, Carroll, and Mammen (2003). However, the asymptotic variances of $\widehat{M}(x,z)$, $\widehat{\alpha}_{P_1}(x)$ and $\widehat{\alpha}_{P_2}(z)$ reflect the lack of knowledge of the link function H through the appearance of the function q in the denominator, which by Assumption I is bounded away from zero and depends on the scale normalization z_0, and the conditional variance $\sigma_r^2(x,z)$ of Y. They can be consistently estimated from the estimates of $r(x,z_0)$, $q(r,z_0)$ in steps 1 and 2, and $\sigma_r^2(x,z)$. For example, if P_i, $l=1,2$, are empirical distribution functions, the standard errors of $\widehat{\alpha}_{P_1}(X_i)$ and $\widehat{\alpha}_{P_2}(Z_i)$ can be computed as

$$\psi^1(k_1)\,\widehat{\sigma}_r^2 n^{-1}\sum_{j=1}^{n}\left[\widehat{f}_W(X_i,Z_j)\,\widehat{q}^2(r(X_i,Z_j),z_0)\right]^{-1}\widehat{f}_Z(Z_j), \text{ and}$$

$$\psi^2(k_1)\,\widehat{\sigma}_r^2 n^{-1}\sum_{j=1}^{n}\left[\widehat{f}_W(X_j,Z_i)\,\widehat{q}^2(r(X_j,Z_i),z_0)\right]^{-1}\widehat{f}_X(X_j)$$

respectively, in which $\psi^l(k_1)\equiv\left[\mathbf{M}_r^{-1}\Gamma_r^l\mathbf{M}_r^{-1}\right]_{0,0}$ for $l=1,2$, $\widehat{f}_W$, $\widehat{f}_X$ and $\widehat{f}_Z$ are the corresponding kernel estimates of f_W, f_X and f_Z, while $\widehat{\sigma}_r^2=n^{-1}\sum_{i=1}^n\left[Y_i-\widehat{r}(X_i,Z_i)\right]^2$ or $\widehat{\sigma}_r^2=n^{-1}\sum_{i=1}^n[Y_i-\widehat{H}(\widetilde{M}(X_i,Z_i))]^2$.

Since our estimators are based on marginal integration of a function of a preliminary $(d+1)$ −dimensional nonparametric estimator, hence the smoothness of G and F must increase as the dimension of X increases to achieve the rate $n^{-p_1/(2p_1+1)}$, the optimal rate of convergence when G and F have p_1 continuous derivatives, see Stone (1985) and Stone

(1986).

Now consider H. Define $\Psi_{M(x,z)} = \{m : m = G(x) + F(z), (x,z) \in \Psi_x \times \Psi_z\}$. If G and F were known, H could be estimated consistently by a local p^*–polynomial mean regression of Y on $M(X,Z) \equiv G(X) + F(Z)$. Otherwise, H can be estimated with unknown M by replacing $G(X_i)$ and $F(Z_i)$ with estimators in the expression for $M(X_i, Z_i)$. This is a classic generated regressors problem as in Ahn (1995). Denote these by $\widehat{\alpha}_{P_1}(X_i)$ and $\widehat{\alpha}_{P_2}(Z_i)$, with $\widetilde{M_i} \equiv \widehat{\alpha}_{P_1}(X_i) + \widehat{\alpha}_{P_2}(Z_i) - \widetilde{c}$ and $M_i \equiv \alpha_{P_1}(X_i) + \alpha_{P_2}(Z_i) - c$. Let $h_\dagger = \max(h_1^{p_1+1}, h_2^{p_2+1}, h_1^{p_1}h_2)$, then $\max\limits_{1\leq j\leq n} \parallel \widetilde{M_j} - M_j \parallel = O_p(\nu_{\dagger n})$, where $\nu_{\dagger n} = n^{-1/2}h_1^{-d/2}\sqrt{\ln n} + h_\dagger$.

We also make the following additional assumption:

Assumption F:

(F1) The kernel k_* is bounded, symmetric about zero, with compact support $[-c_*, c_*]$ and satisfies the property that $\int_{\Re} k_*(u)\,du = 1$. The functions $H_{*j} = u^j k_*(u)$ for all j with $0 \leq j \leq 2p_* + 1$ are Lipschitz continuous. The matrix $\mathbf{M}_H$, defined in the Appendix, is nonsingular.

(F2) Let f_M be the density of $M(X,Z)$, which is assumed to exist, to inherit the smoothness properties of M and f_W and to be bounded away from zero on its compact support.

(F3) The bandwidth sequence h_* goes to zero as $n \to \infty$, and satisfies the following conditions:

(i) $nh_*^{2(p_*+1)+1} \to c \in [0,\infty)$, $nh_*h_\dagger^2 \to c \in [0,\infty)$,

(ii) $n^{1/2}h_1^d h_*^{3/2}/\ln n \to \infty$, and $n^{1/2}h_\dagger^2 h_*^{-3/2} \to 0$.

Assumptions (F1) to (F3) are similar to those in Assumption E. As before, Assumption (F3)(ii) implies that $\left(h_*^{-1}\nu_{\dagger n}\right)^2 = o(n^{-1/2}h_*^{-1/2})$ and also that $\left(h_*^{-1}\nu_{\dagger n}\right) = o(1)$. Assumption (F3) imposes restrictions on the rate at which $h_* \to 0$ as $n \to \infty$. They ensure that no contributions to the asymptotic variance of $\widehat{H}$ are made by previous estimation stages. Let $\sigma_H^2(m) = E\left[\varepsilon_r^2 \middle| M(X,Z) = m\right]$, then we have the following theorem:

Theorem 2.4.3 *Suppose that Assumption I holds, then, under Assumption E and F, there exists a bounded continuous function $\mathcal{B}_H(\cdot)$, such that*

$$\sqrt{nh_*}\left(\widehat{H}(m) - H(m) - \mathcal{B}_H(m)\right) \xrightarrow{d} N\left(0, \frac{\sigma_H^2(m)}{f_M(m)}\left[\mathbf{M}_H^{-1}\Gamma_H\mathbf{M}_H^{-1}\right]_{0,0}\right),$$

for $m \in \Psi_{M(x,z)}$, where $[\mathbf{A}]_{0,0}$ means the upper-left element of matrix $\mathbf{A}$.

Proof. The proof of this theorem, along with definitions of each component, is given in the Appendix. ■

When $p_* = 1$, h_* admits the rate $n^{-1/5}$ when h_1 and h_2 are chosen as suggested above when $d = 1$, as it is done in the application and simulations in Section 2.5. In which case, $\mathcal{B}_H(\cdot)$ simplifies to the standard bias from a univariate local linear regression. Standard errors can be easily computed from the formula above. By evaluating $\widehat{H}$ at each data point, the implied estimator of $\widehat{r}(X_i, Z_i) = \widehat{H}[\widetilde{M}(X_i, Z_i)]$ is $O_p(n^{-1/2}h_1^{-(d-1)/2})$, for large h_1 and d, which can be seen by a straightforward local Taylor–series expansion around $M(X_i, Z_i)$. That is, our proposed methodology has effectively reduced the curse of dimensionality in estimating r by 1 with respect to its fully unrestricted nonparametric counterpart.

2.5 Numerical Results

2.5.1 Simulations

In this section, we describe a small Monte Carlo experiment to study the finite sample properties of the proposed estimator, and compare its performance with that of direct competitors in two leading scenarios: When the link function is known and the case when it is not. Code for these simulations was written in `GAUSS`. The different designs considered below do not reflect any model of interest in economics. They were chosen to emphasize performance issues rather than empirical relevance. In order to simplify things we also restricted our attention to $d = 1$.

Known Link Function

We contrast the performance of our estimator with that of Linton and Nielsen (1995) and Linton and Härdle (1996). Although they are not fully efficient, these alternative estimators use knowledge of the link function. Hence, they provide an appropiate benchmark allowing the performance of our estimator to be compared.

In this case, the experiment was performed as follows: A number, n, of observations (Y, X, Z) were generated from $Y = r(X, Z) + \sigma_r \cdot \varepsilon$, where the distributions of x and z were $U[0, 1]$, and ε was chosen independently of X and Z with a standard normal distribution. For each scenario, $\sigma_r^2 = 1$ and $\sigma_r^2 = 2$, two particular specifications of H, in $r(x, z) =$

$H[M(x,z)]$ where $M(x,z) = G(x) + F(z)$, were used for the same G and F.

$$\begin{aligned} G(x) &= (1/2)\sin(2\pi x) \\ F(z) &= -2z^2 + 2z - 1/3 \\ H[m] &= m \qquad (2.5.1) \\ H[m] &= \ln\left(m + \sqrt{1+m^2}\right) + 3. \qquad (2.5.2) \end{aligned}$$

The curvature and non–monotonicity of G and F provide a test for the estimators describe in Section 2.3. Notice that neither G nor F is homogeneous and both were chosen such that $E[G(X)] = E[F(Z)] = 0$. Also, at $z_0 = 1/4$, we have $f(z_0) = 1$. To obtain our estimators $\widehat{M}$, $\widehat{G}$, $\widehat{F}$ and $\widehat{H}$, we use the second order Gaussian kernel $k_i(u) = \left(1/\sqrt{2\pi}\right)\exp\left(-u^2/2\right)$, $i = 1, 2, *$. The integral in $\widehat{M}$, step 2 in Section 2.3, was evaluated numerically using the trapezoid method. We also fixed $p_1 = 3$, $p_2 = 1$ and $p_* = 1$. We used the bandwidth $h_1 = cc\widehat{s}_W n^{-1/9}$, where cc is a constant term and $\widehat{s}_W$ is the squared root of the average of the sample variances of X_i and Z_i. Namely, this bandwidth is proportional to the optimal rate for 3rd–order Local Polynomial estimation in the first stage, whereas for simplicity, h_2 was fixed as $3h_1$. The bandwidth h_* was set to follow Silverman's rule ($1.06n^{-1/5}$ times the squared root of the average of the regressors variances). Three different choices of cc were considered: $cc \in \{0.5, 1, 1.5\}$.

Each function was estimated at a 50×50 equally spaced grid in $[0,1] \times [0,1]$ when $n = 150$, and at another 60×60 uniform grid on the same domain when $n = 600$. Two criteria summarizing goodness of fit were calculated, the Integrated Root Mean Squared Error (IRMSE) and Integrated Mean Absolute Error (IMAE), at all grid points and then they were averaged. Tables 2.1 and 2.2 report the median of these averages over 2000 replications for each design, scenario and bandwidth. They also report the results obtained when using the estimators proposed by Linton and Nielsen (1995) when (2.5.1) is used, and Linton and Härdle (1996) when (2.5.2) is used instead, on the first column from the left under each fitting criteria respectively. They were constructed using the same unrestricted first stage nonparametric regression used by our estimator.

As seen in the tables, for either sample size, lack of knowledge of the link function increases the fitting error of our estimator by roughly 5 to 85 percent relative to estimates using that knowledge. For each scenario, the IRMSE and IMAE decline as the sample size is quadrupled for both sets of estimators, at somewhat the same less than $\sqrt{n}$–rate. Larger bandwidths produce superior estimates for all functional components in all designs. Estimates of M in both designs and scenarios are generally less accurate than the others. In the estimation of the additive components, G and F, the fitting criteria for Linton and Nielsen (1995) and Linton and Härdle (1996) estimators are of approximately the same magnitude, while the proposed estimator has a smaller IRMSE and IMAE when estimating

F relative to estimates of G. There does not seem to be a dramatic difference in estimates of H between estimators in both designs. All sets of estimates deteriorate when σ_r is increased.

Unknown Link Function

As it was pointed our earlier, when $d = 1$, model (2.1.1) is nested in the class of models Horowitz (2001) considers. Consequently, it is natural to make a comparison with his estimator in this specific case. We replicated Horowitz (2001) original experiment[1] which is as follows: 1000 observations (Y, X, Z) were generated from, $Y = 1\left(G(X) + F(Z) - \varepsilon > 0\right)$, where $\varepsilon \sim N(0,1)$, $X \sim N(0,16)$ and $Z \sim N(0,16)$, and independent of each other. The functions G, F and H are[2]

$$\begin{aligned} G(x) &= 3\sin\left(\frac{\pi}{3}x^2\right), \\ F(z) &= 3\left[\exp(0.35z) - 1\right], \text{ and} \\ H(m) &= \Phi(m), \end{aligned}$$

where Φ is the standard normal distribution function. This is a binary probit model, where $\Pr(Y = 1 | X = x, Z = z) = \Phi(G(x) + F(z)) \equiv r(x, z)$.

Horowitz (2001) (NP2) used the following fourth and second order kernels to estimate G, F and H:

$$\begin{aligned} K(u) &= \frac{105}{64}\left(1 - 5u^2 + 7u^4 - 3u^6\right) 1(|u| \leq 1), \\ K_H(u) &= \frac{15}{16}\left(1 - u^2\right)^2 1(|u| \leq 1). \end{aligned}$$

The weight functions used to calculate $\widehat{G}$, $\widehat{F}$ and $\widehat{H}$ were $w_2(x) = K_H(x)$, $w_1(z) = (1/2)K_H(z/2)$, and $w_H(x, z) = w_2(x) w_1(z)$ respectively. He also used bandwidths $h_{11} = 6$, $h_{21} = 5$, and $h_H = 3.25$. He chose these bandwidths through Monte Carlo experimentation to approximately minimize the unweighted Integrated Mean Squared Errors of his estimators of G, F and H. The additional bandwidths his estimator needs were set using his suggested rule–of–thumb, $h_{k2} = h_{k1} n^{-1/72}$ for $k = 1, 2$.

We implement our proposed estimator (NP1) for this design, using a second order Gaussian kernel as before, with $p_1 = 0$, $p_2 = 1$, and $p_* = 1$. We also found the optimal bandwidths $h_1 = 0.925$, $h_2 = 2.5$ and $h_* = 0.2$ for this design, by Monte Carlo experimentation as Horowitz (2001) did.

[1] The computer code we wrote to implement Horowitz (2001) estimator, was not fast enough to conduct large scale simulations as before.

[2] In Horowitz (2001) notation: $F \equiv f_1$, $G \equiv f_2$, and $H \equiv G$, with $x^1 \equiv z$, $x^2 \equiv x$ and $v \equiv m$.

Figure 2.1 shows the standardized Q–Q plots of both set of estimators at different points well in the interior of the support of each function. These points were chosen sufficiently far from the boundary of the data to avoid boundary effects for both estimators. These plots were based on 300 replications. We observe that the normal approximation of our estimator for G and F are better than Horowitz's at the chosen points. Similar results (not presented) hold for other points well in the interior of the support of (X, Z) for G and F. On the other hand, the normal approximation of our estimator for H is similar to Horowitz's for low values of m only, while it outperforms Horowitz's for higher values.

Finally, Figure 2.2 displays a visualization of the resulting output of 5000 replications of a fourth design using only our procedure. Data was generated as before with the same G and F, but $H[m] = 1 + (16/7)\,m$, with $\sigma_r^2 = 1$. Other information was set accordingly, for example $n = 400$, $h_1 = 0.15$, $h_2 = 0.7$, $p_1 = 3$, and $p_2 = 1$. The white plane and dashed lines represent medians of simulations, and gray planes and dotted lines represent 90% simulation envelopes.

2.5.2 Generalized Homothetic Production Function Estimation

Let y be the log output of a firm and $(\widetilde{x}, z)$ be a vector of inputs. Starting with Shephard (1953) and Shephard (1970), many parametric production function models of the form $y = r^*(\widetilde{x}, z) + \varepsilon_{r^*}$ have been estimated that impose either linear homogeneity or homotheticity for the function r. In the homogenous case, corresponding to known $H(m) = m$, many models have been proposed, see Bairam (1994) and Chung (1994) for parametric examples, and Tripathi and Kim (2003) and Tripathi (1998) for fully nonparametric options. Zellner and Ryu (1998) provides empirical comparisons of a large number of different homothetic production functional forms. In the nonparametric framework[3], Lewbel and Linton (2006) presents an estimator for a homothetically separable function r^*.

However, a more general definition of homogeneous and homothetic functions is given below:

Definition 2.5.1 *A function $M^* : \Psi_w \subset \Re^{d+1} \to \Re$ is said to be generalized homogeneous on Ψ_w if and only if the equation $M^*(\lambda w) = g(\lambda)\, M^*(w)$ holds for all $(\lambda, w) \in \Re_{++} \times \Psi_w$ such that $\lambda w \in \Psi_w$. The function $g : \Re_{++} \to \Re_{++}$ is such that $g(1) = 1$ and $\partial g(\lambda)/\partial\lambda > 0$ for all λ.*

[3]Other examples of nonparametric estimators include Varian (1984) and Primont and Primont (1994). Also, Hanoch and Rothschild (1972) discussed a test to verify whether a homothetic production function exists that could, without statistical errors, generate a given data set. Although these papers do not assume a parametric functional form, by assumption they have no statistical errors. Consequently, they have also no associated distribution theory.

Definition 2.5.2 *A function $r^*: \Psi_w \subset \Re^{d+1} \to \Re$ is said to be generalized homothetic on Ψ_w if and only if $r^*(w) = H[M^*(w)]$, where $H: \Re \to \Re$ is a strictly monotonic function and M^* is generalized homogeneous on Ψ_w.*

It is clear from Definitions 2.5.1 and 2.5.2, that homogeneity of degree κ and homotheticity are the special case in which the function g takes the functional form $g(\lambda) = \lambda^\kappa$. Given a generalized homothetic production function we have

$$\begin{aligned} r^*(\tilde{x}, z) &= H[M^*(\tilde{x}, z)] = H\left[M^*(\tilde{x}/z, 1)\, g(1/z)^{-1}\right] \\ &= H[G(x)\, F(z)] = H[M(x, z)] \equiv r(x, z), \end{aligned} \tag{2.5.3}$$

where $x = \tilde{x}/z$ and $F(z) = 1/g(1/z)$. When H is assumed known and equal to the identity function, Tripathi and Kim (2003) and Tripathi (1998) use the assumption that $M(x, z)$ is a homogeneous function of degree one, i.e. $F(z) = 1/z$, in order to identify the model and achieve dimensionality reduction. Lewbel and Linton (2006) used the same functional assumption regarding F but with an unknown strictly monotonic link function H. In the contrary, the proposed estimator in this chapter could easily be implemented in order to identify M, G, F and H in models such as (2.5.3), i.e. $y = r(x, z) + \varepsilon_r$, without imposing any such parametric specification of F, but exploiting the partial separability of M with respect to z instead along with the fact that $f(z) > 0$[4]. For the same reasons, it does also reduce the dimensionality by 1 as explained earlier.

We have built an `R` package, `JLLprod`, which along with its manual can be freely downloaded from the author's website. After installation, the user also has access to a production data set from the Ecuadorian economy in 2002, and will be able to reproduce the information presented in this section. We then use it in order to estimate generalized homothetic production functions for four industries in mainland China[5] in two time periods, 1995 and 2001. For each firm in every industry, we observe the net value of real fixed assets K, the number of employees L, and Y defined as the log of value–added real output. K and Y are measured in thousands of Yuan converted to the base year 2000 using a general price deflator for the Chinese economy. For details regarding the collection and construction of this data set, see Jefferson, Hu, Guan, and Yu (2003).

We consider both nonparametric and parametric estimates of the production function $r(k, L) \in \mathcal{P}$, which is a set of smooth production functions, and $k = K/L$ as in (2.5.3). To eliminate extreme outliers in both sets of estimates, we sort the data by k and remove the top and bottom 2.5% of observations in each industry and year. Both regressors were also

[4] As $\partial g(\lambda)/\partial\lambda > 0$, and $\lambda = z^{-1}$, it follows that $F(z)$ is strictly increasing, i.e. $f(z) = \partial F(z)/\partial z > 0$ over its entire domain.

[5] Package `JLLprod` also contains production data of 406 firms in the Petroleum, Chemical and Plastics industries in Ecuador in 2002.

normalized by their respective median prior to regression.

Parametric

Consider a general production function (P1) in which log output $Y = r_{\psi_{P1}}(k, L) + \varepsilon_r$, where

$$
\begin{aligned}
r_{\psi_{P1}}(k, L) = \theta_0 + \theta_1 \ln(k) + \theta_2 \ln(L+\gamma) + \theta_3 [\ln(k)]^2 \\
+ \theta_4 \ln(k) \ln(L+\gamma) + \theta_5 [\ln(L+\gamma)]^2 ,
\end{aligned}
\tag{2.5.4}
$$

and $\psi_{P1} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^{\top}$. When $2\theta_1\theta_5 - \theta_2\theta_4 = 0$ and $\theta_1^2\theta_5 - \theta_2^2\theta_3 = 0$, this general model nests the following generalized homothetic production function (P2) specification,

$$
\begin{aligned}
M(k, L) &= k^{\alpha}(L+\gamma) \\
r_{\psi_{P2}}(k, L) &= H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2 ,
\end{aligned}
\tag{2.5.5}
$$

where $\psi_{P2} = (\alpha, \beta_0, \beta_1, \beta_2, \gamma)^{\top}$. Furthermore, if we also impose a third parameter restriction[6], $\gamma = 0$, we obtain the homothetic Translog production function (P3) of Christensen, Jorgenson, and Lau (1973) as a special case as well, i.e.

$$
\begin{aligned}
M(k, L) &= k^{\alpha} L \\
r_{\psi_{P3}}(k, L) &= H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2 ,
\end{aligned}
\tag{2.5.6}
$$

where $\psi_{P3} = (\alpha, \beta_0, \beta_1, \beta_2)^{\top}$.

Figure 2.3 shows isoquants for P2 with $\psi_{P2} = (1/2, 10, 1/2, 1, \gamma)^{\top}$, where $\gamma = -1, 0, +1$. At any level of output, these isoquants are steeper at high levels of k for negative γ than for positive γ. However, as in the homothetic case, $\gamma = 0$, the slopes of their level surfaces are constant along rays through the origin. This important property is preserved by this more general specification.

Fitting these models by Nonlinear Least Squares in each year yields the parameter estimates reported in Tables 2.3–2.5 (Heteroskedasticity robust standard errors are in parentheses).

[6] If we were to impose this restriction alone, (2.5.4) reduces to the ordinary unrestricted Translog production function.

Specification Test

Two sets of parametric restrictions are tested on model (2.5.4) for each year and industry. In order to assess whether model (2.5.4) may be further simplified by (2.5.5),

$$H_0 : 2\theta_1\theta_5 - \theta_2\theta_4 = 0;$$
$$\theta_1^2\theta_5 - \theta_2^2\theta_3 = 0$$

is tested by means of a Wald statistic, W_{12} which is distributed under H_0 as $\chi_{(2)}$. A further simplification, (2.5.6), is also tested by a Wald statistic, W_{13}, which under the

$$H_0 : 2\theta_1\theta_5 - \theta_2\theta_4 = 0;$$
$$\theta_1^2\theta_5 - \theta_2^2\theta_3 = 0;$$
$$\gamma = 0,$$

is distributed as $\chi_{(3)}$. The results of these tests are presented below.

Industry	1995				2001			
	W_{12}	p-value	W_{13}	p-value	W_{12}	p-value	W_{13}	p-value
Chemical	1.280	0.527	2.244	0.523	17.286	0.000	1,095	0.000
Iron	8.834	0.012	14.261	0.003	2.272	0.321	2.343	0.504
Petroleum	1.790	0.409	3.076	0.380	0.791	0.673	0.813	0.846
Transportation	1.735	0.420	1.997	0.573	7.980	0.019	8.252	0.041

Models (2.5.5) and (2.5.6) are valid parametric simplifications of the general production function (2.5.4), except for the iron industry in 1995 and the chemical and transportation industries in 2001.

The suitability of the parametric Generalized Homothetic and Translog production function fits, $r_{\widehat{\psi}_{P2}}(k, L)$ and $r_{\widehat{\psi}_{P3}}(k, L)$, in these industries may be also judged by the use of a residual based test. For this purpose, we decide to employ the test proposed by Zheng (1996) for the hypothesis

$$H_0 : r \in \overline{\mathcal{P}}\{r \in \mathcal{P} | \, r = r_{\psi_{Pl}} \text{ for some } \psi_{Pl}\}.$$

For $l = 2, 3$, their test statistics are given by

$$U_{Pl} = \frac{1}{\lambda^2 n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(Y_i - r_{\widehat{\psi}_{Pl}}(k_i, L_i)\right) \left(Y_j - r_{\widehat{\psi}_{Pl}}(k_j, L_j)\right) K\left(\frac{k_j - k_i}{\lambda}\right) K\left(\frac{L_j - L_i}{\lambda}\right), \tag{2.5.7}$$

with kernel $K(\cdot)$, the Gaussian kernel here, and bandwidth λ, set equal to h_1 in all cases. Given some regularity conditions, under the null hypothesis that the parametric specifications are correct,

$$n\lambda U_{Pl} \sim N\left(0, 2\int K^2(u)\,du \int \left[\sigma_r^2(k,L)\,p(k,L)\right]^2 dkdL\right), \tag{2.5.8}$$

by replacing integrals by sums and unknown functions by their nonparametric estimates in (2.5.7) and (2.5.8), we obtain the following test results:

Industry	1995			2001		
	λ	U_{P2}	p-value	λ	U_{P2}	p-value
Chemical	5.125	-0.7698	0.7793	2.25	-0.4728	0.6818
Iron	4.250	-0.7532	0.7743	4	-0.7367	0.7693
Petroleum	2.750	-0.8117	0.7915	11	-0.7325	0.7681
Transportation	1.750	-0.6519	0.7428	4.37	-0.7124	0.7619

	λ	U_{P3}	p-value	λ	U_{P3}	p-value
Chemical	5.125	-0.7721	0.7800	2.25	-0.4130	0.6602
Iron	4.250	-0.7195	0.7641	4	-0.7417	0.7709
Petroleum	2.750	-0.8088	0.7907	11	-0.7327	0.7681
Transportation	1.750	-0.6503	0.7422	4.37	-0.7065	0.7601

We fail to reject both H_0 for all industries in both years at any level of significance. In all cases, test results are not altered by the choice of smoothing parameter λ. Both sets of results justify the use of both models as sensible parametric simplifications of the data[7] against which we may compare our more flexible specification. Other kernel–based specification tests are Bierens (1990), Härdle and Mammen (1993), Gozalo (1993) and Horowitz and Spokoiny (2001), for example.

Nonparametric

Figures 2.4 to 2.11 show generalized homothetic nonparametric estimates $\widehat{M}(k,L)$, $\widehat{G}(k)$, $\widehat{F}(L)$ and $\widehat{H}(M)$ for both years. For each industry and year, we use local quadratic regression with a Gaussian kernel and bandwidths, h_1, given by a standard unrestricted leave–one–out cross validation method for regression functions. In the second stage, we set bandwidth h_2 to be the same in local linear regressions across industries and time. We also choose the

[7]Although the appropriateness of these parametric models may change through time, see Konishi and Nishiyama (2002).

location and scale normalizations through experimentation to obtain estimated surfaces $\widehat{M}$ with approximately the same range, yielding the following normalizations:

Industry	1995			2001		
	n	$\ln L_0$	r_0	n	$\ln L_0$	r_0
Chemical	1560	3.40	7	1637	3.06	7.0
Iron	376	−0.37	7	341	4.06	8.0
Petroleum	93	2.73	7	119	2.27	8.5
Transportation	989	3.44	7	1230	4.04	7.5

The nonparametric fits of the generalized homogeneous component, $\widehat{M}$, shown in Figures 2.4 and 2.8, are quite similar. They are both increasing in k and L with ranges varying more with labor than with respect to capital to labor ratios, as we would expect[8]. Nonparametric estimates of the functions G and F are different to the parametric Translog model estimates (P3) in Figures 2.5, 2.6, 2.9 and 2.10[9], but they are roughly similar to parametric generalized homothetic model (P2) at low levels of L. They are all strictly increasing in their arguments, but show quite a bit more curvature, departing most markedly from the parametric models for F in 1995 and G in 2001 for most industries. Comparing the nonparametric estimator of F, in Figures 2.6 and 2.10 , with the parametric ones also provides a quick reference to check for the presence of homotheticity in the data set. If homotheticity were present, i.e. $F(L) = L$, all curves would be close to each other, as it happens for the chemical and transportation industry in 1995 and petroleum and transportation industries in 2001. In any case, they are all strictly increasing functions in labor, implying a generalized homogeneous structure for M as conjectured. Figures 2.7 and 2.11 show parametric and nonparametric fits of the unknown link function H, obtained by a local linear regression of $\widehat{r}$ on $\widehat{M}$ with a normal kernel and bandwidth, h_*, given by Silverman's rule. They also show fits from the unconstrained estimator of the function r used in the construction of our estimator in the first stage for each (k, L) for which $\widehat{M}$ was calculated. The nonparametric fits of r and those of H are quite similar in all industries and years, indicating that the imposition of generalized homotheticity is reasonable for these industries. The parametric fits are also broadly similar to the nonparametric ones, but showing more curvature in 2001 for the chemical and iron industries. These also show the design densities at the bottom of each plot.

[8]It was a similar observation by Cobb and Douglas (1928) that motivated the use of homogeneous functions in production theory, see Douglas (1967).

[9]The means of the observed ranges were subtracted from both sets of curves, before plotting.

Specification Test

We are interested in testing our proposed nonparametric generalized homothetic specification within our data set, that is

$$H_0 : r \in \overline{\mathcal{P}}\{r \in \mathcal{P} | \, r = H\left[G\left(k\right) F\left(L\right)\right] \text{ for some } H, \, G \text{ and } F\}\,.$$

Given $\widehat{H}$, $\widehat{G}$ and $\widehat{F}$, the implied restricted estimator of the regression surface is $\widehat{r}\left(k,L\right) = \widehat{H}[\widehat{G}\left(k\right)\widehat{F}\left(L\right)]$. As before, we employ a U–statistic based test as suggested in Fan and Li (1996). That is,

$$U_{NP} = \frac{1}{\lambda^2 n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(Y_i - \widehat{r}\left(k_i, L_i\right)\right)\left(Y_j - \widehat{r}\left(k_j, L_j\right)\right) K\left(\frac{k_j - k_i}{\lambda}\right) K\left(\frac{L_j - L_i}{\lambda}\right),$$

which under the null hypothesis that the generalized homothetic specification proposed in this chapter is correct,

$$n\lambda U_{NP} \sim N\left(0, 2\int K^2\left(u\right) du \int \left[\sigma_r^2\left(k,L\right) p\left(k,L\right)\right]^2 dkdL\right).$$

The results are as follows:

Industry	1995			2001		
	λ	U_{NP}	p-value	λ	U_{NP}	p-value
Chemical	5.125	-0.7702	0.7794	2.25	-0.1448	0.5575
Iron	4.250	-0.7578	0.7757	4	-0.2274	0.5899
Petroleum	2.750	-0.8458	0.8012	11	-0.7334	0.7684
Transportation	1.750	-0.6472	0.7413	4.37	-0.7047	0.7595

As in the parametric case, at all levels of significance, we fail to reject the hypothesis that our specification is a correct nonparametric simplification of the data for all industries and years. As both parametric models are special cases of the transformed partly separable model that was fitted, this result is as expected.

Substitutability and Returns to Scale

Given a generalized homothetic production function $E\left[Y|\, k, L\right] = r\left(k, L\right) = H\left[M\left(k, L\right)\right]$, important properties of production are measures of substitutability of inputs and the elasticity of substitution. A standard measure of the substitutability of inputs for production is the Technical Rate of Substitution, σ^*, defined as the slope of the isoquants in Figure 2.3, that is, $\sigma^* = -\left(dK/dL\right)|_{r(k,L)=r}$, for some constant level of output r. For an arbitrary production function $r\left(k, L\right)$, consider an alternative measure of input substitutability given by

$$T(k,L) \equiv \ln\left(\frac{\sigma^*}{k}\right)$$
$$= \ln\left(\frac{\partial r(k,L)}{\partial \ln L}\right) - \ln\left(\frac{\partial r(k,L)}{\partial \ln k}\right). \tag{2.5.9}$$

Moreover, if $r(k,L)$ is generalized homothetic (NP), $T(k,L) = \ln(\partial \ln F(L)/\partial \ln L) - \ln(\partial \ln G(k)/\partial \ln k)$. For parametric model (P2), this is $T(k,L) = \ln(L/(L+\gamma)) - \ln(\alpha)$, and $T(k,L) = -\ln(\alpha)$ for the Translog model (P3). We use, for the nonparametric model (NP), the approximation $\widehat{T}(k,L) = \ln[(\ln \widehat{F}(L_j) - \ln \widehat{F}(L_{j-1}))/(\ln L_j - \ln L_{j-1})] - \ln[(\ln \widehat{G}(k_j) - \ln \widehat{G}(k_{j-1}))/(\ln k_j - \ln k_{j-1})]$ after ordering the estimation grid points j, and approximations for (P2) and (P3) were obtained by replacing unknown quantities with their parametric estimates. Table 2.6 provides their averages along with standard deviations in parenthesis.

Another property of production that is empirically important, is economies of scale, defined as $\varepsilon^*(K,L) = (\partial r^*(cK,cL)/\partial \ln c)|_{c=1}$, which by (2.5.3), it simplifies to $\varepsilon(k,L) = \partial r(k,L)/\partial \ln L$. If $r(k,L)$ is generalized homothetic, then $\varepsilon(k,L) = RTS(M(k,L),L)$, where

$$RTS(M,L) = \frac{\partial H(M)}{\partial \ln M}\frac{\partial \ln F(L)}{\partial \ln L}. \tag{2.5.10}$$

For model (P2), this is $RTS(M,L) = [\beta_1 + 2\beta_2 \ln(M)](L/(L+\gamma))$, and $RTS(M,L) = \beta_1 + 2\beta_2 \ln(M)$ for the Translog model (P3). They were calculated by replacing the unknown parameters with their respective parametric estimates. In the nonparametric model this measure is estimated as $\widehat{RTS}(M,L) = [(\widehat{H}(\widehat{M}_j) - \widehat{H}(\widehat{M}_{j-1}))/(\ln \widehat{M}_j - \ln \widehat{M}_{j-1})] \times [(\ln \widehat{F}(L_j) - \ln \widehat{F}(L_{j-1}))/(\ln L_j - \ln L_{j-1})]$, using the same ordering as before. Table 2.7 provides summary statistics for all 4 industries in both years.

Calculating these measures in our data set have generated mixed results. Both parametric estimates have similar $T(k,L)$ in average each year, but roughly differ from those predicted by the nonparametric fit, which show a sizeable increase in 2001 relative to their values in 1995 for all sectors but the petroleum industry (which may be caused by the small number of observations for this industry in 1995, $n = 93$). This industry is also the only one for which these three averages coincide in 2001, because of the closeness of the parametric models to the nonparametric fit in Figures 2.9 and 2.10. However, all models show a reduction in economies of scales for all industries between 1995 and 2001. Although average increasing returns to scale, $RTS(M,L) > 1$, are predicted for some sectors in 1995, none is present in 2001. In fact, the chemical and iron industries seem to have decreasing returns to scale, $RTS(M,L) < 1$, in 2001, while the remaining sectors report approximately constant returns to scale, $RTS(M,L) \simeq 1$, in the same year.

In view of our various tests and the shape of the nonparametric estimates of G and F, the findings for $T(k, L)$ may simply be the result of misspecification while constructing the estimated averages in (2.5.10) for both parametric models. Economically, homotheticity (P3) generalizes the idea that pure economic profit will be zero. Since this situation is descriptive of the long–run equilibrium under perfect competition, another possible explanation may be the substantial ownership reform during this period of time. While many larger firms in the Chinese industrial sectors may have been state-owned in 1995, many other enterprizes were open to foreign capitals after 1995 and so could have substantially restructured, and others may have been created as well, thereby enhancing their productivity and increasing competition. For instance, we believe that this along with the hetereogenity regarding different firms specializing in different products in each industry, may explain why combining them into a single cross section might then create the appearance of decreasing returns on average in all models for two industries in 2001. However, these changes over time may more generally be due to changes in technology, demand, and other aspects of China's increasing economic liberalization and growth over this time period.

Another possible explanation is the likelihood that firms with positive productivity shocks may respond by using more inputs, i.e. endogeneity. In the next section, we explain how our estimator can be adapted in order to deal with this potential problem in a more general framework.

2.6 Conclusion and Extensions

We provide a general nonparametric estimator for a transformed partly additive or multiplicative separable model. This type of functional structure is shared by many popular empirical models implied by economic theory. An estimation algorithm is also proposed that does not require any maximization or matching. The resulting estimators are shown to have pointwise asymptotically normal distributions. Their rates of convergence are faster than those of a fully nonparametric alternative. We also provide an empirical application of our proposed methodology to estimation and testing of generalized homothetic production functions. We conclude by describing the following natural extensions.

Additional Regressors

Consider identification of $G(x)$ and $F(z)$, in the model $r(x, z, w) = H[M(x, z), w] \equiv H[G(x) + F(z), w]$, where H is strictly monotonic (and therefore invertible) on its first element, $(x, z) \in \Re^{d+1}$, and $w \in \Psi_W \subseteq \Re^{d_w}$ is a vector of additional regressors. It is straight forward to extend Theorem 2.2.1 or Corollary 2.2.2 in these cases. Specifically, let $s(x,z,w) \equiv \partial r(x,z,w)/\partial z$, and define the function $q(t,z,w)$ by $q(t,z,w) = E[s(X,Z,W)|\, r(X,Z,W) = t, Z = z, W = w]$. Then, the desired identification is achieved by replacing $q(t,z_0)$ in (2.2.2) or (2.2.3) by $q(t,z_0,w)$. For example, in the additive case, $s(x,z,w) = h[M(x,z), w]$, where h now represents the first derivative of H with respect to its first argument, and consequently $q(r,z) = h\left[H^{-1}(r,w), w\right] f(z)$. It follows that

$$\begin{aligned}
\int_{r_{0,w}}^{r(x,z,w)} \frac{dt}{q(t,z_0,w)} &= \int_{r_{0,w}}^{r(x,z,w)} \frac{dt}{h\left[H^{-1}(t,w), w\right] f(z_0)} \\
&= \int_{H^{-1}[r_{0,w},w]}^{H^{-1}[r(x,z,w),w]} \frac{h(m,w)\, dm}{h(m,w) f(z_0)} \\
&= \left(H^{-1}\left[r(x,z,w), w\right] - H^{-1}\left[r_{0,w}, w\right]\right)(1/f(z_0)) \\
&= \left(H^{-1}\left[H\left[M(x,z), w\right], w\right]\right) \equiv M(x,z),
\end{aligned}$$

where the second equality follows from the change of variables $m = H^{-1}(t,w)$, so $dt = h(m,w)\, dm$, and the last equality follows after assuming that $f(z_0) = 1$ and that $r_0 = H[0,w]$ for all w. This result holds for all $w \in \Psi_W$ and (x,z), so it holds in expectation replacing w with $\widetilde{W}$, thereby yielding

$$M(x,z) = E\left[\int_{r_{0,\widetilde{W}}}^{r(x,z,\widetilde{W})} \frac{dt}{q(t,z_0,\widetilde{W})}\right].$$

A consistent estimator of $M(x,z)$ and therefore, by virtue of marginal integration, of $G(x)$ and $F(z)$, is then given by

$$\widehat{M}(x,z) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{r_0}^{\widehat{r}(x,z,W_i)} \frac{dt}{\widehat{q}(t,z_0,W_i)}\right], \tag{2.6.1}$$

and a consistent estimator of h is then given by a nonparametric regression of $\widehat{r}(x,z,w)$ on $(\widetilde{M}(x,z), w)$, as before. The asymptotic properties of these estimators can be analyzed using similar tools as in Lewbel and Linton (2006).

Endogenous Regressors

Now consider estimation of $M(x,z) \equiv G(x) + F(z)$ in the model $y = H^*[M(x,z), \varepsilon]$ where ε is now unobserved and H^* is strictly monotonic in its first argument. If $\varepsilon \perp\!\!\!\perp (X,Z)$, then $r(x,z) = E[Y|\, X = x, Z = z] = H[M(x,z)]$, and our estimator can be applied. However, when some of the covariates (X,Z) are endogenous, and so correlated with ε, estimation of $M(x,z)$ is still possible, under the following conditions. For an observed vector T of exogenous covariates[10], which may include exogenous elements of (X,Z), define $m_x(t) = E[X|T=t]$, $U_x = X - m_x(t)$, $m_z(t) = E[Z|T=t]$, $U_z = Z - m_z(t)$ and let $U = (U_x, U_z)$. Then by construction $\varepsilon|X,Z,T \sim \varepsilon|U,T$. Define $r(x,z,u) \equiv E[Y|X=x, Z=z, U=u]$ and $H[M(x,z),u] = E[H^*[M(x,z),\varepsilon]|X=x, Z=z, T=t, U=u]$. If we then assume that

$$\varepsilon|U,T \sim \varepsilon|U \tag{2.6.2}$$

the form of endogeneity analyzed in the control function models of Blundell and Powell (2003)[11], it then follows that $r(x,z,u) = H[M(x,z),u]$. If U were observed, then the estimator proposed in (2.6.1) could be used by redefining W as U. Otherwise, U must be estimated. That is, we first estimate $\widehat{m}_x(T_i)$ and $\widehat{m}_z(T_i)$ by nonparametric regressions of X and Z on T respectively. We compute $\widehat{r}$ as a nonparametric regression of Y on $(X, Z, \widehat{U}_i)$. Then we construct (2.6.1) by replacing W_i everywhere with $\widehat{U}_i = (\widehat{U}_{x,i}, \widehat{U}_{z,i})$, where $\widehat{U}_{x,i} = X_i - \widehat{m}_x(T_i)$, $\widehat{U}_{z,i} = Z_i - \widehat{m}_z(T_i)$. Consistency of the resulting estimator of the functions M and H will follow from uniform consistency of the nonparametric estimators involved.

Recovery of the function H^* will in general require some additional structure. Once $M(x,z)$ is known, it can be treated as an observed endogenous regressor, and estimation of H^* (or any identifiable functional of H^* that are of applied interest) then reduces to estimation of a nonparametric triangular system. Examples of estimators of such systems are Blundell and Powell (2003), Imbens and Newey (2002) and Chesher (2001).

Further Testing

In production theory, homotheticity of production functions can be assessed by comparing the estimated component $\widehat{F}$ with the parametric model, $F(L) = L$. The assumed separability may also be tested by comparing the unrestricted nonparametric estimator $\widehat{r}(k,L)$ with the implied estimator for r given by the proposed structure, that is, $\widehat{r}(\widehat{M}(k,L))$. Such tests can be performed as in Gozalo and Linton (2001), by using asymptotic critical values or by direct implementation of their bootstrap procedure. Nonetheless, their theoretical justification in our framework would require considerable further work, and it remains a topic of future research.

[10] In production theory, they could include investment as in Olley and Pakes (1996), or intermediate inputs as suggested by Levinsohn and Petrin (2003).

[11] Assumption (2.6.2) also yields a nonparametric triangular system similar to Newey, Powell, and Vella (1999) and Imbens and Newey (2002)

Appendix

2.A Main Proofs

Preliminaries

We use the notation as well as the general approach introduced by Masry (1996b). For the sample $\{Y_i, X_i, Z_i\}_{i=1}^{n}$, let $W_i = \left(X_i^{\top}, Z_i\right)^{\top}$ so we obtained the p_1-th order local polynomial regression of Y_i on W_i by minimizing

$$Q_{r,n}(\theta) = n^{-1} h_1^{-(d+1)} \sum_{i=1}^{n} K_1\left(\frac{W_i - w}{h_1}\right) \left[Y_i - \sum_{0 \leq |\mathrm{j}| \leq p_1} \theta_{\mathrm{j}} \left(W_i - w\right)^{\mathrm{j}}\right]^2, \tag{2.A.1}$$

where the first element in θ denotes the minimizing intercept of (2.A.1), θ_0, and

$$\theta_{\mathrm{j}} = \frac{1}{\mathrm{j}!} \frac{\partial^{|\mathrm{j}|} r(w)}{\partial^{j_1} w_1 \cdots \partial^{j_d} w_d \partial^{j_{d+1}} w_{d+1}}.$$

We also use the following conventions:

$$\mathrm{j} = (j_1, \ldots, j_d, j_{d+1})^{\top}, \quad \mathrm{j}! = j_1! \times \ldots \times j_d \times j_{d+1}!, \quad |\mathrm{j}| = \sum_{k=1}^{d+1} j_k$$

$$a^{\mathrm{j}} = a_1^{j_1} \times \ldots \times a_d^{j_d} \times a_{d+1}^{j_{d+1}}$$

$$\sum_{0 \leq |\mathrm{j}| \leq p_1} = \sum_{k=0}^{p_1} \underset{j_1 + \ldots + j_d + j_{d+1} = k}{\sum_{j_1=0}^{k} \cdots \sum_{j_d=0}^{k} \sum_{j_{d+1}=0}^{k}}$$

where $w = (x^{\top}, z)^{\top}$. Let $N_{r,(l)} = (l + k - 1)!/\left(l!\,(k-1)!\right)$ be the number of distinct k–tuples j with $|\mathrm{j}| = l$, where $k = d + 1$. After arranging them in the corresponding lexicographical order, we let ϕ_l^{-1} denote this one–to–one map. For each j with $0 \leq |\mathrm{j}| \leq 2p_1$, let

$$\mu_{\mathbf{j}}(K_1) = \int_{\Re^{d+1}} u^{\mathbf{j}} K_1(u)\,du,$$

$$\gamma_{\mathbf{j}}(K_1) = \int_{\Re^{d+1}} u^{\mathbf{j}} K_1^2(u)\,du,$$

$$\gamma^1_{\mathbf{k},\mathbf{l}}(K_1) = \int_{\Re^d}\int_{\Re}\int_{\Re} (u_d, u_1)^{\mathbf{k}} (u_d, \widetilde{u}_1)^{\mathbf{l}} K_1(u_d, u_1) K_1(u_d, \widetilde{u}_1)\,du_1 d\widetilde{u}_1, \text{ and}$$

$$\gamma^2_{\mathbf{k},\mathbf{l}}(K_1) = \int_{\Re}\int_{\Re^d}\int_{\Re^d} (u_d, u_1)^{\mathbf{k}} (\widetilde{u}_d, u_1)^{\mathbf{l}} K_1(u_d, u_1) K_1(\widetilde{u}_d, u_1)\,du_d d\widetilde{u}_d,$$

where u_d and u_1 represent the first d and last element of the $d+1$ vector u respectively. Define the $N_r \times N_r$ dimensional matrices $\mathbf{M}_r$ and $\mathbf{\Gamma}_r$, and the $N_r \times N_{r,(p_1+1)}$ matrix $\mathbf{B}_r$ by

$$\mathbf{M}_r = \begin{bmatrix} \mathbf{M}_{r;0,0} & \mathbf{M}_{r;0,1} & \cdots & \mathbf{M}_{r;0,p_1} \\ \mathbf{M}_{r;1,0} & \mathbf{M}_{r;1,1} & \cdots & \mathbf{M}_{r;1,p_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{r;p_1,0} & \mathbf{M}_{r;p_1,1} & \cdots & \mathbf{M}_{r;p_1,p_1} \end{bmatrix},$$

$$\mathbf{\Gamma}_r = \begin{bmatrix} \mathbf{\Gamma}_{r;0,0} & \mathbf{\Gamma}_{r;0,1} & \cdots & \mathbf{\Gamma}_{r;0,p_1} \\ \mathbf{\Gamma}_{r;1,0} & \mathbf{\Gamma}_{r;1,1} & \cdots & \mathbf{\Gamma}_{r;1,p_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{r;p_1,0} & \mathbf{\Gamma}_{r;p_1,1} & \cdots & \mathbf{\Gamma}_{r;p_1,p_1} \end{bmatrix}, \mathbf{B}_r = \begin{bmatrix} \mathbf{M}_{r;0,p_1+1} \\ \mathbf{M}_{r;1,p_1+1} \\ \vdots \\ \mathbf{M}_{r;p_1,p_1+1} \end{bmatrix} \tag{2.A.2}$$

where $N_r = \sum_{l=0}^{p_1} N_{r,(l)}$, $\mathbf{M}_{r;i,j}$ and $\mathbf{\Gamma}_{r;i,j}$ are $N_{r,(i)} \times N_{r,(j)}$ dimensional matrices whose (l,m) elements are $\mu_{\phi_i(l)+\phi_j(m)}$ and $\gamma_{\phi_i(l),\phi_j(m)}$ respectively. $\mathbf{\Gamma}^1_r$ and $\mathbf{\Gamma}^2_r$ are defined similarly by the $N_{r,(i)} \times N_{r,(j)}$ matrices $\mathbf{\Gamma}^1_{r;i,j}$, $\mathbf{\Gamma}^2_{r;i,j}$, whose (l,m) elements are given by $\gamma^1_{\phi_i(l),\phi_j(m)}$ and $\gamma^2_{\phi_i(l),\phi_j(m)}$ respectively. The elements of $\mathbf{M}_r = \mathbf{M}_r(K_1, p_1)$ and $\mathbf{B}_r = \mathbf{B}_r(K_1, p_1)$ are simply multivariate moments of the kernel K_1.

Similarly, for the generated sub-sample set $\{\widehat{s}(X_i, Z_i), \widehat{r}(X_i, Z_i), Z_i\}_{i=1}^n$, an estimator of the function q, defined as $q(t,z) = E[S|\, r(X,Z) = t, Z = z]$, is obtained by the intercept of the following minimizing problem,

$$Q_{q,n}(\theta) = n^{-1} h_2^{-2} \sum_{i=1}^n K_2\left(\frac{\widehat{V}_i - v}{h_2}\right)\left[\widehat{S}_i - \sum_{0 \le |\mathbf{j}| \le p_2} \theta_{\mathbf{j}} \left(\widehat{V}_i - v\right)^{\mathbf{j}}\right]^2,$$

where $\widehat{V}_i = (\widehat{r}_i, Z_i)^\top$ and $v = (t,z)^\top$, define $V_i = (r_i, Z_i)^\top$ accordingly. Let $N_{q,(l)} = (l+k-1)!/(l! \times (k-1)!)$ be the number of distinct k−tuples $\mathbf{j}$ with $|\mathbf{j}| = l$, where $k = 2$. After arranging them in the corresponding lexicographical order, we let ϕ_l^{-1} denote this one–to–one map. For each $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p_2$, let

$$\mu_{\mathbf{j}}(K_2) = \int_{\Re^2} u^{\mathbf{j}} K_2(u)\, du, \text{ and}$$
$$\gamma_{\mathbf{j}}(K_2) = \int_{\Re^2} u^{\mathbf{j}} K_2^2(u)\, du.$$

Define the $N_q \times N_q$ dimensional matrices $\mathbf{M}_q$ and $\mathbf{\Gamma}_q$, and the $N_q \times N_{q,(p_2+1)}$ matrix $\mathbf{B}_q$ by

$$\mathbf{M}_q = \begin{bmatrix} \mathbf{M}_{q;0,0} & \mathbf{M}_{q;0,1} & \cdots & \mathbf{M}_{q;0,p_2} \\ \mathbf{M}_{q;1,0} & \mathbf{M}_{q;1,1} & \cdots & \mathbf{M}_{q;1,p_2} \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{q;p_2,0} & \mathbf{M}_{q;p_2,1} & \cdots & \mathbf{M}_{q;p_2,p_2} \end{bmatrix},$$
$$\mathbf{\Gamma}_q = \begin{bmatrix} \mathbf{\Gamma}_{q;0,0} & \mathbf{\Gamma}_{q;0,1} & \cdots & \mathbf{\Gamma}_{q;0,p_2} \\ \mathbf{\Gamma}_{q;1,0} & \mathbf{\Gamma}_{q;1,1} & \cdots & \mathbf{\Gamma}_{q;1,p_2} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{q;p_2,0} & \mathbf{\Gamma}_{q;p_2,1} & \cdots & \mathbf{\Gamma}_{q;p_2,p_2} \end{bmatrix}, \mathbf{B}_q = \begin{bmatrix} \mathbf{M}_{q;0,p_2+1} \\ \mathbf{M}_{q;1,p_2+1} \\ \vdots \\ \mathbf{M}_{q;p_2,p_2+1} \end{bmatrix} \tag{2.A.3}$$

where $N_q = \sum_{l=0}^{p_2} N_{q,(l)}$, $\mathbf{M}_{q;j,k}$ and $\mathbf{\Gamma}_{q;j,k}$ are $N_{q,(j)} \times N_{q,(k)}$ dimensional matrices whose (l,m) elements are $\mu_{\phi_{q;j}(l)+\phi_{q;k}(m)}$ and $\gamma_{\phi_{q;j}(l),\phi_{q;k}(m)}$ respectively. The elements of $\mathbf{M}_q = \mathbf{M}_q(K_2,p_2)$ and $\mathbf{B}_q = \mathbf{B}_q(K_2,p_2)$ are simply multivariate moments of the kernel K_2. To facilitate the proof, let $\mathcal{K}_{2,i}(v)$ be a $N_q \times 1$ vector, $\mathcal{K}_{2,i}^{(1)}(v)$ be a $N_q \times 2$ matrix, and $\mathbf{M}_{q,n}(v)$ be a symmetric $N_q \times N_q$ matrix such that

$$\mathcal{K}_{2,i}(v) = \begin{bmatrix} \mathcal{K}_{2,i;0}(v) \\ \mathcal{K}_{2,i;1}(v) \\ \vdots \\ \mathcal{K}_{2,i;p_2}(v) \end{bmatrix}, \mathcal{K}_{2,i}^{(1)}(v) = \begin{bmatrix} \mathcal{K}_{2,i;0}^{(1)}(v) \\ \mathcal{K}_{2,i;1}^{(1)}(v) \\ \vdots \\ \mathcal{K}_{2,i;p_2}^{(1)}(v) \end{bmatrix} \tag{2.A.4}$$
$$\mathbf{M}_{q,n}(v) = \begin{bmatrix} \mathbf{M}_{q,n;0,0}(v) & \mathbf{M}_{q,n;0,1}(v) & \cdots & \mathbf{M}_{q,n;0,p_2}(v) \\ \mathbf{M}_{q,n;1,0}(v) & \mathbf{M}_{q,n;1,1}(v) & \cdots & \mathbf{M}_{q,n;1,p_2}(v) \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{q,n;p_2,0}(v) & \mathbf{M}_{q,n;p_2,1}(v) & \cdots & \mathbf{M}_{q,n;p_2,p_2}(v) \end{bmatrix},$$

where $\mathcal{K}_{2,i;l}(v)$ is a $N_{q,(l)} \times 1$ dimensional subvector whose l^0–th element is given by $[\mathcal{K}_{2,i;l}(v)]_{l^0} = ((V_i - v)/h_2)^{\phi_{q;l}(l^0)} K_2((V_i - v)/h_2)$. The $N_{q,(l)} \times 1$ matrix $\mathcal{K}_{2,i;l}^{(1)}(v)$ has l^0 element being the partial derivative of $[\mathcal{K}_{2,i;l}(t,z)]_{l^0}$ with respect to r, and $\mathbf{M}_{q,n;j,k}(v)$ is a $N_{q,(j)} \times N_{q,(k)}$ dimensional submatrix with the (l,l^0) element given by

$$[\mathbf{M}_{q,n;j,k}(v)]_{l,l^0} = \frac{1}{nh_2^2} \sum_{i=1}^{n} \left(\frac{V_i - v}{h_2}\right)^{\phi_{q;j}(l)+\phi_{q;k}(l^0)} K_2\left(\frac{V_i - v}{h_2}\right).$$

$\widehat{\mathcal{K}}_{2,i}(v)$ and $\widehat{\mathbf{M}}_{q,n}(v)$ are defined similarly as $\mathcal{K}_{2,i}(v)$ and $\mathbf{M}_{q,n}(v)$ respectively, but with the generated regressors $\{\widehat{r}_i\}_{i=1}^n$ in place of the unobserved variables $\{r_i\}_{i=1}^n$. Let us define the functions $\widetilde{\mathcal{K}}_{2,i}(z) = \int h_2^{-1}\mathcal{K}_{2,i}(t,z)\,dt$ and $\zeta(t,z) = \partial\left[f_V(t,z)\,q^2(t,z)\right]^{-1}/\partial t$, which are well defined given Assumptions (E1) and (E2). Thus, by integration by parts, it follows that

$$\begin{aligned}\int_{r_0}^{r(x,z)} h_2^{-1}\mathcal{K}_{2,i}^{(1)}(t,z)\left[f_V(t,z)\,q^2(t,z)\right]^{-1}dt &= \Big\{\mathcal{K}_{2,i}(r,z)\left[f_V(r,z)\,q^2(r,z)\right]^{-1} - \mathcal{K}_{2,i}(r_0,z)\times \\ &\qquad \left[f_V(r_0,z)\,q^2(r_0,z)\right]^{-1}\Big\} - \int_{r_0}^{r(x,z)}\mathcal{K}_{2,i}(t,z)\,\zeta(t,z)\,dt \\ &\equiv \varrho_{i,1}^0 - \varrho_{i,2}^0. \end{aligned} \tag{2.A.5}$$

Similarly, let us define $dQ(t) = 1\,(r_0 \le t \le r(x,z))\,dt$, so we write

$$\int h_2^{-1}\mathcal{K}_{2,i}(t,z)\left[f_V(t,z)\,q^2(t,z)\right]^{-1}dQ(t) \equiv \varrho_{i,1}^1 - \varrho_{i,2}^1,$$

where $\varrho_{i,1}^1$ and $\varrho_{i,2}^1$ are like $\varrho_{i,1}^0$ and $\varrho_{i,2}^0$ in (2.A.5), but with $\mathcal{K}_{2,i}^1(r,z)$ replacing $\mathcal{K}_{2,i}(r,z)$, where $\mathcal{K}_{2,i}^1(r,z) = \int_{-\infty}^{r}\mathcal{K}_{2,i}(s,z)\,ds$, a $N_q \times 1$ vector with well-defined functions as elements by virtue of Assumption (E1). Furthermore, $n^{-1}h_2^2\sum_{i=1}^n\mathcal{K}_{2,i}^1(r,z)$ converges to $\mathbf{M}_{q,0}^1 f_V(r,z)$ in mean squared, where $\mathbf{M}_{q,0}^1$ is a $N_q \times 1$ vector with l_0 element given by $\int u^{\phi_{q\cdot 1}(l^0)}K_2^1(u)\,du$, and $K_2^1(u) = \int_{-\infty}^{u}K_2(v)\,dv$. Similarly, $n^{-1}h_2^2\sum_{i=1}^n\mathcal{K}_{2,i}(r,z)$ converges in mean squared to $\mathbf{M}_{q,0}^0 f_V(r,z)$.

Let also arrange the $N_{r,(m)}$ and $N_{q,(m)}$ elements of the derivatives

$$D^{\mathbf{m}}r(w) \equiv \frac{\partial^m r(w)}{\partial^{m_1}w_1,\ldots,\partial^{m_k}w_k},\quad D^{\mathbf{m}}q(v) \equiv \frac{\partial^m q(v)}{\partial^{m_1}v_1,\ldots,\partial^{m_k}v_k},\quad \text{for } |\mathbf{m}| = m$$

as the $N_{r,(m)} \times 1$ and $N_{q,(m)} \times 1$ column vectors $r^{(m)}(w)$ and $q^{(m)}(v)$ in the lexicographical order mentioned above.

Let $\iota_1 = (1,0,\ldots,0)^\top \in \Re^{N_r}$ and $\iota_1^* = (0,1,0,\ldots,0)^\top \in \Re^{N_r}$, then by equation (2.13) (page 574) and Corollary 2(ii) (page 580) in Masry (1996a), we write

$$\begin{aligned}\widehat{r}(w) - r(w) &= \iota_1^\top\left[\mathbf{M}_r f(w)\right]^{-1}\{1 + o_p(1)\} \\ &\quad\times\left\{n^{-1}h_1^{-(d+1)}\sum_{j=1}^n\mathcal{K}_{1,j}(w)\left[\varepsilon_{r,j} + \sum_{|\mathbf{k}|=p_1+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}r(w)\,(W_i - w)^{\mathbf{k}}\right] + \gamma_n(w)\right\},\end{aligned} \tag{2.A.6}$$

$$\widehat{s}(w) - s(w) = h_1^{-1}\iota_1^{*\top}\left[\mathbf{M}_r f(w)\right]^{-1}\{1+o_p(1)\}$$
$$\times\left\{n^{-1}h_1^{-(d+1)}\sum_{j=1}^{n}\mathcal{K}_{1,j}(w)\left[\varepsilon_{r,j} + \sum_{|\mathbf{k}|=p_1+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}r(w)(W_i - w)^{\mathbf{k}}\right] + \gamma_n(w)\right\} \tag{2.A.7}$$

uniformly in w, where

$$\gamma_n(w) \equiv (p_1+1)\,n^{-1}h_1^{-(d+1)}\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}\mathcal{K}_{1,j}(w)(W_j - w)^{\mathbf{k}}$$
$$\times\int_0^1\{D^{\mathbf{k}}r(w+\tau(W_i - w)) - D^{\mathbf{k}}r(w)\}(1-\tau)^{p_1}\,d\tau.$$

As before $\mathcal{K}_{1,i}(w)$, a $N_r\times 1$ dimensional vector, is defined analogously as $\mathcal{K}_{2,i}(v)$ in (2.A.4), with a $N_{r,(l)}\times 1$ dimensional subvector with l^0–th element given by $[\mathcal{K}_{1,i;l}(w)]_{l^0} = ((W_i - w)/h_1)^{\phi_{r;l}(l^0)}K_1((W_i - w)/h_1)$, such that $n^{-1}h_1^{-(d+1)}\sum_{j=1}^{n}\mathcal{K}_{1,j}(w)$ converges in mean squared to $\mathbf{M}_{r,0}f_W(w)$. Define $\gamma(w) = E[\gamma_n(w)]$, then by Proposition 2 (page 581) and by Theorem 4 (page 582) in Masry (1996a), it follows that

$$\sup_w|\gamma(w)| = o(h_1^{p_1+1}),$$
$$\sup_w|h_1^{-(p_1+1)}\gamma_n(w) - \gamma(w)| = h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}). \tag{2.A.8}$$

Let

$$\beta_n(w) \equiv n^{-1}h_1^{-(d+1)}\sum_{j=1}^{n}\mathcal{K}_{1,j}(w)\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}D^{\mathbf{k}}r(w)(W_i - w)^{\mathbf{k}},\ \text{and}$$
$$\beta(w) = \mathbf{B}_r r^{(p_1+1)}(w)f_W(w),$$

then by Theorem 2 (page 579) in Masry (1996a), it follows that

$$\sup_w|h_1^{-(p_1+1)}\beta_n(w) - \beta(w)| = O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}). \tag{2.A.9}$$

For the set $\{Y_i,\widetilde{M}_i\}_{i=1}^n$, as discussed in the main text, an estimator of the function H is obtained by the intercept of the following minimizing problem

$$Q_{H,n}(\theta) = n^{-1}h_*^{-1}\sum_{i=1}^{n}k_*\left(\frac{\widetilde{M}_i - m}{h_*}\right)\left[Y_i - \sum_{0\le j\le p_*}\theta_j\left(\widetilde{M}_i - m\right)^j\right]^2.$$

Because this is a simple univariate nonparametric regression, its associated matrices $\mathbf{M}_H$, $\mathbf{M}_{H,0}^0$, $\mathbf{\Gamma}_H$, $\mathbf{B}_H$, $\mathbf{M}_{H,n}(m)$, $\widehat{\mathbf{M}}_{H,n}(m)$, and vector $\mathcal{K}_{*,i;l}(m)$ have simpler forms. They are as those previously described but replacing the responses by Y_i and the conditioning variables by M_i or $\widetilde{M}_i$ accordingly.

Proof of Corollary 2.2.2

As before, given Assumption I*, it follows that $s(x,z) = h[G(x)F(z)]G(x)f(z)$, consequently $q(t,z_0) = h\left[H^{-1}(t)\right]H^{-1}(t)\left[f(z_0)/F(z_0)\right]$, and using the change of variables $m = H^{-1}(t)$, after noticing that $h\left[H^{-1}(t)\right] = h(m)$ and $dt = h(m)\,dm$, we obtain

$$\begin{aligned}
\int_{r_1}^{r(x,z)} \frac{dt}{q(t,z_0)} &= \int_{r_1}^{r(x,z)} \frac{F(z_0)}{h\left[H^{-1}(t)\right]H^{-1}(t)\,f(z_0)}dt \\
&= \int_{H^{-1}(r_1)}^{H^{-1}(r(x,z))} \frac{F(z_0)}{h(m)\,mf(z_0)}h(m)\,dm \\
&= \left[\frac{F(z_0)}{f(z_0)}\right]\left[\ln\left(H^{-1}\left[r(x,z)\right]\right) - \ln\left(H^{-1}\left[r_1\right]\right)\right] \\
&= \ln\left(M(x,z)\right) \equiv \ln\left(G(x)F(z)\right).
\end{aligned}$$

This proves the result.

Proof of Theorem 2.4.1

Rearranging terms, we have

$$\begin{aligned}
\widehat{M}(x,z) - M(x,z) &= \int_{r_0}^{\widehat{r}(x,z)} \frac{dt}{\widehat{q}(t,z_0)} - \int_{r_0}^{r(x,z)} \frac{dt}{q(t,z_0)} \\
&= \left(\int_{r_0}^{\widehat{r}(x,z)} - \int_{r_0}^{r(x,z)}\right)\frac{dt}{q(t,z_0)} + \int_{r_0}^{\widehat{r}(x,z)}\left(\frac{\widehat{q}(t,z_0) - q(t,z_0)}{\widehat{q}(t,z_0)\,q(t,z_0)}\right)dt.
\end{aligned}$$

By mean value expansions of the first term, in the last equality above, and after some manipulation we obtain,

$$\widehat{M}(x,z) - M(x,z) \doteq \frac{1}{q(r,z_0)}(\widehat{r}(x,z) - r(x,z)) + \int_{r_0}^{r(x,z)} \frac{\widehat{q}(t,z_0) - q(t,z_0)}{q^2(t,z_0)}dt \quad (2.A.10)$$

$$+ \int_{r(x,z)}^{\widehat{r}(x,z)} \frac{\widehat{q}(t,z_0) - q(t,z_0)}{q^2(t,z_0)}dt - \int_{r_0}^{\widehat{r}(x,z)} \frac{(\widehat{q}(t,z_0) - q(t,z_0))^2}{\widehat{q}(t,z_0)\, q^2(t,z_0)}dt \quad (2.A.11)$$

$$\doteq \mathcal{M}_{1,n}(x,z) + \mathcal{M}_{2,n}(x,z) + \mathcal{R}_{M,n}(x,z). \quad (2.A.12)$$

The terms in (2.A.10), $\mathcal{M}_{1,n}(x,z)$ and $\mathcal{M}_{2,n}(x,z)$, are linear in the estimation error from the two nonparametric regressions, while the remaining terms in (2.A.11), $\mathcal{R}_{Mn}(x,z)$, are both quadratic in such errors, and thus they will be shown to be of smaller order. $\mathcal{M}_{1,n}(x,z)$ is just a constant times the estimation error of $\widehat{r}(x,z)$, the unconstrained first–stage nonparametric estimator of $r(x,z)$, and under Assumption E, it can be analyzed directly using Theorem 4 (page 94) in Masry (1996b), given that $q(r(x,z),z) > 0$ over $\Psi_x \times \Psi_z$. That is,

$$\sqrt{nh_1^{d+1}}\left(\mathcal{M}_{1,n}(x,z) - h_1^{p_1+1}\mathcal{B}_4(x,z)\right) \xrightarrow{d} N\left[0, \frac{\sigma_r^2(x,z)}{q^2(r,z_0)\, f_W(x,z)}\left[\mathbf{M}_r^{-1}\Gamma_r\mathbf{M}_r^{-1}\right]_{0,0}\right],$$

$$\mathcal{B}_4(x,z) = \left[\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(x,z)\right]_{0,0} q^{-1}(r,z_0) \quad (2.A.13)$$

where $[A]_{0,0}$ is the upper–left element of matrix A. In order to analyze the second term, $\mathcal{M}_{2,n}(x,z)$, we first notice that for any two symmetric nonsingular matrices A_1 and A_2, we have that $A_1^{-1} - A_2^{-1} = A_2^{-1}(A_2 - A_1)A_1^{-1}$, which implies

$$\begin{aligned}
\frac{\widehat{q}(t,z) - q(t,z)}{q^2(t,z)} &= \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widetilde{V}_{q,n}(v) + \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{B}_{q,n}(v) \\
&= \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{V}_{q,n}(v) + \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{V}_{q,n}^*(v) \\
&+ \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{B}_{q,n}(v) \\
&= \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{V}_{q,n}(v) + \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{V}_{q,n}^*(v) \\
&+ \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{B}_{q,n}(v) \\
&- \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v) - f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{V}_{q,n}(v) \\
&- \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v) - f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{V}_{q,n}^*(v) \\
&- \iota_2^\top\left[f_V(v)\, q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v) - f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{B}_{q,n}(v) \\
&\equiv T_{q,n,1}(v) + T_{q,n,2}(v) + T_{q,n,3}(v) - T_{q,n,4}(v) - T_{q,n,5}(v) - T_{q,n,6}(v)
\end{aligned}$$

where $\mathbf{M}_q$ is defined in (2.A.3). We have also defined $\widetilde{V}_{q,n}(v) = \widehat{V}_{q,n}(v) + \widehat{V}^*_{q,n}(v)$, where the $N_q \times 1$ vectors $\widehat{V}_{q,n}(v)$, $\widehat{V}^*_{q,n}(v)$, and $\widehat{B}_{q,n}(v)$ are

$$\widehat{V}_{q,n}(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\varepsilon_{q,i},$$
$$\widehat{V}^*_{q,n}(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,[\widehat{S}_i - S_i],$$
$$\widehat{B}_{q,n}(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\widehat{\Delta}_{q,i}(v)\,, \text{ and}$$
$$\widehat{\Delta}_{q,i}(v) \equiv q(\widehat{V}_i) - \sum_{0\leq|\mathbf{m}|\leq p_2}\frac{1}{\mathbf{m}!}(D^{\mathbf{m}}q)(v)\,(\widehat{V}_i - v)^{\mathbf{m}}.$$

Consequently,

$$\mathcal{M}_{2n}(x,z) = \mathcal{T}_{q,n,1}(x,z) + \mathcal{T}_{q,n,2}(x,z) + \mathcal{T}_{q,n,3}(x,z) + \mathcal{R}_{q,n}(x,z)\,,$$

where $\mathcal{T}_{q,n,l}(x,z) = \int T_{q,n,l}(t,z_0)\,dQ(t)$ for $l = 1,2,3$ and $dQ(t) = 1\,(r_0 \leq t \leq r(x,z))\,dt$. These terms, along with the remainder $\mathcal{R}_{q,n}(x,z) = \sum_{l=4}^{6}\int T_{q,n,l}(t,z_0)\,dQ(t)$ are dealt with in Lemmas 2.B.1 to 2.B.4, from which we conclude that

$$\begin{aligned}
\mathcal{M}_{2n}(x,z) &= h_1^{p_1+1}\iota_2^\top\mathbf{M}_q^{-1}\mathbf{M}_{q,0}^0\iota_1^\top\mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)g_q(X,Z)\middle|\,r(X,z_0)=r, Z=z_0\right]}{q^2(r,z_0)}\right.\\
&\qquad\left.-\frac{E\left[r^{(p_1+1)}(X,Z)g_q(X,Z)\middle|\,r(X,z_0)=r_0, Z=z_0\right]}{q^2(r_0,z_0)}\right]\\
&\quad + h_1^{p_1}h_2\iota_2^\top\mathbf{M}_q^{-1}\mathbf{M}_{q,0}^1\iota_1^\top\mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)\middle|\,r(X,z_0)=r, Z=z_0\right]}{q^2(r,z_0)}\right.\\
&\qquad\left.-\frac{E\left[r^{(p_1+1)}(X,Z)\middle|\,r(X,z_0)=r_0, Z=z_0\right]}{q^2(r_0,z_0)}\right]\\
&\quad + h_2^{p_2+1}\iota_2^\top\mathbf{M}_q^{-1}\mathbf{B}_q\int_{r_0}^{r(x,z)}\frac{q^{(p_2+1)}(t,z_0)}{q^2(t,z_0)}dt + o_p(n^{-1/2}h_1^{-(d+1)/2})\\
&= h_1^{p_1+1}\mathcal{B}_1(x,z) + h_1^{p_1}h_2\mathcal{B}_2(x,z) + h_2^{p_2+1}\mathcal{B}_3(x,z) + o_p(n^{-1/2}h_1^{-(d+1)/2}).
\end{aligned} \tag{2.A.14}$$

Finally, the last term in (2.A.12), $\mathcal{R}_{M,n}(x,z) = O_p(\nu_{1n})\,O_p(h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n}) + O_p((h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n})^2)$, by Theorem 6 (page 594) in Masry (1996a) and Lemma 2.B.5. Therefore, it follows from Assumption (E5) that $\mathcal{R}_{M,n}(x,z) = o_p(n^{-1/2}h_1^{(d+1)/2})$. By grouping terms, $\mathcal{B}_M(x,z) \equiv h_1^{p_1+1}\mathcal{B}_1(x,z) + h_1^{p_1}h_2\mathcal{B}_2(x,z) + h_2^{p_2+1}\mathcal{B}_3(x,z) + h_1^{p_1+1}\mathcal{B}_4(x,z)$, we conclude the proof of the theorem.

Proof of Theorem 2.4.3

As before, we write

$$\begin{aligned}
\widehat{H}(m) - H(m) &= \iota_*^\top \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widetilde{V}_{H,n}(m) + \iota_*^\top \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m) \\
&= \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \widetilde{V}_{H,n}(m) + \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \widehat{B}_{H,n}(m) \\
&- \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right] \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widetilde{V}_{H,n}(m) \\
&- \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right] \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m) \\
&= \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \widehat{V}_{H,n}(m) + \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \widehat{V}_{H,n}^*(m) \\
&+ \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \widehat{B}_{H,n}(m) \\
&- \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right] \widehat{\mathbf{M}}_{H,n}^{-1}(b)\widehat{V}_{H,n}(b) \\
&- \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right] \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{V}_{H,n}^*(m) \\
&- \iota_*^\top \left[f_M(m)\mathbf{M}_H\right]^{-1} \left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right] \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m) \\
&\equiv T_{H,n,1}(m) + T_{H,n,2}(m) + T_{H,n,3}(m) - T_{H,n,4}(m) - T_{H,n,5}(m) - T_{H,n,6}(m),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{V}_{H,n}(m) &\equiv \widehat{V}_{H,n}(m) + \widehat{V}_{H,n}^*(m), \\
\widehat{V}_{H,n}(m) &= n^{-1}h_*^{-1}\sum_{i=1}^n \widehat{\mathcal{K}}_{*,i}(m)\varepsilon_{r,i}, \\
\widehat{V}_{H,n}^*(m) &= n^{-1}h_*^{-1}\sum_{i=1}^n \widehat{\mathcal{K}}_{*,i}(m)[H(M_i) - H(\widetilde{M}_i)], \text{ and} \\
\widehat{B}_{H,n}(m) &= n^{-1}h_*^{-1}\sum_{i=1}^n \widehat{\mathcal{K}}_{*,i}(m)\widehat{\Delta}_{H,i}(m), \text{ with} \\
\widehat{\Delta}_{H,i}(m) &\equiv H(\widehat{M}_i) - \sum_{0\le j\le p_*}\frac{1}{j!}(\partial^j H\,(m)\,/\partial m^j)(\widehat{M}_i - m)^j.
\end{aligned}$$

We analyze the properties of $T_{H,n,l}(b)$, $l = 1,\ldots,6$ in Lemmas 2.B.7 to 2.B.10, which show that $T_{H,n,1}(m) = O_p(n^{-1/2}h_*^{-1/2})$ and that $T_{H,n,2}(m) \xrightarrow{p} \mathcal{B}_{H2}(m)$, $T_{H,n,3}(m) \xrightarrow{p} \mathcal{B}_{H3}(m)$, where

$$\begin{aligned}
\mathcal{B}_{H2}(m) &\equiv -\iota_*^\top \mathbf{M}_H^{-1}\mathbf{M}_{H,0}^0 E\left[H^{(1)}(M(X,Z))\,\beta(X,Z)\middle|\, H(M(X,Z)) = m\right], \\
\mathcal{B}_{H3}(m) &\equiv h_*^{p_*+1}\iota_*^\top \mathbf{M}_H^{-1}\mathbf{B}_H H^{(p_*+1)}(m),
\end{aligned}$$

with $\beta(w) \equiv \int \mathcal{B}(x,z)\,dP_1(z) + \int \mathcal{B}(x,z)\,dP_2(x) + \int\int \mathcal{B}(x,z)\,dP_1(z)\,dP_2(x)$ which is $O(h_\dagger)$ by construction. By defining $\mathcal{B}_H(m) \equiv \mathcal{B}_{H2}(m) + \mathcal{B}_{H3}(m)$, the proof is completed.

2.B Technical Lemmas

Lemma 2.B.1 *Under Assumption E, we have*

$$\sup_{t,z} |T_{q,n,1}(t,z)| = O_p\left(h_2^{-1}\nu_{1n} + n^{-1/2}h_2^{-1}\sqrt{\ln n}\right), \text{ and} \tag{2.B.1}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,1}(t,z)\,dt = O_p\left(n^{-1/2}h_1^{-(d+1)/2}\right). \tag{2.B.2}$$

Proof. We may rewrite

$$\widehat{V}_{q,n}(t,z) = nh_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\,\varepsilon_{q,i} + nh_2^{-2}\sum_{i=1}^{n}[\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)]\varepsilon_{q,i}.$$

Then, by Theorem 5 (page 593) in Masry (1996a), it follows that the first term is

$$nh_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\,\varepsilon_{q,i} = O_p\left(n^{-1/2}h_2^{-1}\sqrt{\ln n}\right).$$

In the other hand, after a Taylor–series expansion, the second term is bounded by

$$\begin{aligned}
&nh_2^{-2}\sum_{i=1}^{n}[\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)]\varepsilon_{q,i} \\
&\leq \{n^{-1}h_2^{-3}\sum_{i=1}^{n}|\mathcal{K}_{2,i}^{(1)}(t,z)|\,|\varepsilon_{q,i}|\}\max_{1\leq i\leq n}|\widehat{r}_i - r_i| + O_p\left(h_2^{-2}\nu_{1n}^2\right) \\
&= O_p(h_2^{-1}\nu_{1n}) + o_p(n^{-1/2}h_1^{-(d+1)/2}),
\end{aligned}$$

by Assumption (E5(ii)). After collecting terms, (2.B.1) follows.

By using (2.A.6), we further write

$$\widehat{V}_{q,n}(v) - V_{q,n}(v) = [V_{q,n,b}(v) + V_{q,n,c}(v) + V_{q,n,d}(v)]\,\{1 + o_p(1)\} + o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$V_{q,n,b}(v) \equiv n^{-2}h_1^{-(d+1)}h_2^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\alpha}_n(W_i, W_j; v),$$

$$V_{q,n,c}(v) \equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,\varepsilon_{q,i}\widetilde{\beta}_n(W_i), \text{ and}$$

$$V_{q,n,d}(v) \equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,\varepsilon_{q,i}\widetilde{\gamma}_n(W_i);$$

and, we define

$$\widetilde{\alpha}_n(W_i, W_j; v) \equiv \mathcal{K}_{2,i}^{(1)}(v)\,\iota_1^{\top}[\mathbf{M}_r f(W_i)]^{-1}\mathcal{K}_{1,j}(W_i)\,\varepsilon_{q,i}\varepsilon_{r,j},$$

$$\widetilde{\beta}_n(w) \equiv n^{-1}h_1^{-(d+1)}\iota_1^{\top}[\mathbf{M}_r f(w)]^{-1}\sum_{j=1}^{n}\mathcal{K}_{1,j}(w)\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}D^{\mathbf{k}}r(w)(W_i - w)^{\mathbf{k}}, \text{ and}$$

$$\widetilde{\gamma}_n(w) \equiv n^{-1}h_1^{-(d+1)}\iota_1^{\top}[\mathbf{M}_r f(w)]^{-1}\gamma_n(w).$$

Thus, we have

$$\int T_{q,n,1}(t,z)\,dQ(t) = \mathcal{T}_{q,n,1a} + (\mathcal{T}_{q,n,1b} + \mathcal{T}_{q,n,1c} + \mathcal{T}_{q,n,1d})\{1 + o_p(1)\} + o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$\mathcal{T}_{q,n,1a} = \frac{1}{nh_2}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\mathcal{K}_{2,i}(v)\left[f_V(v)\,q^2(v)\right]^{-1}dt,$$

$$\mathcal{T}_{q,n,1b} = \frac{1}{n^2h_1^{(d+1)}h_2^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\widetilde{\alpha}_n(W_i, W_j; v)\left[f_V(v)\,q^2(v)\right]^{-1}dt,$$

$$\mathcal{T}_{q,n,1c} = \frac{1}{nh_2^2}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,\widetilde{\beta}_n(W_i)\left[f_V(v)\,q^2(v)\right]^{-1}dt, \text{ and}$$

$$\mathcal{T}_{q,n,1d} = \frac{1}{nh_2^2}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,\widetilde{\gamma}_n(W_i)\left[f_V(v)\,q^2(v)\right]^{-1}dt.$$

Firstly, by the Law of Iterated Expectations, notice that $E[\mathcal{T}_{q,n,1a}] = 0$. While using representation (2.A.5), we are able to rewrite

$$\begin{aligned}\mathcal{T}_{q,n,1a} &= n^{-1}h_2^{-1}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\varrho_{i,1}^{1} + n^{-1}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}h_2^{-1}\varrho_{i,2}^{1}\\ &\equiv \mathcal{T}_{q,n,1a}^{(I)} + \mathcal{T}_{q,n,1a}^{(II)}.\end{aligned}$$

By another change of variable and integration by parts, it is not difficult to see that $\mathcal{T}_{q,n,1a}^{(II)} = O_p\left(n^{-1/2}h_2\right)$ which is clearly $o_p(n^{-1/2}h_1^{-(d+1)/2})$. Moreover, $\mathcal{T}_{q,n,1a}^{(I)}$ satisfies the Linderberg-Feller Central Limit Theorem by virtue of Assumption E (see Härdle (1990)), thus $\mathcal{T}_{q,n,1a}^{(I)} = O_p\left(n^{-1/2}\right)$ and we conclude that $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,1a} = o_p(1)$.

Now, under Assumptions (E1) – (E5), it is straightforward to extend the proof of Lemmas 3.1 (page 24) and 3.3 (page 26) in Lewbel and Linton (1999) to show that

$$\begin{aligned}\mathcal{T}_{q,n,1b} &= O_p(n^{-1}h_1^{-(d+1)/2}h_2^{-1}),\\ &= o_p(n^{-1/2}h_1^{-(d+1)/2}).\end{aligned}$$

Let $\widetilde{\beta}(w) = \iota_1^{\top}\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(w)$, then

$$\sup_{w}|h_1^{-(p_1+1)}\widetilde{\beta}_n(w) - \widetilde{\beta}(w)| = O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}),$$

by (2.A.9). Therefore, we write (recall $v = (t, z)$)

$$\begin{aligned}&\mathcal{T}_{q,n,1c} \equiv \mathcal{T}_{q,n,1c}^{(I)} + \mathcal{T}_{q,n,1c}^{(II)}, \text{ where}\\ &\mathcal{T}_{q,n,1c}^{(I)} = h_1^{p_1+1}n^{-1}h_2^{-2}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\int h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,\widetilde{\beta}(W_i)\left[f_V(v)\,q^2(v)\right]^{-1}dQ(t)\text{, and}\\ &\mathcal{T}_{q,n,1c}^{(II)} = h_1^{p_1+1}n^{-1}h_2^{-2}\sum_{i=1}^{n}\varepsilon_{q,i}\iota_2^{\top}\mathbf{M}_q^{-1}\\ &\qquad\times\int h_2^{-1}\mathcal{K}_{2,i}^{(1)}(v)\,(h_1^{-(p_1+1)}\widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i))[f_V(v)\,q^2(v)]^{-1}dQ(t)\,.\end{aligned}$$

Recall $\varepsilon_{q,i} = g_q(W_i) + \eta_i$ with $E[\eta_i|\,W_i] = 0$, then we further write

$$\begin{aligned}\mathcal{T}_{q,n,1c}^{(I)} &\equiv \mathcal{T}_{q,n,1c}^{(I-a)} + \mathcal{T}_{q,n,1c}^{(I-b)}\text{, where}\\ \mathcal{T}_{q,n,1c}^{(I-a)} &= h_1^{p_1+1}n^{-1}h_2^{-2}\sum_{i=1}^{n}g_q(W_i)\,\iota_2^{\top}\mathbf{M}_q^{-1}\varrho_{i,1}^{0}\widetilde{\beta}(W_i)\\ &\quad - h_1^{p_1+1}n^{-1}h_2^{-1}\sum_{i=1}^{n}g_q(W_i)\,\iota_2^{\top}\mathbf{M}_q^{-1}h_2^{-1}\varrho_{i,2}^{0}\widetilde{\beta}(W_i)\,,\end{aligned}$$

and $\mathcal{T}_{q,n,1c}^{(I-b)}$ is like $\mathcal{T}_{q,n,1c}^{(I-a)}$, but with η_i replacing $g_q(W_i)$. It follows by Bochner's Lemma that

$$\begin{aligned}
\mathcal{T}_{q,n,1c}^{(I-a)} &= h_1^{p_1+1}\iota_2^\top \mathbf{M}_q^{-1}\mathbf{M}_{q,0}^0 \left[\frac{E\left[\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(X,Z) g_q(X,Z) \middle| r(X,z_0)=r, Z=z_0\right]}{q^2(r,z_0)}\right. \\
&\quad \left. - \frac{E\left[\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(X,Z) g_q(X,Z) \middle| r(X,z_0)=r_0, Z=z_0\right]}{q^2(r_0,z_0)}\right] + o_p(n^{-1/2}h_1^{-(d+1)/2}) \\
&= h_1^{p_1+1}\mathcal{B}_1(x,z) + o_p(n^{-1/2}h_1^{-(d+1)/2}), \text{ by Assumption (E5(iii)).}
\end{aligned}$$

Similarly, by construction $\mathcal{T}_{q,n,1c}^{(I-b)}$ has mean zero and by the Cauchy–Schwarz inequality,

$$\begin{aligned}
|\mathcal{T}_{q,n,1c}^{(I-b)}| &= O_p(h_1^{p_1+1}n^{-1/2}h_2^{-1}) + O_p(h_1^{p_1+1}n^{-1/2}) \\
&= o_p(n^{-1/2}h_1^{-(d+1)/2}),
\end{aligned}$$

by Assumption (E5). With regards to $\mathcal{T}_{q,n,1c}^{(II)}$, this term may be written as

$$\mathcal{T}_{q,n,1c}^{(II)} = \mathcal{T}_{q,n,1c}^{(II-a)} + \mathcal{T}_{q,n,1c}^{(II-b)},$$

which are like $\mathcal{T}_{q,n,1c}^{(I-a)}$ and $\mathcal{T}_{q,n,1c}^{(I-b)}$, but with $h_1^{-(p_1+1)}\widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i)$ replacing $\widetilde{\beta}(W_i)$. Then by using similar arguments as above, we can show that

$$\begin{aligned}
\mathcal{T}_{q,n,1c}^{(II-a)} &= O_p(h_1^{p_1+1})O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) + O_p(h_1^{p_1+1}h_2)O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \\
&= o_p(h_1^{p_1+1}h_2) + o_p(h_1^{p_1+1}h_2^2), \text{ by Assumption (E5(ii)),} \\
&= o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii)).}
\end{aligned}$$

Define $\widetilde{\gamma}(w) = E[\widetilde{\gamma}_n(w)]$, then by result (2.A.8), it follows that

$$\begin{aligned}
&\sup_w |\widetilde{\gamma}(w)| = o(h_1^{p_1+1}), \text{ and} \\
&|h_1^{-(p_1+1)}\widetilde{\gamma}_n(w) - \widetilde{\gamma}(w)| = h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \text{ uniformly over } w.
\end{aligned}$$

Therefore, we write

$$\mathcal{T}_{q,n,1d} = \mathcal{T}_{q,n,1d}^{(I)} + \mathcal{T}_{q,n,1d}^{(II)},$$

where $\mathcal{T}_{q,n,1d}^{(I)}$ and $\mathcal{T}_{q,n,1d}^{(II)}$ are like $\mathcal{T}_{q,n,1c}^{(I)}$ and $\mathcal{T}_{q,n,1c}^{(II)}$, but with $\widetilde{\gamma}(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\gamma}_n(W_i) - \widetilde{\gamma}(W_i)$ replacing $\widetilde{\beta}(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i)$ respectively. Then by the Cauchy–Schwarz inequality,

$$|\mathcal{T}_{q,n,1d}^{(I)}| = h_1^{p_1+1}o_p(n^{-1/2}h_1^{-(d+1)/2}) + h_1^{p_1}h_2 o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii)).}$$

Similarly, by Assumption (E5(ii)),

$$\mathcal{T}_{q,n,1d}^{(II)} = o_p(n^{-1/2} h_1^{-(d+1)/2}).$$

Thus, $\sqrt{nh_1^{d+1}} \mathcal{T}_{q,n,1d} = o_p(1)$. ■

Lemma 2.B.2 *Under Assumption E, we have*

$$\sup_{t,z} |T_{q,n,2}(t,z)| = O_p\left(h_1^{-1}\nu_{1n}\right), \text{ and} \tag{2.B.3}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,2}(t,z)\,dt = O_p\left(n^{-1/2} h_1^{-(d+1)/2}\right). \tag{2.B.4}$$

Proof. Let $\widehat{S}_i - S_i = |\widehat{D}^{|1,0|} r(X_i, Z_i) - D^{|1,0|} r(X_i, Z_i)|$. Then, by Theorem 6 (page 594) in Masry (1996a),

$$\max_{1 \le i \le n} |\widehat{S}_i - S_i| = O_p\left(h_1^{-1}\nu_{1n}\right).$$

We now write

$$\widehat{V}_{q,n}^{*}(v) = nh_2^{-2} \sum_{i=1}^{n} \mathcal{K}_{2,i}(v)(\widehat{S}_i - S_i) + nh_2^{-2} \sum_{i=1}^{n} [\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)](\widehat{S}_i - S_i).$$

The first term is clearly

$$nh_2^{-2} \sum_{i=1}^{n} \mathcal{K}_{2,i}(v)(\widehat{S}_i - S_i) = O_p\left(h_1^{-1}\nu_{1n}\right) \text{ uniformly in } v.$$

The second term, after a Taylor–series expansion, is

$$\begin{aligned}
& nh_2^{-2} \sum_{i=1}^{n} [\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)](\widehat{S}_i - S_i) \\
& \le \{n^{-1} h_2^{-3} \sum_{i=1}^{n} |\mathcal{K}_{2,i}^{(1)}(t,z)|\} \max_{1 \le i \le n} |\widehat{r}_i - r_i| \max_{1 \le i \le n} \left|\widehat{S}_i - S_i\right| + O_p\left(h_1^{-1}\nu_{1n}\right) O_p\left(h_2^{-2}\nu_{1n}^2\right), \\
& = O_p\left(h_2^{-1}\nu_{1n}\right) O_p\left(h_1^{-1}\nu_{1n}\right) + O_p\left(h_1^{-1}\nu_{1n}\right) O_p\left(h_2^{-2}\nu_{1n}^2\right), \text{ by Assumption (E5(ii))}, \\
& = o_p(n^{-1/2} h_1^{-(d+1)/2}).
\end{aligned}$$

Therefore, (2.B.3) follows immediately.

We write

$$\widehat{V}_{q,n}^{*}(v) = [V_{q,n,a}^{*}(v) + V_{q,n,b}^{*}(v) + V_{q,n,c}^{*}(v)]\{1 + o_p(1)\} + o_p(n^{-1/2} h_1^{-(d+1)/2})$$

by using (2.A.7), where

$$V^*_{q,n,a}(v) \equiv n^{-2}h_1^{-(d+2)}h_2^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\alpha}^*_n(W_i, W_j; v, w),$$

$$V^*_{q,n,b}(v) \equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\widetilde{\beta}^*_n(W_i), \text{ and}$$

$$V^*_{q,n,c}(v) \equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\widetilde{\gamma}^*_n(W_i)$$

with

$$\begin{aligned}
\widetilde{\alpha}^*_n(W_i, W_j; v) &\equiv \mathcal{K}_{2,i}(v)\,\iota_1^{*\top}[\mathbf{M}_r f_W(W_i)]^{-1}\mathcal{K}_{1,j}(W_i)\,\varepsilon_{r,j},\\
\widetilde{\beta}^*_n(w) &\equiv n^{-1}h_1^{-(d+1)}\iota_1^{*\top}[\mathbf{M}_r f_W(w)]^{-1}\sum_{j=1}^{n}\mathcal{K}_{1,j}(w)\\
&\quad\times\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}D^{\mathbf{k}}r(w)\,(W_i-w)^{\mathbf{k}}, \text{ and}\\
\widetilde{\gamma}^*_n(w) &\equiv n^{-1}h_1^{-(d+1)}\iota_1^{*\top}[\mathbf{M}_r f_W(w)]^{-1}\gamma_n(w).
\end{aligned}$$

Thus, we have

$$\int T_{q,n,2}(t,z)\,dQ(t) = (\mathcal{T}_{q,n,2a}+\mathcal{T}_{q,n,2b}+\mathcal{T}_{q,n,2c})\{1+o_p(1)\}+o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$\begin{aligned}
\mathcal{T}_{q,n,2a} &= \frac{h_2}{n^2h_1^{(d+2)}h_2^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\widetilde{\alpha}^*_n(W_i, W_j; v)\left[f_V(v)\,q^2(v)\right]^{-1}dt,\\
\mathcal{T}_{q,n,2b} &= \frac{h_2}{nh_1h_2^2}\sum_{i=1}^{n}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\mathcal{K}_{2,i}(v)\,\widetilde{\beta}^*_n(W_i)\left[f_V(v)\,q^2(v)\right]^{-1}dt, \text{ and}\\
\mathcal{T}_{q,n,2c} &= \frac{h_2}{nh_1h_2^2}\sum_{i=1}^{n}\iota_2^{\top}\mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)}h_2^{-1}\mathcal{K}_{2,i}(v)\,\widetilde{\gamma}^*_n(W_i)\left[f_V(v)\,q^2(v)\right]^{-1}dt.
\end{aligned}$$

Using (2.A.5),

$$\begin{aligned}
&\int h_2^{-1}\widetilde{\alpha}^*_n(W_i, W_j; v)\left[f_V(v)\,q^2(v)\right]^{-1}dQ(t)\\
&\equiv \alpha^*_{n,1}(W_i, W_j; r, r_0, z_0) - \alpha^*_{n,2}(W_i, W_j; r, r_0, z_0),
\end{aligned}$$

where

$$\alpha_{n,1}^{*}\left(W_i, W_j; r, r_0, z_0\right) \equiv \varrho_{i,1}^{1} \iota_1^{*\top} \left[\mathbf{M}_r f_W\left(W_i\right)\right]^{-1} \mathcal{K}_{1,j}\left(W_i\right) \varepsilon_{r,j}, \text{ and}$$
$$\alpha_{n,2}^{*}\left(W_i, W_j; r, r_0, z_0\right) \equiv \varrho_{i,2}^{1} \iota_1^{*\top} \left[\mathbf{M}_r f_W\left(W_i\right)\right]^{-1} \mathcal{K}_{1,j}\left(W_i\right) \varepsilon_{r,j}.$$

By the Law of Iterated Expectations, $E\left[\alpha_{n,1}^{*}\right] = E\left[\alpha_{n,2}^{*}\right] = 0$, and by applying a second order U-statistic theory for random samples (e.g. Powell, Stock, and Stoker (1989)), it is not difficult but lengthy to show that

$$\mathcal{T}_{q,n,2a} = O_p(n^{-1} h_1^{-(d+2)/2}) + O_p(n^{-1} h_1^{-(d+2)/2} h_2).$$

Thus, by Assumption (E5), $\sqrt{n h_1^{d+1}} \mathcal{T}_{q,n,2a} = o_p(1)$.

By (2.A.9),

$$\sup_{w} |h_1^{-(p_1+1)} \widetilde{\beta}_n^{*}(w) - \widetilde{\beta}^{*}(w)| = O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n}),$$

where $\widetilde{\beta}^{*}(w) = \iota_1^{*\top} \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(w)$. Therefore, we write

$$\begin{aligned}
\mathcal{T}_{q,n,2b} &= \mathcal{T}_{q,n,2b}^{(I)} + \mathcal{T}_{q,n,2b}^{(II)}, \text{ where} \\
\mathcal{T}_{q,n,2b}^{(I)} &= h_1^{p_1} n^{-1} h_2^{-1} \sum_{i=1}^{n} \iota_2^{\top} \mathbf{M}_q^{-1} \int h_2^{-1} \mathcal{K}_{2,i}(v) \widetilde{\beta}^{*}(W_i) \left[f_V(v) q^2(v)\right]^{-1} dQ(t), \\
\mathcal{T}_{q,n,2b}^{(II)} &= h_1^{p_1} n^{-1} h_2^{-1} \sum_{i=1}^{n} \iota_2^{\top} \mathbf{M}_q^{-1} \\
&\quad \times \int h_2^{-1} \mathcal{K}_{2,i}(v) (h_1^{-(p_1+1)} \widetilde{\beta}_n^{*}(W_i) - \widetilde{\beta}^{*}(W_i)) [f_V(v) q^2(v)]^{-1} dQ(t).
\end{aligned}$$

Then, by using representation (2.A.5), we may further write

$$\mathcal{T}_{q,n,2b}^{(I)} = n^{-1} h_2^{-1} h_1^{p_1} \sum_{i=1}^{n} \iota_2^{\top} \mathbf{M}_q^{-1} \varrho_{i,1}^{1} \widetilde{\beta}^{*}(W_i) - n^{-1} h_1^{p_1} \sum_{i=1}^{n} \iota_2^{\top} \mathbf{M}_q^{-1} h_2^{-1} \varrho_{i,2}^{1} \widetilde{\beta}^{*}(W_i).$$

The right–hand side of the above expression converges in mean squared to (by Bochner's Lemma)

$$\begin{aligned}
\mathcal{T}_{q,n,2b}^{(I)} &= h_1^{p_1} h_2 \iota_2^{\top} \mathbf{M}_q^{-1} \mathbf{M}_{q,0}^{1} \left[\frac{E\left[\iota_1^{\top} \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(X,Z) \middle| r(X, z_0) = r, Z = z_0\right]}{q^2(r, z_0)} \right. \\
&\quad \left. - \frac{E\left[\iota_1^{\top} \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(X,Z) \middle| r(X, z_0) = r_0, Z = z_0\right]}{q^2(r_0, z_0)} \right] + o_p(n^{-1/2} h_1^{-(d+1)/2}) \\
&= h_1^{p_1} h_2 \mathcal{B}_2(x, z) + o_p(n^{-1/2} h_1^{-(d+1)/2}), \text{ by Assumption (E5(iii))}.
\end{aligned}$$

Furthermore, $\mathcal{T}^{(II)}_{q,n,2b}$ is like $\mathcal{T}^{(I)}_{q,n,2b}$, but with $h_1^{-(p_1+1)}\widetilde{\beta}^*_n(w) - \widetilde{\beta}^*(w)$ replacing $\widetilde{\beta}^*(W_i)$. Then, it follows by Cauchy–Schwarz inequality, that

$$\begin{aligned}|\mathcal{T}^{(II)}_{q,n,2b}| &= O_p(h_1^{p_1}h_2)O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) + O_p(h_1^{p_1}h_2^2)O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}),\\ &= O_p(h_1^{p_1}h_2)o_p(h_2) + O_p(h_1^{p_1}h_2^2)o_p(h_2), \text{ by Assumption (E5(ii))},\\ &= o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by virtue of Assumption (E5(iii)).}\end{aligned}$$

Consequently, $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,2b} = O_p(1)$.

Define $\widetilde{\gamma}^*(w) = E[\widetilde{\gamma}^*_n(w)]$, then by result (2.A.8), it follows that

$$\begin{aligned}\sup_w|\widetilde{\gamma}^*(w)| &= o(h_1^{p_1+1}), \text{ and}\\ |h_1^{-(p_1+1)}\widetilde{\gamma}^*_n(w) - \widetilde{\gamma}^*(w)| &= h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \text{ uniformly in } w.\end{aligned}$$

Therefore, we write

$$\mathcal{T}_{q,n,2c} = \mathcal{T}^{(I)}_{q,n,2c} + \mathcal{T}^{(II)}_{q,n,2c},$$

where these two terms are like $\mathcal{T}^{(I)}_{q,n,2b}$ and $\mathcal{T}^{(II)}_{q,n,2b}$, but with $\widetilde{\gamma}^*(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\gamma}^*_n(W_i) - \widetilde{\gamma}^*(W_i)$ replacing $\widetilde{\beta}^*(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\beta}^*_n(W_i) - \widetilde{\beta}^*(W_i)$ respectively. Then by the Cauchy–Schwarz inequality,

$$\begin{aligned}|\mathcal{T}^{(I)}_{q,n,2c}| &= O_p(h_1^{p_1}h_2)o_p(n^{-1/2}h_1^{-(d+1)/2})\\ &\quad + O_p(h_1^{p_1}h_2^2)o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii)).}\end{aligned}$$

Similarly, by Assumption (E5(ii)) and (E5(iii)),

$$\begin{aligned}|\mathcal{T}^{(II)}_{q,n,2c}| &= O_p(h_1^{p_1}h_2)h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n})\\ &\quad + O_p(h_1^{p_1}h_2^2)h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n})\\ &= o_p(n^{-1/2}h_1^{-(d+1)/2}).\end{aligned}$$

Thus, $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,2c} = o_p(1)$. ■

Lemma 2.B.3 *Under Assumption E, we have*

$$\sup_{t,z}|T_{q,n,3}(t,z)| = O_p\left(h_2^{p_2+1}\right), \text{ and} \tag{2.B.5}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,3}(t,z)\,dt = O_p\left(n^{-1/2}h_1^{-(d+1)/2}\right). \tag{2.B.6}$$

Proof. Define

$$\Delta_{q,i}(v) = q(V_i) - \sum_{0 \le |\mathbf{k}| \le p_2} \frac{1}{\mathbf{k}!} D^{\mathbf{k}} q(v)(V_i - v)^{\mathbf{k}},$$

and $\widehat{\Delta}_{q,i}(v)$ is like $\Delta_{q,i}(v)$ but with $\widehat{V}_i$ in place of $V_i = (r(X_i, Z_i), Z_i)^{\top}$. Then by Assumption (E4),

$$\Delta_{q,i}(v) = \sum_{|\mathbf{k}| = p_2+1} \frac{1}{\mathbf{k}!} D^{\mathbf{k}} q(v^*)(V_i - v)^{\mathbf{k}}$$

for some v^* that lies between V_i and v, also

$$\widehat{\Delta}_{q,i}(v) = \sum_{|\mathbf{k}| = p_2+1} \frac{1}{\mathbf{k}!} D^{\mathbf{k}} q(\widehat{v}^*)(V_i - v)^{\mathbf{k}},$$

where $\widehat{v}^*$ lies between $\widehat{V}_i$ and v. It is also clear that $\|\widehat{v}^* - v^*\| = O_p(\nu_{1n})$ and $|\Delta_{q,i}(v)| = O_p(h_2^{p_2+1})$ for $\|V_i - v\| \le ch_2$. These observations along with Assumption (E5(ii)) imply that

$$\begin{aligned}
\widehat{\Delta}_{q,i}(v) - \Delta_{q,i}(v) &= \sum_{|\mathbf{k}| = p_2+1} \frac{1}{\mathbf{k}!} D^{\mathbf{k}} q(\widehat{v}^*)\left[(\widehat{V}_i - v)^{\mathbf{k}} - (V_i - v)^{\mathbf{k}}\right] \\
&\quad + \sum_{|\mathbf{k}| = p_2+1} \frac{1}{\mathbf{k}!}\left[D^{\mathbf{k}} q(\widehat{v}^*) - D^{\mathbf{k}} q(v^*)\right](V_i - v)^{\mathbf{k}}, \\
&= O_p(h_2^{p_2}\nu_{1n}) = o_p(n^{-1/2}h_1^{-(d+1)/2}),
\end{aligned}$$

uniformly in v and i such that $\|V_i - v\| \le ch_2$. So we conclude that

$$|\widehat{\Delta}_{q,i}(v)| = O_p(h_2^{p_2+1}) + o_p(n^{-1/2}h_1^{-(d+1)/2})$$

uniformly in v and i for $\|V_i - v\| \le ch_2$.

We now write

$$\left|\widehat{\mathbf{B}}_{q,n}(t,z) - \mathbf{B}_{q,n}(t,z)\right| \le \left\{\frac{1}{nh_2^2}\sum_{i=1}^{n} |\mathcal{K}_{2,i}(t,z)|\right\} \max_{1 \le i \le n} \sup_{v} |\widehat{\Delta}_{q,i}(v) - \Delta_{q,i}(v)| \tag{2.B.7}$$

$$+ \left\{\frac{1}{nh_2^2}\sum_{i=1}^{n} \left|\widehat{\mathcal{K}}_{2,i}(t,z) - \mathcal{K}_{2,i}(t,z)\right|\right\} \max_{1 \le i \le n} \sup_{v} |\widehat{\Delta}_{q,i}(v)|. \tag{2.B.8}$$

It is clear that (2.B.7) is $o_p(n^{-1/2}h_1^{-(d+1)/2})$. The order in probability of (2.B.8), after a

Taylor–series expansion, is given by

$$
\begin{aligned}
&\{n^{-1}h_2^{-3}\sum_{i=1}^{n}|\mathcal{K}_{2,i}^{(1)}(t,z)|\}\max_{1\le i\le n}|\widehat{r}_i-r_i|\max_{1\le i\le n}\sup_v|\widehat{\Delta}_{q,i}(v)|+O_p\left(h_2^{-2}\nu_{1n}^2\right)\max_{1\le i\le n}\sup_v|\widehat{\Delta}_{q,i}(v)|\\
&=O_p(h_2^{-1}\nu_{1n})O_p(h_2^{p_2+1})+O_p(h_2^{-2}\nu_{1n}^2)O_p(h_2^{p_2+1})\\
&=o_p(n^{-1/2}h_1^{-(d+1)/2}),\text{ by Assumption (E5(ii))}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sup_v|\widehat{\mathbf{B}}_{q,n}(v)|&\le\sup_v|\mathbf{B}_{q,n}(v)|+\sup_v|\widehat{\mathbf{B}}_{q,n}(v)-\mathbf{B}_{q,n}(v)|\\
&=O_p(h_2^{p_2+1})+O_p(h_2^{p_2+1}\nu_{1n}),
\end{aligned}
$$

proving (2.B.5). Furthermore, we rewrite

$$
\begin{aligned}
T_{q,n,3}(t,z_0)&=h_2^{p_2+1}\iota_2^{\top}\mathbf{M}_q^{-1}\mathbf{B}_q q^{(p_2+1)}(t,z_0)\,q^{-2}(t,z_0)\\
&+h_2^{p_2+1}\iota_2^{\top}\mathbf{M}_q^{-1}[\mathbf{B}_{q,n}(t,z_0)-\mathbf{B}_q f_V(t,z_0)]q^{(p_2+1)}(t,z_0)\,q^{-2}(t,z_0)\\
&+o_p(n^{-1/2}h_1^{-(d+1)/2}).
\end{aligned}
$$

Clearly, the first term of the above equation is $O_p(h_2^{p_2+1})$, and the second is $o_p(h_2^{p_2+1})$ by Corollary 2 (page 580) in Masry (1996a). Then,

$$
\begin{aligned}
\int_{r_0}^{r(x,z)}T_{q,n,3}(t,z)\,dt&=h_2^{p_2+1}i_2^{\top}\mathbf{M}_q^{-1}\mathbf{B}_q\int_{r_0}^{r(x,z)}\frac{q^{(p_2+1)}(t,z_0)}{q^2(t,z_0)}dt+o_p(n^{-1/2}h_1^{-(d+1)/2})\\
&=h_2^{p_2+1}\mathcal{B}_3(x,z)+o_p(n^{-1/2}h_1^{-(d+1)/2})
\end{aligned}
$$

follows. ■

Lemma 2.B.4 *Under Assumption E, we have*

$$
\begin{aligned}
\mathcal{R}_{q,n}(x,z)&=\int T_{q,n,4}(t,z_0)\,dQ(t)+\int T_{q,n,5}(t,z_0)\,dQ(t)+\int T_{q,n,6}(t,z_0)\,dQ(t)\\
&=o_p(n^{-1/2}h_1^{-(d+1)/2}).
\end{aligned}
$$

Proof. A typical element of $\widehat{\mathbf{M}}_{q,n}(v)-\mathbf{M}_{q,n}(v)$ is given by

$$
\begin{aligned}
&[\widehat{\mathbf{M}}_{q,n,j,k}(v)]_{l,l^0}-[\mathbf{M}_{q,n,j,k}(v)]_{l,l^0}\\
&=\frac{1}{nh_2^2}\sum_{i=1}^{n}\left[\left(\frac{\widehat{V}_i-v}{h_2}\right)^{\phi_{q;j}(l)+\phi_{q;k}(l^0)}K_2\left(\frac{\widehat{V}_i-v}{h_2}\right)-\left(\frac{V_i-v}{h_2}\right)^{\phi_{q;j}(l)+\phi_{q;k}(l^0)}K_2\left(\frac{V_i-v}{h_2}\right)\right].
\end{aligned}
$$

After a Taylor–series expansion of the last expression at V_i, it is not difficult to show that

$$\sup_{t,z}|[\widehat{\mathbf{M}}_{q,n,j,k}(t,z)]_{l,l^0} - [\mathbf{M}_{q,n,j,k}(t,z)]_{l,l^0}| = O_p\left(h_2^{-1}\nu_{1n}\right).$$

By the triangle inequality, we have

$$\sup_{t,z}|\widehat{\mathbf{M}}_{q,n}(t,z) - f_V(t,z)\mathbf{M}_q| \leq \sup_{t,z}|\widehat{\mathbf{M}}_{q,n}(t,z) - \mathbf{M}_{q,n}(t,z)| + \sup_{t,z}|\mathbf{M}_{q,n}(t,z) - f_V(t,z)\mathbf{M}_q|.$$

The first term of the right–hand side of the inequality is $O_p\left(h_2^{-1}\nu_{1n}\right) = o_p(1)$, while the second is, by Corollary 2 (page 580) in Masry (1996a), $O_p(n^{-1/2}h_2^{-1}\sqrt{\ln n} + h_2) = o_p(1)$. Furthermore, by Assumption (E1), $\widehat{\mathbf{M}}_{q,n}^{-1}(v) = O_p(1)$ with probability approaching one. Therefore, results (2.B.2), (2.B.4) and (2.B.6) imply that

$$\int T_{q,n,4}(t,z_0)\,dQ(t) = o_p(1)\,O_p(n^{-1/2}h_1^{-(d+1)/2}),$$
$$\int T_{q,n,5}(t,z_0)\,dQ(t) = o_p(1)\,O_p(n^{-1/2}h_1^{-(d+1)/2}), \text{ and}$$
$$\int T_{q,n,6}(t,z_0)\,dQ(t) = o_p(1)\,O_p(n^{-1/2}h_1^{-(d+1)/2}),$$

respectively. ■

Lemma 2.B.5 *Under Assumption E, we have*

$$\sup_{t,z}|\widehat{q}(t,z) - q(t,z)| = O_p\left(h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n}\right).$$

Proof. This result follows from (2.B.1), (2.B.3), (2.B.5) and Lemma 2.B.4. ■

Lemma 2.B.6 *Let Assumptions E and F hold, then the estimators $\widehat{\alpha}_{P_1}(x)$ and $\widehat{\alpha}_{P_2}(z)$ satisfies the following asymptotic expansions:*

$$\begin{aligned}\widehat{\alpha}_{P_1}(x) - \alpha_{P_1}(x) &= \iota_1^\top \mathbf{M}_r^{-1}\{1 + o_p(1)\} \\ &\quad\times\left\{n^{-1}h_1^{-d}\sum_{j=1}^{n}\left[\int h_1^{-1}\mathcal{K}_{1,j}(x,z)\frac{dP_1(z)}{q(r,z_0)\,f_W(x,z)}\right]\varepsilon_{r,j}\right\} \\ &\quad+\int \mathcal{B}(x,z)\,dP_1(z) + R_{P_1,n}(x),\end{aligned}$$

$$
\begin{aligned}
\widehat{\alpha}_{P_2}(z) - \alpha_{P_2}(z) &= \iota_1^\top \mathbf{M}_r^{-1}\{1 + o_p(1)\} \\
&\quad \times \left\{ n^{-1} h_1^{-1} \sum_{j=1}^{n} \left[\int h_1^{-d} \mathcal{K}_{1,j}(x,z) \frac{dP_2(x)}{q(r,z_0) f_W(x,z)} \right] \varepsilon_{r,j} \right\} \\
&\quad + \int \mathcal{B}(x,z)\, dP_2(x) + R_{P_2,n}(z),
\end{aligned}
$$

where $o_p(1)$'s are uniformly in x and z, and the remainder terms $R_{P_1,n}(x)$, and $R_{P_2,n}(z)$ satisfy

$$
\begin{aligned}
&\sup_x |R_{P_1,n}(x)| = o_p(n^{-1/2} h_*^{-1/2}), \text{ and} \\
&\sup_z |R_{P_2,n}(z)| = o_p(n^{-1/2} h_*^{-1/2}) \text{ respectively.}
\end{aligned}
$$

Proof. This result follows from Lemmas 2.B.1–2.B.5 and Assumption (F3). ■

Lemma 2.B.7 *Let Assumptions E and F hold, then*

$$
\sqrt{nh_*} T_{H,n,1}(m) \xrightarrow{d} N\left(0, \frac{\sigma_H^2(m)}{f_M(m)} \left[\mathbf{M}_H^{-1} \Gamma_H \mathbf{M}_H^{-1}\right]_{0,0}\right).
$$

Proof. Let $V_{H,n}(m) = nh_*^{-1} \sum_{i=1}^{n} \mathcal{K}_{*,i}(m)\, \varepsilon_{r,i}$, then we have

$$
\begin{aligned}
\widehat{V}_{H,n}(m) - V_{H,n}(m) &= \frac{1}{nh_*} \sum_{i=1}^{n} [\widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m)] \varepsilon_{r,i} \\
&= \frac{1}{nh_*^2} \sum_{i=1}^{n} \mathcal{K}_{*,i}^{(1)}(m)\, (\widetilde{M}_i - M_i) \varepsilon_{r,i} + O_p\left(h_*^{-2} \nu_{\dagger n}^2\right) \\
&= \frac{1}{nh_*^2} \sum_{i=1}^{n} \mathcal{K}_{*,i}^{(1)}(m) \{[\widehat{\alpha}_{P_1}(X_i) - \alpha_{P_1}(X_i)] \\
&\qquad\qquad + [\widehat{\alpha}_{P_2}(Z_i) - \alpha_{P_2}(Z_i)] - [\widetilde{c}_i - c_i]\} \varepsilon_{r,i} \\
&\quad + o_p(n^{-1/2} h_*^{(d+1)/2}),
\end{aligned}
$$

where the second equality follows from a Taylor–series expansion and Assumption (F3), and the last by Assumption (F3(ii)). So,

$$
\begin{aligned}
\widehat{T}_{H,n}(m) &= T_{H,n,a}(m) + T_{H,n,b}(m)\{1 + o(1)\} \\
&\quad + T_{H,n,c}(m) + T_{H,n,d}(m) + o_p(n^{-1/2} h_*^{-1/2}),
\end{aligned}
$$

where

$$T_{H,n,a}(m) = \frac{1}{nh_*}\sum_{i=1}^{n} \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}(m)\,\varepsilon_{r,i}$$

$$T_{H,n,b}(m) = \frac{1}{n^2h_*^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \alpha_n(\xi_i,\xi_j;m), \tag{2.B.9}$$

$$T_{H,n,c}(m) = \frac{1}{nh_*^2}\sum_{i=1}^{n} \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}\beta(W_i), \tag{2.B.10}$$

$$T_{H,n,d}(m) = \frac{1}{nh_*^2}\sum_{i=1}^{n} \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}R_n(W_i). \tag{2.B.11}$$

We now discuss the properties of each term above.

Firstly, let $\mathcal{F}_W$ and $\mathcal{F}_H$ be the sigma algebras generated by $W^\top = (X^\top, Z)$ and $r(W) = H[M(W)]$ respectively, then by the tower property of conditional expectations, i.e. Theorem (34.3) in Billingsley (1986), we have $E[\varepsilon_{r,i}|\, r(W_i)] = 0$, which implies that $E[\varsigma_i] = 0$ by the Law of Iterated Expectations, where

$$\varsigma_i = h_*^{-1}\iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)\,\varepsilon_{r,i}.$$

Therefore, by Theorem 4 (page 94) in Masry (1996b), it follows that

$$\begin{aligned}\sqrt{nh_*}T_{H,n}(m) &= \frac{1}{n^{1/2}h_*^{1/2}}\sum_{i=1}^{n}\varsigma_i \\ &\xrightarrow{d} N\left(0, \frac{\sigma_M^2(m)}{f_M(m)}\left[\mathbf{M}_H^{-1}\Gamma_H\mathbf{M}_H^{-1}\right]_{0,0}\right),\end{aligned}$$

where $\sigma_H^2(b) = E[\varepsilon_{r,i}^2|B(W_i) = b]$. The term α_n in (2.B.9) may be written as

$$\begin{aligned}\alpha_n(\xi_i,\xi_j;b) &\equiv \alpha_n^I(\xi_i,\xi_j;b) + \alpha_n^{II}(\xi_i,\xi_j;b) - \alpha_n^{III}(\xi_i,\xi_j;b)\text{, with}\\ \alpha_n^I(\xi_i,\xi_j;m) &\equiv h_1^{-d}\mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}\int h_1^{-1}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)}{q(r_i,z_0)\,f_W(W_i)}\varepsilon_{r,j},\\ \alpha_n^{II}(\xi_i,\xi_j;m) &\equiv h_1^{-1}\mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}\int h_1^{-d}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_2(X_i)}{q(r_i,z_0)\,f_W(W_i)}\varepsilon_{r,j},\\ \alpha_n^{III}(\xi_i,\xi_j;m) &\equiv \mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}\int\int h_1^{-(d+1)}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)\,dP_2(X_i)}{q(r_i,z_0)\,f_W(W_i)}\varepsilon_{r,j},\end{aligned}$$

where $\xi_i \equiv (W_i^\top, \varepsilon_{r,i})^\top$. Thus, by applying a second order U-statistic theory for random samples (e.g. Powell, Stock, and Stoker (1989)), we can show that under Assumptions E and F,

$$T_{H,n,b}(m) = O_p(n^{-1}h_1^{-d/2}h_*^{-3/2}) + O_p(n^{-1}h_1^{-1/2}h_*^{-3/2}) + O_p(n^{-1}h_*^{-3/2}).$$

Consequently, by Assumption (F3(ii)),

$$\begin{aligned}\sqrt{nh_*}T_{H,n,b}(m) &= \sqrt{nh_*}O_p(n^{-1}h_1^{-d/2}h_*^{-3/2})\\ &+ \sqrt{nh_*}O_p(n^{-1}h_1^{-1/2}h_*^{-3/2}) + \sqrt{nh_*}O_p(n^{-1}h_*^{-3/2})\\ &= O_p(n^{-1/2}h_1^{-d/2}h_*^{-1}) + O_p(n^{-1/2}h_1^{-1/2}h_*^{-1}) + O_p(n^{-1/2}h_*^{-1}) = o_p(1).\end{aligned}$$

Similarly,

$$\begin{aligned}\beta(w) &\equiv \int \mathcal{B}(x,z)\,dP_1(z) + \int \mathcal{B}(x,z)\,dP_2(x) + \int\int \mathcal{B}(x,z)\,dP_1(z)\,dP_2(x),\\ &= O(h_\dagger) \text{ by construction.}\end{aligned}$$

Then, by Assumption (F3(ii)), (2.B.10) is $h_\dagger O_p(n^{-1/2}h_*^{-3/2})$, so

$$\sqrt{nh_*}T_{H,n,c}(m) = O_p\left(h_\dagger h_*^{-1}\right) = o_p(1).$$

Finally, the term (2.B.11) satisfies

$$\begin{aligned}&\left|n^{-1}h_*^{-2}\sum_{i=1}^{n}\iota_*^\top\left[f_M(m)\mathbf{M}_H\right]^{-1}\mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}R_n(W_i)\right|\\ &\le 3\times\sup_x\left|R_{P_1,n}(x)\right|\left\{n^{-1}h_*^{-2}\sum_{i=1}^{n}\left|\iota_*^\top\left[f_M(m)\mathbf{M}_H\right]^{-1}\mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}\right|\right\},\end{aligned}$$

and the right–hand side is $o_p(n^{-1/2}h_*^{-1/2})$ by Lemma 2.B.6, completing the proof. ■

Lemma 2.B.8 *Let Assumptions E and F hold, then*

$$T_{H,n,2}(m) = O_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. Firstly, $\max_{1\le i\le n}|H(M_i) - H(\widetilde{M}_i)| = O_p(\nu_{\dagger n})$ by Assumption (I2). Let

$$V_{H,n}^*(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\mathcal{K}_{*,i}(m)[H(M_i) - H(\widetilde{M}_i)],$$

then after a Taylor–series expansion,

$$
\begin{aligned}
\widehat{V}^*_{H,n}(m) - V^*_{H,n}(m) &= n^{-1}h_*^{-1}\sum_{i=1}^{n}\left\{\widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m)\right\}[H(M_i) - H(\widetilde{M}_i)] \\
&= n^{-1}h_*^{-2}\sum_{i=1}^{n}\mathcal{K}^{(1)}_{*,i}(m)[\widetilde{M}_i - M_i][H(M_i) - H(\widetilde{M}_i)] + O_p(h_*^{-2}\nu_{\dagger n}^3) \\
&= O_p(h_*^{-1}\nu_{\dagger n}^2) + O_p(h_*^{-2}\nu_{\dagger n}^3) = o_p(n^{-1/2}h_*^{-1/2}),
\end{aligned}
$$

by Assumption (F3(ii)). Consequently, after a further Taylor–series expansion,

$$
\begin{aligned}
T_{H,n,2}(m) &= n^{-1}h_*^{-1}\sum_{i=1}^{n}\mathcal{K}_{*,i}(m)[H(M_i) - H(\widetilde{M}_i)] + o_p(n^{-1/2}h_*^{-1/2}) \\
&= -n^{-1}h_*^{-1}\sum_{i=1}^{n}\mathcal{K}_{*,i}(m)\frac{\partial H(M_i)}{\partial b}[\widetilde{M}_i - M_i] + O_p\left(\nu_{\dagger n}^2\right) + o_p(n^{-1/2}h_*^{-1/2}) \\
&= -n^{-1}h_*^{-1}\sum_{i=1}^{n}\mathcal{K}_{*,i}(m)H^{(1)}(M_i)[\widetilde{M}_i - M_i] + o_p(n^{-1/2}h_*^{-1/2}),
\end{aligned}
$$

where the last equality follows from Assumption (F3(ii)). Therefore, by Lemma 2.B.6, we have

$$
\begin{aligned}
T_{H,n,2}(m) &= T_{H,n,2a}(m)\{1 + o_p(1)\} \\
&\quad + T_{H,n,2b}(m) + T_{H,n,2c}(m) + o_p(n^{-1/2}h_*^{-1/2}),
\end{aligned}
$$

where

$$
\begin{aligned}
T_{H,n,2a}(m) &= -\frac{1}{n^2h_*}\sum_{i=1}^{n}\sum_{j=1}^{n}\iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\alpha_n^*(W_i, W_j; m), \\
T_{H,n,2b}(m) &= -\frac{1}{nh_*}\sum_{i=1}^{n}\iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)H^{(1)}(M_i)\beta(W_i), \\
T_{H,n,2c}(m) &= -\frac{1}{nh_*}\sum_{i=1}^{n}\iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)H^{(1)}(M_i)R_n(W_i),
\end{aligned}
$$

with $\beta(\cdot)$, and $R_n(\cdot)$ defined as in Lemma 2.B.7, and

$$
\begin{aligned}
\alpha_n^*(W_i, W_j; m) &\equiv \alpha_n^{*I}(W_i, W_j; m) + \alpha_n^{*II}(W_i, W_j; m) - \alpha_n^{*III}(W_i, W_j; m), \text{ with} \\
\alpha_n^{*I}(W_i, W_j; m) &\equiv h_1^{-d}\mathcal{K}_{*,i}(m)\int h_1^{-1}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j}, \\
\alpha_n^{*II}(W_i, W_j; m) &\equiv h_1^{-1}\mathcal{K}_{*,i}(m)\int h_1^{-d}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_2(X_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j}, \\
\alpha_n^{*III}(W_i, W_j; m) &\equiv \mathcal{K}_{*,i}(m)\int\int h_1^{-(d+1)}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)\,dP_2(X_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j}.
\end{aligned}
$$

Again, by applying a second order U–statistic theory (e.g. Powell, Stock, and Stoker (1989)), we can show that under Assumptions E and F,

$$\begin{aligned} T_{H,n,2a}(m) &= O_p(n^{-1}h_1^{-d/2}h_*^{-1/2}) + O_p(n^{-1}h_1^{-1/2}h_*^{-1/2}) + O_p(n^{-1}h_*^{-1/2}) \\ &= o_p(n^{-1/2}h_*^{-1/2}). \end{aligned}$$

By Bochner's Lemma and Lemma 2.B.6, it follows that $T_{H,n,2b}(m)$ converges in mean squared to

$$\mathcal{B}_{H2}(m) \equiv -\iota_*^\top \mathbf{M}_H^{-1}\mathbf{M}_{H,0}^0 E\left[H^{(1)}(M(X,Z))\,\beta(X,Z)\,\middle|\, H(M(X,Z)) = m\right],$$

which is $O(h_\dagger)$. Finally, the last term is bounded by

$$|T_{H,n,2c}(m)| \leq \max_{1\leq i\leq n}|R_n(W_i)| \max_{1\leq i\leq n}\left|H^{(1)}(M_i)\right| \left\{n^{-1}h_*^{-1}\sum_{i=1}^{n}\left|\iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)\right|\right\},$$

where the right–hand side is $o_p(n^{-1/2}h_*^{-1/2})$. ∎

Lemma 2.B.9 *Let Assumptions E and F hold, then*

$$T_{H,n,3}(m) = O_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. As in Lemma 2.B.3, by Assumption (E4), we write

$$\begin{aligned} \widehat{\Delta}_{H,i}(m) - \Delta_{H,i}(m) = &\sum_{j=p_*+1}\frac{1}{j!}\left(\partial^j H(\widetilde{m}^*)/\partial m^j\right)\left[(\widehat{M}_i - m)^j - (M_i - m)^j\right] \\ &+ \sum_{j=p_*+1}\frac{1}{j!}\left[\partial^j H(\widetilde{m}^*)/\partial m^j - \partial^j H(m^*)/\partial m^j\right](M_i - m)^j, \quad (2.B.12) \end{aligned}$$

where $(\widetilde{m}^*, m^*)$ lie between $(\widetilde{M}_i, M_i)$ and m, such that $\|\widetilde{m}^* - m^*\| = O_p(\nu_{\dagger n})$. For $\|M_i - m\| \leq ch_*$, $|\Delta_{H,i}(m)| = O_p(h_*^{p_*+1})$. These observations imply that (2.B.12) is $O_p(h_*^{p_*}\nu_{\dagger n})$, which by Assumption (F3(ii)), is $o_p(n^{-1/2}h_*^{-1/2})$. Therefore, we conclude that

$$|\widehat{\Delta}_{H,i}(m)| = O_p(h_*^{p_*+1}) + o_p(n^{-1/2}h_*^{-1/2})$$

uniformly in m and i for $\|M_i - m\| \leq ch_*$. Now, we write by the triangle inequality

$$\begin{aligned} \left|\widehat{\mathbf{B}}_{H,n}(m) - \mathbf{B}_{H,n}(m)\right| \leq &\left\{\frac{1}{nh_*}\sum_{i=1}^{n}|\mathcal{K}_{*,i}(m)|\right\}\max_{1\leq i\leq n}\sup_b|\widehat{\Delta}_{H,i}(m) - \Delta_{H,i}(m)| \\ &+ \left\{\frac{1}{nh_*}\sum_{i=1}^{n}\left|\widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m)\right|\right\}\max_{1\leq i\leq n}\sup_b|\widehat{\Delta}_{H,i}(m)|. \end{aligned}$$

The sup of the first term is $o_p(n^{-1/2}h_*^{-1/2})$, and by Taylor–series expansion, the second term is

$$\frac{1}{nh_*^2}\sum_{i=1}^{n}\left|\mathcal{K}_{*,i}^{(1)}(m)\right|\left|\widetilde{M}_i - M_i\right| |\widehat{\Delta}_{H,i}(m)| + O_p((h_*^{-1}\nu_{\dagger n})^2)\max_{1\le i\le n}\sup_b|\widehat{\Delta}_{H,i}(m)|$$
$$= O_p\left(h_*^{-1}\nu_{\dagger n}\right)O_p(h_*^{p_*+1}) + O_p((h_*^{-1}\nu_{\dagger n})^2)O_p(h_*^{p_*+1})$$
$$= o_p(n^{-1/2}h_*^{-1/2}) \text{ uniformly in } m \text{ by Assumption (F3).}$$

Therefore, we conclude that

$$\widehat{\mathbf{B}}_{H,n}(m) = \mathbf{B}_{H,n}(m) + o_p(n^{-1/2}h_*^{-1/2})$$

uniformly in m, and by Kolmogorov's Law of Large numbers, it follows that

$$T_{H,n,3}(m) = h_*^{p_*+1}\iota_*^{\top}\mathbf{M}_H^{-1}\mathbf{B}_H H^{(p_*+1)}(m) + o_p(n^{-1/2}h_*^{-1/2}),$$

as required. ■

Lemma 2.B.10 *Let Assumptions E and F hold, then*

$$\mathcal{R}_{H,n}(m) = T_{H,n,4}(m) + T_{H,n,5}(m) + T_{H,n,6}(m)$$
$$= o_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. For a typical element of $\widehat{\mathbf{M}}_{H,n}(m) - \mathbf{M}_{H,n}(m)$ is given by

$$[\widehat{\mathbf{M}}_{H,n,j,k}(m)]_{l,l^0} - [\mathbf{M}_{H,n,j,k}(m)]_{l,l^0}$$
$$= \frac{1}{nh_*}\sum_{i=1}^{n}\left[\left(\frac{\widetilde{M}_i - m}{h_*}\right)^{\phi_{H;j}(l)+\phi_{H;k}(l^0)} K_2\left(\frac{\widetilde{M}_i - m}{h_*}\right) - \left(\frac{M_i - m}{h_*}\right)^{\phi_{H;j}(l)+\phi_{H;k}(l^0)} K_2\left(\frac{M_i - m}{h_*}\right)\right].$$

After expanding the last expression at M_i, it is not difficult to show that

$$\sup_m|[\widehat{\mathbf{M}}_{H,n,j,k}(m)]_{l,l^0} - [\mathbf{M}_{H,n,j,k}(m)]_{l,l^0}| = O_p\left(h_*^{-1}\nu_{\dagger n}\right).$$

By the triangle inequality, we have

$$\begin{aligned}\sup_b|\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H| &\le \sup_b|\widehat{\mathbf{M}}_{H,n}(m) - \mathbf{M}_{H,n}(m)| \\ &\quad + \sup_b|\mathbf{M}_{H,n}(m) - f_M(m)\mathbf{M}_H| \\ &= O_p\left(h_*^{-1}\nu_{\dagger n}\right) + O_p(n^{-1/2}h_*^{-1/2}\sqrt{\ln n} + h_*) = o_p(1),\end{aligned}$$

where the last equality follows from Assumption (F3(ii)) and Corollary 2 (page 580) in Masry (1996a). Furthermore, by Assumption (F1), $\widehat{\mathbf{M}}_{H,n}^{-1}(m) = O_p(1)$ with probability approaching one. Therefore, Lemmas 2.B.7–2.B.9 imply

$$\begin{aligned}T_{H,n,4}(m) &= o_p(1)\,O_p(n^{-1/2}h_*^{-1/2}),\\ T_{H,n,5}(m) &= o_p(1)\,O_p(n^{-1/2}h_*^{-1/2}), \text{ and}\\ T_{H,n,6}(m) &= o_p(1)\,O_p(n^{-1/2}h_1^{-1/2}),\end{aligned}$$

respectively. ■

2.C Tables & Figures

Table 2.1: Median of Monte Carlo fit criteria over grid for Design 1.

$\widehat{M}$ vs M		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2491	0.4526	0.2027	0.3119	0.3511	0.6744	0.2855	0.4795
	600	0.1299	0.2101	0.1067	0.1669	0.1826	0.3230	0.1493	0.2412
1	150	0.2491	0.4073	0.2028	0.2942	0.3512	0.6486	0.2854	0.4606
	600	0.1299	0.2026	0.1067	0.1592	0.1827	0.3028	0.1494	0.2264
1.5	150	0.2470	0.3888	0.2004	0.2854	0.3471	0.6292	0.2821	0.4405
	600	0.1250	0.1860	0.1024	0.1473	0.1752	0.2631	0.1434	0.2025

$\widehat{G}$ vs G		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1600	0.2863	0.1301	0.2420	0.2243	0.4374	0.1833	0.3415
	600	0.0854	0.1546	0.0697	0.1334	0.1192	0.2173	0.0972	0.1849
1	150	0.1600	0.2651	0.1301	0.2267	0.2242	0.3821	0.1832	0.3116
	600	0.0855	0.1467	0.0696	0.1247	0.1191	0.2028	0.0973	0.1731
1.5	150	0.1583	0.2558	0.1291	0.2170	0.2223	0.3988	0.1808	0.3198
	600	0.0820	0.1401	0.0671	0.1201	0.1136	0.1890	0.0926	0.1626

$\widehat{F}$ vs F		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1573	0.2359	0.1289	0.1878	0.2220	0.3223	0.1826	0.2532
	600	0.0815	0.1118	0.0661	0.0917	0.1153	0.1744	0.0937	0.1428
1	150	0.1571	0.2113	0.1289	0.1709	0.2221	0.2978	0.1823	0.2314
	600	0.0817	0.1050	0.0662	0.0865	0.1154	0.1624	0.0937	0.1323
1.5	150	0.1554	0.2028	0.1279	0.1608	0.2195	0.2849	0.1801	0.2262
	600	0.0775	0.0944	0.0628	0.0768	0.1094	0.1401	0.0886	0.1124

$\widehat{H}$ vs H		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2491	0.2874	0.2027	0.2241	0.3511	0.3939	0.2855	0.3089
	600	0.1299	0.1607	0.1067	0.1263	0.1826	0.2275	0.1493	0.1786
1	150	0.2491	0.2876	0.2028	0.2232	0.3512	0.3958	0.2854	0.3114
	600	0.1299	0.1607	0.1067	0.1262	0.1827	0.2270	0.1494	0.1784
1.5	150	0.2470	0.2816	0.2004	0.2178	0.3471	0.3908	0.2821	0.3044
	600	0.1250	0.1474	0.1024	0.1142	0.1752	0.2082	0.1434	0.1611

Table 2.2: Median of Monte Carlo fit criteria over grid for Design 2.

$\widehat{M}$ vs M		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2836	0.5179	0.2297	0.3518	0.4210	0.7313	0.3378	0.5354
	600	0.1421	0.2325	0.1148	0.1800	0.2023	0.3690	0.1643	0.2644
1	150	0.2834	0.4762	0.2297	0.3393	0.4206	0.7035	0.3382	0.5139
	600	0.1422	0.2172	0.1149	0.1710	0.2023	0.3334	0.1640	0.2454
1.5	150	0.2804	0.4545	0.2268	0.3245	0.4155	0.6845	0.3344	0.5064
	600	0.1363	0.1994	0.1109	0.1574	0.1923	0.2877	0.1566	0.2187

$\widehat{G}$ vs G		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1801	0.3243	0.1461	0.2702	0.2661	0.4954	0.2141	0.4090
	600	0.0925	0.1696	0.0752	0.1426	0.1301	0.2391	0.1058	0.1999
1	150	0.1800	0.3052	0.1461	0.2559	0.2663	0.4605	0.2141	0.3677
	600	0.0925	0.1588	0.0752	0.1334	0.1302	0.2218	0.1061	0.1842
1.5	150	0.1778	0.2939	0.1442	0.2495	0.2614	0.4745	0.2112	0.3775
	600	0.0887	0.1491	0.0723	0.1260	0.1243	0.2025	0.1014	0.1732

$\widehat{F}$ vs F		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1801	0.2553	0.1440	0.2027	0.2679	0.3525	0.2119	0.2812
	600	0.0908	0.1261	0.0741	0.1034	0.1300	0.1926	0.1061	0.1533
1	150	0.1800	0.2371	0.1439	0.1922	0.2678	0.3215	0.2124	0.2601
	600	0.0909	0.1153	0.0742	0.0934	0.1301	0.1734	0.1060	0.1392
1.5	150	0.1774	0.2272	0.1410	0.1822	0.2625	0.3105	0.2086	0.2533
	600	0.0861	0.1038	0.0701	0.0847	0.1229	0.1557	0.0999	0.1262

$\widehat{H}$ vs H		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2582	0.2875	0.2098	0.2240	0.3717	0.3910	0.3030	0.3083
	600	0.1321	0.1625	0.1069	0.1272	0.1869	0.2298	0.1525	0.1794
1	150	0.2582	0.2870	0.2099	0.2228	0.3717	0.3948	0.3031	0.3101
	600	0.1323	0.1626	0.1070	0.1273	0.1870	0.2283	0.1526	0.1788
1.5	150	0.2556	0.2818	0.2077	0.2183	0.3673	0.3890	0.2992	0.3038
	600	0.1263	0.1483	0.1032	0.1146	0.1777	0.2087	0.1457	0.1608

Table 2.3: Parametric General Production Function Estimates (P1)

Industry	θ_0		θ_1		θ_2		θ_3		θ_4		θ_5		γ	
	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001
Chemical	9.622	9.771	0.452	0.498	0.932	0.799	0.064	0.092	0.047	0.001	0.053	0.050	-0.028	-0.022
	(0.043)	(0.036)	(0.041)	(0.032)	(0.056)	(0.028)	(0.046)	(0.022)	(0.046)	(0.025)	(0.027)	(0.015)	(0.038)	(0.001)
Iron	10.148	10.191	0.835	0.434	1.028	1.280	0.053	0.027	-0.162	0.251	0.060	-0.073	-0.032	0.301
	(0.122)	(0.390)	(0.113)	(0.133)	(0.146)	(0.370)	(0.116)	(0.071)	(0.069)	(0.105)	(0.047)	(0.085)	(0.098)	(0.253)
Petroleum	11.090	11.337	0.724	0.809	1.419	1.252	-0.224	0.060	0.272	0.030	-0.120	-0.079	0.061	0.120
	(0.217)	(0.290)	(0.183)	(0.101)	(0.339	(0.440	(0.186)	(0.071)	(0.175)	(0.098)	(0.114)	(0.160)	(0.118)	(0.194)
Transportation	8.963	9.630	0.493	0.493	1.538	1.071	-0.005	0.172	0.259	0.116	-0.084	0.023	0.546	0.118
	(0.588)	(0.150)	(0.147)	(0.048)	(0.512	(0.165	(0.070)	(0.037)	(0.117)	(0.053)	(0.117)	(0.049)	(0.389)	(0.124)

Table 2.4: Parametric Generalized Homothetic Estimates (P2)

Industry	α		β_0		β_1		β_2		γ	
	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.500	0.605	9.609	9.736	0.926	0.877	0.053	0.024	-0.025	0.077
(Chemical)	(0.050)	(0.064)	(0.037)	(0.070)	(0.055)	(0.072)	(0.021)	(0.018)	(0.029)	(0.058)
Smelting and processing	0.563	0.653	9.634	10.518	1.570	0.816	-0.066	0.059	0.399	-0.019
of ferrous metals (Iron)	(0.130)	(0.112)	(0.417)	(0.082)	(0.329)	(0.071)	(0.058)	(0.020)	(0.246)	(0.028)
Petroleum processing	0.823	0.783	10.860	11.352	1.099	1.014	0.013	0.033	-0.054	0.004
(Petroleum)	(0.177)	(0.155)	(0.102)	(0.128)	(0.128)	(0.131)	(0.031)	(0.032)	(0.026)	(0.034)
Transportation equipment	0.689	0.601	9.605	9.819	0.923	0.911	0.075	0.073	0.017	0.021
(Transportation)	(0.092)	(0.061)	(0.090)	(0.060)	(0.089)	(0.059)	(0.027)	(0.019)	(0.079)	(0.043)

Table 2.5: Parametric Translog Estimates (P3)

Industry	α		β_0		β_1		β_2	
	1995	2001	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.479	0.696	9.585	9.815	0.961	0.783	0.045	0.036
(Chemical)	(0.041)	(0.040)	(0.026)	(0.032)	(0.035)	(0.028)	(0.019)	(0.013)
Smelting and processing	0.932	0.621	10.262	10.499	1.025	0.847	0.017	0.054
of ferrous metals (Iron)	(0.095)	(0.098)	(0.064)	(0.079)	(0.055)	(0.051)	(0.019)	(0.018)
Petroleum processing	0.674	0.795	10.810	11.351	1.271	1.001	-0.020	0.036
(Petroleum)	(0.121)	(0.125)	(0.100)	(0.125)	(0.093)	(0.081)	(0.031)	(0.028)
Transportation equipment	0.705	0.626	9.623	9.839	0.905	0.883	0.078	0.079
(Transportation)	(0.062)	(0.045)	(0.038)	(0.038)	(0.046)	(0.031)	(0.024)	(0.016)

Table 2.6: Average Substitutability, $T(k, L)$.

Industry	P2		P3		NP	
	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.728	0.384	0.738	0.390	0.795	1.328
(Chemical)	(0.042)	(0.135)	–	–	(0.135)	(0.762)
Smelting and processing	0.163	0.474	0.081	0.480	0.383	1.188
of ferrous metals (Iron)	(0.376)	(0.108)	–	–	(0.510)	(0.244)
Petroleum processing	0.286	0.234	0.442	0.228	0.451	0.255
(Petroleum)	(0.165)	(0.023)	–	–	(0.259)	(0.158)
Transportation equipment	0.351	0.473	0.352	0.470	0.147	0.834
(Transportation)	(0.020)	(0.049)	–	–	(0.321)	(0.842)

Table 2.7: Average Return to Scale, $RTS(M, L)$.

Industry	P2		P3		NP	
	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.962	0.796	0.968	0.799	1.016	0.752
(Chemical)	(0.074)	(0.120)	(0.073)	(0.077)	(0.166)	(0.181)
Smelting and processing	1.186	0.881	1.035	0.881	1.034	0.828
of ferrous metals (Iron)	(0.428)	(0.178)	(0.049)	(0.179)	0.373)	(0.201)
Petroleum processing	1.231	1.020	1.258	1.016	1.162	0.946
(Petroleum)	(0.279)	(0.124)	(0.060)	(0.126)	(0.173)	(0.147)
Transportation equipment	0.934	0.901	0.932	0.896	0.944	0.964
(Transportation)	(0.153)	(0.183)	(0.152)	(0.182)	(0.291)	(0.328)

Figure 2.1: $Q-Q$ plots for G, F and H.

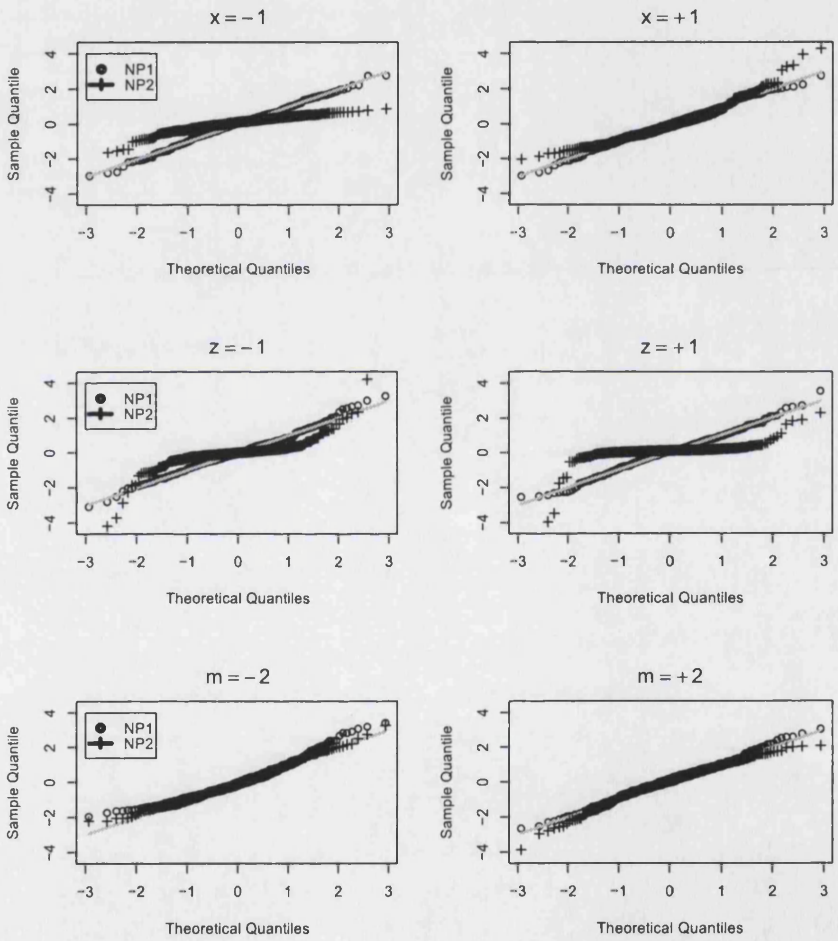
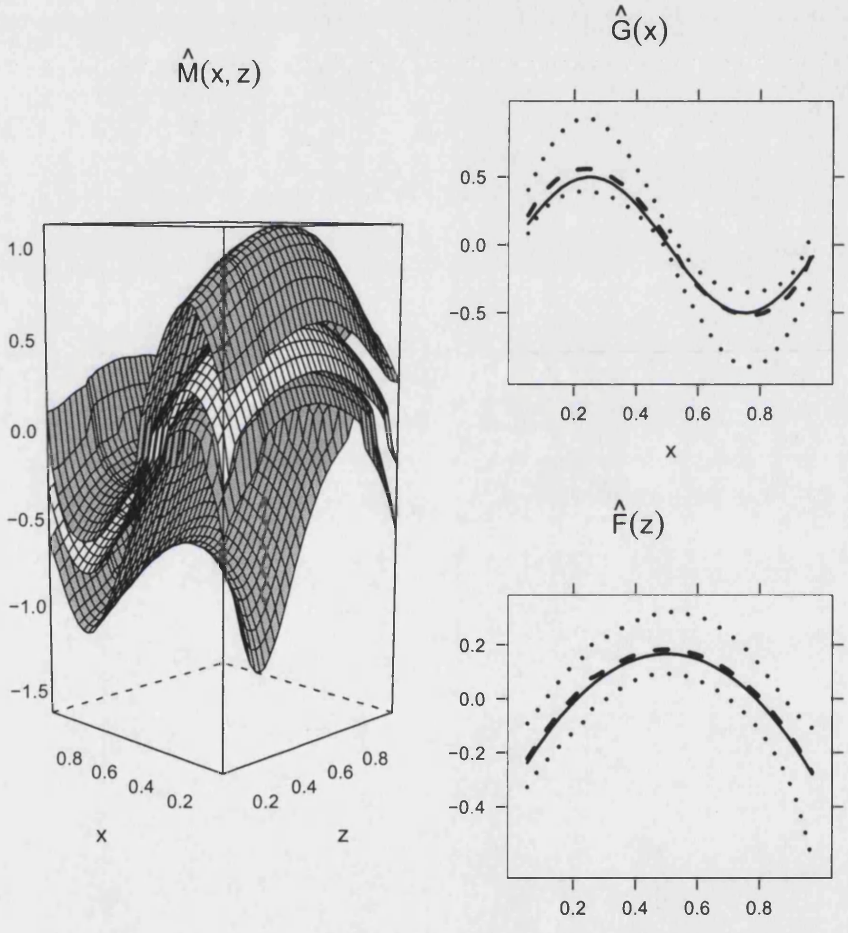


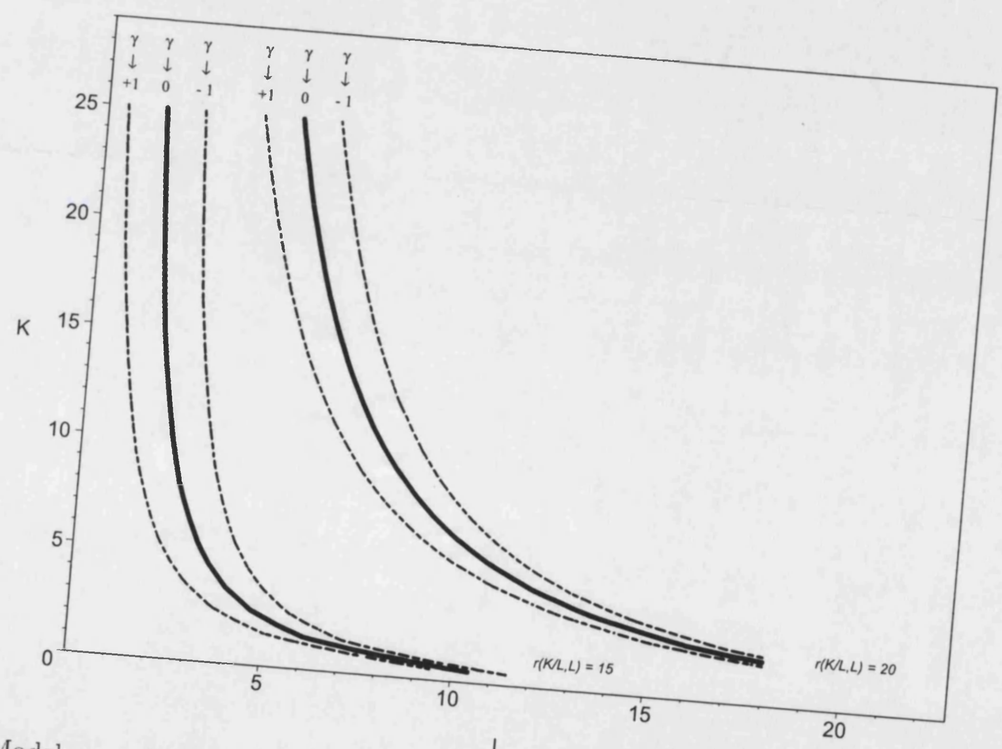
Figure 2.2: Simulation Envelopes for M, G and F.



[a] $h_1 = 0.15$, $h_2 = 0.7$, $p_1 = 3$, $p_2 = 1$, and $n = 400$.

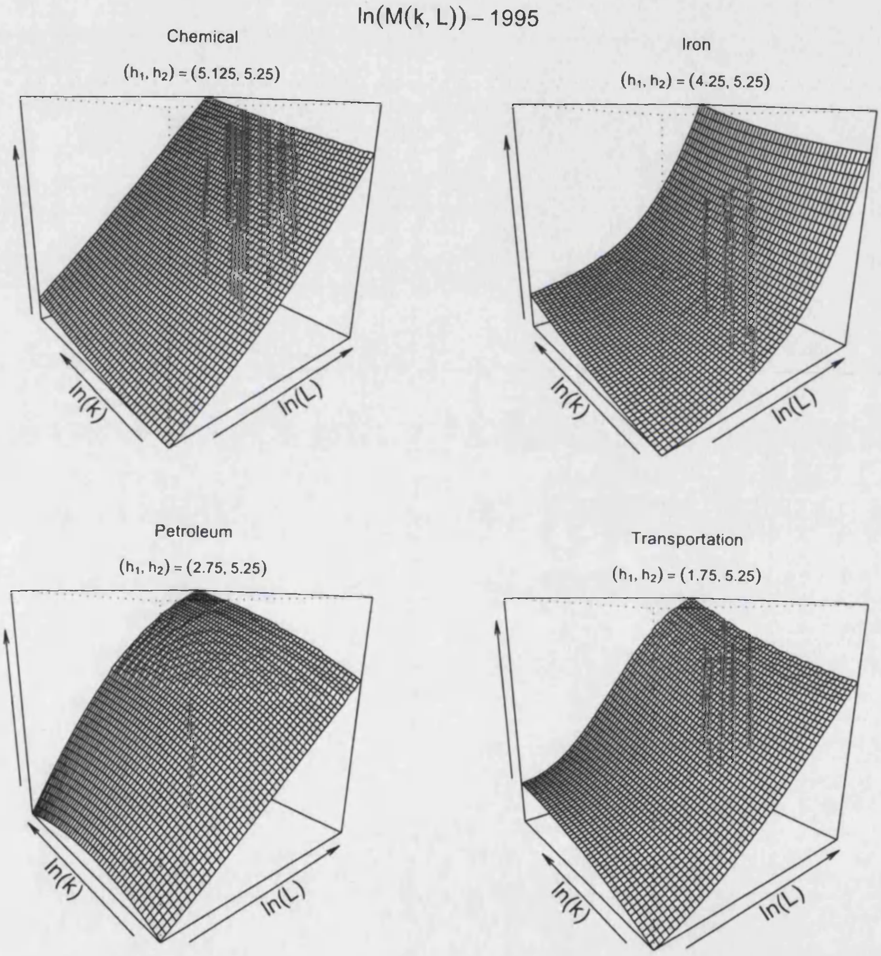
[b] White planes and dashed lines represent medians. Gray planes and dotted lines represent 90% simulation envelopes.

Figure 2.3: Generalized Homothetic Translog Isoquants.



[a] Model: $r_{\psi_{P2}}(k, L) = 10 + (1/2)\ln(M) + [\ln(M)]^2$, with $M(k, L) = k^{1/2}(L + \gamma)$

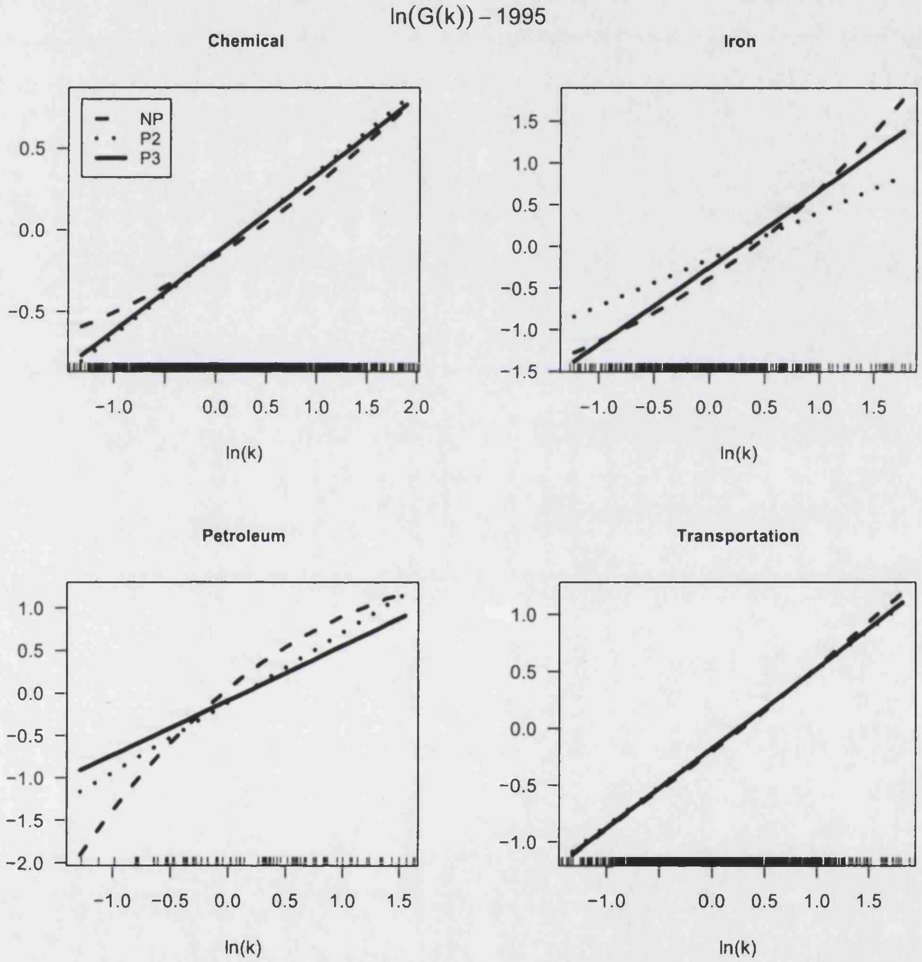
Figure 2.4: Generalized Homogeneous M (1995)



[a] Raw Chemical Materials and Chemical Products: 1560 plants; Iron and Steel: 376 plants; Petroleum Processing and Coking: 93 plants; Transportation Equipment Manufacturing: 989 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

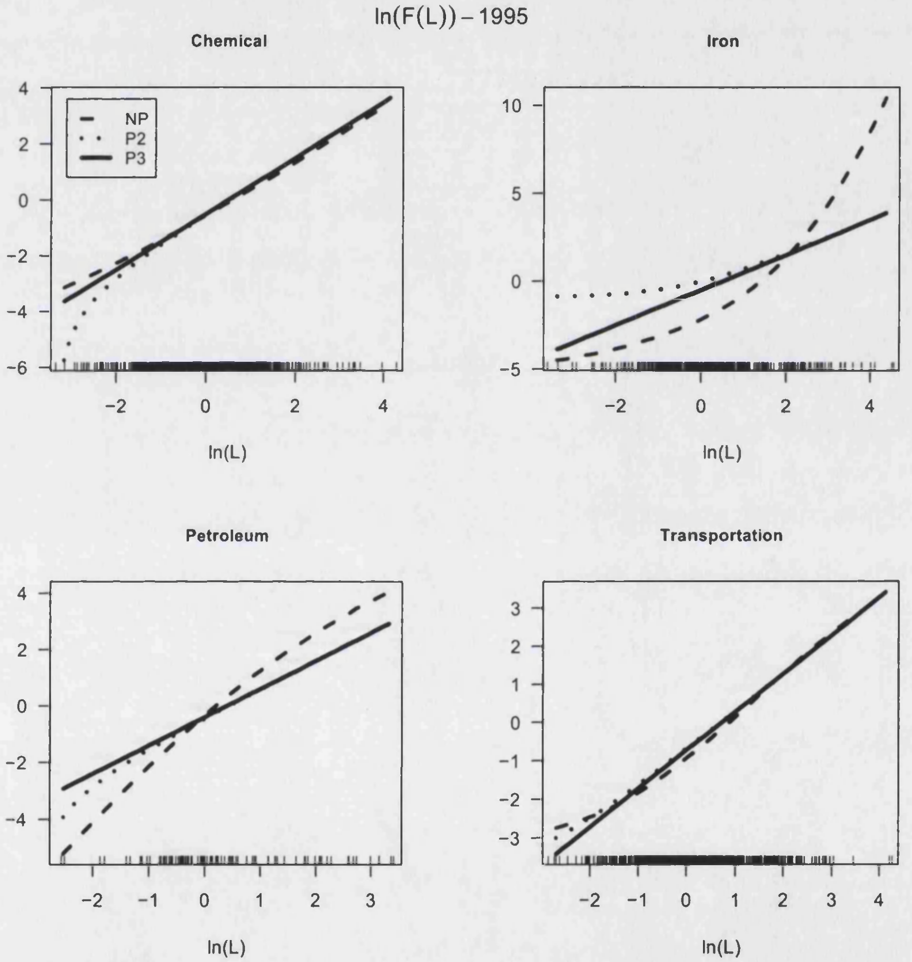
Figure 2.5: Generalized Homogeneous Component G (1995)



[a] Raw Chemical Materials and Chemical Products: 1560 plants; Iron and Steel: 376 plants; Petroleum Processing and Coking: 93 plants; Transportation Equipment Manufacturing: 989 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

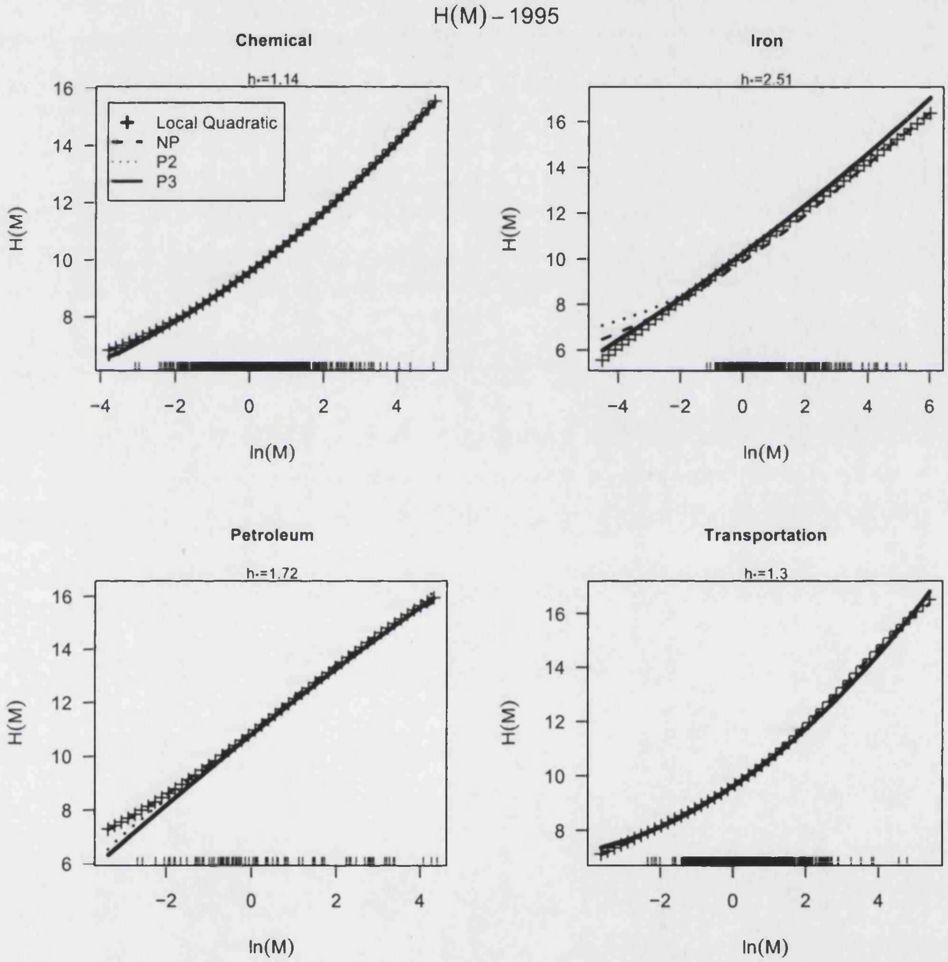
Figure 2.6: Generalized Homogeneous Component F (1995)



[a] Raw Chemical Materials and Chemical Products: 1560 plants; Iron and Steel: 376 plants; Petroleum Processing and Coking: 93 plants; Transportation Equipment Manufacturing: 989 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

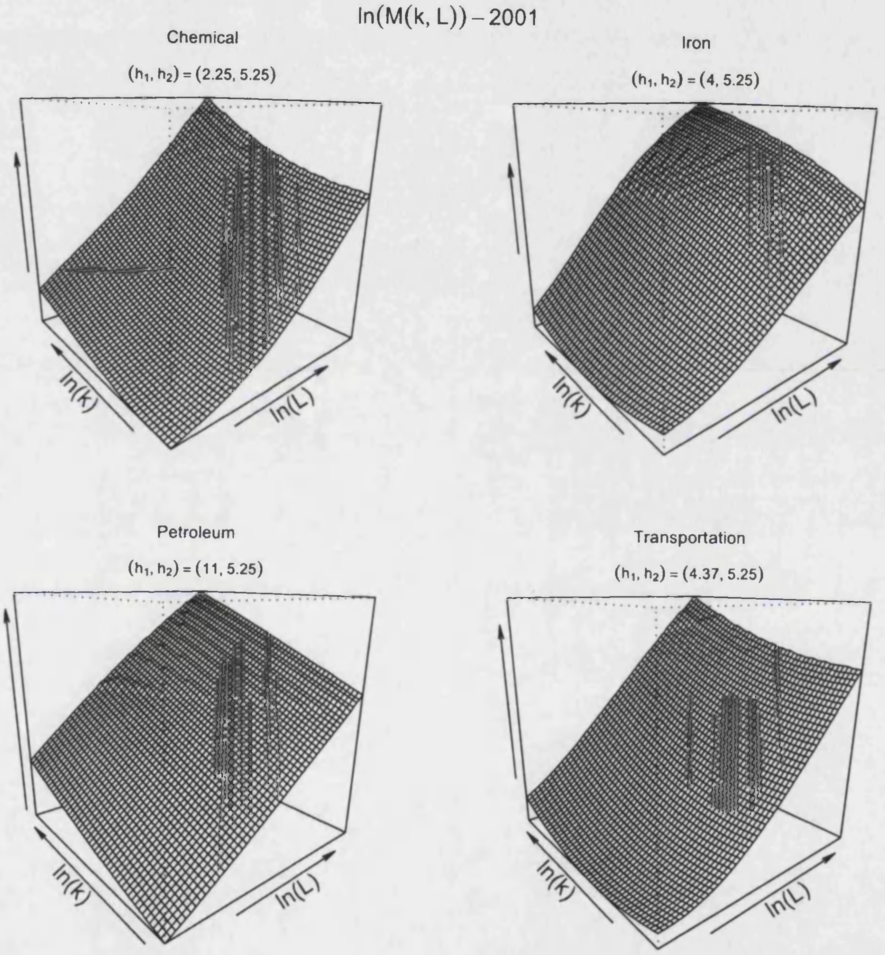
Figure 2.7: Strictly Monotonic Component H (1995)



[a] Raw Chemical Materials and Chemical Products: 1560 plants; Iron and Steel: 376 plants; Petroleum Processing and Coking: 93 plants; Transportation Equipment Manufacturing: 989 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

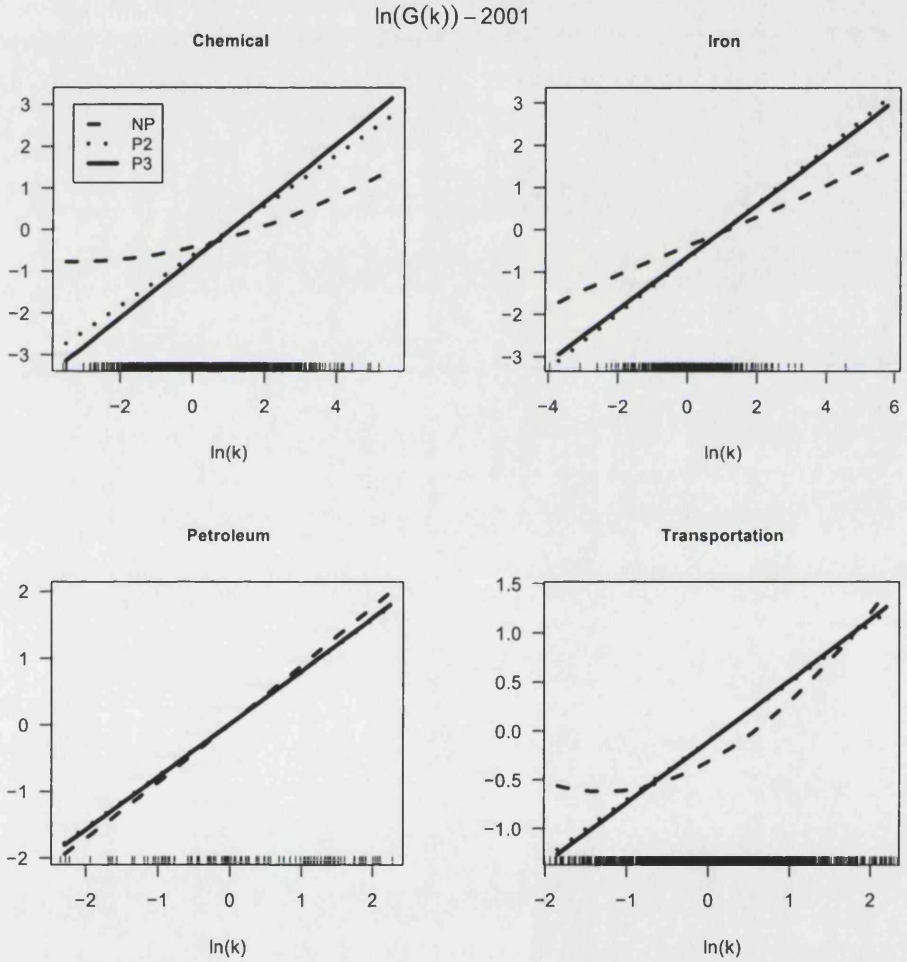
Figure 2.8: Generalized Homogeneous M (2001)



[a] Raw Chemical Materials and Chemical Products: 1637 plants; Iron and Steel: 341 plants; Petroleum Processing and Coking: 119 plants; Transportation Equipment Manufacturing: 1230 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

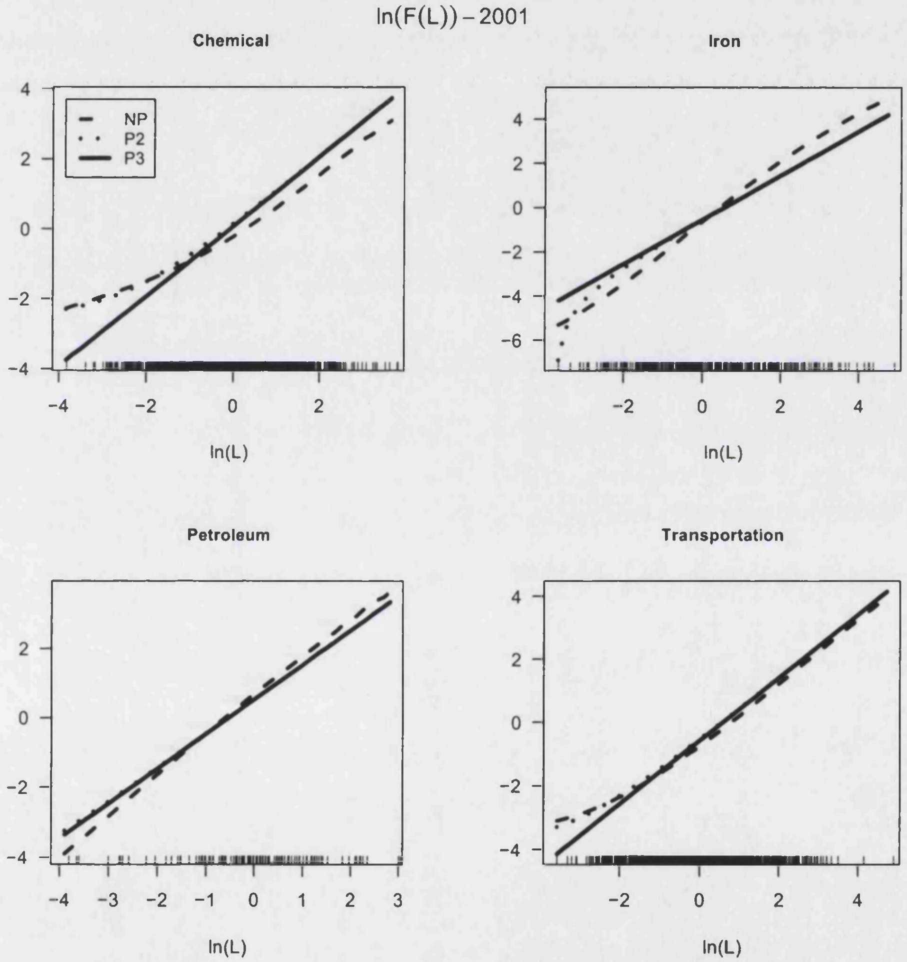
Figure 2.9: Generalized Homogeneous Component G (2001)



[a] Raw Chemical Materials and Chemical Products: 1637 plants; Iron and Steel: 341 plants; Petroleum Processing and Coking: 119 plants; Transportation Equipment Manufacturing: 1230 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

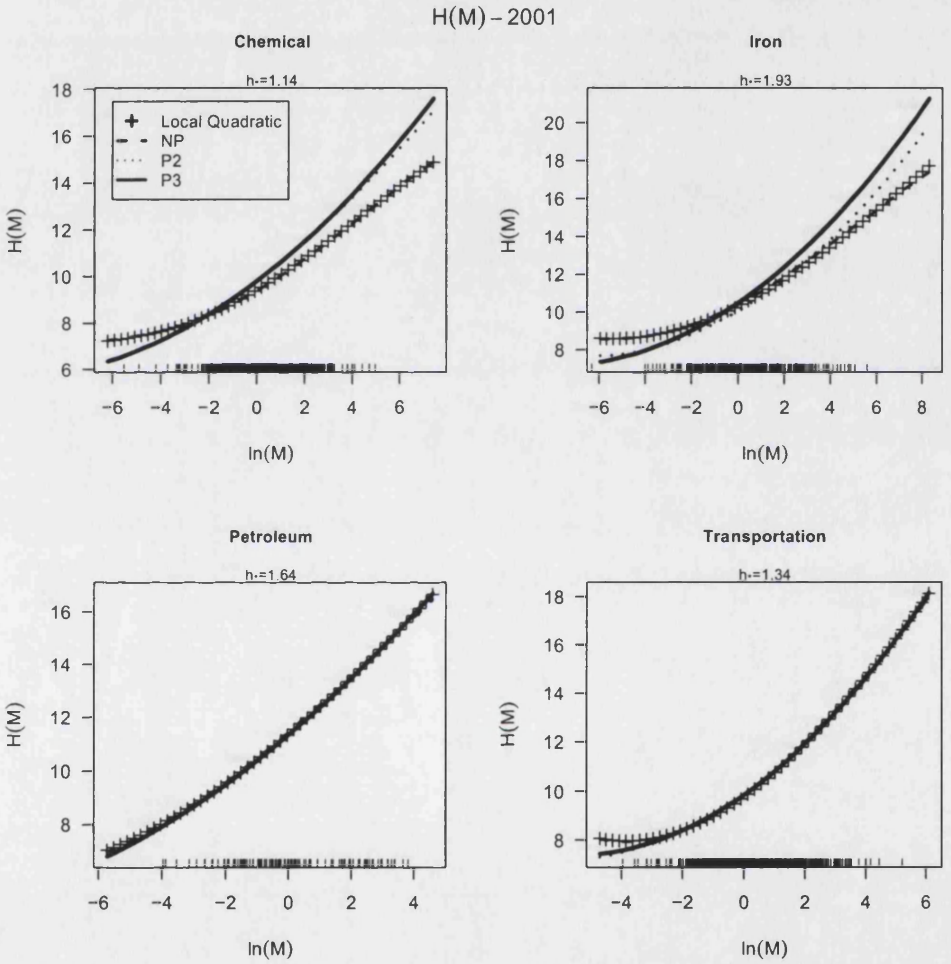
Figure 2.10: Generalized Homogeneous Component F (2001)



[a] Raw Chemical Materials and Chemical Products: 1637 plants; Iron and Steel: 341 plants; Petroleum Processing and Coking: 119 plants; Transportation Equipment Manufacturing: 1230 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

Figure 2.11: Strictly Monotonic Component H (2001)



[a] Raw Chemical Materials and Chemical Products: 1637 plants; Iron and Steel: 341 plants; Petroleum Processing and Coking: 119 plants; Transportation Equipment Manufacturing: 1230 plants.

[b] Data Source: Jefferson, Hu, Guan, and Yu (2003).

Chapter 3

Efficiency Bounds in Semiparametric Models defined by Moment Restrictions using an Estimated Conditional Probability Density

3.1 Introduction

The main objective of this chapter is to derive efficiency bounds (minimum asymptotic variance) for estimating some unique finite–dimensional parameter π_0 in an important class of econometric models. These models satisfy the following unconditional moment restriction:

$$E\left[\mathbf{g}\left(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2;\pi_0,f_{\mathbf{w}_1|\mathbf{w}_2}\right)\right]=\mathbf{0}, \tag{3.1.1}$$
$$\mathbf{g}\left(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2;\pi_0,f_{\mathbf{w}_1|\mathbf{w}_2}\right)=\frac{\mathbf{m}\left(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2;\pi_0\right)}{f_{\mathbf{w}_1|\mathbf{w}_2}\left(\mathbf{w}_1|\,\mathbf{w}_2\right)}-\mathbf{s}\left(\pi_0\right),$$

where $\mathbf{m}(\cdot)$ and $\mathbf{s}(\cdot)$ are known vector–valued functions of an observed random vector $\left(\mathbf{y},\mathbf{w}_1^{\top},\mathbf{w}_2^{\top}\right)$ with joint density $f_{\mathbf{y},\mathbf{w}_1,\mathbf{w}_2}\left(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2\right)$. The function $f_{\mathbf{w}_1|\mathbf{w}_2}\left(\mathbf{w}_1|\,\mathbf{w}_2\right)$, although unknown, represents the conditional probability density of $\mathbf{w}_1$ given $\mathbf{w}_2$.

Models defined by restrictions such as (3.1.1) are examples of distribution–free[1] models, where the parameter space contains a finite–dimensional component, π_0, as well as an

[1] Those in which the distribution of unobserved error terms is unknown.

infinite–dimensional one, $f_{\mathbf{w}_1|\mathbf{w}_2}$. Chamberlain (1992) and Ai and Chen (2003) obtained the semiparametric efficiency bound for regular estimators of the finite–dimensional component in such models. Our analysis is different from theirs in that they did not impose any particular structure for the unknown function in their moment restrictions, such unknown function happens to be a conditional density function in (3.1.1). Here, we make full use of this extra restriction in our calculations. A commonality within the literature of semiparametric efficiency bounds derivation is that this type of extra information would typically lead to different conclusions[2].

Unconditional moment restrictions such as (3.1.1) are induced by the identification of Limited Dependent Variable[3] (LDV) models. We illustrate this using three leading examples:

Model 1: (General Limited Dependent Variable Model) Let $\left(y, v, \mathbf{x}^{\top}, \mathbf{z}^{\top}\right)$ be an observed random vector generated from the model $y = L\left(\alpha_0 v + \mathbf{x}^{\top}\beta_0 + \varepsilon\right)$, where L is a known or unknown function, and ε is an unobserved error term. Lewbel (1998) showed that the following unconditional moment is sufficient to identify $\left(\alpha_0, \beta_0^{\top}\right)$:

$$E\left[\begin{array}{c} \alpha_0 v\theta(y)/f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x},\mathbf{z}) \\ \mathbf{z}\alpha_0^2 v\theta(y)/f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x},\mathbf{z}) \end{array}\right] = \left[\begin{array}{l} \kappa_1 \\ \kappa_2 E[\mathbf{z}] - \kappa_1 E\left[\mathbf{z}\mathbf{x}^{\top}\right]\beta_0 \end{array}\right], \tag{3.1.2}$$

where θ is a known real–valued function, with constants $\kappa_1 = \int_{-\infty}^{\infty}\theta\left[L(a)\right]da$ and $\kappa_2 = \int_{-\infty}^{\infty} a\theta\left[L(a)\right]da$. The 'special' regressor v is assumed to have large support and, conditional on $(\mathbf{x}^{\top}, \mathbf{z}^{\top})$, not to affect the distribution of ε. The error ε is also assumed to be uncorrelated with variables $\mathbf{z}$, i.e. $E[\mathbf{z}\varepsilon] = \mathbf{0}$.

Model 2: (Selection or Treatment Model) Let $\left(y, d, v, \mathbf{x}^{\top}, \mathbf{z}^{\top}\right)$ be a random vector from a data generating process (dgp) where $y = y^* \times d$, with $d = 1\left(0 \le v + M^* \le c\right)$, for some known $c < \infty$, and unobserved random scalars y^* and M^*. Furthermore, for some finitely–parameterized (by $\theta \in \Theta$) known vector–value function $\Psi(\cdot)$, the unconditional moment $E\left[\Psi\left(y^*, v, \mathbf{x}^{\top}, \mathbf{z}^{\top}; \theta_0\right)\right] = 0$ holds. Then Lewbel (2006) showed that the following unconditional moment is also satisfied:

$$E\left[\Psi\left(y, v, \mathbf{x}^{\top}, \mathbf{z}^{\top}; \theta_0\right) d/f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x},\mathbf{z})\right] = \mathbf{0}, \tag{3.1.3}$$

for the same unknown $\theta_0 \in \Theta$, which is assumed to be unique. The 'special' variable v is again assumed to have large support and, conditional on $(\mathbf{x}^{\top}, \mathbf{z}^{\top})^{\top}$, not to affect the distribution of M^*.

[2]See Tripathi (2000) for an example where information that the unknown function has certain shape may lead to substantial asymptotic efficiency gains in estimating the parameter of interest.

[3]Loosely speaking, LDV models are where the data generating process (dgp) induces a probability distribution on the realized observations that differs from the underlying distribution, for which inferences are to be made.

Model 3: (Programme Evaluation) Let $\left(y, d, \mathbf{x}^{\top}\right)$ be an observation generated from a dgp where $y = y_1 d + y_0 (1-d)$, d equals one if an observation is treated and zero otherwise. Hirano, Imbens, and Ridder (2003) used an 'unconfoundedness' assumption, $d \perp\!\!\!\perp (y_1, y_0) | \mathbf{x}$, and the conditional probability, $p(\mathbf{x}) = \Pr(d = 1 | \mathbf{x})$, (also known as the propensity score) to construct a weighted moment condition

$$E\left[h(\mathbf{x})\left(\frac{yd}{p(\mathbf{x})} - \frac{y(1-d)}{1-p(\mathbf{x})} - \tau_0\right)\right] = 0, \tag{3.1.4}$$

to identify average treatment effects for different subgroups of the population by carefully choosing the weighting function h. If $h(\mathbf{x}) = 1$, the moment restriction above identifies $\tau_0 = E[y_1 - y_0]$, the average effect of the treatment in the population and if $h(\mathbf{x}) = p(\mathbf{x})$, it identifies $\tau_0 = E[y_1 - y_0 | d = 1]$, the effect on the treated.

A simple look at (3.1.2), (3.1.3), and (3.1.4) is sufficient to realize that all these unconditional moments are nested in (3.1.1). In particular, by setting $\mathbf{y} = y$, $\mathbf{w}_1 = v$, and $\mathbf{w}_2 = (\mathbf{x}^{\top}, \mathbf{z}^{\top})^{\top}$ we obtain Model 1. Likewise, setting $\mathbf{y} = (y, d)^{\top}$, $\mathbf{w}_1 = v$, and $\mathbf{w}_2 = (\mathbf{x}^{\top}, \mathbf{z}^{\top})^{\top}$ gives us Model 2, and setting $\mathbf{y} = y$, $\mathbf{w}_1 = d$, and $\mathbf{w}_2 = \mathbf{x}$ will give us Model 3. That is, Models 1, 2 and 3 belong to the same class[4].

The regularity conditions behind the existence of moments (3.1.2), (3.1.3), and (3.1.4) are similar in nature. They are explained in great detail by their proponents: Lewbel (1998), Lewbel (2006), and Hirano, Imbens, and Ridder (2003) respectively. For example the 'special' regressors, v in Models 1 and 2, and d in Model 3, are assumed to be redundant in explaining the realized y once conditioned on $\mathbf{x}$ and $\mathbf{z}$. They are also assumed to have some special properties ensuring that moments such as (3.1.1) exist. For instance, the random variable v is assumed to have a conditional distribution that is absolutely continuous with respect to a Lebesgue measure with nondegenerate Radon-Nikodym conditional density $f(v | \mathbf{x}, \mathbf{z})$, which is bounded away from zero[5]. The programme participation indicator d, in Model 3, is such that $0 < p(\mathbf{x}) < 1$ almost everywhere on the support of $\mathbf{x}$.

The advantage of looking at an object such as (3.1.1) is that it will allow us to unify the efficiency–bound–derivation theory for Models 1, 2, and 3. Similar to the lower bound for Generalized Method of Moment (GMM) estimators, the derived efficiency bound in this chapter shares its form, and it has the desireable property of being easy to compute.

[4] Other examples of models defined by this type of unconditional moment restrictions are Lewbel (2000b), Honoré and Lewbel (2002), and Khan and Lewbel (2006).

[5] The use of density functions has a long history of aiding identification in LDV models. Examples range from the fully parametric maximum likelihood technique assuming normality of errors, to the use of likelihood–based semiparametric estimators such as Cosslett (1983), Gabler, Laisney, and Lechner (1993), Gozalo and Linton (2000) in the Binary Choice models. Other density–based estimators in Censored linear regression models are for example Buckley and James (1979), Horowitz (1986), Moon (1989), Horowitz (1988), Powell, Stock, and Stoker (1989) and Ichimura (1993), among others.

The bound is also sharp. Estimators that use plug–in kernel estimators of the conditional density in (3.1.2), and (3.1.3) are shown to achieve these bounds. On the other hand, for Model 3, we are also able to show that the efficiency bound for any regular estimator of τ_0 (which is based on (3.1.4)) coincides with that derived by Hahn (1998). Hirano, Imbens, and Ridder (2003) proved that a semiparametric estimator based on (3.1.4) achieves Hahn (1998)'s bound. They also found that using an estimate rather than the true propensity score is efficient.

Semiparametric estimators of LDV models are attractive because they can remain consistent in situations where a fully parametric estimator is not. However, this robustness property comes at a price in terms of asymptotic variance. In this sense, semiparametric efficiency bounds are of fundamental importance because they quantify the efficiency loss that can result from the use of a semiparametric, rather than a parametric estimator. These bounds are also used here to prove that using an estimate rather than the true conditional density in (3.1.1) is more efficient. The result presented in this paper also generalizes earlier results of Hirano, Imbens, and Ridder (2003) and Magnac and Maurin (2004). We further explore this in a small Monte Carlo experiment, where we also compare the performance of three different kernel–based estimators and assess their relative efficiency.

The chapter is organized as follows: We derive the semiparametric efficiency bound of regular estimators of π_0 based on a particular functional form of (3.1.1) in Section 3.2 where we also prove that estimators that use the true conditional density are inefficient compared with those using an estimated conditional density instead. Section 3.3 applies the efficiency bound to Models 1, 2 and 3, and shows that the estimator proposed by Lewbel (1998), and Lewbel (2006) are semiparametric efficient. A direct link with Hirano, Imbens, and Ridder (2003) is also discussed. Section 3.4 describes the results of a simulation experiment and Section 3.5 concludes.

3.2 Efficiency and Information

The framework is as follows: The information on one unit in a random sample is contained in a $d \times 1$ vector $\mathbf{u}^{\top} = \left(\mathbf{y}^{\top}, \mathbf{w}_1^{\top}, \mathbf{w}_2^{\top}\right) \in \mathbf{U}$ with unknown density $f\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2\right)$ with respect to a dominating measure $\mu = \mu_{\mathbf{y}} \times \mu_{\mathbf{w}_1} \times \mu_{\mathbf{w}_2} = \left(\otimes_{i=1}^{d_y} \mu_{y_i}\right) \times \left(\otimes_{i=1}^{d_1} \mu_{w_{1i}}\right) \times \left(\otimes_{i=1}^{d_2} \mu_{w_{2i}}\right)$, where $d = d_y + d_1 + d_2$. Since $\mathbf{u}$ can have discrete components, the μ_{y_i}'s, $\mu_{w_{1i}}$'s and $\mu_{w_{2i}}$'s need not be Lebesgue measures. Furthermore, f can be decomposed into conditional probability densities with respect to measures $\mu_{\mathbf{y}}$, $\mu_{\mathbf{w}_1}$and $\mu_{\mathbf{w}_2}$; that is, if $\mathbf{g}$: $\mathbf{U} \rightarrow \Re^q$ is an integrable function, then

$$\int \mathbf{g}(\mathbf{u}) f(\mathbf{u}) \mu(d\mathbf{u}) = \int\int\int \mathbf{g}(\mathbf{u}) f_{\mathbf{y}|\mathbf{w}}(\mathbf{y}|\,\mathbf{w}) \mu_{\mathbf{y}}(d\mathbf{y}) \times$$
$$f_{\mathbf{w}_1|\mathbf{w}_2}(\mathbf{w}_1|\,\mathbf{w}_2) \mu_{\mathbf{w}_1}(d\mathbf{w}_1) f_{\mathbf{w}_2}(\mathbf{w}_2) \mu_{\mathbf{w}_2}(d\mathbf{w}_2), \quad (3.2.1)$$

where $f_{\mathbf{y}|\mathbf{w}}$, $f_{\mathbf{w}_1|\mathbf{w}_2}$ and $f_{\mathbf{w}_2}$ are the implied densities with respect to the conditional measures $\mu_{\mathbf{y}}$, $\mu_{\mathbf{w}_1}$ and $\mu_{\mathbf{w}_2}$ respectively. For a given open set $\Pi \subset \Re^p$, define functions $\mathbf{m}$: $\mathbf{U} \times \Pi \rightarrow \Re^q$ and s: $\Pi \rightarrow \Re^q$ such that for each $\pi \in \Pi$, $\mathbf{m}(\cdot;\pi)$: $\mathbf{U} \rightarrow \Re^q$ is measurable, and for each $\mathbf{u} \in \mathbf{U}$, $\nabla_\pi \mathbf{m}(\mathbf{u};\pi)$ and $\nabla_\pi \mathbf{s}(\pi)$ are continuous mappings on Π. The restrictions on $f_{\mathbf{y}|\mathbf{w}}$, $f_{\mathbf{w}_1|\mathbf{w}_2}$ and $f_{\mathbf{w}_2}$ also involve the following unconditional moment:

$$E\left[\frac{\mathbf{m}(\mathbf{u};\pi_0)}{f_{\mathbf{w}_1|\mathbf{w}_2}(\mathbf{w}_1|\,\mathbf{w}_2)} - \mathbf{s}(\pi_0)\right] = 0, \quad (3.2.2)$$

where π_0 is some point in Π, and $f_{\mathbf{w}_1|\mathbf{w}_2}$ is the true (conditional) density of $\mathbf{w}_1$ given $\mathbf{w}_2$. This notation will become clearer in the proof of the following theorem.

Theorem 3.2.1 *Assume that the true distribution of* $\mathbf{u}$ *satisfies* (3.2.2) *for a unique value* $\pi_0 \in \Pi$, *where* $\Pi \subset \mathbb{R}^p$, *and that the matrix* $E\left[\mathbf{g}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)\mathbf{g}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)^\top \middle|\, \mathbf{w}_1, \mathbf{w}_2\right]$ *is finite and nonsingular with probability one, then the semiparametric efficiency bound for regular estimators of* π_0 *is given by*

$$(\mathbf{M}_0^\top \widetilde{\mathbf{\Omega}}_0^{-1} \mathbf{M}_0)^{-1}, \quad (3.2.3)$$

where

$$\mathbf{g}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right) = \frac{\mathbf{m}(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2;\pi_0)}{f_{\mathbf{w}_1|\mathbf{w}_2}(\mathbf{w}_1|\,\mathbf{w}_2)} - \mathbf{s}(\pi_0),$$
$$\widetilde{\mathbf{g}}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right) = \mathbf{g} - E(\mathbf{g}|\,\mathbf{w}_1,\mathbf{w}_2) + E(\mathbf{g}|\,\mathbf{w}_2),$$
$$\mathbf{M}_0 = E\left[\nabla_\pi \mathbf{g}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)\right],$$
$$\widetilde{\mathbf{\Omega}}_0 = E\left[\widetilde{\mathbf{g}}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)\widetilde{\mathbf{g}}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)^\top\right].$$

Proof. See Appendix. ■

These matrices can easily be estimated by replacing population moments with their sample counterparts and unknown functions with their nonparametric estimates. The function $-\mathbf{M}_0\widetilde{\mathbf{g}}\left(\mathbf{u};\pi_0, f_{\mathbf{w}_1|\mathbf{w}_2}\right)$ is an influence function, as Hampel (1974) defined.

It is worth noticing that the bound (3.2.3) characterizes the variance of an optimal regular estimator of π_0 defined by a set of assumed unconditional moment restrictions. Therefore, they can be interpreted as in the standard GMM case. These restrictions are not assumed to produce estimators of π_0 that are optimal under different identification criteria other than (3.2.2). This means that it is possible to improve upon this estimator by using (or assuming) extra information regarding the dgp, as in the generic GMM framework. However, we do not pursue this here.

Relative Efficiency

By the assumptions of the theorem above, $\mathbf{\Omega}_0 = E\left[\mathbf{g}\mathbf{g}^\top\right]$ is of full rank, q. If the conditional density $f_{\mathbf{w}_1|\mathbf{w}_2}$ was known, hence so it is no longer in the parameter space, the lower bound ($l.b.$) for the variance of regular estimators for π_0 based on the unconditional moment restriction (3.2.2) is given by that of the generic GMM bound, $(\mathbf{M}_0^\top \mathbf{\Omega}_0^{-1}\mathbf{M}_0)^{-1}$. However, we have previously calculated the bound when such density is unknown and it is given by $(\mathbf{M}_0^\top \widetilde{\mathbf{\Omega}}_0^{-1}\mathbf{M}_0)^{-1}$. The relationship between these two bounds is given by the following corollary:

Corollary 3.2.2 *The variance of the estimate of π_0, defined by the unconditional moment condition* (3.2.2), *when $f_{\mathbf{w}_1|\mathbf{w}_2}$ is known, is not smaller than that when $f_{\mathbf{w}_1|\mathbf{w}_2}$ is estimated.*

Proof. See Appendix. ■

Lewbel (2000b) first noticed this property through a Monte Carlo experiment of a binary choice estimator constructed using a moment condition such as (3.2.2). This conjecture was later proved by Magnac and Maurin (2004) for this particular case. They suggested the result may be better understood by using similar arguments to the ones proposed by Crépon, Kramarz, and Trognon (1998). Hirano, Imbens, and Ridder (2003) also report similar findings for their estimator based on (3.1.4). Using a similar canvas as in (3.1.4) in a missing data imputation model, Wang, Linton, and Härdle (2004) found that some of their estimators remained consistent although $\mathbf{m}$ may have been incorrectly specified.

As long as the conditional density can be consistently estimated when it is unknown, this knowledge does not affect the identification or estimation of the parameter of interest but its efficiency. This would highlight the fact that knowledge of $f_{\mathbf{w}_1|\mathbf{w}_2}$ is ancillary for consistent estimation of econometric models based on the above unconditional moment restriction. However, conjectures from Corollary 3.2.2 should be taken with caution. This 'loss' of efficiency while using the true density is among the class of consistent estimators that uses (3.2.2) as the only source of identification. The corresponding result does not pertain to estimators that would efficiently incorporate this new information. Therefore, Corollary 3.2.2 can be viewed as the consequence of the fact that estimators based on (3.2.2) do not necessarily make optimal use of this extra knowledge. Although it is rarely the case that a researcher has partial observability of the dgp in a data set, if such conditional density were known, the information may efficiently be used in this estimation framework while constructing the sample counterpart of (3.2.1).

A variety of efficiency results have also been found and estimators proposed in the literature for many LDV models. Examples are Chamberlain (1986), Cosslett (1987), Klein and Spady (1993) in the Binary Choice model, Newey and Powell (1990), Lai and Ying (1992), Kim and Lai (2000), and Cosslett (2004) among others in the Censored or Truncated regression framework. With the exception of Newey and Powell (1990), all these estimators assume some sort of conditional or unconditional independence between unobservables and regressors.

3.3 Some Efficiency Bounds

In the following, we use an approach suggested by Magnac and Maurin (2004) for a vector-valued parameter of interest. If the parameter of interest is a vector β_0 that appears everywhere in (3.2.2) as $\pi_\beta = E\left[\mathbf{z}\mathbf{x}^\top\right]\beta_0$, we consider the calculation of the efficiency bound in two steps. Firstly, we calculate the efficiency bound at π_β, call it V_β, then under the standard regularity conditions, by Newey and McFadden (1994), the variance matrix of the corresponding β_0 is given by $\left(E\left[\mathbf{z}\mathbf{x}^\top\right]V_\beta^{-1}E\left[\mathbf{z}\mathbf{x}^\top\right]^\top\right)^{-1}$

3.3.1 Model 1

Define $\mathbf{m} = (m_1, \mathbf{m}_2)^\top$ as $m_1(y, v, \mathbf{x}, \mathbf{z}; \pi_0) = \alpha_0\theta(y)$ and $\mathbf{m}_2(y, v, \mathbf{x}, \mathbf{z}; \pi_0) = \mathbf{z}\alpha_0^2 v\theta(y)$. Also assume that the function L is known. Lewbel (1998) proved that under some conditions regarding a 'special' regressor v and its conditional distribution, the following moment restrictions hold,

$$E\begin{bmatrix} m_1/f(v|\,\mathbf{x},\mathbf{z}) \\ \mathbf{m}_2/f(v|\,\mathbf{x},\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \kappa_1 \\ \kappa_2 E[\mathbf{z}] - \kappa_1 E\left[\mathbf{z}\mathbf{x}^\top\right]\beta_0 \end{bmatrix}. \tag{3.3.1}$$

We are interested in the calculation of the efficiency bound at $\left(\alpha_0, \beta_0^\top\right)^\top$. Let our parameter of interest be $\pi_0^\top = \left(\pi_{0;1}, \pi_{0;2}^\top\right)$ where $\pi_{0;1} = \alpha_0$ and $\pi_{0;2} = E\left[\mathbf{z}\mathbf{x}^\top\right]\beta_0$. Then the resulting lower bound for the variance matrix is given by $\left(\mathbf{M}_0^{-1}\right)\widetilde{\boldsymbol{\Omega}}_0\left(\mathbf{M}_0^{-1}\right)^\top = E\left(\mathbf{Q}\mathbf{Q}^\top\right)$ as in Theorem 3.2.1, for a nonsingular matrix $\mathbf{Q} = \mathbf{M}_0^{-1}\left[\widetilde{\mathbf{g}} - E\left(\widetilde{\mathbf{g}}|\,v,\mathbf{x},\mathbf{z}\right) + E\left(\widetilde{\mathbf{g}}|\,\mathbf{x},\mathbf{z}\right)\right]$ and $\widetilde{\mathbf{m}} =$

$\mathbf{m}(y, v, \mathbf{x}, \mathbf{z}; \pi_0) f^{-1}_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})$. Using (3.3.1), it is not difficult to show that

$$\mathbf{g}\left(y, v, \mathbf{x}, \mathbf{z}; \pi_0, f_{v|\mathbf{x},\mathbf{z}}\right) = \begin{bmatrix} \pi_{0;1}\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) - \kappa_1 \\ \mathbf{z}(\pi_{0;1})^2 v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) - \kappa_2 E[\mathbf{z}] + \kappa_1 \pi_{0;2} \end{bmatrix},$$

$$E[\nabla_\pi \mathbf{g}] = \begin{bmatrix} \kappa_1/\pi_{0;1} & \mathbf{0}^\top \\ E[\mathbf{z}]\frac{2\kappa_2}{\pi_{0;1}} - \pi_{0;2}\frac{2\kappa_1}{\pi_{0;1}} & \kappa_1 I_{(q-1)} \end{bmatrix} = \mathbf{M}_0,$$

$$\mathbf{M}_0^{-1} = \begin{bmatrix} \pi_{0;1}/\kappa_1 & \mathbf{0}^\top \\ \pi_{0;2}\frac{2}{\kappa_1} - E[\mathbf{z}]\frac{2\kappa_2}{\kappa_1^2} & \kappa_1^{-1} I_{(q-1)} \end{bmatrix}, \text{ and}$$

$$\mathbf{M}_0^{-1}\widetilde{\mathbf{m}} = \begin{bmatrix} \frac{(\pi_{0;1})^2\theta(y)}{\kappa_1 f(v|\mathbf{x},\mathbf{z})} \\ \pi_{0;2}\frac{2\pi_{0;1}\theta(y)}{\kappa_1 f(v|\mathbf{x},\mathbf{z})} - E[\mathbf{z}]\frac{2\kappa_2\pi_{0;1}\theta(y)}{\kappa_1^2 f(v|\mathbf{x},\mathbf{z})} + \mathbf{z}\frac{(\pi_{0;1})^2 v\theta(y)}{\kappa_1 f(v|\mathbf{x},\mathbf{z})} \end{bmatrix},$$

where $\mathbf{0}$ is a $q \times 1$ vector of zeros and $I_{(n)}$ is the identity matrix of order n. Some straightforward algebra gives the efficiency bounds ($l.b.$) evaluated α_0 and β_0 as

$$l.b.(\widehat{\alpha}) = E\left[\gamma_\alpha^2\right]$$
$$l.b.\left(\widehat{\beta}\right) = \left(E\left[\mathbf{z}\mathbf{x}^\top\right]\Psi_\beta^{-1} E\left[\mathbf{z}\mathbf{x}^\top\right]^\top\right)^{-1},$$

where

$$\gamma_\alpha = \frac{\alpha_0^2\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})} - \alpha_0 - E\left[\frac{\alpha_0^2\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})}\middle| v, \mathbf{x}, \mathbf{z}\right] + E\left[\frac{\alpha_0^2\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})}\middle| \mathbf{x}, \mathbf{z}\right]$$
$$\gamma_{\beta_1} = \left[E\left[\mathbf{z}\mathbf{x}^\top\right]\beta_0\frac{2}{\alpha_0} - E[\mathbf{z}]\frac{2\kappa_2}{\kappa_1\alpha_0}\right]\gamma_\alpha$$
$$\gamma_{\beta_2} = \mathbf{z}\frac{\alpha_0^2 v\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})} - \left[\mathbf{z}\frac{\kappa_2}{\kappa_1} - \mathbf{z}\mathbf{x}^\top\beta_0\right] - E\left[\mathbf{z}\frac{\alpha_0^2 v\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})}\middle| v, \mathbf{x}, \mathbf{z}\right] + E\left[\mathbf{z}\frac{\alpha_0^2 v\theta(y)}{\kappa_1 f(v|\,\mathbf{x},\mathbf{z})}\middle| \mathbf{x}, \mathbf{z}\right]$$

and $\gamma_\beta = \gamma_{\beta_1} + \gamma_{\beta_2}$ with $\Psi_\beta = E\left[\gamma_\beta\gamma_\beta^\top\right]$.

For simplicity, let's assume that $E[\varepsilon] = 0$, $E[\mathbf{x}] = 0$, $E[\mathbf{z}] = 0$, $E[\mathbf{z}\varepsilon] = 0$ and $E\left[\mathbf{z}\mathbf{x}^\top\right] = I$ (the vector of instruments $\mathbf{z}$ has the same number of elements as the vector $\mathbf{x}$). In this case, $l.b.(\widehat{\alpha}) = E\left[\gamma_\alpha^2\right]$ as before but $l.b.(\widehat{\beta}) = \Psi_{\beta^*} = E\left[\gamma_{\beta^*}\gamma_{\beta^*}^\top\right]$ with $\gamma_{\beta^*} = \gamma_{\beta_1^*} + \gamma_{\beta_2}$ and $\gamma_{\beta_1^*} = \beta_0(2/\alpha_0)\gamma_\alpha$. Then, it is straightforward to verify that the asymptotic variance of the estimator of (α_0, β_0) given in Lewbel (1998) (equations (3.9) and (3.10) on page 113) is equal to the lower bound we have just obtained as a special case.

For the case where L and hence κ_1 and κ_2 are not known, we may only identify the parameters $(\pi_{0;1}, \pi_{0;2}) = (\kappa_1/\alpha_0, \beta_0/\alpha_0)$ given the above assumptions. That is,

$$\begin{aligned}
\mathbf{g}\left(y, v, \mathbf{x}, \mathbf{z}; \pi_0, f_{v|\mathbf{x},\mathbf{z}}\right) &= \begin{bmatrix} \theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) - \pi_{0;1} \\ \mathbf{z}v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) + \pi_{0;2}\pi_{0;1} \end{bmatrix}, \\
E\left[\nabla_\pi \mathbf{g}\right] &= \begin{bmatrix} -1 & \mathbf{0}^\top \\ \pi_{0;2} & \pi_{0;1} I_{(q-1)} \end{bmatrix} = \mathbf{M}_0, \\
\mathbf{M}_0^{-1} &= \begin{bmatrix} -1 & \mathbf{0}^\top \\ \pi_{0;2}\pi_{0;1}^{-1} & \pi_{0;1}^{-1} I_{(q-1)} \end{bmatrix}, \text{ and} \\
\mathbf{M}_0^{-1}\widetilde{\mathbf{m}} &= \begin{bmatrix} -\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) \\ \pi_{0;2}\pi_{0;1}^{-1}\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) + \mathbf{z}\pi_{0;1}^{-1} v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) \end{bmatrix},
\end{aligned}$$

where $\mathbf{x}$ (and therefore $\mathbf{z}$) does not include a constant as the location is not identified. In this special case, the efficiency bound for regular estimators of $\widehat{\beta/\alpha}$ is given by

$$\begin{aligned}
l.b.(\widehat{\beta/\alpha}) &= E\left[\widetilde{r}\widetilde{r}^\top\right], \text{ where} \\
&\widetilde{r} = \pi_{0;2}\pi_{0;1}^{-1} q_2 + \mathbf{q}\pi_{0;1}^{-1}, \text{ with} \\
&q_2 = \theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) \\
&\quad - \pi_{0;1} + E\left[\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z})\middle|\, v, \mathbf{x}, \mathbf{z}\right] - E\left[\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z})\middle|\, \mathbf{x}, \mathbf{z}\right], \\
&\mathbf{q} = \mathbf{z}v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z}) \\
&\quad + \pi_{0;2}\pi_{0;1} + E\left[\mathbf{z}v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z})\middle|\, v, \mathbf{x}, \mathbf{z}\right] - E\left[\mathbf{z}v\theta(y) f^{-1}(v|\,\mathbf{x}, \mathbf{z})\middle|\, \mathbf{x}, \mathbf{z}\right],
\end{aligned}$$

as on page 113 in Lewbel (1998). Therefore, Lewbel (1998)'s estimator, which uses a ratio of two kernel density estimates as an estimator for $f^{-1}(v|\,\mathbf{x}, \mathbf{z})$, is semiparametric efficient. This result seems to be new in the literature.

3.3.2 Model 2

It is clear from our earlier discussion that the sample or treatment model above is an example of (3.1.1) where $\mathbf{m}\left(y, d, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0\right) = \boldsymbol{\Psi}\left(y, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0\right) \times d$, $\mathbf{s}(\pi_0) = \mathbf{0}$, with $\pi_0 \equiv \theta_0$. If $\boldsymbol{\Psi}(\cdot; \pi)$ is continuously differentiable in π, it then follows from Theorem 3.2.1 that the semiparametric efficiency bounds for regular estimators of π_0 is given by

$$\left(E\left[\nabla_\pi \frac{\boldsymbol{\Psi}\left(y, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0\right) \times d}{f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})}\right]^\top \widetilde{\boldsymbol{\Sigma}}_0^{-1} E\left[\nabla_\pi \frac{\boldsymbol{\Psi}\left(y, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0\right) \times d}{f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})}\right]\right)^{-1},$$

where $\widetilde{\Sigma}_0 = E\left[\widetilde{\mathbf{g}}\left(y, d, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0, f_{v|\mathbf{x},\mathbf{z}}\right) \widetilde{\mathbf{g}}\left(y, d, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0, f_{v|\mathbf{x},\mathbf{z}}\right)^\top\right]$, and

$$\widetilde{\mathbf{g}}\left(y, d, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0, f_{v|\mathbf{x},\mathbf{z}}\right) = \frac{\mathbf{m}}{f_{v|\mathbf{x},\mathbf{z}}} - E\left(\frac{\mathbf{m}}{f_{v|\mathbf{x},\mathbf{z}}}\middle| v, \mathbf{x}, \mathbf{z}\right) + E\left(\frac{\mathbf{m}}{f_{v|\mathbf{x},\mathbf{z}}}\middle| \mathbf{x}, \mathbf{z}\right),$$

with $\mathbf{m} \equiv \boldsymbol{\Psi}\left(y, v, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0\right) \times d$.

As an illustration, consider the 'example model' in Lewbel (2006), where the object of interest is $\pi_0 \equiv E[y^*]$ or equivalently, it is defined by $E[\Psi(y^*; \pi_0)] = 0$, where $\Psi(y^*; \pi_0) = y^* - \pi_0$. It then follows from Theorem 1 in Lewbel (2006) that,

$$E\left[\frac{\Psi(y; \pi_0)\, d}{f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})}\right] = 0$$

also defines π_0 uniquely. It is simple to show that,

$$l.b.(\pi_0) = \gamma_0^{-2} Var\left(m - E(m|\, v, \mathbf{x}, \mathbf{z}) + E(m|\,\mathbf{x}, \mathbf{z})\right),$$

where $\gamma_0 \equiv E\left[d/f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})\right]$, and $m\left(y, d, \mathbf{x}^\top, \mathbf{z}^\top; \pi_0; f_{v|\mathbf{x},\mathbf{z}}\right) = \Psi(y; \pi_0)\, d/f_{v|\mathbf{x},\mathbf{z}}(v|\,\mathbf{x}, \mathbf{z})$. Therefore, Lewbel (2006)'s estimators are semiparametric efficient.

3.3.3 Model 3

We are interested in the calculation of semiparametric efficiency bounds of estimators based on the following weighted moment restriction as proposed by: Hirano, Imbens, and Ridder (2003)

$$E\left[g\left(y, d, \mathbf{x}; \pi_0, p(\mathbf{x})\right)\right] = 0, \tag{3.3.2}$$

where

$$g\left(y, d, \mathbf{x}; \pi_0, p(\mathbf{x})\right) = h(\mathbf{x})\left(yd/p(\mathbf{x}) - y(1-d)/(1-p(\mathbf{x})) - \pi_0\right),$$

and π_0 represents their weighted average treatment effect. Given the unconfoundedness assumption along with a regularity condition, $0 < p(\mathbf{x}) < 1$, ensuring the existence of the above moment, it follows by the law of iterated expectations that

$$\begin{aligned}
E\left[h(\mathbf{x}) Y_1|\, D, \mathbf{X} = \mathbf{x}\right] &= E\left[h(\mathbf{x}) Y_1|\,\mathbf{X} = \mathbf{x}\right] = h(\mathbf{x})\mu_1(\mathbf{x}), \\
E\left[h(\mathbf{x}) Y_0|\, D, \mathbf{X} = \mathbf{x}\right] &= E\left[h(\mathbf{x}) Y_0|\,\mathbf{X} = \mathbf{x}\right] = h(\mathbf{x})\mu_0(\mathbf{x}), \\
E\left[h(\mathbf{x}) YD|\,\mathbf{X} = \mathbf{x}\right] &= h(\mathbf{x})\mu_1(\mathbf{x})\, p(\mathbf{x}), \text{ and} \\
E\left[h(\mathbf{x}) Y(1-D)|\,\mathbf{X} = \mathbf{x}\right] &= h(\mathbf{x})\mu_0(\mathbf{x})\left(1 - p(\mathbf{x})\right).
\end{aligned}$$

We can again apply our general result and conclude that the efficiency bound is given by

$$
\begin{aligned}
& V_\pi \\
&= \frac{1}{E\left[h\left(\mathbf{x}\right)\right]^2} E\left[\left(h\left(\mathbf{x}\right)\left(\frac{YD}{p\left(\mathbf{x}\right)} - \frac{Y\left(1-D\right)}{1-p\left(\mathbf{x}\right)} - \pi_0\right)\right.\right. \\
&\left.\left. -E\left[h\left(\mathbf{x}\right)\left(\frac{YD}{p\left(\mathbf{x}\right)} - \frac{Y\left(1-D\right)}{1-p\left(\mathbf{x}\right)}\right) \middle| D, \mathbf{X}=\mathbf{x}\right] + E\left[h\left(\mathbf{x}\right)\left(\frac{YD}{p\left(\mathbf{x}\right)} - \frac{Y\left(1-D\right)}{1-p\left(\mathbf{x}\right)}\right) \middle| \mathbf{X}=\mathbf{x}\right]\right)^2\right] \\
&= \frac{1}{E\left[h\left(\mathbf{x}\right)\right]^2} E\left[\left(h\left(\mathbf{x}\right)\left(\frac{YD}{p\left(\mathbf{x}\right)} - \frac{Y\left(1-D\right)}{1-p\left(\mathbf{x}\right)} - \pi_0\right)\right.\right. \\
&\left.\left. -h\left(\mathbf{x}\right)\left(\frac{\mu_1\left(\mathbf{x}\right)D}{p\left(\mathbf{x}\right)} - \frac{\mu_0\left(\mathbf{x}\right)\left(1-D\right)}{1-p\left(\mathbf{x}\right)}\right) + h\left(\mathbf{x}\right)\left(\frac{\mu_1\left(\mathbf{x}\right)p\left(\mathbf{x}\right)}{p\left(\mathbf{x}\right)} - \frac{\mu_0\left(\mathbf{x}\right)\left(1-p\left(\mathbf{x}\right)\right)}{1-p\left(\mathbf{x}\right)}\right)\right)^2\right] \\
&= \frac{1}{E\left[h\left(\mathbf{x}\right)\right]^2} E\left[h\left(\mathbf{x}\right)^2\left(\left(\frac{YD}{p\left(\mathbf{x}\right)} - \frac{Y\left(1-D\right)}{1-p\left(\mathbf{x}\right)} - \pi_0\right) - \left(\frac{\mu_1\left(\mathbf{x}\right)}{p\left(\mathbf{x}\right)} + \frac{\mu_0\left(\mathbf{x}\right)}{1-p\left(\mathbf{x}\right)}\right)\left(D-p\left(\mathbf{x}\right)\right)\right)^2\right].
\end{aligned}
$$

This efficiency bound corresponds to that of Theorem 4 described in Hirano, Imbens, and Ridder (2003). Using a series estimate for $p(\mathbf{x})$, they also showed that the bound was sharp for the Average Treatment Effect (ATE), $E[y_1 - y_0]$, when $h(\mathbf{x}) = 1$ and the Treatment on the Treated Effect (TTE), $E[y_1 - y_0 | d = 1]$, when $h(\mathbf{x}) = p(\mathbf{x})$. Interestingly though, we have found that their estimator is semiparametric efficient among regular estimators of π_0 defined by (3.3.2). It turns out that these bounds are the same as those originally obtained by Hahn (1998) for ATE and TTE. This explicit link has not been previously noticed in the literature.

3.4 A Monte Carlo Investigation

The finite sample performance of estimators based on (3.1.1) have been widely explored by their proponents under different controlled circumstances and scenarios. Comparisons to other semiparametric estimators have also been performed, as well as studies regarding their sensitivity to underlying identification assumptions. In order to study the theoretical properties relating to their efficiency discussed in this chapter, we construct a small Monte Carlo experiment with three further goals in mind. Firstly, we are interested in measuring the efficiency loss when using a semiparametric instead of a fully parametric estimator. This is done in a variety of scenarios. Secondly, we verify the theoretical predictions of Section 3.2. Finally, we also contrast three different kernel–based estimators of (3.1.1).

The simulated latent variable is generated as

$$y^* = v + \beta_1 + \beta_2 x + \varepsilon,$$

for scalar random variables v, x and ε. The variable ε is generated as $N(0,1)$ and x was generated independently of ε as $Beta(a_1,b_1) - 1/4$ with probability $1/2$ and $Beta(a_2,b_2) + 1/4$ with probability $1/2$ minus $1/2$, where the shape parameters are $(a_1,b_1) = (2,4)$ and $(a_2,b_2) = (4,2)$. This mixture of *Betas* is designed to yield a distribution that is both bimodal and symmetric around zero with variance one. Three different scenarios are considered, based on the relation between the special regressor v and the scalar covariate x:

$$
\begin{aligned}
v|\,x &\sim N(0,6) && \text{(Design 1)}\\
v|\,x &\sim N(x,6) && \text{(Design 2)}\\
v|\,x &\sim N\left(x, 6\left(1+x^2\right)\right). && \text{(Design 3)}
\end{aligned}
$$

The first design assumes independence among regressors, while the second makes v mean dependent on x. Design 3 is similar to 2 but with added increasing heteroskedasticity with respect to x.

We also consider three different latent variable specifications, $L(\cdot)$, for the observed $y = L(y^*)$: A binary choice model (LDV I), $L(y^*) = 1(y^* > 0)$, a censored regression model (LDV II), $L(y^*) = y^* \times 1(y^* > 0)$ with $\pi_0 = (\beta_1, \beta_2)$ where $\beta_1 = 1$ and $\beta_2 = 1/2$, and finally an ordered choice model[6] (LDV III), $L(y^*) = \sum_{j=1}^{3} j \cdot 1(\alpha_{j-1} \leq y^* < \alpha_j)$ with $\pi_0 = (\alpha_1, \alpha_2, \beta_2)$ where $\beta_1 = 0$, $\beta_2 = 1$, $\alpha_0 = -\infty$, $\alpha_1 = -3/2$, $\alpha_2 = 3/2$ and $\alpha_3 = +\infty$. The censored regression model, LDV II, corresponds to model 1 discussed in the text. We implement the related estimator by using knowledge of the L function along with $\theta(y) = y^2 \exp\left(-y^2\right)$, as suggested in Lewbel (1998). Models LDV I and LDV III were not discussed in the main text, but are consider in this experiment because they provide valid examples of the generality of the results under discussion. Also, we notice that these Limited Dependent Variable models require different moment restrictions for semiparametric identification.

We consider the well–known kernel smoother, here denoted by $\widehat{f}_{NW}$, proposed by Rosenblatt (1969), the Local Linear estimator, $\widehat{f}_{LL}$, proposed by Fan, Yao, and Tong (1996), and the Two-steps estimator, $\widehat{f}_{2S}$, recently suggested by Hansen (2004). Although bandwidth selection methods are readily available for them, as we will explain in the next chapter, $\sqrt{N}$–semiparametric estimators require undersmoothing relative to optimal pointwise convergence of their nonparametric component. Although, we could use the 'plug–in' optimal–bandwidth estimator (see Chapter 4) when calculating $\widehat{f}_{NW}$, the optimal bandwidths when using $\widehat{f}_{LL}$ or $\widehat{f}_{2S}$ will be generally different. It is for this reason that we have taken a simplified and sensible approach for bandwidth selection. For each sample size $N = 200$,

[6] As discussed in Section 6 (page 161) in Lewbel (2000b). Apart from the experiment presented in this chapter, Stewart (2005) also provides Monte Carlo evidence with regards to its performance and compares it against other semiparametric competitors. However, they do not use any kernel-based estimator as we do here, but instead they employ the ordered data estimator of Lewbel and Schennach (2005) in place of the inverse of the unknown conditional density.

400 and 600, in 2000 replications, we calculate the semiparametric estimator using each of these estimators for the nonparametric component. We repeat this over the same grid of bandwidths, in which their semiparametric optimal bandwidths were known to lie through experimentation. Then we report the results for the fixed bandwidth that yield the smallest combined root mean squared error for each estimator. Although unfeasible in a real application, this procedure provides a fair framework in which the performance of these semiparametric estimators can be compared. The same second order kernel and also the same bandwidth is used in all dimensions or steps for all three estimators. In these simulations, we also calculate the optimally–weighted semiparametric estimator using the true conditional density, f, and the maximum likelihood estimator (MLE), which is not only consistent but also the most efficient estimator in all designs and models.

The results are presented in Tables 3.1 to 3.9. As would be expected, for all designs and models, the variance of semiparametric estimators based on (3.1.1) are bounded below by the variance of the MLE. In view of Corollary 3.2.2, they are also bounded above by the variance of the estimator that uses the true f. The loss in efficiency varies among models for a given design. For example, using Design 1 in Tables 3.1, 3.2 and 3.3, the average loss ranges between 3–8% while using an estimated $\widehat{f}$ compared to a loss of around 32% using the true density for $\widehat{\beta}_2$ in LDV I. These relative losses, with respect to the MLE, become more dramatic in LDV II, with substantial losses of 47–57 % and 125% respectively. For the same design, Table 3.3 reports efficiency losses for estimation of α_1 and α_2 up to 5 times when using the true density in place of an estimated one. These losses seem to slightly increase in all models when using Designs 2 and 3 for all sample sizes.

Biases are sizeable when using any $\widehat{f}$. Although they tend to decrease as the sample size increases, for a small sample size, $N = 200$, in Design 2 they can be as high as 26% while using $\widehat{f}_{2S}$ in LDV II (Table 3.5) for example, leads to an overall to reasonably higher $RMSE$. Likewise the MLE, the semiparametric estimator using f is virtually mean unbiased in all models and designs. Nevertheless, their bigger variances lead significantly to the highest root mean squared error among all the estimators under consideration. It is also the case that the associated measures of fitting criterias, $RMSE$ and MAE, are bounded below by that of the MLE and above by that of the estimator that uses knowledge of f. For all estimators, these fitting criteria tend to deteriorate as the dependence between v and x intensifies in all models.

With reference to the relative performance of $\widehat{f}_{NW}$, $\widehat{f}_{LL}$ andr $\widehat{f}_{2S}$, Tables 3.1 and 3.3 show no clear ranking in their performance. For LDV I and LDV III, $\widehat{f}_{NW}$ seems to produce smaller $RMSE$ in estimating intercepts and using $\widehat{f}_{LL}$ produces better estimates of β_2 in LDV III and of β_1 in LDV II. Nonetheless, Tables 3.5, 3.8, 3.6 and 3.9 (Designs 2 and 3) show that the use of the Local Linear estimator of Fan, Yao, and Tong (1996), produces smaller root mean squared errors than its competitors for all parameters in LDV II and LDV

III. Two reasons may explain this improvement. Firstly, the associated pointwise bias of $\widehat{f}_{LL}$ does not depend on the unknown probability density function of x as $\widehat{f}_{NW}$, and $\widehat{f}_{2S}$ do, and therefore makes it design adaptive (see Fan (1992)). Secondly, it is also known that the asymptotic bias and variance of $\widehat{f}_{LL}$ are of the same order of magnitude in the interior as well as near the boundary of the support of x, while this is not the case for $\widehat{f}_{NW}$, and $\widehat{f}_{2S}$. This later result is likely to produce fewer outliers arising from near–zero $\widehat{f}$ values, and might be alleviated with trimming. In our experiments however, no trimming is performed. It might also be possible that modifications, proposed by Hyndman, Bashtannyk, and Grunwald (1996) and De Gooijer and Zerom (2003), to the Nadaraya–Watson estimator may help to mitigate the problem. In conclusion, at least in theory, the Local Linear estimator's design adaptation and its immunity from boundary effects makes it a more attractive choice than the others. Modifications by Hyndman and Yao (2002) may also further improve its performance. It is also true that the performance of $\widehat{f}_{2S}$ is always dominated by either $\widehat{f}_{NW}$, or $\widehat{f}_{LL}$ in terms of $RMSE$.

3.5 Conclusion

We derive the semiparametric efficiency bound for a class of estimators, which are based on unconditional moment restrictions that involve weighting by the inverse of a conditional probability function. The efficiency bound resembles that of Chamberlain (1987). It is easy to compute and its form is shown to mimic that of the standard GMM. In this unifying framework, we also prove that the asymptotic variance of these estimators, when the conditional probability is known, is not smaller than when it is unknown. These findings generalize those of Magnac and Maurin (2004), as they are simply special cases of these more general results. An explicit link is made with Hirano, Imbens, and Ridder (2003), as we are able to reproduce their bound using our calculations. We show this as well as prove that the estimator for a general LDV model, proposed by Lewbel (1998), is semiparametric efficient, a finding that seem to be new in the literature. Similarly, we prove that estimators in Lewbel (2006) are also semiparametric efficient.

A small Monte Carlo experiment is performed in which we do not only confirm the validity of our results, but also compare three different kernel–based estimators. We find evidence that the use of the Local Linear estimator of Fan, Yao, and Tong (1996) outperforms the other two in certain cases.

Appendix

3.A Main Proofs

Although, the procedure described in Newey (1990) has dominated most of the literature of semiparametric efficiency bound derivation, we have taken a different but equivalent approach to calculating such bounds. Using some standard Hilbert space theory, Severini and Tripathi (2001) presents a simplified approach to computing efficiency bounds in semiparametric models. In our case, this approach greatly simplifies the derivation. The reader is encouraged to consult Severini and Tripathi (2001) for a wider explanation on the steps discussed in the proofs below.

NOTATION: In what follows, $L^2(S_{\mathbf{w}};\mu_{\mathbf{w}})$ represents the set of all real-valued functions on $S_{\mathbf{w}}$ which are square integrable with respect to the $\mu_{\mathbf{w}}$ measure. $L^2(S_{\mathbf{w}};\mathbf{w})$ also denotes the set of all real-valued functions on $S_{\mathbf{w}}$ which are square integrable with respect to the probability distribution of $\mathbf{w}$. We also use the symbol $E_{\mathbf{w}}$ to denote integrals with respect to the distribution of $\mathbf{w}$, and $E[\cdot|\,\mathbf{w}]$ as conditional expectation given $\mathbf{w}$.

Proof of Theorem 3.2.1

The random vector $\mathbf{u}=(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2)$ has unknown density function $f(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2)$ (with respect to a dominating measure $\mu=\mu_{\mathbf{y}}\times\mu_{\mathbf{w}_1}\times\mu_{\mathbf{w}_2}$) which is rewritten as

$$\begin{aligned} f(\mathbf{y},\mathbf{w}_1,\mathbf{w}_2) &= f_{\mathbf{y}|\mathbf{w}}(\mathbf{y}|\,\mathbf{w}_1,\mathbf{w}_2)\,f_{\mathbf{w}_1|\mathbf{w}_2}(\mathbf{w}_1|\,\mathbf{w}_2)\,f_{\mathbf{w}_2}(\mathbf{w}_2) \\ &= \psi_0^2(\mathbf{y}|\,\mathbf{w}_1,\mathbf{w}_2)\,\phi_{0;1}^2(\mathbf{w}_1|\,\mathbf{w}_2)\,\phi_{0;2}^2(\mathbf{w}_2), \end{aligned}$$

where $\psi_0\in\Psi$, $\phi_{0;1}\in\Phi_1$, $\phi_{0;2}\in\Phi_2$ and

$$\Psi=\left\{\psi\in S_{\mathbf{y}}\times S_{\mathbf{w}_1}\times S_{\mathbf{w}_2}\to\mathbb{R},\ \psi^2(\mathbf{y}|\,\mathbf{w})\geq 0,\ \text{bounded},\ \int_{S_{\mathbf{y}}}\psi^2(\mathbf{y}|\,\mathbf{w})\,\mu_{\mathbf{y}}(d\mathbf{y})=1\right\},$$

$$\Phi_1 = \left\{ \phi_1 \in L^2\left(S_{\mathbf{w}_1};\mu_{\mathbf{w}_1}\right), \text{ moment (3.2.2) exists}, \int\limits_{S_{\mathbf{w}_1}} \phi_1^2\left(\mathbf{w}_1|\,\mathbf{w}_2\right)\mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) = 1 \right\},$$

$$\Phi_2 = \left\{ \phi_2 \in L^2\left(S_{\mathbf{w}_2};\mu_{\mathbf{w}_2}\right), \phi_2^2\left(\mathbf{w}_2\right) > 0, \text{ bounded}, \int\limits_{S_{\mathbf{w}_2}} \phi_2^2\left(\mathbf{w}\right)\mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right) = 1 \right\}.$$

The idea is to think of Φ_1 as imposing some restrictions on elements of $L^2\left(S_{\mathbf{w}_1};\mu_{\mathbf{w}_1}\right)$ such that (3.2.2) exists. These restrictions will be motivated by the functional form of $\mathbf{m}\left(\mathbf{u};\pi_0,\phi_{0;1}^2\right)$ as well as the underlying assumptions regarding the subvector $\mathbf{w}_1$. The assumption that matrix $E\left[\mathbf{g}\left(\mathbf{u};\pi_0,\phi_{0;1}^2\right)\mathbf{g}\left(\mathbf{u};\pi_0,\phi_{0;1}^2\right)^{\top}\right]$ exists and is not singular, implies that none of the components of $\widetilde{\mathbf{m}}\left(\mathbf{u};\pi_0,\phi_{0;1}^2\right) - \mathbf{s}\left(\pi_0\right)$ are redundant, i.e. linearly dependent, where $\widetilde{\mathbf{m}}\left(\mathbf{u};\pi_0,\phi_{0;1}^2\right) = \mathbf{m}\left(\mathbf{u};\pi_0\right)/\phi_{0;1}^2$.

We want to calculate the efficiency bound for estimating $\pi_0 \in \mathbb{R}^p$, our parameter of interest. To simplify this problem, we look at a real–valued function, consider estimating $\rho\left(\psi_0,\phi_{0;1},\phi_{0;2}\right) = \mathbf{c}^{\top}\pi_0$ as our 'structural' parameter of interest, where $\mathbf{c} \in \mathbb{R}^p$ is arbitrary.

As described in Section 2 in Severini and Tripathi (2001) we begin by parameterizing $\psi_0, \phi_{0;1}$ and $\phi_{0;2}$ as a one–dimensional subproblem. For some $t_0 > 0$ let $t \longmapsto (\psi_t,\phi_{t;1},\phi_{t;2},\pi_t)$ be a curve from $[0,t_0]$ into $\Psi \times \Phi_1 \times \Phi_2 \times \mathbb{R}^p$ which passes through $(\psi_0,\phi_{0;1},\phi_{0;2},\pi_0)$ at $t = 0$. Let the tangent space to $\Psi \times \Phi_1 \times \Phi_2 \times \mathbb{R}^p$ at the true value $(\psi_0,\phi_{0;1},\phi_{0;2})$ be denoted by $\overline{lin\ T\left(\Psi \times \Phi_1 \times \Phi_2 \times \mathbb{R}^p, (\psi_0,\phi_{0;1},\phi_{0;2})\right)}$. This tangent space is the smallest linear space which is closed in the L^2-norm and which contains all $(\dot{\psi},\dot{\phi}_1,\dot{\phi}_2) \in L^2(S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2};\mu_{\mathbf{y}} \times \mu_{\mathbf{w}_1} \times \mu_{\mathbf{w}_2})$ that are tangent to $\Psi \times \Phi_1 \times \Phi_2 \times \mathbb{R}^p$ at $(\psi_0,\phi_{0;1},\phi_{0;2})$.

As shown[7] in Severini and Tripathi (2001), the tangent space is the product of $\overline{lin\ T\left(\Psi,\psi_0\right)}$,

[7] In fact, they showed this for $\mu_{\mathbf{y}}$ and $\mu_{\mathbf{w}_2}$ Lebesgue measures. These sets can be made more specific in a given LDV model, at the expense of additional notation and complexity without altering the final result.

$\overline{lin\ T\left(\Phi_1, \phi_{0;1}\right)}$ and $\overline{lin\ T\left(\Phi_2, \phi_{0;2}\right)}$ where

$$\overline{lin\ T\left(\Psi, \psi_0\right)} = \left\{ \dot{\psi} \in L^2\left(S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}; \mu_{\mathbf{y}} \times \mathbf{w}\right), \right.$$

$$\left. \int_{S_{\mathbf{y}}} \dot{\psi}\psi_0\left(\mathbf{y}|\,\mathbf{w}\right)\mu_{\mathbf{y}}\left(d\mathbf{y}\right) = 0 \text{ for almost all } \mathbf{w} \right\}$$

$$\overline{lin\ T\left(\Phi_2, \phi_{0;2}\right)} = \left\{ \dot{\phi}_2 \in L^2\left(S_{\mathbf{w}_2}; \mu_{\mathbf{w}_2}\right), \int_{S_{\mathbf{w}_2}} \dot{\phi}_2\phi_{0;2}\left(\mathbf{w}_2\right)\mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right) = 0 \right\}.$$

We notice that there is no need to define explicitly the set $\overline{lin\ T\left(\Phi_1, \phi_{0;1}\right)}$ since ϕ_1^* will soon be shown to be zero, which is always an element of the tangent space, whatever it may be. This enforces our earlier claim that $\phi_{0;1}$ is ancillary to π_0. They also provided the framework in which we define, for any $(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)$ and $(\dot{\psi}', \dot{\phi}_1', \dot{\phi}_2')$ elements of the tangent space, the Fisher information inner product $\langle\cdot,\cdot\rangle_F$ and the corresponding norm $\|\cdot\|_F$ as

$$< (\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2), (\dot{\psi}', \dot{\phi}_1', \dot{\phi}_2') >_F = 4\int_{S_{\mathbf{y}}} E_{\mathbf{w}}(\dot{\psi}\dot{\psi}')\mu_{\mathbf{y}}\left(d\mathbf{y}\right) + 4E_{\mathbf{w}_2}\left(\int_{S_{\mathbf{w}_1}} \dot{\phi}_1\dot{\phi}_1'\mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right)\right)$$
$$+ 4\int_{S_{\mathbf{w}_2}} \dot{\phi}_2\dot{\phi}_2'\mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right) \tag{3.A.1}$$

$$\left\|(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)\right\|_F^2 =< (\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2), (\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2) >_F .$$

However, not every $(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2) \in \overline{lin\ T\left(\Psi, \psi_0\right)} \times \overline{lin\ T\left(\Phi_1, \phi_{0;1}\right)} \times \overline{lin\ T\left(\Phi_2, \phi_{0;2}\right)}$ may be used to calculate $\langle\cdot,\cdot\rangle_F$. After differentiating the unconditional moment conditions,

$$E\left[\widetilde{\mathbf{m}}\left(\mathbf{u}; \pi_t, \phi_{t;1}^2\right) - \mathbf{s}\left(\pi_t\right)\right] = E\left[\mathbf{g}\left(\mathbf{u}; \pi_0, \phi_{0;1}^2\right)\right],$$

with respect to t, we can notice that only those $(\dot{\phi}_1, \dot{\psi}, \dot{\phi}_2)$ may be used that also satisfy

$$\mathbf{0} = \int_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \nabla_\pi \left[\left(\mathbf{m}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0\right) - \mathbf{s}\left(\pi_0\right) \phi_{0;1}^2 \right) \right] \psi_0^2 \phi_{0;2}^2 \mu_{\mathbf{y}}\left(d\mathbf{y}\right) \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right) \dot{\pi}$$
$$+ 2 \int_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \left\{ \mathbf{m}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0\right) \left(\psi_0 \dot{\psi} \phi_{0;2}^2 + \psi_0^2 \phi_{0;2} \dot{\phi}_2 \right) \right.$$
$$\left. - \mathbf{s}\left(\pi_0\right) \left(\psi_0 \dot{\psi} \phi_{0;1}^2 \phi_{0;2}^2 + \psi_0^2 \phi_{0;1} \dot{\phi}_1 \phi_{0;2}^2 + \psi_0^2 \phi_{0;1}^2 \phi_{0;2} \dot{\phi}_2 \right) \right\} \mu_{\mathbf{y}}\left(d\mathbf{y}\right) \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right)$$
$$\mathbf{0} = \int_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \left\{ \nabla_\pi \left[\mathbf{m}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0\right) \left(\phi_{0;1}\right)^{-2} - \mathbf{s}\left(\pi_0\right) \right] \phi_{0;1}^2 \psi_0^2 \phi_{0;2}^2 \dot{\pi} \right.$$
$$\left. + 2\mathbf{m}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0\right) \left(\psi_0 \dot{\psi} \phi_{0;2}^2 + \psi_0^2 \phi_{0;2} \dot{\phi}_2 \right) \right\} \mu_{\mathbf{y}}\left(d\mathbf{y}\right) \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right)$$
$$\mathbf{0} = \left[E \nabla_\pi \mathbf{g}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0, \phi_{0;1}^2\right) \right] \dot{\pi} + 2 \int_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \widetilde{\mathbf{m}}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0, \phi_{0;1}^2\right) \phi_{0;1}^2 \times$$
$$\left(\psi_0 \dot{\psi} \phi_{0;2}^2 + \psi_0^2 \phi_{0;2} \dot{\phi}_2 \right) \mu_{\mathbf{y}}\left(d\mathbf{y}\right) \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right).$$

The second equality follows from the fact that for any $(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)$ in the tangent space, $\int_{S_{\mathbf{y}}} \psi_0 \dot{\psi} \mu_{\mathbf{y}}\left(d\mathbf{y}\right) = 0$ for almost all $\mathbf{w}$, $\int_{S_{\mathbf{w}_1}} \phi_{0;1} \dot{\phi}_1 \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) = 0$ and $\int_{S_{\mathbf{w}_2}} \phi_{0;2} \dot{\phi}_2 \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right) = 0$. Since $q > p$, i.e. π_0 may be overidentified, the above equation will generally have a nonunique solution in $\dot{\pi}$. A sufficient condition for π_0 to be locally identified (see Rothenberg (1971)) is that we find a nonstochastic full rank $q \times q$ matrix $\mathbf{W}$ such that the matrix $\left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1}$ exist, where $\mathbf{M}_0 = E\left[\nabla_\pi \mathbf{g}\left(\mathbf{u}; \pi_0, \phi_{0;1}^2\right)\right]$,

$$\dot{\pi} = -2 \int_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \widetilde{\mathbf{m}}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0, \phi_{0;1}^2\right) \phi_{0;1}^2 \left(\psi_0 \dot{\psi} \phi_{0;2}^2 + \psi_0^2 \phi_{0;2} \dot{\phi}_2 \right)$$
$$\times \mu_{\mathbf{y}}\left(d\mathbf{y}\right) \mu_{\mathbf{w}_1}\left(d\mathbf{w}_1\right) \mu_{\mathbf{w}_2}\left(d\mathbf{w}_2\right).$$

The tangent vector $(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)$ used to calculate $\dot{\pi}$ also has to satisfy $\nabla\rho(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2) = \mathbf{c}^\top \dot{\pi}$.

It is clear that $\nabla\rho$ is a linear functional on the product tangent space and that

$$\left(\overline{lin\, T\left(\Psi, \psi_0\right)} \times \overline{lin\, T\left(\Phi_1, \phi_{0;1}\right)} \times \overline{lin\, T\left(\Phi_2, \phi_{0;2}\right)}, \langle\cdot,\cdot\rangle \right)$$

is a Hilbert space. Then the Riesz–Fréchet theorem implies that for all $(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)$ in the tangent space there exists a unique $(\psi^*, \phi_1^*, \phi_2^*) \in \overline{lin\, T\left(\Psi, \psi_0\right)} \times \overline{lin\, T\left(\Phi_1, \phi_{0;1}\right)} \times$

$\overline{lin\ T(\Phi_2, \phi_{0;2})}$ such that

$$\nabla\rho(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2) = <(\psi^*, \phi_1^*, \phi_2^*), (\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)>_F$$

Firstly, notice that $\nabla\rho(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2)$ can be rewritten as

$$\begin{aligned}
&\nabla\rho(\dot{\psi}, \dot{\phi}_1, \dot{\phi}_2) \\
&= -2 \int\limits_{S_{\mathbf{y}} \times S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \Big\{ \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \widetilde{\mathbf{m}} \times \\
&\qquad\qquad \phi_{0;1}^2 \left(\psi_0 \dot{\psi} \phi_{0;2}^2 + \psi_0^2 \phi_{0;2} \dot{\phi}_2 \right) \Big\} \mu_{\mathbf{y}}(d\mathbf{y})\, \mu_{\mathbf{w}_1}(d\mathbf{w}_1)\, \mu_{\mathbf{w}_2}(d\mathbf{w}_2) \\
&= -2 \int\limits_{S_{\mathbf{y}}} \left\{ \int\limits_{S_{\mathbf{w}_1} \times S_{\mathbf{w}_2}} \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \widetilde{\mathbf{m}}\left(\mathbf{u}; \pi_0, \phi_{0;1}^2\right) \times \right. \\
&\qquad\qquad \left. \psi_0 \dot{\psi} \phi_{0;1}^2 \phi_{0;2}^2 \right\} \mu_{\mathbf{w}_1}(d\mathbf{w}_1)\, \mu_{\mathbf{w}_2}(d\mathbf{w}_2)\, \mu_{\mathbf{y}}(d\mathbf{y}) \\
&\quad - 2 \int\limits_{S_{\mathbf{w}_2}} \left\{ \int\limits_{S_{\mathbf{y}} \times S_{\mathbf{w}_1}} \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \widetilde{\mathbf{m}}\left(\mathbf{u}; \pi_0, \phi_{0;1}^2\right) \times \right. \\
&\qquad\qquad \left. \psi_0^2 \phi_{0;1}^2 \mu_{\mathbf{y}}(d\mathbf{y})\, \mu_{\mathbf{w}_1}(d\mathbf{w}_1) \right\} \phi_{0;2} \dot{\phi}_2 \mu_{\mathbf{w}_2}(d\mathbf{w}_2).
\end{aligned}$$

Comparing the above expression with (3.A.1) and using the fact that $\int_{S_{\mathbf{y}}} \psi_0 \dot{\psi} \mu_{\mathbf{y}}(d\mathbf{y}) = 0$ for almost all $\mathbf{w}$ and $\int_{S_{\mathbf{w}_2}} \phi_{0;2} \dot{\phi}_2 \mu_{\mathbf{w}_2}(d\mathbf{w}_2) = 0$ for any $\dot{\psi}$ and $\dot{\phi}_2$ in the tangent space, it is possible to deduce $(\psi^*, \phi_1^*, \phi_2^*)$ as

$$\begin{aligned}
\psi^*(\mathbf{y}|\, \mathbf{w}_1, \mathbf{w}_2) &= -\frac{1}{2} \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \left[\widetilde{\mathbf{m}} - E(\widetilde{\mathbf{m}}|\, \mathbf{w}_1, \mathbf{w}_2)\right] \psi_0 \\
\phi_1^*(\mathbf{w}_1|\, \mathbf{w}_2) &= 0 \\
\phi_2^*(\mathbf{w}_2) &= -\frac{1}{2} \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \left[E(\widetilde{\mathbf{m}}|\, \mathbf{w}_2) - E(\widetilde{\mathbf{m}})\right] \phi_{0;2}.
\end{aligned}$$

It is straightforward to verify that the proposed $(\psi^*, \phi_1^*, \phi_2^*)$ belongs to the tangent space, i.e. $\int_{S_{\mathbf{y}}} \psi_0 \psi^* \mu_{\mathbf{y}}(d\mathbf{y}) = 0$ and $\int_{S_{\mathbf{w}_2}} \phi_{0;2} \phi_2^* \mu_{\mathbf{w}_2}(d\mathbf{w}_2) = 0$. Therefore, from the Riesz–Fréchet theorem we can conclude that $\nabla\rho$ is not only linear but also continuous which implies that our object of interest $\rho(\psi_0, \phi_{0;1}, \phi_{0;2})$ is pathwise differentiable.

As shown in Severini and Tripathi (2001), we can use $(\psi^*, \phi_1^*, \phi_2^*)$ with the efficiency bound for the asymptotic variance of regular estimators of $\mathbf{c}^\top \pi_0$ now given by

$$\begin{aligned}
&\|(\psi^*, \phi_1^*, \phi_2^*)\|_F^2 \\
&= \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \times \\
&\mathbf{W} E\left[(\tilde{\mathbf{m}} - E(\tilde{\mathbf{m}} | \mathbf{w}_1, \mathbf{w}_2))(\tilde{\mathbf{m}} - E(\tilde{\mathbf{m}} | \mathbf{w}_1, \mathbf{w}_2))^\top\right] \mathbf{W}^\top \mathbf{M}_0 \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{c} \\
&+ \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \times \\
&\mathbf{W} E\left[(E(\tilde{\mathbf{m}} | \mathbf{w}_2) - E(\tilde{\mathbf{m}}))(E(\tilde{\mathbf{m}} | \mathbf{w}_2) - E(\tilde{\mathbf{m}}))^\top\right] \mathbf{W}^\top \mathbf{M}_0 \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{c} \\
&= \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} E\left[(\tilde{\mathbf{m}} - E(\tilde{\mathbf{m}}) - E(\tilde{\mathbf{m}} | \mathbf{w}_1, \mathbf{w}_2) + E(\tilde{\mathbf{m}} | \mathbf{w}_2)) \times\right. \\
&\left.(\tilde{\mathbf{m}} - E(\tilde{\mathbf{m}}) - E(\tilde{\mathbf{m}} | \mathbf{w}_1, \mathbf{w}_2) + E(\tilde{\mathbf{m}} | \mathbf{w}_2))^\top\right] \mathbf{W}^\top \mathbf{M}_0 \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{c} \\
&= \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \tilde{\boldsymbol{\Omega}}_0 \mathbf{W}^\top \mathbf{M}_0 \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{c}.
\end{aligned}$$

However, this result is not satisfactory, because the lower bound depends upon $\mathbf{W}$, the auxiliary matrix used to solve for $\dot{\pi}$. Since by assumption, $\tilde{\boldsymbol{\Omega}}_0$ is nonsingular, and using Hansen (1982)'s trick, it is possible to show that

$$l.b.\left(\mathbf{c}^\top \hat{\pi}_n\right) = \mathbf{c}^\top \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{M}_0^\top \mathbf{W} \tilde{\boldsymbol{\Omega}}_0 \mathbf{W}^\top \mathbf{M}_0 \left(\mathbf{M}_0^\top \mathbf{W} \mathbf{M}_0\right)^{-1} \mathbf{c} \geq \mathbf{c}^\top (\mathbf{M}_0^\top \tilde{\boldsymbol{\Omega}}_0^{-1} \mathbf{M}_0)^{-1} \mathbf{c}.$$

Therefore the efficiency bound for regular estimators of $\mathbf{c}^\top \pi_0$ is given by $\mathbf{c}^\top (\mathbf{M}_0^\top \tilde{\boldsymbol{\Omega}}_0^{-1} \mathbf{M}_0)^{-1} \mathbf{c}$. Because $\mathbf{c} \in \Re^p$ was arbitrary, we conclude that the efficiency bound for regular estimators of π_0 is $(\mathbf{M}_0^\top \tilde{\boldsymbol{\Omega}}_0^{-1} \mathbf{M}_0)^{-1}$. ■

Proof of Corollary 3.2.2

We prove that $(\mathbf{M}_0^\top \mathbf{\Omega}_0^{-1}\mathbf{M}_0)^{-1} - (\mathbf{M}_0^\top \widetilde{\mathbf{\Omega}}_0^{-1}\mathbf{M}_0)^{-1}$ is a positive semi-definite matrix, which is equivalent to prove that $\mathbf{M}_0^\top \mathbf{\Omega}_0^{-1}\mathbf{M}_0 - \mathbf{M}_0^\top \widetilde{\mathbf{\Omega}}_0^{-1}\mathbf{M}_0 = \mathbf{M}_0^\top (\mathbf{\Omega}_0^{-1} - \widetilde{\mathbf{\Omega}}_0^{-1})\mathbf{M}_0$ is negative semi-definite or that $\Delta_0 = \mathbf{\Omega}_0 - \widetilde{\mathbf{\Omega}}_0$ is positive semi–definite. Recall

$$\begin{aligned}\mathbf{\Omega}_0 &= E\left[\mathbf{g}\mathbf{g}^\top\right]\\ \widetilde{\mathbf{\Omega}}_0 &= E\left[\widetilde{\mathbf{g}}\widetilde{\mathbf{g}}^\top\right],\end{aligned}$$

where $\mathbf{g} = \mathbf{g}\left(\mathbf{y}, \mathbf{w}_1, \mathbf{w}_2; \pi_0\right)$, $\widetilde{\mathbf{g}} = \mathbf{g} - E\left(\mathbf{g}|\, \mathbf{w}_1, \mathbf{w}_2\right) + E\left(\mathbf{g}|\, \mathbf{w}_2\right) = \mathbf{g} - \mathbf{h}$ and $\mathbf{h} = E\left[\mathbf{g}|\, \mathbf{w}_1, \mathbf{w}_2\right] - E\left[\mathbf{g}|\, \mathbf{w}_2\right]$. Then it follows that

$$\begin{aligned}\Delta &= E\left[\mathbf{g}\mathbf{g}^\top - (\mathbf{g}-\mathbf{h})(\mathbf{g}-\mathbf{h})^\top\right]\\ &= E\left[\mathbf{g}\mathbf{h}^\top + \mathbf{h}\mathbf{g}^\top - \mathbf{h}\mathbf{h}^\top\right]\\ &= E\left[E\left[\mathbf{g}|\, \mathbf{w}_1, \mathbf{w}_2\right]\mathbf{h}^\top + \mathbf{h}E\left[\mathbf{g}^\top \middle|\, \mathbf{w}_1, \mathbf{w}_2\right] - \mathbf{h}\mathbf{h}^\top\right]\\ &= E\left[\mathbf{h}\mathbf{h}^\top + E\left[\mathbf{g}|\, \mathbf{w}_2\right]\mathbf{h}^\top + \mathbf{h}E\left[\mathbf{g}^\top \middle|\, \mathbf{w}_2\right]\right]\\ &= E\left[\mathbf{h}\mathbf{h}^\top\right],\end{aligned}$$

where the last equality follows after noticing that $E\left[E\left[\mathbf{g}|\, \mathbf{w}_2\right]\mathbf{h}^\top\right] = E\left[E\left[\mathbf{g}|\, \mathbf{w}_2\right]E\left[\mathbf{h}^\top|\, \mathbf{w}_2\right]\right] = 0$. As $\Delta = E\left[\mathbf{h}\mathbf{h}^\top\right]$ is positive semi–definite, the result follows. ■

3.B Tables

Table 3.1: Monte Carlo results for Design 1

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	-0.0125	0.3116	0.3118	0.2491	-0.0031	0.1423	0.1423	0.1130
	$f(v\|x)$	-0.0096	0.4028	0.4028	0.3198	-0.0106	0.1865	0.1868	0.1497
	$\widehat{f}_{NW}(v\|x)$	-0.0365	0.3244	0.3264	0.2623	-0.0498	0.1498	0.1578	0.1262
	$\widehat{f}_{LL}(v\|x)$	-0.0516	0.3129	0.3170	0.2557	-0.0649	0.1493	0.1627	0.1297
	$\widehat{f}_{2S}(v\|x)$	-0.0354	0.3375	0.3393	0.2722	-0.0522	0.1548	0.1634	0.1306
400	MLE	-0.0074	0.2252	0.2252	0.1798	0.0003	0.1002	0.1002	0.0798
	$f(v\|x)$	-0.0028	0.2955	0.2954	0.2340	-0.0015	0.1315	0.1315	0.1046
	$\widehat{f}_{NW}(v\|x)$	-0.0318	0.2313	0.2334	0.1857	-0.0494	0.1029	0.1141	0.0917
	$\widehat{f}_{LL}(v\|x)$	-0.0260	0.2344	0.2357	0.1879	-0.0253	0.1071	0.1100	0.0876
	$\widehat{f}_{2S}(v\|x)$	-0.0296	0.2381	0.2399	0.1915	-0.0508	0.1066	0.1181	0.0948
600	MLE	-0.0051	0.1852	0.1852	0.1470	-0.0022	0.0822	0.0822	0.0662
	$f(v\|x)$	-0.0025	0.2469	0.2469	0.1979	-0.0019	0.1089	0.1089	0.0878
	$\widehat{f}_{NW}(v\|x)$	-0.0212	0.1956	0.1967	0.1575	-0.0335	0.0859	0.0922	0.0747
	$\widehat{f}_{LL}(v\|x)$	-0.0306	0.1912	0.1936	0.1556	-0.0425	0.0856	0.0955	0.0773
	$\widehat{f}_{2S}(v\|x)$	-0.0206	0.2009	0.2019	0.1614	-0.0341	0.0884	0.0947	0.0765

[a] Binary Choice Model (LDV I): $y = 1(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.2: Monte Carlo results for Design 1

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	-0.0063	0.1749	0.1750	0.1404	-0.0027	0.0805	0.0805	0.0640
	$f(v\|x)$	0.0008	0.3890	0.3889	0.3077	-0.0075	0.1729	0.1730	0.1361
	$\widehat{f}_{NW}(v\|x)$	-0.1980	0.2212	0.2968	0.2441	0.0521	0.0986	0.1115	0.0891
	$\widehat{f}_{LL}(v\|x)$	-0.2258	0.2061	0.3057	0.2555	0.0566	0.0943	0.1100	0.0881
	$\widehat{f}_{2S}(v\|x)$	-0.2004	0.2239	0.3004	0.2476	0.0511	0.0995	0.1118	0.0891
400	MLE	-0.0019	0.1242	0.1242	0.0979	-0.0009	0.0568	0.0568	0.0456
	$f(v\|x)$	0.0023	0.2793	0.2792	0.2215	-0.0029	0.1233	0.1233	0.0977
	$\widehat{f}_{NW}(v\|x)$	-0.1123	0.2023	0.2314	0.1867	0.0312	0.0910	0.0962	0.0764
	$\widehat{f}_{LL}(v\|x)$	-0.1392	0.1924	0.2374	0.1931	0.0353	0.0883	0.0951	0.0754
	$\widehat{f}_{2S}(v\|x)$	-0.1136	0.2054	0.2347	0.1905	0.0303	0.0908	0.0957	0.0758
600	MLE	-0.0028	0.1004	0.1004	0.0797	-0.0010	0.0462	0.0462	0.0368
	$f(v\|x)$	-0.0008	0.2288	0.2287	0.1811	-0.0034	0.1005	0.1005	0.0790
	$\widehat{f}_{NW}(v\|x)$	-0.0808	0.1756	0.1932	0.1546	0.0199	0.0795	0.0819	0.0650
	$\widehat{f}_{LL}(v\|x)$	-0.1031	0.1686	0.1976	0.1593	0.0233	0.0780	0.0814	0.0644
	$\widehat{f}_{2S}(v\|x)$	-0.0803	0.1775	0.1948	0.1554	0.0203	0.0796	0.0821	0.0655

[a] Censored Regression Model (LDV II): $y = [v + \beta_1 + \beta_2 x + \varepsilon] \times 1(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.
[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.3: Monte Carlo results for Design 1

		Slope: β_2				Threshold: α_1				Threshold: α_2			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0030	0.2359	0.2359	0.1859	-0.0083	0.1448	0.1450	0.1158	0.0050	0.1479	0.1479	0.1184
	$f(v\|x)$	-0.0121	0.3717	0.3718	0.2954	0.0028	0.2101	0.2101	0.1665	0.0058	0.2137	0.2138	0.1715
	$\widehat{f}_{NW}(v\|x)$	-0.0404	0.2813	0.2842	0.2249	0.0324	0.1581	0.1613	0.1288	-0.0322	0.1626	0.1657	0.1323
	$\widehat{f}_{LL}(v\|x)$	-0.0373	0.2707	0.2732	0.2159	0.0181	0.1623	0.1633	0.1304	-0.0180	0.1665	0.1675	0.1337
	$\widehat{f}_{2S}(v\|x)$	-0.0425	0.2920	0.2950	0.2336	0.0356	0.1636	0.1674	0.1339	-0.0349	0.1680	0.1715	0.1364
400	MLE	-0.0028	0.1655	0.1655	0.1306	-0.0041	0.1019	0.1020	0.0811	0.0014	0.1056	0.1056	0.0843
	$f(v\|x)$	-0.0133	0.2608	0.2611	0.2078	0.0021	0.1523	0.1523	0.1218	0.0019	0.1513	0.1513	0.1212
	$\widehat{f}_{NW}(v\|x)$	-0.0336	0.1889	0.1918	0.1526	0.0302	0.1099	0.1139	0.0917	-0.0307	0.1134	0.1175	0.0939
	$\widehat{f}_{LL}(v\|x)$	-0.0297	0.1870	0.1893	0.1497	0.0153	0.1142	0.1152	0.0927	-0.0163	0.1169	0.1180	0.0945
	$\widehat{f}_{2S}(v\|x)$	-0.0357	0.1950	0.1982	0.1573	0.0317	0.1131	0.1175	0.0947	-0.0323	0.1156	0.1200	0.0961
600	MLE	0.0004	0.1343	0.1343	0.1062	-0.0032	0.0849	0.0849	0.0675	-0.0018	0.0867	0.0867	0.0696
	$f(v\|x)$	-0.0078	0.2161	0.2162	0.1730	0.0042	0.1276	0.1276	0.1014	0.0004	0.1241	0.1241	0.1004
	$\widehat{f}_{NW}(v\|x)$	-0.0185	0.1554	0.1564	0.1255	0.0179	0.0922	0.0939	0.0754	-0.0201	0.0932	0.0953	0.0762
	$\widehat{f}_{LL}(v\|x)$	-0.0149	0.1542	0.1549	0.1239	0.0061	0.0956	0.0958	0.0767	-0.0081	0.0958	0.0961	0.0772
	$\widehat{f}_{2S}(v\|x)$	-0.0194	0.1600	0.1612	0.1292	0.0192	0.0940	0.0959	0.0768	-0.0208	0.0946	0.0968	0.0774

[a] Ordered Response Model (LDV III): $y = \sum_{j=1}^{3} j \cdot 1\left(\alpha_{j-1} \leq v + \beta_1 + \beta_2 x + \varepsilon < \alpha_j\right)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined $RMSE$ over 40 fixed grid points.

Table 3.4: Monte Carlo results for Design 2

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0091	0.3339	0.3340	0.2665	0.0011	0.1475	0.1475	0.1160
	$f(v\vert x)$	0.0009	0.4359	0.4358	0.3464	0.0004	0.1979	0.1978	0.1577
	$\widehat{f}_{NW}(v\vert x)$	-0.0626	0.3462	0.3517	0.2803	-0.0512	0.1518	0.1601	0.1278
	$\widehat{f}_{LL}(v\vert x)$	-0.0588	0.3378	0.3428	0.2732	-0.0624	0.1532	0.1654	0.1320
	$\widehat{f}_{2S}(v\vert x)$	-0.0423	0.3665	0.3689	0.2941	-0.0508	0.1590	0.1669	0.1344
400	MLE	0.0085	0.2355	0.2356	0.1882	0.0011	0.1007	0.1007	0.0800
	$f(v\vert x)$	0.0025	0.3012	0.3011	0.2392	-0.0011	0.1366	0.1366	0.1094
	$\widehat{f}_{NW}(v\vert x)$	-0.0427	0.2474	0.2510	0.2008	-0.0243	0.1049	0.1076	0.0860
	$\widehat{f}_{LL}(v\vert x)$	-0.0252	0.2457	0.2469	0.1984	-0.0263	0.1072	0.1103	0.0881
	$\widehat{f}_{2S}(v\vert x)$	-0.0149	0.2567	0.2571	0.2059	-0.0221	0.1086	0.1108	0.0889
600	MLE	0.0063	0.1922	0.1923	0.1534	0.0012	0.0840	0.0840	0.0675
	$f(v\vert x)$	0.0015	0.2487	0.2486	0.1988	-0.0001	0.1121	0.1121	0.0894
	$\widehat{f}_{NW}(v\vert x)$	-0.0354	0.2002	0.2032	0.1624	-0.0346	0.0863	0.0930	0.0747
	$\widehat{f}_{LL}(v\vert x)$	-0.0171	0.2038	0.2045	0.1639	-0.0145	0.0900	0.0911	0.0731
	$\widehat{f}_{2S}(v\vert x)$	-0.0226	0.2070	0.2082	0.1664	-0.0343	0.0882	0.0947	0.0763

[a] Binary Choice Model (LDV I): $y = 1(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined $RMSE$ over 40 fixed grid points.

Table 3.5: Monte Carlo results for Design 2

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0054	0.1801	0.1802	0.1432	-0.0043	0.0832	0.0833	0.0661
	$f(v\|x)$	0.0036	0.4186	0.4186	0.3295	0.0020	0.1790	0.1790	0.1401
	$\widehat{f}_{NW}(v\|x)$	-0.1055	0.3225	0.3393	0.2733	0.0159	0.1408	0.1417	0.1112
	$\widehat{f}_{LL}(v\|x)$	-0.1370	0.3027	0.3322	0.2707	0.0208	0.1358	0.1374	0.1076
	$\widehat{f}_{2S}(v\|x)$	-0.2629	0.2236	0.3451	0.2935	0.0560	0.0967	0.1118	0.0893
400	MLE	-0.0002	0.1278	0.1277	0.1026	-0.0027	0.0583	0.0584	0.0463
	$f(v\|x)$	-0.0030	0.2821	0.2820	0.2218	0.0016	0.1273	0.1273	0.1008
	$\widehat{f}_{NW}(v\|x)$	-0.0671	0.2358	0.2451	0.1965	0.0112	0.1079	0.1084	0.0857
	$\widehat{f}_{LL}(v\|x)$	-0.0762	0.2297	0.2419	0.1938	0.0136	0.1067	0.1076	0.0848
	$\widehat{f}_{2S}(v\|x)$	-0.0531	0.2451	0.2507	0.1981	0.0107	0.1110	0.1115	0.0875
600	MLE	0.0012	0.1036	0.1035	0.0821	-0.0011	0.0471	0.0471	0.0378
	$f(v\|x)$	-0.0032	0.2318	0.2317	0.1848	0.0018	0.1051	0.1051	0.0831
	$\widehat{f}_{NW}(v\|x)$	-0.0547	0.1987	0.2060	0.1649	0.0085	0.0904	0.0908	0.0724
	$\widehat{f}_{LL}(v\|x)$	-0.0319	0.2025	0.2049	0.1631	0.0059	0.0928	0.0930	0.0738
	$\widehat{f}_{2S}(v\|x)$	-0.0376	0.2059	0.2092	0.1668	0.0075	0.0914	0.0917	0.0730

[a] Censored Regression Model (LDV II): $y = [v + \beta_1 + \beta_2 x + \varepsilon] \times 1\,(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.
[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.6: Monte Carlo results for Design 2

		Slope: β_2				Threshold: α_1				Threshold: α_2			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0114	0.2472	0.2474	0.1970	-0.0003	0.1544	0.1544	0.1228	0.0030	0.1563	0.1563	0.1242
	$f(v\|x)$	-0.0070	0.3919	0.3919	0.3136	0.0052	0.2274	0.2274	0.1832	-0.0015	0.2274	0.2274	0.1808
	$\widehat{f}_{NW}(v\|x)$	-0.1360	0.2736	0.3055	0.2464	0.0914	0.1661	0.1896	0.1523	-0.0892	0.1643	0.1869	0.1512
	$\widehat{f}_{LL}(v\|x)$	-0.0593	0.2860	0.2920	0.2348	0.0304	0.1770	0.1796	0.1423	-0.0259	0.1759	0.1777	0.1415
	$\widehat{f}_{2S}(v\|x)$	-0.0607	0.3116	0.3174	0.2541	0.0468	0.1823	0.1882	0.1503	-0.0446	0.1790	0.1845	0.1476
400	MLE	0.0058	0.1750	0.1751	0.1383	0.0018	0.1079	0.1079	0.0863	0.0013	0.1081	0.1080	0.0859
	$f(v\|x)$	-0.0049	0.2820	0.2820	0.2263	0.0065	0.1616	0.1617	0.1296	-0.0008	0.1622	0.1622	0.1288
	$\widehat{f}_{NW}(v\|x)$	-0.0965	0.2009	0.2229	0.1787	0.0501	0.1175	0.1277	0.1027	-0.0459	0.1174	0.1260	0.1009
	$\widehat{f}_{LL}(v\|x)$	-0.0436	0.2013	0.2059	0.1645	0.0245	0.1218	0.1243	0.0999	-0.0186	0.1222	0.1235	0.0987
	$\widehat{f}_{2S}(v\|x)$	-0.0478	0.2138	0.2191	0.1747	0.0429	0.1236	0.1308	0.1051	-0.0382	0.1230	0.1287	0.1030
600	MLE	0.0064	0.1422	0.1424	0.1136	-0.0014	0.0883	0.0883	0.0699	0.0017	0.0881	0.0881	0.0697
	$f(v\|x)$	-0.0029	0.2307	0.2306	0.1836	0.0044	0.1325	0.1325	0.1064	0.0023	0.1314	0.1314	0.1043
	$\widehat{f}_{NW}(v\|x)$	-0.0815	0.1657	0.1846	0.1484	0.0333	0.0969	0.1025	0.0823	-0.0304	0.0959	0.1006	0.0807
	$\widehat{f}_{LL}(v\|x)$	-0.0285	0.1656	0.1680	0.1336	0.0100	0.1006	0.1011	0.0808	-0.0062	0.1005	0.1006	0.0802
	$\widehat{f}_{2S}(v\|x)$	-0.0298	0.1753	0.1778	0.1408	0.0243	0.1012	0.1040	0.0835	-0.0216	0.0999	0.1022	0.0820

[a] Ordered Response Model (LDV III): $y = \sum_{j=1}^{3} j \cdot 1\left(\alpha_{j-1} \leq v + \beta_1 + \beta_2 x + \varepsilon < \alpha_j\right)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.7: Monte Carlo results for Design 3

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0291	0.3306	0.3318	0.2641	0.0091	0.1517	0.1519	0.1210
	$f(v\|x)$	0.0241	0.4463	0.4468	0.3583	0.0018	0.2053	0.2053	0.1623
	$\widehat{f}_{NW}(v\|x)$	-0.0535	0.3506	0.3546	0.2822	-0.0492	0.1582	0.1656	0.1325
	$\widehat{f}_{LL}(v\|x)$	-0.0521	0.3479	0.3517	0.2792	-0.0520	0.1603	0.1685	0.1351
	$\widehat{f}_{2S}(v\|x)$	-0.0255	0.3690	0.3698	0.2968	-0.0477	0.1672	0.1739	0.1390
400	MLE	0.0192	0.2368	0.2375	0.1893	0.0056	0.1064	0.1065	0.0846
	$f(v\|x)$	0.0149	0.3169	0.3171	0.2552	0.0036	0.1451	0.1451	0.1149
	$\widehat{f}_{NW}(v\|x)$	-0.0403	0.2442	0.2475	0.1988	-0.0538	0.1106	0.1230	0.0992
	$\widehat{f}_{LL}(v\|x)$	-0.0184	0.2540	0.2546	0.2035	-0.0163	0.1153	0.1164	0.0928
	$\widehat{f}_{2S}(v\|x)$	-0.0246	0.2568	0.2579	0.2071	-0.0525	0.1155	0.1268	0.1018
600	MLE	0.0142	0.1941	0.1946	0.1543	0.0033	0.0849	0.0849	0.0674
	$f(v\|x)$	0.0089	0.2581	0.2582	0.2073	0.0020	0.1174	0.1174	0.0936
	$\widehat{f}_{NW}(v\|x)$	-0.0334	0.2023	0.2050	0.1641	-0.0368	0.0885	0.0958	0.0770
	$\widehat{f}_{LL}(v\|x)$	-0.0131	0.2098	0.2101	0.1684	-0.0063	0.0923	0.0925	0.0736
	$\widehat{f}_{2S}(v\|x)$	-0.0163	0.2097	0.2103	0.1683	-0.0351	0.0925	0.0989	0.0789

[a] Binary Choice Model (LDV I): $y = 1(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.8: Monte Carlo results for Design 3

		Slope: β_2				Intercept: β_1			
N	*Est.*	*Bias*	*SD*	*RMSE*	*MAE*	*Bias*	*SD*	*RMSE*	*MAE*
200	MLE	0.0034	0.1801	0.1801	0.1428	0.0011	0.0821	0.0821	0.0654
	$f(v\|x)$	0.0170	0.4041	0.4044	0.3200	0.0057	0.1827	0.1828	0.1447
	$\widehat{f}_{NW}(v\|x)$	-0.1000	0.3182	0.3334	0.2680	0.0265	0.1429	0.1453	0.1154
	$\widehat{f}_{LL}(v\|x)$	-0.1357	0.2966	0.3261	0.2637	0.0307	0.1380	0.1413	0.1123
	$\widehat{f}_{2S}(v\|x)$	-0.0913	0.3204	0.3331	0.2672	0.0252	0.1452	0.1473	0.1165
400	MLE	0.0038	0.1302	0.1303	0.1036	0.0004	0.0582	0.0582	0.0461
	$f(v\|x)$	0.0125	0.2905	0.2907	0.2291	0.0043	0.1321	0.1321	0.1056
	$\widehat{f}_{NW}(v\|x)$	-0.0564	0.2425	0.2489	0.1978	0.0157	0.1092	0.1103	0.0881
	$\widehat{f}_{LL}(v\|x)$	-0.0712	0.2376	0.2480	0.1981	0.0177	0.1082	0.1096	0.0875
	$\widehat{f}_{2S}(v\|x)$	-0.0398	0.2490	0.2521	0.1992	0.0154	0.1108	0.1119	0.0892
600	MLE	0.0021	0.1073	0.1073	0.0852	-0.0002	0.0464	0.0464	0.0373
	$f(v\|x)$	0.0073	0.2371	0.2372	0.1870	0.0025	0.1044	0.1044	0.0824
	$\widehat{f}_{NW}(v\|x)$	-0.0427	0.2017	0.2061	0.1632	0.0098	0.0883	0.0888	0.0710
	$\widehat{f}_{LL}(v\|x)$	-0.0472	0.2003	0.2058	0.1645	0.0109	0.0878	0.0884	0.0705
	$\widehat{f}_{2S}(v\|x)$	-0.0228	0.2078	0.2090	0.1651	0.0094	0.0895	0.0900	0.0720

[a] Censored Regression Model (LDV II): $y = [v + \beta_1 + \beta_2 x + \varepsilon] \times 1\,(v + \beta_1 + \beta_2 x + \varepsilon > 0)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined *RMSE* over 40 fixed grid points.

Table 3.9: Monte Carlo results for Design 3

		Slope: β_2				Threshold: α_1				Threshold: α_2			
N	$Est.$	$Bias$	SD	$RMSE$	MAE	$Bias$	SD	$RMSE$	MAE	$Bias$	SD	$RMSE$	MAE
200	MLE	0.0154	0.2496	0.2500	0.1997	-0.0074	0.1540	0.1541	0.1230	0.0033	0.1569	0.1569	0.1237
	$f(v\|x)$	0.0157	0.4092	0.4094	0.3289	-0.0131	0.2347	0.2350	0.1867	-0.0031	0.2352	0.2351	0.1871
	$\widehat{f}_{NW}(v\|x)$	-0.1192	0.2856	0.3094	0.2502	0.0836	0.1635	0.1836	0.1485	-0.0888	0.1673	0.1894	0.1511
	$\widehat{f}_{LL}(v\|x)$	-0.0734	0.2873	0.2965	0.2383	0.0358	0.1717	0.1754	0.1410	-0.0433	0.1771	0.1823	0.1445
	$\widehat{f}_{2S}(v\|x)$	-0.0404	0.3178	0.3202	0.2557	0.0375	0.1782	0.1820	0.1461	-0.0453	0.1810	0.1865	0.1482
400	MLE	0.0096	0.1760	0.1762	0.1385	-0.0057	0.1077	0.1078	0.0868	0.0045	0.1099	0.1100	0.0868
	$f(v\|x)$	0.0092	0.2950	0.2951	0.2370	-0.0063	0.1635	0.1636	0.1306	0.0058	0.1646	0.1647	0.1312
	$\widehat{f}_{NW}(v\|x)$	-0.0840	0.2040	0.2206	0.1756	0.0434	0.1155	0.1234	0.0987	-0.0419	0.1172	0.1245	0.0992
	$\widehat{f}_{LL}(v\|x)$	-0.0373	0.2060	0.2093	0.1654	0.0040	0.1204	0.1204	0.0961	-0.0040	0.1230	0.1230	0.0974
	$\widehat{f}_{2S}(v\|x)$	-0.0348	0.2173	0.2200	0.1750	0.0340	0.1214	0.1261	0.1002	-0.0348	0.1236	0.1284	0.1015
600	MLE	0.0086	0.1446	0.1448	0.1141	-0.0069	0.0895	0.0898	0.0719	0.0034	0.0908	0.0909	0.0722
	$f(v\|x)$	0.0072	0.2458	0.2458	0.1976	-0.0078	0.1348	0.1350	0.1084	0.0053	0.1368	0.1369	0.1077
	$\widehat{f}_{NW}(v\|x)$	-0.0725	0.1683	0.1832	0.1459	0.0274	0.0977	0.1014	0.0808	-0.0285	0.0971	0.1012	0.0805
	$\widehat{f}_{LL}(v\|x)$	-0.0402	0.1651	0.1699	0.1341	0.0119	0.1001	0.1008	0.0799	-0.0132	0.1003	0.1012	0.0804
	$\widehat{f}_{2S}(v\|x)$	-0.0196	0.1786	0.1796	0.1443	0.0173	0.1012	0.1027	0.0822	-0.0194	0.1021	0.1039	0.0819

[a] Ordered Response Model (LDV III): $y = \sum_{j=1}^{3} j \cdot 1\left(\alpha_{j-1} \leq v + \beta_1 + \beta_2 x + \varepsilon < \alpha_j\right)$.

[b] Results are based on 2000 replications. For each semiparametric estimator, optimal bandwidths were chosen by minimizing the simulated combined $RMSE$ over 40 fixed grid points.

Chapter 4

Optimal Bandwidth Choice for Estimation of Inverse Conditional–Density–Weighted Expectations

4.1 Introduction

An important class of semiparametric estimators, first proposed by Lewbel (1998), involves the use of kernel–based nonparametric estimates in place of the true conditional density in objects of the form

$$\eta = E\left[\frac{\omega}{f_{V|\mathbf{U}}(V|\,\mathbf{U})}\right], \tag{4.1.1}$$

where $\{\omega^{\top}, V, \mathbf{U}^{\top}\}$ is a random vector, and $f_{V|\mathbf{U}}(\cdot)$ denotes the conditional density function of a scalar continuous random variable V given the random subvector $\mathbf{U}$. This conditional density function is assumed to be estimated here by the ratio of kernel estimators for $f_{V\mathbf{U}}(\cdot)$ and $f_{\mathbf{U}}(\cdot)$, the joint and marginal densities of $(V, \mathbf{U}^{\top})$ and $(\mathbf{U})$ respectively.

For Limited Dependent Variable models, examples of estimators belonging to this class are Lewbel (1998), Lewbel (2000b), Honoré and Lewbel (2002), and Khan and Lewbel (2006). Results derived in this chapter are directly applicable to these estimators. Specifically, if one has a random sample $\{\omega_i^{\top}, v_i, \mathbf{u}_i^{\top}\}$ from the joint distribution of $\{\omega^{\top}, V, \mathbf{U}^{\top}\}$ for $i = 1, \ldots, N$, implementation of any of these estimators requires choosing the numerical value of a bandwidth parameter, h, for the nonparametric kernel estimator of $f_{V|\mathbf{U}}(\cdot)$ in (4.1.1). This chapter discusses formally how to perform this selection. Given that the asymptotic (first–order) distribution of this semiparametric estimator, $\widetilde{\eta}(h)$, of (4.1.1) does not depend on the bandwidth[1] h, any optimal bandwidth formula must be based on a higher–order approximation to such distribution. Technically, such approximations become more complex in the presence of stochastic denominators in a simple 'plug–in' semiparametric estimator of (4.1.1) as explained above. Therefore, we take an alternative approach. We first show that $\widetilde{\eta}(h)$ is asymptotically equivalent to a linear combination of functions of U–statistics, which we call its 'asymptotic representation', $\widehat{\eta}(h)$, and does not have a stochastic denominator. This asymptotic representation includes functions of a U–statistic of order one (a simple sample average), and two data dependent (via the bandwidth parameter h) second–order U–statistics. Finally, we find a formula for the optimal bandwidth that minimizes (with respect to h) the leading terms of an asymptotic approximation to

$$E\left[\|\widehat{\eta}(h)-\eta\|^2\right],$$

where $\|\cdot\|$ is the standard Euclidean[2] norm.

Related calculations to the ones derived here can be found in the literature of bandwidth selection for average derivative estimation, see e.g. Härdle, Hart, Marron, and Tsybakov (1992), Härdle and Tsybakov (1993) and Powell and Stoker (1996). Our results are different from theirs in that the optimal bandwidth for semiparametric kernel estimators of (4.1.1) can be chosen on the basis of bias alone. In particular, we show that the leading terms in the Mean Squared Error (MSE) are two biases. One is attributed to the pointwise 'smoothing' bias of the kernel density estimator used, and the other to its variance. Linton (1991) called the latter 'degrees–of–freedom' bias. Similar results were found by Jones and Sheather (1991) for the kernel–based integrated squared density derivatives estimator of Hall and Marron (1987), and by Ichimura and Linton (2005) for a kernel–based implementation of Hirano, Imbens, and Ridder (2003)'s estimators of treatment effects. Linton (1991) discussed a similar result for the variance estimator in the presence of unknown mean. Furthermore, unlike the standard case in average derivative estimation[3], semiparametric estimation of (4.1.1) could include discrete elements (specifically in $\mathbf{U}$) through its nonparametric component without the need of additional conditions. We explain this extension in greater detail in our discussion below.

One of the main conclusion from this chapter is that the derived asymptotically optimal bandwidth, $h_{\mathbf{opt}}$, must shrink more rapidly to zero than it would be for optimal pointwise kernel estimation of $f_{V|\mathbf{U}}(\cdot)$, i.e. estimating this function at a point. In this sense, 'asymp-

[1] See Lewbel (1998), Lewbel (2000a) Lewbel (2000b), Honoré and Lewbel (2002), and Khan and Lewbel (2006) for precise derivations.

[2] Similarly, we could replace $\|\mathbf{a}\|$ everywhere in this chapter by $\|\mathbf{a}\|_{\mathbf{W}}=\mathbf{a}^{\top}\mathbf{W}\mathbf{a}$, where $\mathbf{W}$ is any positive semidefinite weighting matrix. The results will not change.

[3] Horowitz and Härdle (1996) adapted the average derivative estimator to allow for some discrete components. This requires additional conditions than in the standard case.

totic undersmoothing' is necessary for $\sqrt{N}$–consistent estimation of (4.1.1). This feature is explained in the unifying theory of Goldstein and Messer (1992), whose main focus was to highlight differences in the conditions of limiting theory between nonparametric and semiparametric estimation, but did not address the issue of bandwidth selection for particular applications such as the one discussed here.

The remainder of the chapter is organized as follows: Section 4.2, presents the notation and assumptions used throughout the chapter. In this section, we analyze the sensitivity of kernel–based semiparametric estimator of (4.1.1) to the choice of bandwidth, and its kernel's order via a second–order asymptotic expansion of its MSE. We also make explicit the difference between 'nonparametric' and 'semiparametric' optimal bandwidths. Section 4.3 discusses how to exploit the asymptotic representation of $\widetilde{\eta}(h)$ in order to construct a simple estimator of the optimal bandwidth. We also prove its consistency. In Section 4.4, a Monte Carlo experiment is performed to assess the small sample behavior of the proposed 'plug–in' estimator of the optimal bandwidth. We also compare its performance against other reference rules proposed in the literature for estimation of the nonparametric component $f_{V|\mathbf{U}}(\cdot)$. Section 4.5 examines how results in Section 4.2 can be extended to cases when some components of $\mathbf{U}$ are discrete, and outlines how a bootstrap procedure for bandwidth selection, shown to work for average derivatives, can be adapted to work in our framework. Section 4.6 summarizes and gives concluding remarks. All proofs are presented in the Appendix.

4.2 Asymptotic Mean Square Error

Firstly, we introduce some notation and definitions that will aid the latter discussion.

4.2.1 Framework

We assume that each observation in a data set, $\{\omega_i^\top, v_i, \mathbf{u}_i^\top\}$, is an independently, identically distributed draw from the joint distribution of $\{\omega^\top, V, \mathbf{U}^\top\}$ for $i = 1, \ldots, N$, where $\mathbf{U}$ is a $d-1$ vector, V is a scalar, and ω another $\dim(\omega) \times 1$ observed vector of random variables or known functions of random variables. The distributions of $\mathbf{U}$ and $(V, \mathbf{U}^\top)$ are absolutely continuous with respect to some Lebesgue measures, with Radon–Nikodym densities $f_{\mathbf{U}}(\mathbf{u})$ and $f_{V\mathbf{U}}(v, \mathbf{u})$ with bounded supports $\Omega_{\mathbf{U}}$ and $\Omega_{V\mathbf{U}}$ respectively.

For a bandwidth sequence $h \equiv h(N) \to 0$ and $N \to \infty$, the nonparametric estimators of the unknown densities $f_{\mathbf{U}}(\mathbf{u})$ and $f_{V\mathbf{U}}(v,\mathbf{u})$ used here are the well known kernel smoothers:

$$\widehat{f}_{\mathbf{U}}(\mathbf{u}_i;h) \equiv \frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{h^{d-1}}\mathcal{K}\left(\frac{\mathbf{u}_j-\mathbf{u}_i}{h}\right), \text{ and} \tag{4.2.1}$$

$$\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h) \equiv \frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{1}{h^{d}}W\left(\frac{v_j-v_i}{h}\right)\mathcal{K}\left(\frac{\mathbf{u}_j-\mathbf{u}_i}{h}\right) \tag{4.2.2}$$

respectively. Here

$$\mathcal{K}(x_1,\ldots,x_{d-1}) = \prod_{j=1}^{d-1}K(x_j),\ x=(x_1,\ldots,x_{d-1})\in\Re^{d-1},$$

where K and W are one–dimensional bounded symmetric kernel functions that integrate to one. We have also used the 'leave–one–out' paradigm in the construction of our smoothers above. A natural estimator[4] for $f_{V\mathbf{U}}(v_i|\,\mathbf{u}_i)$ is then given by $\widehat{f}_{V|\mathbf{U}}(v_i|\,\mathbf{u}_i;h) = \widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)\,/\widehat{f}_{\mathbf{U}}(\mathbf{u}_i;h)$, and its inverse can be estimated by

$$\widehat{l}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h) = \frac{\widehat{f}_{\mathbf{U}}(\mathbf{u}_i;h)}{\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)},$$

and an estimator of $\eta \equiv E\left[\omega/f(v|\,\mathbf{u})\right]$ is then given by

$$\widetilde{\eta}(h) = N^{-1}\sum_{i=1}^{N}\omega_i\widehat{l}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)\,. \tag{4.2.3}$$

As previously noted, this estimator is technically inconvenient to handle given the presence of the stochastic denominator in $\widehat{l}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)$. Therefore, we also define an asymptotic representation which will be the basis of our analysis below,

$$\widehat{\eta}(h) = N^{-1}\sum_{i=1}^{N}\omega_i\widehat{L}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)\,, \tag{4.2.4}$$

where

$$\begin{aligned}\widehat{L}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h) = {} & \frac{f_{\mathbf{U}i}}{f_{V\mathbf{U}i}} + 2\frac{\widehat{f}_{\mathbf{U}i}}{f_{V\mathbf{U}i}} - 2\frac{f_{\mathbf{U}i}\widehat{f}_{V\mathbf{U}i}}{f^2_{V\mathbf{U}i}}\\ & - \frac{\widehat{f}_{V\mathbf{U}i}\widehat{f}_{\mathbf{U}i}}{f^2_{V\mathbf{U}i}} + \frac{f_{\mathbf{U}i}\widehat{f}^2_{V\mathbf{U}i}}{f^3_{V\mathbf{U}i}},\end{aligned} \tag{4.2.5}$$

and $\widehat{f}_{\mathbf{U}i} \equiv \widehat{f}_{\mathbf{U}}(\mathbf{u}_i;h)$, $\widehat{f}_{V\mathbf{U}i} \equiv \widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i;h)$.

[4]This estimator was first proposed by Rosenblatt (1969), for the case $d = 2$, and later analyzed by Hyndman, Bashtannyk, and Grunwald (1996).

Now, let us define the following quantities:

$$\begin{aligned}
\widehat{\delta}_1 &\equiv N^{-1}\sum_{i=1}^{N}\omega_i\left(f_{\mathbf{U}i}/f_{V\mathbf{U}i}\right),\\
\widehat{\delta}_2\left(h\right) &\equiv N^{-1}\sum_{i=1}^{N}\omega_i(\widehat{f}_{\mathbf{U}i}/f_{V\mathbf{U}i}),\\
\widehat{\delta}_3\left(h\right) &\equiv N^{-1}\sum_{i=1}^{N}\omega_i(f_{\mathbf{U}i}\widehat{f}_{V\mathbf{U}i}/f^2_{V\mathbf{U}i}),\\
\widehat{\delta}_4\left(h\right) &\equiv N^{-1}\sum_{i=1}^{N}\omega_i\left(\widehat{f}_{V\mathbf{U}i}\widehat{f}_{\mathbf{U}i}/f^2_{V\mathbf{U}i}\right),\\
\widehat{\delta}_5\left(h\right) &\equiv N^{-1}\sum_{i=1}^{N}\omega_i\left(f_{\mathbf{U}i}\widehat{f}^2_{V\mathbf{U}i}/f^3_{V\mathbf{U}i}\right).
\end{aligned}$$

It then follows that

$$\widehat{\eta}\left(h\right) = \widehat{\delta}_1 + 2\widehat{\delta}_2\left(h\right) - 2\widehat{\delta}_3\left(h\right) - \widehat{\delta}_4\left(h\right) + \widehat{\delta}_5\left(h\right), \tag{4.2.6}$$

That is, $\widehat{\eta}\left(h\right)$ can be written as a linear combination of functions of certain U–statistics. In particular, $\widehat{\delta}_2\left(h\right)$ and $\widehat{\delta}_3\left(h\right)$ are generic second–order U–statistics:

$$\begin{aligned}
\widehat{\delta}_2\left(h\right) &= \binom{N}{2}^{-1}\sum_{i<j}\frac{\varpi_{2i}+\varpi_{2j}}{2h^{d-1}}\mathcal{K}\left(\frac{\mathbf{u}_i-\mathbf{u}_j}{h}\right)\\
&\equiv \binom{N}{2}^{-1}\sum_{i<j}p_2\left(\mathbf{t}_{2i},\mathbf{t}_{2j};h\right), \text{ and}\\
\widehat{\delta}_3\left(h\right) &= \binom{N}{2}^{-1}\sum_{i<j}\frac{\varpi_{3i}+\varpi_{3j}}{2h^{d}}W\left(\frac{v_i-v_j}{h}\right)\mathcal{K}\left(\frac{\mathbf{u}_i-\mathbf{u}_j}{h}\right)\\
&\equiv \binom{N}{2}^{-1}\sum_{i<j}p_3\left(\mathbf{t}_{3i},\mathbf{t}_{3j};h\right),
\end{aligned}$$

where $\mathbf{t}^{\top}_{3\tau i} = \left(\varpi^{\top}_{3\tau i}, v_i, \mathbf{u}^{\top}_i\right)$, and $\mathbf{t}^{\top}_{3\tau i} = \left(\varpi^{\top}_{3\tau i}, v_i, \mathbf{u}^{\top}_i\right)$, with $\varpi_{2i} \equiv \omega_i/f_{V\mathbf{U}i}$, and $\varpi_{3i} \equiv \omega_i f_{\mathbf{U}i}/f^2_{V\mathbf{U}i}$ respectively. By simple inspection, we notice that these U–statistics 'kernel' functions $p_2\left(\cdot\right)$ and $p_3\left(\cdot\right)$ are symmetric – that is, $p_2\left(\mathbf{t}_{2i},\mathbf{t}_{2j};h\right) = p_2\left(\mathbf{t}_{2j},\mathbf{t}_{2i};h\right)$ and $p_3\left(\mathbf{t}_{3i},\mathbf{t}_{3j};h\right) = p_3\left(\mathbf{t}_{3j},\mathbf{t}_{3i};h\right)$. Powell, Stock, and Stoker (1989) derived first–order limiting theory for this type of linear functions that involves data–dependent (via the bandwidth parameter h) U–statistics. Similarly, we also define $\varpi_{4i} \equiv \omega_i/f^2_{V\mathbf{U}i}$, and $\varpi_{5i} \equiv \omega_i f_{\mathbf{U}i}/f^3_{V\mathbf{U}i}$.

It then follows, under conditions explained below, that

$$\begin{aligned}\eta &= \lim_{h\to 0} E\left[\widehat{\eta}\left(h\right)\right] \\ &= E\left[\widehat{\delta}_1\right] + 2\times \lim_{h\to 0} E\left[\widehat{\delta}_2\left(h\right)\right] - 2\times \lim_{h\to 0} E\left[\widehat{\delta}_3\left(h\right)\right] - \lim_{h\to 0} E\left[\widehat{\delta}_4\left(h\right)\right] + \lim_{h\to 0} E\left[\widehat{\delta}_5\left(h\right)\right] \\ &= \eta + 2\eta - 2\eta - \eta + \eta.\end{aligned}$$

Also, notice that by construction

$$\widetilde{\eta}\left(h\right) - \widehat{\eta}\left(h\right) = \widehat{\vartheta}\left(h\right),$$

where $\widehat{\vartheta}\left(h\right) = N^{-1}\sum_{i=1}^{N}(\widehat{\vartheta}_{1i}\left(h\right) - \widehat{\vartheta}_{2i}\left(h\right))\omega_i$, with

$$\begin{aligned}&\widehat{\vartheta}_{1i}\left(h\right) \equiv (\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i})^2(\widehat{f}_{\mathbf{U}i} - f_{\mathbf{U}i})/(f_{V\mathbf{U}i}^2\widehat{f}_{V\mathbf{U}i}), \text{ and} \\ &\widehat{\vartheta}_{2i}\left(h\right) \equiv (\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i})^3 f_{\mathbf{U}i}/(f_{V\mathbf{U}i}^3\widehat{f}_{V\mathbf{U}i}).\end{aligned}$$

4.2.2 Sensitivity Analysis

The objective of this chapter is to characterize the optimal bandwidth h_{opt} for computing η. Towards that end, we make the following assumptions:

Assumption A:

(A1) The kernels $W : [-1,1] \to \Re$, and $K : [-1,1] \to \Re$ are bounded, continuously differentiable, symmetric such that $\int W\left(c\right) dc = \int K\left(c\right) dc = 1$.

(A2) Kernels $W\left(c\right)$, and $K\left(c\right)$ have order P, that is, there exists a positive integer $P \geq 2$ such that $\int c^j W\left(c\right) dc = \int c^j K\left(c\right) dc = 0$, $j = 1, \ldots, P-1$, $\int c^P W\left(c\right) dc = d_W \neq 0$ and $\int c^P K\left(c\right) dc = d_K \neq 0$.

(A3) The continuous density functions $f_{\mathbf{u}}\left(\mathbf{u}\right)$, $f_{V\mathbf{U}}(v, \mathbf{u})$ exist and are bounded away from zero. The functions $\pi_1\left(\mathbf{u}\right) = E\left[\varpi_1 \middle| \mathbf{U} = \mathbf{u}\right]$, $\pi_2\left(\mathbf{u}\right) = E\left[\varpi_2 \middle| \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\pi}_1\left(v, \mathbf{u}\right) = E\left[\varpi_1 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\pi}_2\left(v, \mathbf{u}\right) = E\left[\varpi_2 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\pi_3\left(v, \mathbf{u}\right) = E[\varpi_3 | V = v, \mathbf{U} = \mathbf{u}]$, $\pi_4\left(v, \mathbf{u}\right) = E[\varpi_4 | V = v, \mathbf{U} = \mathbf{u}]$, $\pi_5\left(v, \mathbf{u}\right) = E[\varpi_5 | V = v, \mathbf{U} = \mathbf{u}]$ exist and have bounded continuous partial derivatives up to the order P on their compact supports $\Omega_{\mathbf{U}} \equiv \prod_{j=1}^{d-1}\left[\underline{U}_j, \overline{U}_j\right]$ and $\Omega_{v\mathbf{u}} \equiv \left[\underline{V}, \overline{V}\right] \times \Omega_{\mathbf{U}}$ respectively, for $-\infty < \underline{V} < \overline{V} < \infty$, and $-\infty < \underline{U}_j < \overline{U}_j < \infty$, for $j = 1, \ldots, d-1$.

(A4) $\sup_{\Omega_{V\mathbf{U}}} \|\omega\| < \infty$, and $E\left[\|\omega\|^{\epsilon} \middle| V = v, \mathbf{U} = \mathbf{u}\right]$ has bounded continuous partial derivatives up to order P on their compact support, for $\epsilon = 1, 2, 3, 4$

(A5) $h_{\mathbf{opt}} \propto N^{-1/(P+d)}$.

Assumptions (A1)–(A3) are standard conditions when using kernel smoothers ensuring the regularity of W, $\mathcal{K}$, $f_{V\mathbf{U}}$, and $f_{\mathbf{U}}$. Assumption (A4) will facilitate the proofs and can be relaxed at the expense of more complicated mathematics. The last Assumption, (A5), predefines the optimal rate of $h_{\mathbf{opt}}$, which is derived below. The following Lemma guarantees that the MSE–expansion of $\widehat{\eta}(h)$ is equivalent to that of $\widetilde{\eta}(h)$ up to the third power.

Lemma 4.2.1 *(Asymptotic Representation) Under Assumptions* (A1)–(A5),

$$\sqrt{N}\widehat{\vartheta}(h) = o_p\left(N^{-(P-d)/(P+d)}\right), \text{ as } N \to \infty.$$

Proof. See Appendix. ■

This Lemma guarantees the asymptotic equivalence between $\widehat{\eta}(\cdot)$ and $\widetilde{\eta}(\cdot)$, which means that (4.2.3) may be replaced by (4.2.4) for purpose of this analysis. We make an additional technical assumption before we state the main result of this chapter:

(A6) The vectors of errors $\varepsilon_1 = \varpi_1 - \pi_1(\mathbf{u})$, $\widetilde{\varepsilon}_1 = \varpi_1 - \widetilde{\pi}_1(v, \mathbf{u})$, $\varepsilon_2 = \varpi_2 - \pi_2(\mathbf{u})$, $\widetilde{\varepsilon}_2 = \varpi_2 - \widetilde{\pi}_2(v, \mathbf{u})$, $\varepsilon_3 = \varpi_3 - \pi_3(v, \mathbf{u})$, $\varepsilon_4 = \varpi_4 - \pi_4(v, \mathbf{u})$, and $\varepsilon_5 = \varpi_5 - \pi_5(v, \mathbf{u})$ are such that $\sigma_1^2(\mathbf{u}) = E\left[\varepsilon_1^{\top}\varepsilon_1 \middle| \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\sigma}_1^2(v, \mathbf{u}) = E\left[\widetilde{\varepsilon}_1^{\top}\widetilde{\varepsilon}_1 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\sigma_2^2(\mathbf{u}) = E\left[\varepsilon_2^{\top}\varepsilon_2 \middle| \mathbf{U} = \mathbf{u}\right]$, $\sigma_3^2(v, \mathbf{u}) = E\left[\varepsilon_3^{\top}\varepsilon_3 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\sigma_4^2(v, \mathbf{u}) = E[\varepsilon_4^{\top}\varepsilon_4 | V = v, \mathbf{U} = \mathbf{u}]$, $\sigma_5^2(v, \mathbf{u}) = E\left[\varepsilon_5^{\top}\varepsilon_5 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\sigma_{12}(\mathbf{u}) = E\left[\varepsilon_1^{\top}\varepsilon_2 \middle| \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\sigma}_{13}(v, \mathbf{u}) = E[\widetilde{\varepsilon}_1^{\top}\varepsilon_3 | V = v, \mathbf{U} = \mathbf{u}]$, $\widetilde{\sigma}_{14}(v, \mathbf{u}) = E\left[\widetilde{\varepsilon}_1^{\top}\varepsilon_4 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\sigma}_{15}(v, \mathbf{u}) = E[\widetilde{\varepsilon}_1^{\top}\varepsilon_5 | V = v, \mathbf{U} = \mathbf{u}]$, $\widetilde{\sigma}_{23}(v, \mathbf{u}) = E\left[\widetilde{\varepsilon}_2^{\top}\varepsilon_3 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\widetilde{\sigma}_{24}(v, \mathbf{u}) = E[\widetilde{\varepsilon}_2^{\top}\varepsilon_4 | V = v, \mathbf{U} = \mathbf{u}]$, $\widetilde{\sigma}_{25}(v, \mathbf{u}) = E\left[\widetilde{\varepsilon}_2^{\top}\varepsilon_5 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\sigma_{34}(v, \mathbf{u}) = E\left[\varepsilon_3^{\top}\varepsilon_4 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, $\sigma_{35}(v, \mathbf{u}) = E\left[\varepsilon_3^{\top}\varepsilon_5 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$, and $\sigma_{45}(v, \mathbf{u}) = E\left[\varepsilon_4^{\top}\varepsilon_5 \middle| V = v, \mathbf{U} = \mathbf{u}\right]$ are bounded on their respective compact supports $\Omega_{\mathbf{U}}$ and $\Omega_{V\mathbf{U}}$.

We now formulate the Mean Square Error of $\widehat{\eta}(h)$ for η, in terms of the dominant components in an asymptotic expansion.

Theorem 4.2.2 *If Assumptions* (A1)–(A3), *and* (A6), *hold, then*

$$E\left[\|\widehat{\eta}(h) - \eta\|^2\right] = O\left(N^{-1}\right) + \left\|\mathfrak{B}_1 h^P + \mathfrak{B}_2 N^{-1} h^{-d}\right\|^2 + O\left(\frac{h^P}{N} + \frac{1}{N^2}\right) + o\left(\frac{h^P}{N} + \frac{1}{N^2 h^{2d}} + h^{2P}\right), \tag{4.2.7}$$

as $N \to \infty$, *where*

$$\mathfrak{B}_1 = \int \pi_2(\mathbf{u}) S_K(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) \, d\mathbf{u} - \int \pi_3(v, \mathbf{u}) S_{WK}(v, \mathbf{u}) f_{V\mathbf{U}}(v, \mathbf{u}) \, dv d\mathbf{u}, \tag{4.2.8}$$

$$\mathfrak{B}_2 = C_{WK} \int \pi_3(v, \mathbf{u}) f_{V\mathbf{U}}(v, \mathbf{u}) \, dv d\mathbf{u}, \tag{4.2.9}$$

and

$$C_{WK} = \left[\int W^2(c)\, dc\right] \left[\int K^2(c)\, dc\right]^{d-1},$$

$$S_K(\mathbf{u}) = d_K \frac{1}{P} \sum_{j=1}^{d-1} \frac{\partial^P f_{\mathbf{U}}(\mathbf{u})}{\partial u_j^P},$$

$$S_{WK}(v, \mathbf{u}) = \frac{1}{P} \left[d_W \frac{\partial^P f_{V\mathbf{U}}(v, \mathbf{u})}{\partial v^P} + d_K \sum_{j=1}^{d-1} \frac{\partial^P f_{V\mathbf{U}}(v, \mathbf{u})}{\partial u_j^P} \right].$$

Proof. See Appendix. ■

The first bias, $\mathfrak{B}_1$, is related to the 'smoothing' bias of the kernel smoother used, while the second bias, $\mathfrak{B}_2$, comes from its pointwise variance. This 'degrees-of-freedom' bias dominates the $O\left(N^{-2}h^{-d}\right)$ variance term that would otherwise appear in the expansion (see Powell and Stoker (1996) for such calculation).

Optimization

The result of Theorem 4.2.2 can be used to optimize the performance of (4.2.4) with respect to bandwidth choice and order of kernel.

Choice of h

The asymptotically optimal bandwidth is obtained by minimizing (4.2.7) on the basis of h. This is achieved when

$$h_{\mathbf{opt}} = C_0 \times N^{-1/(P+d)}, \tag{4.2.10}$$

where C_0 is a proportionality constant. The choice of bandwidth equates the leading orders of both biases, $\mathfrak{B}_1 h^P$ and $\mathfrak{B}_2 N^{-1} h^{-d}$. By choosing this bandwidth, we have

$$\begin{aligned} E\left[\|\widehat{\eta} - \eta\|^2\right] &= O\left(N^{-1}\right) \\ &+ \left\|\mathfrak{B}_1 C_0^P + \mathfrak{B}_2 C_0^{-d}\right\|^2 N^{-2P/(P+d)} \\ &+ O\left(N^{-2}\right) + o(N^{-2P/(P+d)}), \text{ as } N \to \infty. \end{aligned} \tag{4.2.11}$$

That is, the decreasing rate of the best bandwidth optimizing second–order terms is of order $N^{-1/(P+d)}$, which results in an optimal MSE–rate of convergence of N^{-1}. In comparison with the leading term, the second term in (4.2.11) is not small in general, since their ratio is $O(N^{-(P-d)/(P+d)})$. This means that very large values of N are needed before its influence eventually disappears.

Interestingly, unlike other semiparametric estimators (see Hall and Marron (1987), and Linton (1995)) the use of 'leave–one–out' estimators ((4.2.1) and (4.2.2)) has not fully eliminated the 'degrees–of–freedom' bias[5] of order $O(N^{-2}h^{-2d})$.

Choice of P

If we believe that $f_{V\mathbf{U}}$, $f_{\mathbf{U}}$, $\pi_2(\mathbf{u})$ and $\pi_3(v,\mathbf{u})$ are infinitely many times continuously differentiable, it follows from (4.2.11) that the best $O\left(N^{-1}\right)$ rate of the MSE, is not attained unless $P > d$. For example, in the case $d = 2$, we must use $P > 2$. As Assumption (A2) permits, a higher value for P must be chosen for larger values of d. In this sense, the use of oscillating higher–order kernels guarantees the best rate of convergence.

4.2.3 'Nonparametric' vs 'Semiparametric' Optimal Bandwidths

For the case $d = 2$, the asymptotic properties of the kernel estimator $\widehat{f}_{V|\mathbf{U}}(v|\,\mathbf{u};h)$, used in (4.2.3), were first derived by Hyndman, Bashtannyk, and Grunwald (1996), and discussed further by Chen, Linton, and Robinson (2001). When $d > 2$, it follows from their results that the Integrated MSE–minimizing optimal bandwidth is

$$h^{+}_{\mathbf{opt}} \propto N^{-1/(2P+d)}. \tag{4.2.12}$$

A direct comparison with (4.2.10) indicates that in the semiparametric case, the optimal bandwidth shrinks to 0 at a faster rate of its nonparametric component's optimal bandwidth $h^{+}_{\mathbf{opt}}$. This phenomenon is known as 'asymptotic undersmoothing'. Other semiparametric estimators sharing this feature are Robinson (1988), Powell, Stock, and Stoker (1989), Härdle and Stoker (1989), and Härdle, Hart, Marron, and Tsybakov (1992), among others.

It should also be noticed that this comparison does not imply that $h_{\mathbf{opt}}$ is numerically smaller than $h^{+}_{\mathbf{opt}}$ in any particular case or sample size. Particularly, let A_0 be the propor-

[5] Ichimura and Linton (2005) proposed an explicit bias correction mechanism that indeed 'knocked' this term out, allowing for a smaller MSE for Hirano, Imbens, and Ridder (2003)'s estimator. This method can be easily adapted to our framework.

tionality constant[6] in (4.2.12). It then follows that

$$\begin{aligned}
h_{\mathbf{opt}} &= D_0 E_N h_{\mathbf{opt}}^{+}, \text{ where} \\
D_0 &= C_0 / A_0, \\
E_N &= N^{-P/[(2P+d)(P+d)]}.
\end{aligned}$$

D_0 is an adjustment factor that depends on the underlying structure of the bias and variance of $\widehat{f}_{V|\mathbf{U}}$ and $\widetilde{\eta}$. In general, $D_0 \lessgtr 1$, and it does not vary with sample size. On the other hand, the adjustment term for sample size, E_N is always less than 1 when $N \geq 2$. Therefore, whether $h_{\mathbf{opt}}$ is larger or smaller than $h_{\mathbf{opt}}^{+}$ will depend on C_0 being larger or smaller than A_0, and this, in turn, relies on the bias and variance structure in a particular application.

4.3 Optimal 'plug–in' Bandwidth Estimator

If we knew $\mathfrak{B}_1$ and $\mathfrak{B}_2$ in (4.2.11), we can define C_0 (and therefore $h_{\mathbf{opt}}$) by the following minimization problem:

$$C_0 = \underset{C_0 \in \Re^{++}}{\arg\min} \left\| \mathfrak{B}_1 C_0^P + \mathfrak{B}_2 \frac{1}{C_0^d} \right\|^2 .$$

As these quantities are unknown in general, a feasible procedure will be to replace them by consistent estimators based on empirical implementations of (4.2.8) and (4.2.9). These estimators for the 'smoothing' and 'degrees–of–freedom' bias terms are denoted here as $\widehat{\mathfrak{B}}_1$, and $\widehat{\mathfrak{B}}_2$, respectively. However, with a sample of practical size, any kernel–based estimator may be affected by boundary effects which are endemic in kernel density estimation, see Silverman (1986). In view of Assumption (A3), this technicality is resolved here by using a known asymptotic trimming function, $a_\tau(v, \mathbf{u})$ in their construction[7], that is

$$\widehat{\mathfrak{B}}_1(h_0) = \frac{\widetilde{\eta}(\Delta h_0) - \widetilde{\eta}(h_0)}{h_0^P(\Delta^P - 1)}, \tag{4.3.1}$$

$$\widehat{\mathfrak{B}}_2(h_*) = \frac{C_{W\mathcal{K}}}{N} \sum_{i=1}^{N} \widehat{\varpi}_{*3\tau i}, \tag{4.3.2}$$

where $\widehat{\varpi}_{*3\tau i} = w_i a_\tau(v_i, \mathbf{u}_i) \widehat{f}_{*\mathbf{U}i} / \widehat{f}^2_{*V\mathbf{U}i}$, $\varpi_{3\tau i} = w_i a_\tau(v_i, \mathbf{u}_i) f_{\mathbf{U}i} / f^2_{V\mathbf{U}i}$, and Δ is a known constant which is greater than 1. Here we have used $\widehat{f}_{*V\mathbf{U}i} \equiv \widehat{f}_{V\mathbf{U}}(v_i, \mathbf{u}_i; h_*)$, $\widehat{f}_{*\mathbf{U}i} \equiv \widehat{f}_{\mathbf{U}}(\mathbf{u}_i; h_*)$, $f_{V\mathbf{U}i} \equiv f_{V\mathbf{U}}(v_i, \mathbf{u}_i)$, and $f_{\mathbf{U}i} \equiv f_{\mathbf{U}}(\mathbf{u}_i)$ in order to ease notation, where $\widehat{f}_{\mathbf{U}}(\cdot)$, and $\widehat{f}_{V\mathbf{U}}(\cdot)$ are given by (4.2.1), and (4.2.2) respectively. The estimator $\widetilde{\eta}(\cdot)$ is like (4.2.3), after replacing ω_i by $\omega_i a_\tau(v_i, \mathbf{u}_i)$ everywhere. The estimator (4.3.1) is similar to the average

[6] See Bashtannyk and Hyndman (2001) and Chen, Linton, and Robinson (2001) for derivations.

[7] Another possibility would be to use boundary kernels, see Gasser, Müller, and Mammitzsch (1985). Fernandes and Monteiro (2005) derived the asymptotic behavior of asymmetric kernel functionals.

derivative as proposed by Powell and Stoker (1996). The quantities h_* and h_0 are pilot bandwidths which must be chosen beforehand.

The type of asymptotic trimming used here is that proposed by Lewbel (2000a):

$$a_\tau (v_i, \mathbf{u}_i) = 1 \left(v_i \in [\underline{V} + \tau, \overline{V} - \tau]\right) \prod_{j=1}^{d-1} 1 \left(\mathbf{u}_i^{[j]} \in [\underline{U}_j + \tau, \overline{U}_j - \tau]\right), \tag{4.3.3}$$

where $1(\cdot)$ is the indicator function that equals 1 if its argument is true and zero otherwise, the values $\underline{U}_j$ and $\overline{U}_j$, were defined in Assumption (A3), and $\mathbf{u}^{[j]}$ refers to the j–th element of the vector $\mathbf{u}$. By using this type of trimming in the construction of our estimators (4.3.1) and (4.3.2), we set to zero all terms in these averages that have observations within a distance τ of the boundary of the support where bias of the kernel estimators are of a different order than for the interior points. The following assumption guarantees that the trimming induced bias goes to zero rapidly, so the consistency of the estimators are not affected.

(A7) The value τ is such that $h_0/\tau \to 0$, and $N\tau^2 \to 0$ as $N \to \infty$.

This trimming has a disadvantage in that it requires knowledge of the support of $(v, \mathbf{u})$. Nevertheless, this support could be estimated in practice. For example, Khan and Lewbel (2006) proposed a data–dependent trimming function, by replacing $\underline{U}_j$, $\overline{U}_j$, $\underline{V}$, and $\overline{V}$ in (4.3.3), by the observed maximums and minimums from a sample of N observations of the corresponding variables. They showed that this data–dependent feasible trimming function is asymptotically equivalent to (4.3.3). Their result is applicable in situations where the boundary of the support is unknown, and τ equals the bandwidth used in the kernel estimators above.

Consequently, the optimal bandwidth is estimated as

$$\begin{aligned} \widehat{h}_{\mathbf{opt}} &= \widehat{C}_0 \times N^{-1/(P+d)}, \text{ where} \\ \widehat{C}_0 &= \underset{C_0 \in \Re^{++}}{\arg\min} \left\| \widehat{\mathfrak{B}}_1 C_0^P + \widehat{\mathfrak{B}}_2 \frac{1}{C_0^d} \right\|^2 . \end{aligned} \tag{4.3.4}$$

An interesting characteristic of estimators (4.3.1) and (4.3.2) is that they do not require estimation of higher order derivatives of unknown functions. This feature makes their calculation computationally very simple. Likewise, the minimization problem in (4.3.4) is also computationally straightforward, because it only involves a univariate numerical search over strictly positive real numbers. The consistency[8] of this procedure is ensured by the following proposition:

Proposition 4.3.1 *Let Assumptions* (A1)-(A3), (A4), (A6) *and* (A7) *hold. If* $h_* \to 0$, $h_0 \to 0$, *with* $Nh_*^d \to \infty$, *and* $Nh_0^{2P+d} \to \infty$ *as* $N \to \infty$, *then*

$$\widehat{\mathfrak{B}}_1(h_0) \xrightarrow{p} \mathfrak{B}_1,$$
$$\widehat{\mathfrak{B}}_2(h_*) \xrightarrow{p} \mathfrak{B}_2.$$

Proof. See Appendix. ∎

An important part of this estimator of the optimal bandwidth is the choice of pilot bandwidths h_*, h_0, constant Δ, and trimming parameter τ. Given the conditions on h_*, an obvious way of choosing this bandwidth would be by standard cross–validation methods[9], see Silverman (1986); or using a reference rule for kernel–based conditional density estimators, see Section 4.4. The resulting bandwidth, $\widehat{h}_*$, would be of order $N^{-1/(2P+d)}$. We can then set $h_0 = \widehat{h}_* \times N^{\delta}$, where $0 < \delta < 1/(2P+d)$. As a result, only $\Delta > 1$ and $\tau \geq 0$ are left to be chosen. In practice, for a fixed number of observations, a feasible approach would be to fix the value τ, and choose a high value of Δ and then decrease it until $\widehat{\mathfrak{B}}_1$ does not vary significantly.

A technical proviso explained by Powell and Stoker (1996), for the estimated optimal bandwidth of the average derivative estimator, is also applicable in this framework. That is, we have not shown that Assumption A will guarantee the proposed 'plug–in' estimator $\widetilde{\eta}(\widehat{h}_{\mathbf{opt}})$ is asymptotically equivalent to $\widetilde{\eta}(h_{\mathbf{opt}})$. Firstly, the calculation of $\widetilde{\eta}$ itself would be subject to some trimming with a fixed–size sample. Doing this alone will increase the MSE, by the square of the trimming–induced bias. Secondly, the (stochastic) bandwidth $\widehat{h}_{\mathbf{opt}}$ was calculated using the same data as it is used in the construction of $\widetilde{\eta}$. All the calculations used to derive the asymptotic MSE expansion in Theorem 4.2.2 implicitly assume a fixed rather than a stochastic value of h. From this, it does not immediately follow that $\widehat{h}_{\mathbf{opt}}$ will be of the same order as $h_{\mathbf{opt}}$. Powell and Stoker (1996) discussed possible solutions to this problem in the framework of density–weighted average derivative estimators.

[8] An alternative estimator for $\mathfrak{B}_2$ is given by

$$\binom{N}{2}\sum_{i<j}\left(\frac{\widehat{\varpi}_{*3\tau i}\widehat{f}_{*V\mathbf{U}i}^{-1} + \widehat{\varpi}_{*3\tau j}\widehat{f}_{*V\mathbf{U}j}^{-1}}{4h_0^d}\right) W^2\left(\frac{v_i - v_j}{h_0}\right)\mathcal{K}^2\left(\frac{\mathbf{u}_i - \mathbf{u}_j}{h_0}\right),$$

and its consistency can be proven by the exact same arguments used in this section.

[9] Wand and Jones (1995), chapters 3 and 4, described in great detail many other (computationally simpler) bandwidth selection procedures that could be used instead.

4.4 Monte Carlo Experiments

This section reports the results of a small–scale Monte Carlo investigation of the finite sample behavior of our proposed 'plug–in' estimator for the optimal bandwidth, and the behavior of the associated estimated η's. Samples were generated from a two–dimensional random variable (V, U) having a bivariate normal distribution doubly truncated with respect to both variables. The joint distribution is given by

$$f_{VU}(v,u) = g(v,u)/G, \quad \underline{v} \le v \le \overline{v},\ \underline{u} \le u \le \overline{u},$$

where

$$g(v,u) = \frac{1}{2\pi\sigma_v\sigma_u\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{v-\mu_v}{\sigma_v}\right)^2 - 2\rho\left(\frac{v-\mu_v}{\sigma_v}\right)\left(\frac{u-\mu_u}{\sigma_u}\right) + \left(\frac{u-\mu_u}{\sigma_u}\right)^2\right]\right\},$$

and

$$G = \int_{\underline{u}}^{\overline{u}}\int_{\underline{v}}^{\overline{v}} g(v,u)\, dv du.$$

The marginal density of U is then given by

$$f_U(u) = h(u)/G, \quad \underline{u} \le u \le \overline{u},$$

where

$$h(u) = \frac{1}{\sigma_u}\phi\left(\frac{u-\mu_u}{\sigma_u}\right)\left[\Phi\left(\left(\left(\frac{\overline{v}-\mu_v}{\sigma_v}\right) - \rho\left(\frac{u-\mu_u}{\sigma_u}\right)\right)/\sqrt{1-\rho^2}\right) - \Phi\left(\left(\left(\frac{\underline{v}-\mu_v}{\sigma_v}\right) - \rho\left(\frac{u-\mu_u}{\sigma_u}\right)\right)/\sqrt{1-\rho^2}\right)\right],$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the density and cumulative distribution of a standard normal random variable. The object of interest in this simulation is

$$\eta = E\left[1/f_{V|U}(V|\,U)\right],$$

where

$$f_{V|U}(v|\,u) = g(v,u)/h(u).$$

For simplicity, we set $\underline{v} = \underline{u} = -3$, $\overline{v} = \overline{u} = 3$, $\mu_v = 0$, $\sigma_v^2 = \sigma_u^2 = 6$, and consider 3 designs

$$\begin{aligned}&\text{Design 1: } \mu_u = 0;\\ &\text{Design 2: } \mu_u = 1;\\ &\text{Design 3: } \mu_u = 2.\end{aligned}$$

For each design, we consider 2 cases based on possible values for ρ: (a) $\rho = 0$, and (b) $\rho = 1/4$. Their associated joint, conditional and marginal densities can be visualized in Figures 4.1, 4.2, and 4.3. These designs were chosen so that their associated marginal densities are bounded well above zero at the boundary of their support. A similar property is displayed by their conditional densities.

We set $P = 2$, and set $W \equiv K$ to be a gaussian second–order kernel. Their associated constants are $d_K = 1$, and $C_K = 1/2\sqrt{\pi}$.

Reference Rules

Preliminary bandwidths employed in this simulation study are based on the following three assumptions underlying the joint distribution of (V, U):

(R1) $f_{VU}(v,u) \equiv g(u,v)$, with $\underline{v} = \underline{u} = -\infty$, $\overline{v} = \overline{u} = +\infty$, $\mu_v = \mu_u = u$, and $\sigma_v^2 = \sigma_u^2 = \sigma^2$. Under this assumption, Chen, Linton, and Robinson (2001) calculated

$$A_0 = \sigma \left[\frac{16\pi\sqrt{2}\left(1-\rho^2\right)^{5/2} C_K^2}{\left(15\rho^4 - 50\rho^2 + 39\right) d_K^2} \right]^{1/6} \equiv A_{R1}.$$

(R2) $V|\, U = u \sim N(c + du, (p + qu)^2)$, U is uniform over $[\underline{u}, \overline{u}]$, with $-\infty \le V \le +\infty$. Under this assumption, similar calculations to those in Bashtannyk and Hyndman (2001) shows

$$A_0 = \left[\frac{256 q \sqrt{\pi} C_K^2}{3z\left(4 + w + 8d^2 - 12q^2\right) d_K^2} \right]^{1/6} \equiv A_{R2},$$

where $w = 19q^4 + 4d^4 + 28q^2d^2$, and $z = \left[(p + q\overline{u})^4 - (p + q\underline{u})^4\right] / \left(p + q\overline{u}\right)^4 \left(p + q\underline{u}\right)^4$.

(R3) $f_{VU}(v,u) \equiv g(u,v)$, with $\underline{v} = \underline{u} = -\infty$, $\overline{v} = \overline{u} = +\infty$, $\mu_v = \mu_u = 0$, $\rho = 0$, and $\sigma_v^2 = \sigma^2$. Under this assumption, $f_{V|U}(v|\, u) = f_V(v)$, for which Silverman (1986), pages 45–47, calculated

$$A_0 = \left[\frac{8\sqrt{\pi} C_K}{3 d_K^2} \right]^{1/5} \sigma \equiv A_{R3}.$$

We make these reference rules operational by making A_{R1}, A_{R2}, and A_{R3} vary with each replication. We define these quantities as $\widehat{A}_{R1}$, $\widehat{A}_{R2}$, and $\widehat{A}_{R3}$ respectively. Specifically, let $\{v_i^s, u_i^s\}_{i=1}^{N^s}$ be a size–N^s generated data set at draw s, then $\widehat{A}_{R1}$ is obtained by replacing σ^2, and ρ by $\widehat{\sigma}^2 = N^{-1}\sum_{i=1}^{N}(v_i^s - \overline{v}^s)$, and $\widehat{\rho} = [\widehat{\sigma}^2(N-1)]^{-1}\sum_{i=1}^{N}(v_i^s - \overline{v}^s)(u_i^s - \overline{u}^s)$ respectively. Likewise, $\widehat{A}_{R2}$ is calculated by setting $\underline{u} = \min_{i=1,\ldots,N} u_i^s$, $\overline{u} = \max_{i=1,\ldots,N} u_i^s$, (c, d) as the least squares coefficients from a regression of v on u, and (p, q) as the least squares coefficients from a regression of the squared residuals from the previous regressions on u including a constant term. Similarly, $\widehat{A}_{R3}$ is made operational by replacing σ^2 by $\widehat{\sigma}^2$ as calculated above.

Hence, we calculate $\widetilde{\eta}$ using $\widehat{h}_{R1} = \widehat{A}_{R1}N^{-1/6}$, $\widehat{h}_{R2} = \widehat{A}_{R2}N^{-1/6}$, and $\widehat{h}_{R3} = \widehat{A}_{R3}N^{-1/6}$. Of course, in our designs these bandwidths are neither optimal for η, nor do they have the optimal rate of convergence derived in Section 4.2.2. However, we have chosen them for comparison purposes because of their computational simplicity, as well as the fact that they were the most likely to be chosen by a practitioner prior to the results discussed in this chapter.

We also look at the behavior of $\widetilde{\eta}$ using our 'plug–in' estimator for the optimal bandwidth explained in Section 4.3. We implement this estimator by setting $\widehat{h}_0 = \widehat{h}_{Rl}^* N^{\delta}$, where $\widehat{h}_{Rl}^* \equiv \widehat{h}_{R1}$ for $l = 1, 2, 3$. Other parameters are chosen accordingly and kept constant throughout the experiments, i.e. $\delta = 1/12$, $\tau = 0$ (no trimming) and $\Delta = 2$. The results of 2000 replications are presented in Tables 4.1 to 4.6.

Tables 4.1, 4.3, and 4.5 report the small sample performance of the proposed 'plug–in' estimator for the optimal bandwidth under different conditions. The true optimal bandwidths $h_{\mathbf{opt}}$, are also reported in the first row for each case. As we would expect, higher correlation between V and U entails a larger bandwidth in each design. These results show that the proposed 'plug–in' estimator performs fairly well in all circumstances. This good performance seems not to be affected by the choice of pilot bandwidths, h_* and h_0 in large samples. On the other hand, there is more variation among the bandwidths predicted by the reference rules than among the estimated ones. Numerically, differences among them become more evident when samples sizes are large. The bandwidths' simulated standard deviations increase as we increase the theoretical mean of U. The use of trimming could reduce these variances.

The respective MSE are presented in Tables 4.2, 4.4, and 4.6. We notice that the main component of these simulated MSE is bias instead of variance in each case, as is predicted by the expansion derived here. The use of either the theoretical or estimated optimal bandwidths dominates the use of those predicted by the reference rules in terms of MSE, for all sample sizes, designs and scenarios. The MSE associated to the estimated optimal bandwidths are numerically very close to the simulated theoretical ones.

In our designs, it is also the case that the 'degrees–of–freedom' bias is numerically large, up to 10 times greater than the 'smoothing' bias. Similar calculations for other designs (not presented here) have also shown such a pattern. This lends support to the use of an explicit bias correction mechanism for such term, see for example Ichimura and Linton (2005). This remains a topic for future research.

Finally, Figures 4.4, 4.5 and 4.6 show how close the theoretically optimal bandwidths are to the actual MSE–minimizing bandwidths. The MSE for $\widetilde{\eta}$ are obtained by simulation as functions of a grid of fixed bandwidth parameters. The vertical gray lines represent the optimal bandwidths predicted by Theorem 4.2.2 in each case. Note that even for small sample sizes, the approximation results are very good. However, the quality of the approximation may deteriorate in situations where trimming is necessary.

4.5 Extensions

In this section, we examine the situation in which the conditioning variables, $\mathbf{U}$, have continuous as well as discrete components. In this case, the order of magnitude of the optimal bandwidth only depends on the number of continuously distributed elements of the random vector $\left(V, \mathbf{U}^{\top}\right)$. We also discuss how we could adapt a technique for bandwidth selection proposed by Horowitz (1998) in semiparametric estimation to our bandwidth–selection problem.

4.5.1 Mixed Continuous and Discrete Case

Let us consider the case when the random vector, $\mathbf{U}$, can be partitioned as $\mathbf{U} = \left(\mathbf{U}^{(1)\top}, \mathbf{U}^{(2)\top}\right)$, with $\mathbf{U}^{(1)} \in \Omega_{\mathbf{U}^{(1)}}$, and $\mathbf{U}^{(2)} \in \Omega_{\mathbf{U}^{(2)}}$, where $\Omega_{\mathbf{U}^{(1)}} \subset \Re^{d^{(1)}-1}$, and $\Omega_{\mathbf{U}^{(2)}} \subset \Re^{d^{(2)}}$ is a set with finite number of real points, such that $d^{(1)} + d^{(2)} = d$ with $d^{(1)} \geq 1$ as before. Let $f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}\left(v, \mathbf{u}^{(1)} \middle| \mathbf{u}^{(2)}\right)$ be the probability density of $\left(V, \mathbf{U}^{(1)}\right)$ conditional on $\mathbf{U}^{(2)} = \mathbf{u}^{(2)}$, let $f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}\left(\mathbf{U}^{(1)} \middle| \mathbf{U}^{(2)}\right)$ be the probability density of $\mathbf{U}^{(1)}$ conditional on $\mathbf{U}^{(2)} = \mathbf{u}^{(2)}$, and let $p\left(\mathbf{u}^{(2)}\right)$ be the probability mass that $\mathbf{u}^{(2)} \in \Omega_{\mathbf{U}^{(2)}}$. Then

$$
\begin{aligned}
f_{V\mathbf{U}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) &\equiv f_{V\mathbf{U}}(v, \mathbf{u}) = f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)} \Big| \mathbf{u}^{(2)}) p(\mathbf{u}^{(2)}), \text{ and} \\
f_{\mathbf{U}}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) &\equiv f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)} \Big| \mathbf{u}^{(2)}) p(\mathbf{u}^{(2)}).
\end{aligned}
$$

We also replace (4.2.1) and (4.2.2) with

$$\widehat{f}_{\mathbf{U}}(\mathbf{u}_i^{(1)}, \mathbf{u}_i^{(2)}; h) \equiv \frac{1}{N-1}\sum_{\substack{j=1\\j\neq i}}^{N} \frac{1}{h^{d^{(1)}-1}}\mathcal{K}\left(\frac{\mathbf{u}_j^{(1)}-\mathbf{u}_i^{(1)}}{h}\right) 1\left(\mathbf{u}_j^{(2)}=\mathbf{u}_i^{(2)}\right), \text{ and} \tag{4.5.1}$$

$$\widehat{f}_{V\mathbf{U}}(v_i, \mathbf{u}_i^{(1)}, \mathbf{u}_i^{(2)}; h) \equiv \frac{1}{N-1}\sum_{\substack{j=1\\j\neq i}}^{N} \frac{1}{h^{d^{(1)}}}W\left(\frac{v_j-v_i}{h}\right)\mathcal{K}\left(\frac{\mathbf{u}_j^{(1)}-\mathbf{u}_i^{(1)}}{h}\right) 1\left(\mathbf{u}_j^{(2)}=\mathbf{u}_i^{(2)}\right), \tag{4.5.2}$$

respectively, and recalculate (4.2.3) and (4.2.4). We also redefine $\pi_1(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \pi_1(\mathbf{u})$, $\pi_2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \pi_2(\mathbf{u})$, $\widetilde{\pi}_1(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\pi}_1(v, \mathbf{u})$, $\widetilde{\pi}_2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\pi}_2(v, \mathbf{u})$, $\pi_3(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \pi_3(v, \mathbf{u})$, $\pi_4(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \pi_4(v, \mathbf{u})$, $\pi_5(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \pi_5(v, \mathbf{u})$, $\sigma_1^2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \sigma_1^2(\mathbf{u})$, $\sigma_2^2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \sigma_2^2(\mathbf{u})$, $\sigma_{12}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \sigma_{12}(\mathbf{u})$, $\widetilde{\sigma}_1^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_1^2(v, \mathbf{u})$, $\widetilde{\sigma}_{13}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{13}(v, \mathbf{u})$, $\widetilde{\sigma}_{14}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{14}(v, \mathbf{u})$, $\widetilde{\sigma}_{15}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{15}(v, \mathbf{u})$, $\widetilde{\sigma}_{23}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{23}(v, \mathbf{u})$, $\widetilde{\sigma}_{24}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{24}(v, \mathbf{u})$, $\widetilde{\sigma}_{25}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{25}(v, \mathbf{u})$, $\sigma_3^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \sigma_3^2(v, \mathbf{u})$, $\widetilde{\sigma}_{34}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{34}(v, \mathbf{u})$, $\widetilde{\sigma}_{35}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{35}(v, \mathbf{u})$, $\sigma_4^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})$, $\widetilde{\sigma}_{45}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \widetilde{\sigma}_{45}(v, \mathbf{u})$, and $\sigma_5^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \equiv \sigma_5^2(v, \mathbf{u})$.

In order to extend our results to this mixed case, we need to re–state Assumptions (A3) and (A6) as:

(A3*) $f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\pi_1(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$ and $\pi_2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, understood as functions of $\mathbf{u}^{(1)}$, exist and have bounded continuous partial derivatives up to the order P on $\Omega_{\mathbf{U}^{(1)}} \equiv \prod_{j=1}^{d^{(1)}-1}\left[\underline{U}_j^{(1)}, \overline{U}_j^{(1)}\right]$, where $-\infty < \underline{U}_j^{(1)} < \overline{U}_j^{(1)} < \infty$, for $j = 1, \ldots, d^{(1)}-1$. Furthermore, $f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\widetilde{\pi}_1(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \times f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\widetilde{\pi}_2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\pi_3(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \times f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\pi_4(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \times f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, and $\pi_5(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \times f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$ understood as functions of v and $\mathbf{u}^{(1)}$, exist and have bounded continuous partial derivatives up to the order P on $\Omega_{V\mathbf{U}^{(1)}} \equiv [\underline{V}, \overline{V}] \times \Omega_{\mathbf{U}^{(1)}}$, for $-\infty < \underline{V} < \overline{V} < \infty$. The probability mass function $p(\mathbf{u}^{(2)}) > 0$.

(A6*) The functions, $\sigma_1^2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\sigma_2^2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$ and $\sigma_{12}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)})\, f_{\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(\mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, understood as functions of $\mathbf{u}^{(1)}$, are bounded on their compact support $\Omega_{\mathbf{U}^{(1)}}$. Similarly, $\widetilde{\sigma}_1^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\sigma_3^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\sigma_4^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, $\sigma_5^2(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, and $\widetilde{\sigma}_{lk}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)})f_{V\mathbf{U}^{(1)}|\mathbf{U}^{(2)}}(v, \mathbf{u}^{(1)}|\, \mathbf{u}^{(2)})$, for $\forall l, k =$ 1, 2, 3, 4 such that $l \neq k$, understood as functions of v and $\mathbf{u}^{(1)}$, are bounded on their compact support $\Omega_{V\mathbf{U}^{(1)}}$.

As expected, in this mixed case scenario, similar conditions have to be imposed on the continuous part of the problem, but no new techniques are required in order to prove the following corollary:

Corollary 4.5.1 *Let Assumptions* (A1), (A2) *hold, and Assumptions* (A3*) *and* (A6*) *hold for every* $\mathbf{u}^{(2)} \in \Omega_{\mathbf{u}^{(2)}}$, *then*

$$h^{(1)}_{\mathbf{opt}} = C^{(1)}_0 \times \left(\frac{1}{N}\right)^{1/(P+d^{(1)})}, \text{ where}$$
$$C^{(1)}_0 = \underset{C^{(1)}_0 \in \Re^{++}}{\arg\min} \left\| \mathfrak{B}^{(1)}_1 C^P_0 + \mathfrak{B}^{(1)}_2 \frac{1}{C^d_0} \right\|^2 .$$

and

$$\mathfrak{B}^{(1)}_1 = \sum_{\mathbf{u}^{(2)} \in \Omega_{\mathbf{u}^{(2)}}} \int \pi_2(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) S^{(1)}_K(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) f_{\mathbf{U}}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) d\mathbf{u}^{(1)}$$
$$- \sum_{\mathbf{u}^{(2)} \in \Omega_{\mathbf{u}^{(2)}}} \int \pi_3(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) S^{(1)}_{W\mathcal{K}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) f_{V\mathbf{U}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) dv d\mathbf{u}^{(1)}, \tag{4.5.3}$$
$$\mathfrak{B}^{(1)}_2 = C_{W\mathcal{K}} \sum_{\mathbf{u}^{(2)} \in \Omega_{\mathbf{u}^{(2)}}} \int \pi_3(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) f_{V\mathbf{U}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) dv d\mathbf{u}^{(1)}, \tag{4.5.4}$$

with

$$C_{W\mathcal{K}} = \left[\int W^2(c)\, dc\right] \left[\int K^2(c)\, dc\right]^{d^{(1)}-1},$$
$$S^{(1)}_K(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}) = d_K \frac{1}{P} \sum_{j=1}^{d^{(1)}-1} \frac{\partial^P}{(\partial u^{(1)}_j)^P} f_{\mathbf{u}}(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}),$$
$$S^{(1)}_{W\mathcal{K}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) = \frac{1}{P}\left[d_W \frac{\partial^P}{\partial v^P} f_{V\mathbf{U}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) + d_K \sum_{j=1}^{d^{(1)}-1} \frac{\partial^P}{(\partial u^{(1)}_j)^P} f_{V\mathbf{U}}(v, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}) \right].$$

Proof. See proof of Theorem 4.2.2 in the Appendix. ∎

From this result, and similarly to other nonparametric and semiparametric models (see Delgado and Mora (1995)), we note that the MSE–minimizing rate of bandwidth shrinkage in our case only depends on the number of continuously distributed random variables in our sample from $(V, \mathbf{U})$. The unknown constants (4.5.4) and (4.5.3) can be consistently estimated, by simple extensions of the estimators described in Section 4.3. Likewise, trimming (if needed) should be performed only with respect to the continuously distributed variables, in particular

$$a_\tau(v_i, \mathbf{u}_i^{(1)}) = 1\left(v_i \in \left[\underline{V}+\tau, \overline{V}-\tau\right]\right) \prod_{j=1}^{d^{(1)}-1} 1\left(\mathbf{u}_i^{(1)[j]} \in \left[\underline{U}_j^{(1)}+\tau, \overline{U}_j^{(1)}-\tau\right]\right).$$

4.5.2 Bandwidth Selection using the m–out–of–N Bootstrap

Horowitz (1998), Chapter 2 (page 51), suggested a bootstrap–based method for bandwidth selection in the density–weighted average derivative estimator. This bootstrap technique involves resampling without replacement, and Goh (2004) proved its validity while choosing the asymptotically MSE–minimizing optimal bandwidth of general U–statistics. We could use this technique here because of the fact that (4.2.3) is asymptotically a linear combination of functions of U–statistics (see equation (4.2.4) above).

Some of the notation used earlier is redefined, and new definitions is introduced in order to make the explanation clearer. Firstly, let $\widehat{F}_N$ denote the empirical distribution function of our original random sample $\{\omega_i, v_i, \mathbf{u}_i^\top\}$, for $i = 1, \dots, N$, where $\mathbf{u}_i$ could have a mixed composition as explained in Section 4.5.1. Let $\widehat{F}_N^*$ denote the distribution of the bootstrap sample generated by resampling $m < N$ members of the original sample without replacement. Let's call this sample $\{\omega_i^*, v_i^*, \mathbf{u}_i^{*\top}\}$, for $i = 1, \dots, m$. We also make explicit the dependence of (4.2.3) on these samples, as well as on some fixed bandwidth, h, as

$$\widetilde{\eta}(h; \widehat{F}_N) \equiv \widetilde{\eta}(h), \tag{4.5.5}$$

and similarly $\widetilde{\eta}(h; \widehat{F}_N^*)$ represents (4.2.3) when the bootstrap sample is used in its calculation instead. Let θ be an element of $\Theta \subset \Re^{++}$, a finite set of strictly positive numbers. Then, a bootstrap–based bandwidth selection procedure is carried out in the following way:

Step (1) Calculate $\widetilde{\eta}(cN^{-1/(P+d)}; \widehat{F}_N)$ using the original estimation sample, $\{\omega_i, v_i, \mathbf{u}_i^\top\}_{i=1}^N$, for some known constant c.

Step (2) Resample $m < N$ observations without replacement from the original estimation sample. Calculate $\widetilde{\eta}(\theta m^{-1/(P+d)}; \widehat{F}_N^*)$ using this data, $\{\omega_i^*, v_i^*, \mathbf{u}_i^{*\top}\}_{i=1}^m$, for each $\theta \in \Theta$.

Step (3) Repeat J–times step 2 and call these estimators: $\widetilde{\eta}_j(\theta m^{-1/(P+d)}; \widehat{F}_N^*)$ for $j = 1, \dots, J$. Define $\widehat{\theta}$ to be the solution to the problem

$$\min_{\theta\in\Theta} \frac{1}{J}\sum_{j=1}^{N} \left[\widetilde{\eta}_j(\theta m^{-1/(P+d)}; \widehat{F}_N^*) - \widetilde{\eta}(cN^{-1/(P+d)}; \widehat{F}_N)\right]^2. \tag{4.5.6}$$

Then $\widehat{\theta}$ estimates consistently C_0 in (4.2.10), and the bootstrap estimator of $h_{\mathbf{opt}}$ is given by

$$\widehat{h}^*_{\mathbf{opt}} = \widehat{\theta} N^{-1/(P+d)}.$$

This bandwidth selection mechanism also requires the use of a pilot bandwidth in step 1 as in the 'plug–in' estimator described in Section 4.3. In view of Theorem 4.3.1, we could set $c = \widehat{C}_0$, where $\widehat{C}_0$ is defined in (4.3.4). In practice, we would also minimize (4.5.6) numerically over a suitable grid of values for θ. This is can be rapidly and easily computed.

We do not compare the performance of this procedure against that of the proposed 'plug–in' estimator in Section 4.4, purely because of the computational burden impacting its implementation in a Monte Carlo experiment. Similar difficulties were faced by Goh (2004), who instead used numerical approximations in a limited simulation study with 10 Monte Carlo replications, while setting $J = 100$ for 3 unrelated designs.

4.6 Conclusion

A crucial part of estimators with a nonparametric component is the choice of the smoothing parameter. Our main objective in this chapter is to provide some guidance for choice of bandwidth for a class of semiparametric estimators that employ kernel estimators in the form of inverse–conditional–density weighted averages. By exploiting the fact that these estimators can be asymptotically represented as a linear combination of functions of $U-$statistics, we derive a formula for the optimal bandwidth based on a second–order Mean Squared Error expansion. The derived formula for the optimal bandwidth equates the order of magnitude arising from the squared of the sum of two biases: a 'smoothing bias' and a 'degrees–of–freedom' bias. This formula shows that the optimal bandwidth, for estimating the parameter of interest, must decrease towards zero at a faster rate than the optimal for its nonparametric component. In this sense, asymptotic undersmoothing (as explained in Powell and Stoker (1996)) is needed.

A 'plug–in' estimator of the optimal bandwidth is also constructed exploiting the semiparametric estimator's biases formulae. The problem of random denominators is also addressed in the construction of the proposed estimator through the use of a trimming function. This trimming function, proposed by Lewbel (2000a), is set to give zero–weights in the averages, to observations which are within a certain distance of the boundary of the observed support of the distribution. This estimator is shown to perform fairly well in small samples in a Monte Carlo experiment. We also describe the use of other data–driven bandwidth selectors, such as the bootstrap. In this sense, we also explain how the m–out–of–N bootstrap, proposed by Horowitz (1998), is a viable bandwidth selection alternative, although its validity in our framework is yet to be shown formally, and its performance analyzed in a Monte Carlo experiment. We also discuss how the formula for the optimal bandwidth can be adapted when continuous as well as discrete elements are present in the weighted averages.

Appendix

4.A Main Proofs

Let $\|\cdot\|$ denote Euclidean norm, and let $\langle\cdot,\cdot\rangle$ represent the inner product when applied to vectors. We also use the following results from Masry (1996a) (see Silverman (1978) and Collomb and Härdle (1986) for earlier results):

$$\max_{i=1,\ldots,n}\left|\widehat{f}_{V\mathbf{U}}\left(v_i,\mathbf{u}_i;h\right)-f_{V\mathbf{U}}\left(v_i,\mathbf{u}_i\right)\right|=O_p\left(\sqrt{\frac{\log N}{Nh^d}}+h^P\right), \tag{4.A.1}$$

$$\max_{i=1,\ldots,n}\left|\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i;h\right)-f_{\mathbf{U}}\left(\mathbf{u}_i\right)\right|=O_p\left(\sqrt{\frac{\log N}{Nh^{d-1}}}+h^P\right). \tag{4.A.2}$$

Proof of Lemma 4.2.1

Firstly, from Assumption (A3), it follows that $\widehat{\vartheta}_{2i}\left(h\right)$ is

$$\begin{aligned}\left\|N^{-1/2}\textstyle\sum_{i=1}^N\widehat{\vartheta}_{2i}\left(h\right)\right\| &\le \left(N^{-1/2}\textstyle\sum_{i=1}^N\|\omega_i\|\left|f_{\mathbf{U}i}\right|\right)\left[\min_{i=1,\ldots,n}\left|\widehat{f}_{V\mathbf{U}i}\right|\right]^{-1}\\ &\quad\times\left[\max_{i=1,\ldots,n}\left|\widehat{f}_{V\mathbf{U}i}-f_{V\mathbf{U}i}\right|\right]^3\left[\min_{i=1,\ldots,n}\left|f^3_{V\mathbf{U}i}\right|\right]^{-1}\\ &\le \left(N^{-1/2}\textstyle\sum_{i=1}^N\|\omega_i\|\left|f_{\mathbf{U}i}\right|\right)\left[\min_{i=1,\ldots,n}\left|f_{V\mathbf{U}i}\right|\right]^{-1}\\ &\quad\times\left[\max_{i=1,\ldots,n}\left|\widehat{f}_{V\mathbf{U}i}-f_{V\mathbf{U}i}\right|\right]^3\left[\min_{i=1,\ldots,n}\left|f^3_{V\mathbf{U}i}\right|\right]^{-1}\\ &=O_p\left(\sqrt{N}\right)O_p\left(\left(\sqrt{\frac{\log N}{Nh^d}}+h^P\right)^3\right)=o_p\left(N^{-(P-d)/(P+d)}\right),\end{aligned} \tag{4.A.3}$$

where (4.A.3) follows after observing that

$$\begin{aligned}\inf_{\Omega_{V\mathbf{U}}}\widehat{f}_{V\mathbf{U}}\left(v,\mathbf{u};h\right) &\ge \inf_{\Omega_{V\mathbf{U}}}f_{V\mathbf{U}}\left(v,\mathbf{u}\right)-\sup_{\Omega_{V\mathbf{U}}}\left|\widehat{f}_{V\mathbf{U}}\left(v,\mathbf{u};h\right)-f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\right|\\ &\ge \inf_{\Omega_{V\mathbf{U}}}f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+o_p\left(1\right),\end{aligned}$$

and the last inequality follows from (4.A.1). Finally, by the exact same argument, it also follows that

$$\begin{aligned}\left\|N^{-1/2}\textstyle\sum_{i=1}^{N}\widehat{\vartheta}_{1i}\left(h\right)\right\| &\leq \left(N^{-1/2}\textstyle\sum_{i=1}^{N}\left\|\omega_i\right\|\right)\left[\min_{i=1,\dots,n}\left|f_{V\mathbf{U}i}\right|\right]^{-1} \\ &\quad \left[\max_{i=1,\dots,n}\left|\widehat{f}_{V\mathbf{U}i}-f_{V\mathbf{U}i}\right|\right]^{2}\left[\max_{i=1,\dots,n}\left|\widehat{f}_{\mathbf{U}i}-f_{\mathbf{U}i}\right|\right]\left[\min_{i=1,\dots,n}\left\|f_{V\mathbf{U}i}\right\|^{2}\right] \\ &= o_p\left(N^{-(P-d)/(P+d)}\right),\end{aligned}$$

as required.

Proof of Theorem 4.2.2

This is a long proof. It consists mostly of repetitive steps and calculations. Specifically, we look at the contribution to the MSE from each element on the right–hand side of (4.2.6). Firstly, let us denote

$$\begin{aligned}\delta_1 &= E\left[\varpi_1\right], \\ \delta_2 &= E\left[\pi_2\left(\mathbf{U}\right)f_{\mathbf{U}}\left(\mathbf{U}\right)\right], \\ \delta_3 &= E\left[\pi_3\left(V,\mathbf{U}\right)f_{V\mathbf{U}}\left(V,\mathbf{U}\right)\right], \\ \delta_4 &= E\left[\pi_4\left(V,\mathbf{U}\right)f_{\mathbf{U}}\left(\mathbf{U}\right)f_{V\mathbf{U}}\left(V,\mathbf{U}\right)\right], \text{ and} \\ \delta_5 &= E\left[\pi_5\left(V,\mathbf{U}\right)f_{V\mathbf{U}}^{2}\left(V,\mathbf{U}\right)\right].\end{aligned}$$

Then, by using the definitions in Section 4.2 and the properties of conditional expectations, it follows that

$$\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \eta.$$

In particular, we have for example,

$$\begin{aligned}\delta_3 &= E\left[\pi_3\left(V,\mathbf{U}\right)f_{V\mathbf{U}}\left(V,\mathbf{U}\right)\right] \\ &= E\left[E\left[\left.\frac{\omega f_{\mathbf{U}}\left(\mathbf{U}\right)}{f_{V\mathbf{U}}^{2}\left(V,\mathbf{U}\right)}\right| V=v,\mathbf{U}=\mathbf{u}\right]f_{V\mathbf{U}}\left(V,\mathbf{U}\right)\right] \\ &= E\left[E\left[\left.\frac{\omega f_{\mathbf{U}}\left(\mathbf{U}\right)}{f_{V\mathbf{U}}\left(V,\mathbf{U}\right)}\right| V=v,\mathbf{U}=\mathbf{u}\right]\right] \\ &= E\left[\frac{\omega f_{\mathbf{U}}\left(\mathbf{u}\right)}{f_{V\mathbf{U}}\left(v,\mathbf{u}\right)}\right] = E\left[\frac{\omega}{f_{V|\mathbf{U}}\left(v|\,\mathbf{u}\right)}\right] \\ &= \eta.\end{aligned}$$

Similar results hold for the rest.

Therefore, we can write $E[\|\widehat{\eta}(h)-\eta\|^2]$ as,

$$E\left[\|\widehat{\eta}(h)-\eta\|^2\right] = E\left[\left\|\widehat{\delta}_1-\delta_1\right\|^2\right] \tag{4.A.4}$$
$$+\,4E\left[\left\|\widehat{\delta}_2(h)-\delta_2\right\|^2\right] \tag{4.A.5}$$
$$+\,4E\left[\left\|\widehat{\delta}_3(h)-\delta_3\right\|^2\right] \tag{4.A.6}$$
$$+\,E\left[\left\|\widehat{\delta}_4(h)\right\|^2\right] \tag{4.A.7}$$
$$+\,E\left[\left\|\widehat{\delta}_5(h)\right\|^2\right] \tag{4.A.8}$$
$$+\,4E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_2(h)-\delta_2\right\rangle\right] \tag{4.A.9}$$
$$-\,4E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_3(h)-\delta_3\right\rangle\right] \tag{4.A.10}$$
$$-\,2E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_4(h)-\delta_4\right\rangle\right] \tag{4.A.11}$$
$$+\,2E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_5(h)-\delta_5\right\rangle\right] \tag{4.A.12}$$
$$-\,8E\left[\left\langle\widehat{\delta}_2(h)-\delta_2,\widehat{\delta}_3(h)-\delta_3\right\rangle\right] \tag{4.A.13}$$
$$-\,4E\left[\left\langle\widehat{\delta}_2(h)-\delta_2,\widehat{\delta}_4(h)-\delta_4\right\rangle\right] \tag{4.A.14}$$
$$+\,4E\left[\left\langle\widehat{\delta}_2(h)-\delta_2,\widehat{\delta}_5(h)-\delta_5\right\rangle\right] \tag{4.A.15}$$
$$+\,4E\left[\left\langle\widehat{\delta}_3(h)-\delta_3,\widehat{\delta}_4(h)-\delta_4\right\rangle\right] \tag{4.A.16}$$
$$-\,4E\left[\left\langle\widehat{\delta}_3(h)-\delta_3,\widehat{\delta}_5(h)-\delta_5\right\rangle\right] \tag{4.A.17}$$
$$-\,2E\left[\left\langle\widehat{\delta}_4(h)-\delta_4,\delta_5(h)-\delta_5\right\rangle\right]. \tag{4.A.18}$$

In what follows, we look at each of the 15 terms above.

Term: $E\left[\left\|\widehat{\delta}_1-\delta_1\right\|^2\right]$

Firstly, notice that $E[\|\ \widehat{\delta}_1-\delta_1\ \|^2]=E[\|\ N^{-1}\sum_{i=1}^N\varepsilon_{1i}\ \|^2]+E[\|\ N^{-1}\sum_{i=1}^N\widetilde{\pi}_{1i}-\delta_1\ \|^2]$, and clearly

$$E\left[\varepsilon_{1i}^\top\varepsilon_{1j}\middle|(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)\right]=\begin{cases}\sigma_1^2(V_i,\mathbf{U}_i), & i=j,\\ 0, & i\neq j.\end{cases}$$

so

$$E\left[\left\|\widehat{\delta}_1-\delta_1\right\|^2\right]=O\left(\frac{1}{N}\right). \tag{4.A.19}$$

Terms: $E\left[\left\|\widehat{\delta}_2(h)-\delta_2\right\|^2\right]$ **and** $E\left[\left\|\widehat{\delta}_3(h)-\delta_3\right\|^2\right]$

We have

$$E\left[\left\|\widehat{\delta}_2(h)-\delta_2\right\|^2\right]=E\left[\left\|\frac{1}{N}\sum_{i=1}^N\varepsilon_{2i}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)\right\|^2\right]+E\left[\left\|\frac{1}{N}\sum_{i=1}^N(\zeta_{2i}-E[\zeta_{2i}])\right\|^2\right]$$
$$+\left\|\frac{1}{N}\sum_{i=1}^N E[\zeta_{2i}]-\delta_2\right\|^2, \text{ and}$$
$$E\left[\left\|\widehat{\delta}_3(h)-\delta_3\right\|^2\right]=E\left[\left\|\frac{1}{N}\sum_{i=1}^N\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]+E\left[\left\|\frac{1}{N}\sum_{i=1}^N(\zeta_{3i}-E[\zeta_{3i}])\right\|^2\right]$$
$$+\left\|\frac{1}{N}\sum_{i=1}^N E[\zeta_{3i}]-\delta_3\right\|^2,$$

with $\zeta_{2i}\equiv\pi_2(\mathbf{u}_i)\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)$, and $\zeta_{3i}\equiv\pi_3(v_i,\mathbf{u}_i)\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)$. They are such that $E[\zeta_{2i}]=E[\zeta_{21}]=q_2$, and $E[\zeta_{3i}]=E[\zeta_{31}]=q_3$, where we use the definitions $q_2\equiv E[\pi_2(\mathbf{U}_1)\times\mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1)]$, $q_3\equiv E[\pi_3(V_1,\mathbf{U}_1)W_h(V_2-V_1)\mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1)]$, and notation $W_h(c)=h^{-1}W(c/h)$ and $\mathcal{K}_h(\mathbf{c})=h^{-(d-1)}\mathcal{K}(h^{-1}\mathbf{c})$.

Notice that,

$$E\left[\varepsilon_{2i}^{\top}\varepsilon_{2j}\middle|\mathbf{U}_1,\ldots,\mathbf{U}_N\right]=\begin{cases}\sigma_2^2(\mathbf{U}_i), & i=j,\\ 0, & i\neq j.\end{cases},$$
$$E\left[\varepsilon_{3i}^{\top}\varepsilon_{3j}\middle|(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)\right]=\begin{cases}\sigma_3^2(V_i,\mathbf{U}_i), & i=j,\\ 0, & i\neq j.\end{cases},$$

and we write

$$E\left[\left\|\frac{1}{N}\sum_{i=1}^N\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]=\frac{1}{N^2}\sum_{i=1}^N E\left[\sigma_3^2(v_i,\mathbf{u}_i)\left\|\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]$$
$$=\frac{1}{N(N-1)^2}\int\sigma_3^2(v,\mathbf{u})E\left[\left\|\sum_{t=2}^N W_h(V_t-v)\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right\|^2\right]f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}.$$

We have $E[\|\sum_{t=2}^N W_h(V_t-v)\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\|^2]=(N-1)E(\|W_h(V_1-v)\mathcal{K}_h(\mathbf{U}_1-\mathbf{u})\|^2)+(N-1)(N-2)\|E(W_h(V_1-v)\mathcal{K}_h(\mathbf{U}_1-\mathbf{u}))\|^2$. If follows from Lemmas 4.B.1 and 4.B.2 that

$$E\left[\left\|\sum_{t=2}^{N} W_h\left(V_t-v\right)\mathcal{K}_h\left(\mathbf{U}_t-\mathbf{u}\right)\right\|^2\right]=\left(N-1\right)\left[C_{W\mathcal{K}}f_{V\mathbf{U}}\left(v,\mathbf{u}\right)h^{-d}+\psi_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right]$$
$$+\left(N-1\right)\left(N-2\right)\left\|f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+h^P S_{W\mathcal{K}}\left(v,\mathbf{u}\right)+\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right\|^2.$$

Thence

$$\begin{aligned}E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}\left(v_i,\mathbf{u}_i\right)\right\|^2\right]&=\frac{1}{N}\int\sigma_3^2\left(v,\mathbf{u}\right)f_{V\mathbf{U}}^3\left(v,\mathbf{u}\right)dvd\mathbf{u}\\&+\frac{C_{W\mathcal{K}}}{N^2h^d}\int\sigma_3^2\left(v,\mathbf{u}\right)f_{V\mathbf{U}}^2\left(v,\mathbf{u}\right)dvd\mathbf{u}\\&+O\left(\frac{h^P}{N}+\frac{1}{N^2}\right)+o\left(\frac{1}{N^2h^d}\right),\ N\to\infty,\end{aligned}\tag{4.A.20}$$

by similar arguments, we also show that $E[\|\ N^{-1}\sum_{i=1}^{N}\varepsilon_{2i}\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i\right)\ \|^2]=N^{-1}\int\sigma_2^2\left(\mathbf{u}\right)\times f_{\mathbf{U}}^3\left(\mathbf{u}\right)d\mathbf{u}+N^{-2}h^{-(d-1)}C_K\int\sigma_2^2\left(\mathbf{u}\right)f_{\mathbf{U}}^2\left(\mathbf{u}\right)d\mathbf{u}+O\left(N^{-1}h^P+N^{-2}\right)+o(N^{-2}h^{-(d-1)}),\ N\to\infty$.

Next,

$$\begin{aligned}E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{li}-E\left[\zeta_{li}\right]\right)\right\|^2\right]&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\langle\zeta_{li},\zeta_{lj}\rangle\right]-\left\|E\left[\zeta_{l1}\right]\right\|^2\\&=\frac{1}{N^2}\sum_{i=1}^{N}E\left[\left\|\zeta_{li}\right\|^2\right]+\frac{1}{N^2}\sum_{\substack{i=1\\i\neq j}}^{N}\sum_{j=1}^{N}E\left[\langle\zeta_{li},\zeta_{lj}\rangle\right]-\left\|E\left[\zeta_{l1}\right]\right\|^2\\&=\frac{1}{N}E\left[\left\|\zeta_{l1}\right\|^2\right]+\frac{N-1}{N}E\left[\langle\zeta_{l1},\zeta_{l2}\rangle\right]-\left\|q_l\right\|^2,\end{aligned}\tag{4.A.21}$$

for $l=2,3$.

Now, we have

$$E\left[\|\zeta_{31}\|^2\right] = \frac{1}{(N-1)^2} E\left[\|\pi_3(V_1,\mathbf{U}_1)\|^2 \sum_{t=2}^{N}\sum_{s=2}^{N} W_h(V_t - V_1)\mathcal{K}_h(\mathbf{U}_t - \mathbf{U}_1) \times W_h(V_s - V_1)\mathcal{K}_h(\mathbf{U}_s - \mathbf{U}_1)\right]$$

$$= \frac{1}{(N-1)^2}\int \|\pi_3(v,\mathbf{u})\|^2 E\left[\sum_{t=2}^{N}\sum_{s=2}^{N} W_h(V_t - v)\mathcal{K}_h(\mathbf{U}_t - \mathbf{u}) \times W_h(V_s - v)\mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right] f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u},$$

where

$$E\left[\sum_{t=2}^{N}\sum_{s=2}^{N} W_h(V_t - v)\mathcal{K}_h(\mathbf{U}_t - \mathbf{u}) W_h(V_s - v)\mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right]$$
$$= \sum_{t=2}^{N} E\left[\|W_h(V_t - v)\mathcal{K}_h(\mathbf{U}_t - \mathbf{u})\|^2\right]$$
$$+ \sum_{\substack{t=2\\ t\neq s}}^{N}\sum_{s=2}^{N} E\left[W_h(V_t - v)\mathcal{K}_h(\mathbf{U}_t - \mathbf{u})\right] E\left[W_h(V_s - v)\mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right]$$
$$= (N-1)E\left[\|W_h(V_1 - v)\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\|^2\right] + (N-1)(N-2)\|E\left[W_h(V_1 - v)\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\right]\|^2.$$

It follows from Lemmas 4.B.1 and 4.B.2 that

$$E\left[\sum_{t=2}^{N}\sum_{s=2}^{N} W_h(V_t - v)\mathcal{K}_h(\mathbf{U}_t - \mathbf{u}) W_h(V_s - v)\mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right]$$
$$= (N-1)\left(C_{W\mathcal{K}} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-d} + \psi_{W\mathcal{K}}(h,(v,\mathbf{u}))\right)$$
$$+ (N-1)(N-2)\left\|f_{V\mathbf{U}}(v,\mathbf{u}) + h^P S_{W\mathcal{K}}(v,\mathbf{u}) + \beta_{W\mathcal{K}}(h,(v,\mathbf{u}))\right\|^2.$$

Hence

$$\frac{1}{N}E\left[\|\zeta_{31}\|^2\right] = \frac{1}{N}\int \|\pi_3(v,\mathbf{u})\|^2 f^3_{v\mathbf{u}}(v,\mathbf{u})\,dvd\mathbf{u} + \frac{C_{W\mathcal{K}}}{N^2h^d}\int \|\pi_3(v,\mathbf{u})\|^2 f^2_{v\mathbf{u}}(v,\mathbf{u})\,dvd\mathbf{u}$$
$$+ O\left(N^{-1}h^P + N^{-2}\right) + o\left(N^{-2}h^{-d}\right), \text{ as } N \to \infty, \tag{4.A.22}$$

by Assumption (A3), and the properties of $\psi_{W\mathcal{K}}(h,(v,\mathbf{u}))$ and $\beta_{W\mathcal{K}}(h,(v,\mathbf{u}))$. Similarly, we show that

$$\frac{1}{N}E\left[\|\zeta_{21}\|^2\right] = \frac{1}{N}\int \|\pi_2(\mathbf{u})\|^2 f_{\mathbf{u}}^3(\mathbf{u})\,d\mathbf{u} + \frac{C_{\mathcal{K}}}{N^2h^{d-1}}\int \|\pi_2(\mathbf{u})\|^2 f_{\mathbf{u}}^2(\mathbf{u})\,d\mathbf{u}$$
$$+\,O\left(N^{-1}h^P + N^{-2}\right) + o\left(N^{-2}h^{-(d-1)}\right), \text{ as } N \to \infty. \tag{4.A.23}$$

Now consider the term

$$\frac{N-1}{N}E\left[\langle\zeta_{21},\zeta_{22}\rangle\right] = \frac{1}{N(N-1)}E\Bigg[\langle\pi_2(\mathbf{U}_1),\pi_2(\mathbf{U}_2)\rangle \times\sum_{\substack{t=1\\t\neq1,s\neq2}}^{N}\sum_{s=1}^{N}K_h(\mathbf{U}_t-\mathbf{U}_1)K_h(\mathbf{U}_s-\mathbf{U}_2)\Bigg]$$
$$= \frac{1}{N(N-1)}\sum_{\substack{t=1\\t\neq1,s\neq2}}^{N}\sum_{s=1}^{N}\Delta_{2,ts},$$

where $\Delta_{2,ts} = E\left[\langle\pi_2(\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_1),\pi_2(\mathbf{U}_2)\mathcal{K}_h(\mathbf{U}_s-\mathbf{U}_2)\rangle\right]$. Similarly, $N^{-1}(N-1)\times E\left[\langle\zeta_{31},\zeta_{32}\rangle\right] = N^{-1}(N-1)^{-1}\sum_{\substack{t=1\\t\neq1,s\neq2}}^{N}\sum_{s=1}^{N}\Delta_{3,ts}$, with

$$\Delta_{3,ts} = E[\langle\pi_3(V_1,\mathbf{U}_1)W_h(V_t-V_1)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_1),\pi_3(V_2,\mathbf{U}_2)W_h(V_s-V_2)\mathcal{K}_h(\mathbf{U}_s-\mathbf{U}_2)\rangle].$$

Furthermore, for $l = 2,3$, we write

$$\Delta_{l,ts} = \begin{cases} \mathcal{B}_{l,I} & ; \quad s=t, \\ \mathcal{B}_{l,II} & ; \quad s\neq t,\ t\neq 2,\ s\neq 1, \\ \mathcal{B}_{l,III} & ; \quad s=t\ \&\ t=2,\ s\neq 1 \text{ or } t\neq 2,\ s=1, \\ \mathcal{B}_{l,IV} & ; \quad s\neq t,\ t=2,\ s=1. \end{cases}$$

Here we make the following definitions:

$$\begin{aligned}
\mathcal{B}_{2,I} &= E\left[\|E[\pi_2(\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_1)|\,\mathbf{U}_3]\|^2\right] &;\\
\mathcal{B}_{2,II} &= \|E[\pi_2(\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_1)]\|^2 \equiv \|q_2\|^2 &;\\
\mathcal{B}_{2,III} &= E\left[\langle\pi_2(\mathbf{U}_1),\pi_2(\mathbf{U}_2)\rangle\mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_2)\right] &;\\
\mathcal{B}_{2,IV} &= E\left[\langle\pi_2(\mathbf{U}_1),\pi_2(\mathbf{U}_2)\rangle\|\mathcal{K}_h(\mathbf{U}_1-\mathbf{U}_2)\|^2\right] &;
\end{aligned}$$

$$\begin{aligned}
\mathcal{B}_{3,I} &= E\left[\left\|E\left[\pi_3\left(V_1,\mathbf{U}_1\right)W_h\left(V_3-V_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\right|V_3,\mathbf{U}_3\right]\right\|^2\Big] & ;\\
\mathcal{B}_{3,II} &= \left\|E\left[\pi_3\left(V_1,\mathbf{U}_1\right)W_h\left(V_3-V_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\right]\right\|^2 \equiv \left\|q_3\right\|^2 & ;\\
\mathcal{B}_{3,III} &= E\left[\left\langle\pi_3\left(V_1,\mathbf{U}_1\right),\pi_3\left(V_2,\mathbf{U}_2\right)\right\rangle W_h\left(V_2-V_1\right)\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)\times\right.\\
&\qquad \left.W_h\left(V_3-V_2\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_2\right)\right] & ;\\
\mathcal{B}_{3,IV} &= E\left[\left\langle\pi_3\left(V_1,\mathbf{U}_1\right),\pi_3\left(V_2,\mathbf{U}_2\right)\right\rangle\left\|W_h\left(V_1-V_2\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{U}_2\right)\right\|^2\right], \text{ and}
\end{aligned}$$

Therefore, we are able to write

$$\begin{aligned}
\frac{N-1}{N}E\left[\left\langle\zeta_{21},\zeta_{22}\right\rangle\right] &= \frac{1}{N\left(N-1\right)}\left\{\left(N-2\right)\mathcal{B}_{2,I}+\left(N^2-5N+6\right)\left\|q_2\right\|^2\right.\\
&\qquad\left.+2\left(N-2\right)\mathcal{B}_{2,III}+\mathcal{B}_{2,IV}\right\}, \qquad (4.A.24)\\
\frac{N-1}{N}E\left[\left\langle\zeta_{31},\zeta_{32}\right\rangle\right] &= \frac{1}{N\left(N-1\right)}\left\{\left(N-2\right)\mathcal{B}_{3,I}+\left(N^2-5N+6\right)\left\|q_3\right\|^2\right.\\
&\qquad\left.+2\left(N-2\right)\mathcal{B}_{3,III}+\mathcal{B}_{3,IV}\right\}. \qquad (4.A.25)
\end{aligned}$$

We now show the working of (4.A.25), because (4.A.24)'s is the same. As $N\to\infty$, it follows that

$$\begin{aligned}
\frac{N-1}{N}E\left[\left\langle\zeta_{31},\zeta_{32}\right\rangle\right] &= \frac{1}{N}\left[\mathcal{B}_{3,I}+2\mathcal{B}_{3,III}\right]+\left\|q_3\right\|^2\left[1-\frac{5}{N}\right]+\frac{1}{N^2}\mathcal{B}_{3,IV}\\
&\quad+\frac{1}{N^2}\left[6\left\|q_3\right\|^2-2\mathcal{B}_{3,I}-4\mathcal{B}_{3,III}\right]+o\left(N^{-2}\right)\\
&= \frac{3}{N}\int\left\|\pi_3\left(v,\mathbf{u}\right)\right\|^2 f_{\mathbf{U}}^3\left(v,\mathbf{u}\right)dvd\mathbf{u}+\left\|q_3\right\|^2\left[1-\frac{5}{N}\right]\\
&\quad+\frac{C_{W\mathcal{K}}}{N^2h^d}\int\left\|\pi_3\left(v,\mathbf{u}\right)\right\|^2 f_{V\mathbf{U}}^2\left(v,\mathbf{u}\right)dvd\mathbf{u}\\
&\quad+O\left(N^{-2}\right)+o\left(N^{-2}h^{-d}+N^{-1}h^P\right), \qquad (4.A.26)
\end{aligned}$$

where the last equality follows from Lemmas 4.B.1 and 4.B.6. We now put together (4.A.22), (4.A.26) in (4.A.21) and obtain,

$$\begin{aligned}
E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{3i}-E\left[\zeta_{3i}\right]\right)\right\|^2\right] &= \frac{1}{N}\left[4\int\left\|\pi_3\left(v,\mathbf{u}\right)\right\|^2 f_{V\mathbf{U}}^3\left(v,\mathbf{u}\right)dvd\mathbf{u}-5\left\|\delta_3\right\|^2\right]\\
&\quad+\frac{2C_{W\mathcal{K}}}{N^2h^d}\int\left\|\pi_3\left(v,\mathbf{u}\right)\right\|^2 f_{V\mathbf{U}}^2\left(v,\mathbf{u}\right)dvd\mathbf{u}\\
&\quad+O\left(N^{-2}\right)+o\left(N^{-2}h^{-d}+N^{-1}h^P\right), \text{ as } N\to\infty, \qquad (4.A.27)
\end{aligned}$$

similarly

$$E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{2i}-E\left[\zeta_{2i}\right]\right)\right\|^{2}\right]=\frac{1}{N}\left[4\int\left\|\pi_{2}\left(\mathbf{u}\right)\right\|^{2}f_{\mathbf{U}}^{3}\left(\mathbf{u}\right)d\mathbf{u}-5\left\|\delta_{2}\right\|^{2}\right]$$
$$+\frac{2C_{\mathcal{K}}}{N^{2}h^{d-1}}\int\left\|\pi_{2}\left(\mathbf{u}\right)\right\|^{2}f_{\mathbf{U}}^{2}\left(\mathbf{u}\right)d\mathbf{u}$$
$$+O\left(N^{-2}\right)+o\left(N^{-2}h^{-(d-1)}+N^{-1}h^{P}\right),\text{ as }N\rightarrow\infty,$$

Also, it follows from equations (4.B.4) and (4.B.3) in Lemma 4.B.1, that

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}E\left[\zeta_{3i}\right]-\delta_{3}\right\|^{2}&=\left\|q_{3}-\delta_{3}\right\|^{2}\\&=\left\|h^{P}\int\pi_{3}\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}+\gamma_{WK}\left(h\right)\right\|^{2}\\&=h^{2P}\left\|\int\pi_{3}\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\|^{2}\\&+o\left(h^{2P}\right),\text{ as }N\rightarrow\infty,\end{aligned}\tag{4.A.28}$$

and similarly $N^{-1}\left\|\sum_{i=1}^{N}E\left[\zeta_{2i}\right]-\delta_{2}\right\|^{2}=h^{2P}\left\|\int\pi_{2}\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)d\mathbf{u}\right\|^{2}+o\left(h^{2P}\right)$. Finally, it follows from (4.A.20), (4.A.27) and (4.A.28),

$$\begin{aligned}E\left[\left\|\widehat{\delta}_{3}\left(h\right)-\delta_{3}\right\|^{2}\right]&=h^{2P}\left\|\int\pi_{3}\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\|^{2}+O\left(\frac{1}{N}\right)+O\left(\frac{1}{N^{2}h^{d}}\right)\\&+O\left(\frac{h^{P}}{N}+\frac{1}{N^{2}}\right)+o\left(\frac{h^{P}}{N}+\frac{1}{N^{2}}+\frac{1}{N^{2}h^{d}}+h^{2P}\right),\text{ as }N\rightarrow\infty.\end{aligned}\tag{4.A.29}$$

Similarly

$$\begin{aligned}E\left[\left\|\widehat{\delta}_{2}\left(h\right)-\delta_{2}\right\|^{2}\right]&=h^{2P}\left\|\int\pi_{2}\left(\mathbf{u}\right)S_{K}\left(\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)d\mathbf{u}\right\|^{2}+O\left(\frac{1}{N}\right)+O\left(\frac{1}{N^{2}h^{d-1}}\right)\\&+O\left(\frac{h^{P}}{N}+\frac{1}{N^{2}}\right)+o\left(\frac{h^{P}}{N}+\frac{1}{N^{2}}+\frac{1}{N^{2}h^{d-1}}+h^{2P}\right),\text{ as }N\rightarrow\infty.\end{aligned}\tag{4.A.30}$$

Terms: $E\left[\|\delta_4(h)-\delta_4\|^2\right]$ **and** $E\left[\|\delta_5(h)-\delta_5\|^2\right]$

We have

$$E\left[\left\|\widehat{\delta}_4(h)-\delta_4\right\|^2\right]=E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{4i}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]+E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}(\zeta_{4i}-E[\zeta_{4i}])\right\|^2\right]$$
$$+\left\|\frac{1}{N}\sum_{i=1}^{N}E[\zeta_{4i}]-\delta_4\right\|^2,\text{ and}$$
$$E\left[\left\|\widehat{\delta}_5(h)-\delta_5\right\|^2\right]=E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{5i}\widehat{f}^2_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]+E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}(\zeta_{5i}-E[\zeta_{5i}])\right\|^2\right]$$
$$+\left\|\frac{1}{N}\sum_{i=1}^{N}E[\zeta_{5i}]-\delta_5\right\|^2. \tag{4.A.31}$$

We only show the working for $E[\|\widehat{\delta}_5(h)-\delta_5\|^2]$, the dominant term. We also work with the following expression

$$E\left[\left\|\sum_{t=2}^{N}\sum_{s=2}^{N}W_h(V_t-v)\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})W_h(V_s-v)\mathcal{K}_h(\mathbf{U}_s-\mathbf{u})\right\|^2\right]$$
$$=(N-1)(N-2)(N-3)(N-4)\|E[W_h(V_2-v)\mathcal{K}_h(\mathbf{U}_2-\mathbf{u})]\|^4$$
$$+6(N-1)(N-2)(N-3)E\left[\|W_h(V_2-v)\mathcal{K}_h(\mathbf{U}_2-\mathbf{u})\|^2\right]\|E[W_h(V_3-v)\mathcal{K}_h(\mathbf{U}_3-\mathbf{u})]\|^2$$
$$+3(N-1)(N-2)E\left[\|W_h(V_2-v)\mathcal{K}_h(\mathbf{U}_2-\mathbf{u})\|^2\right]E\left[\|W_h(V_3-v)\mathcal{K}_h(\mathbf{U}_3-\mathbf{u})\|^2\right]$$
$$+4(N-1)(N-2)E\left[W_h^3(V_2-v)\mathcal{K}_h^3(\mathbf{U}_2-\mathbf{u})\right]E[W_h(V_3-v)\mathcal{K}_h(\mathbf{U}_3-\mathbf{u})]$$
$$+(N-1)E\left[\|W_h(V_2-v)\mathcal{K}_h(\mathbf{U}_2-\mathbf{u})\|^4\right].$$

Firstly, Lemmas 4.B.1 and 4.B.2, imply that

$$E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{5i}\widehat{f}^2_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\|^2\right]$$
$$=\frac{1}{N(N-1)^4}\int\sigma_5^2(v,\mathbf{u})E\left[\left\|\sum_{t=2}^{N}W_h(V_t-v)\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right\|^4\right]f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} \tag{4.A.32}$$
$$=O\left(N^{-1}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-3}h^{-2d}\right)+O\left(N^{-4}h^{-3d}\right)$$
$$+O\left(N^{-1}h^{4P}\right)+o\left(N^{-1}+N^{-2}h^{-d}\right).$$

We now analyze the next term,

$$E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{5i}-E\left[\zeta_{5i}\right]\right)\right\|^{2}\right]=\frac{1}{N}E\left[\left\|\zeta_{51}\right\|^{2}\right]+\frac{N-1}{N}E\left[\left\langle\zeta_{51},\zeta_{52}\right\rangle\right]-\left\|E\left[\zeta_{51}\right]\right\|^{2}.$$

The first term in the last equation is like (4.A.32), after replacing $\sigma_5^2(v,\mathbf{u})$ by $\|\pi_5(v,\mathbf{u})\|^2$, and therefore, it is of the same order of magnitude. Now

$$\frac{N-1}{N}E\left[\left\langle\zeta_{51},\zeta_{52}\right\rangle\right]$$

$$=\frac{1}{N(N-1)^3}E\left[\left\langle\pi_5(V_1,\mathbf{U}_1)\sum_{\substack{t=1\\t\neq 1,s\neq 1}}^{N}\sum_{s=1}^{N}W_h(V_t-V_1)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_1)W_h(V_s-V_1)\mathcal{K}_h(\mathbf{U}_s-\mathbf{U}_1)\right.\right.$$

$$\left.\left.,\pi_5(V_2,\mathbf{U}_2)\sum_{\substack{t=1\\t\neq 2,s\neq 2}}^{N}\sum_{s=1}^{N}W_h(V_t-V_2)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_2)W_h(V_s-V_2)\mathcal{K}_h(\mathbf{U}_s-\mathbf{U}_2)\right\rangle\right].$$

Let us introduce the following notation here

$$\begin{aligned}\pi_{5;t}&\equiv\pi_5(V_t,\mathbf{U}_t)\\W_{h;t1}\mathcal{K}_{h;t1}&\equiv W_h(V_t-V_1)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_1),\text{ and}\\W_{h;t2}\mathcal{K}_{h;t2}&\equiv W_h(V_t-V_2)\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_2),\end{aligned}$$

then it follows

$$\frac{N-1}{N}E\left[\left\langle\zeta_{51},\zeta_{52}\right\rangle\right]=\frac{1}{N(N-1)^3}\left[\mathcal{B}_{5,I}+\mathcal{B}_{5,II}+\mathcal{B}_{5,III}+\mathcal{B}_{5,IV}+\mathcal{B}_{5,V}+\mathcal{B}_{5,VI}\right]+\mathcal{B}_{5,IV}$$

where

$$
\begin{aligned}
\mathcal{B}_{5,I} &= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^{4}\mathcal{K}_{h;12}^{4}\right], \\
\mathcal{B}_{5,II} &= (N-2)\,E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^{3}\mathcal{K}_{h;12}^{3} W_{h;32}\mathcal{K}_{h;32}\right], \\
\mathcal{B}_{5,III} &= (N-2)\,E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^{2}\mathcal{K}_{h;12}^{2} W_{h;31}\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32}\right], \\
\mathcal{B}_{5,IV} &= (N-2)\,(N-3)\,E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^{2}\mathcal{K}_{h;12}^{2} W_{h;41}\mathcal{K}_{h;41} W_{h;32}\mathcal{K}_{h;32}\right], \\
\mathcal{B}_{5,V} &= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}\times\right. \\
&\quad \left\{(N-2)\,E\left[W_{h;31}\mathcal{K}_{h;31}W_{h;32}^{2}\mathcal{K}_{h;32}^{2}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)\right] + \\
&\quad + 2\,(N-2)\,(N-3)\,E\left[W_{h;31}\mathcal{K}_{h;31}W_{h;32}\mathcal{K}_{h;32}W_{h;42}\mathcal{K}_{h;42}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)] \\
&\quad + (N-2)\,(N-3)\,E\left[W_{h;41}\mathcal{K}_{h;41}W_{h;32}^{2}\mathcal{K}_{h;32}^{2}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)] \\
&\quad \left. + (N-2)\,(N-3)\,(N-4)\,E\left[W_{h;31}\mathcal{K}_{h;31}W_{h;42}\mathcal{K}_{h;42}W_{h;52}\mathcal{K}_{h;52}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)]\}\right], \\
\mathcal{B}_{5,VI} &= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^{2}\mathcal{K}_{h;12}^{2}\times\right. \\
&\quad \left\{(N-2)\,E\left[W_{h;32}^{2}\mathcal{K}_{h;32}^{2}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)\right] \\
&\quad \left. + (N-2)\,(N-3)\,E\left[W_{h;32}\mathcal{K}_{h;32}W_{h;42}\mathcal{K}_{h;42}\right|(\mathbf{V}_1,\mathbf{U}_1),(\mathbf{V}_2,\mathbf{U}_2)]\}\right], \text{ and} \\
\mathcal{B}_{5,VII} &= \frac{1}{N\,(N-1)^3}E\left[\left\langle \pi_{5;1}\left\|\textstyle\sum_{t=3}^{N}W_{h;t1}\mathcal{K}_{h;t1}\right\|^2, \pi_{5;2}\left\|\textstyle\sum_{s=3}^{N}W_{h;s2}\mathcal{K}_{h;s2}\right\|^2\right\rangle\right].
\end{aligned}
$$

Then by Lemmas 4.B.7, and 4.B.8, this term is simply

$$
\begin{aligned}
\frac{N-1}{N}E\left[\langle \zeta_{51}, \zeta_{52}\rangle\right] &= \|E[\zeta_{51}]\|^2 + O\left(N^{-1}\right) + O\left(N^{-2}h^{-d}\right) + O\left(N^{-3}h^{-2d}\right) \\
&\quad + O\left(N^{-4}h^{-3d}\right) + O\left(N^{-1}h^{P}\right),
\end{aligned}
$$

and conclude that

$$
\begin{aligned}
E\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{5i} - E[\zeta_{5i}]\right)\right\|^2\right] &= O\left(N^{-1}\right) + O\left(N^{-2}\right) + O\left(N^{-2}h^{-d}\right) \\
&\quad + O\left(N^{-1}h^{P}\right) + o\left(N^{-1} + N^{-2}h^{-2d}\right).
\end{aligned}
$$

We now turn our attention to (4.A.31). Firstly,

$$
\begin{aligned}
E[\zeta_{51}] &= \frac{1}{N-1}E\left[\pi_5\left(V_1, \mathbf{U}_1\right)W_h^2\left(V_2 - V_1\right)\mathcal{K}_h^2\left(\mathbf{V}_2 - \mathbf{V}_1\right)\right] \\
&\quad + E\left[\pi_5\left(V_1, \mathbf{U}_1\right)W_h\left(V_2 - V_1\right)\mathcal{K}_h\left(\mathbf{V}_2 - \mathbf{V}_1\right)W_h\left(V_3 - V_1\right)\mathcal{K}_h\left(\mathbf{V}_3 - \mathbf{V}_1\right)\right] \\
&= \frac{C_{W\mathcal{K}}}{Nh^d}\int \pi_5(v,\mathbf{u})\, f_{V\mathbf{U}}^2(v,\mathbf{u})\, dv d\mathbf{u} + 2h^P\left[\int \pi_3(v,\mathbf{u})\, S_{W\mathcal{K}}(v,\mathbf{u})\, f_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u}\right] \\
&\quad + \delta_5 + o\left(N^{-1}h^{-d}\right) + o\left(h^P\right), \text{ and conclude}
\end{aligned}
$$

$$\left\|E\left[\zeta_{51}\right]-\delta_5\right\|^2=\left\|\frac{C_{W\mathcal{K}}}{Nh^d}\int\pi_3\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right.$$
$$\left.+2h^P\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\|^2,$$

because of Lemmas 4.B.5, and 4.B.1, and noticing that

$$\begin{aligned}\pi_5\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)&=E\left[\left.\frac{\omega f_{\mathbf{U}}\left(\mathbf{U}\right)}{f_{V\mathbf{U}}^3\left(V,\mathbf{U}\right)}\right|V=v,\mathbf{U}=\mathbf{u}\right]f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\\&=E\left[\left.\frac{\omega f_{\mathbf{U}}\left(\mathbf{U}\right)}{f_{V\mathbf{U}}^2\left(V,\mathbf{U}\right)}\right|V=v,\mathbf{U}=\mathbf{u}\right]\\&\equiv\pi_3\left(v,\mathbf{u}\right).\end{aligned}$$

In conclusion, we have that

$$E\left[\left\|\widehat{\delta}_5\left(h\right)-\delta_5\right\|^2\right]=\left\|\frac{C_{W\mathcal{K}}}{Nh^d}\int\pi_3\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right.$$
$$\left.+2h^P\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\|^2$$
$$+O\left(N^{-1}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-1}h^{2P}+N^{-2}\right).$$

By the exact same argument, we could also show that

$$E\left[\left\|\widehat{\delta}_4\left(h\right)-\delta_4\right\|^2\right]=h^{2P}\left\|\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)d\mathbf{u}+\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\|^2$$
$$+O\left(N^{-2}h^{2(d-1)}\right)+O\left(N^{-1}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-1}h^{2P}+N^{-2}\right),$$

where the first term of the last equation follows from Lemma 4.B.1.

Terms: $E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_2\left(h\right)-\delta_2\right\rangle\right]$ **and** $E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_3\left(h\right)-\delta_3\right\rangle\right]$

As it was previously done, we have

$$E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_2\left(h\right)-\delta_2\right\rangle\right]=E\left[\left\langle\frac{1}{N}\sum_{i=1}^N\varepsilon_{1i},\frac{1}{N}\sum_{i=1}^N\varepsilon_{2i}\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i\right)\right\rangle\right]$$
$$+E\left[\left\langle\frac{1}{N}\sum_{i=1}^N\left(\pi_{1i}-E\left[\pi_{1i}\right]\right),\frac{1}{N}\sum_{i=1}^N\left(\zeta_{2i}-E\left[\zeta_{2i}\right]\right)\right\rangle\right],$$

$$E\left[\left\langle\widehat{\delta}_1-\delta_1,\widehat{\delta}_3(h)-\delta_3\right\rangle\right]=E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\widetilde{\varepsilon}_{1i},\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right\rangle\right] \tag{4.A.33}$$

$$+E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\pi}_{1i}-E\left[\widetilde{\pi}_{1i}\right]\right),\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{3i}-E\left[\zeta_{3i}\right]\right)\right\rangle\right]. \tag{4.A.34}$$

We show the working for $E\langle\,\widehat{\delta}_1-\delta_1,\widehat{\delta}_3(h)-\delta_3\rangle$ only. Firstly, Lemma 4.B.1 implies that (4.A.33) equals

$$\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N}E\left[\widetilde{\sigma}_{13}(v_i,\mathbf{u}_i)\,\widehat{f}_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right]\\
&=\frac{1}{N(N-1)}\int\widetilde{\sigma}_{13}(v,\mathbf{u})\,E\left[\sum_{t=2}^{N}W_h(V_t-v)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right]f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\\
&=\frac{1}{N}\int\widetilde{\sigma}_{13}(v,\mathbf{u})\,f_{V\mathbf{U}}^2(v,\mathbf{u})\,dvd\mathbf{u}+O\left(h^P/N\right)+o\left(h^P\right),\text{ as }N\to\infty,
\end{aligned}$$

where $E\left[\widetilde{\varepsilon}_{1i}^{\mathsf{T}}\varepsilon_{3j}\,|\,(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)\right]=1(i=j)\,\widetilde{\sigma}_{13}(V_i,\mathbf{U}_i)$, from Assumption (A6). Also

$$\begin{aligned}
E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\pi}_{1i}-E\left[\widetilde{\pi}_{1i}\right]\right),\frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{3i}-E\left[\zeta_{3i}\right]\right)\right\rangle\right]&=\frac{1}{N}E\left[\langle\widetilde{\pi}_{11},\zeta_{31}\rangle\right]\\
&+\frac{N-1}{N}E\left[\langle\widetilde{\pi}_{11},\zeta_{32}\rangle\right]-\langle E\left[\widetilde{\pi}_{11}\right],E\left[\zeta_{31}\right]\rangle.
\end{aligned}$$

The first term equals, $N^{-1}E\left[\langle\widetilde{\pi}_{11},\zeta_{31}\rangle\right]$,

$$\begin{aligned}
&=\frac{1}{N(N-1)}\int\langle\widetilde{\pi}_1(v,\mathbf{u}),\pi_3(v,\mathbf{u})\rangle\,E\left[\sum_{t=2}^{N}W_h(V_t-v)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right]f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\\
&=\frac{1}{N}\int\langle\widetilde{\pi}_1(v,\mathbf{u}),\pi_3(v,\mathbf{u})\rangle\,f_{V\mathbf{U}}^2(v,\mathbf{u})\,dvd\mathbf{u}+O\left(h^P/N\right)+o\left(h^P\right),\text{ as }N\to\infty,
\end{aligned}$$

from Lemma 4.B.1. The second term, $N^{-1}(N-1)\,E\left[\langle\pi_{11},\zeta_{32}\rangle\right]$, equals

$$\begin{aligned}
&\frac{1}{N}E\left[\langle\widetilde{\pi}_1(V_1,\mathbf{U}_1),\pi_3(V_2,\mathbf{U}_2)\rangle\sum_{\substack{t=1\\t\neq2}}^{N}W_h(V_t-V_2)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_2)\right]\\
&=\frac{1}{N}\mathcal{B}_{13,I}+\left(\frac{N-2}{N}\right)\mathcal{B}_{13,II},
\end{aligned}$$

where $\mathcal{B}_{13,I}=E\left[\langle\widetilde{\pi}_1(V_1,\mathbf{U}_1),\pi_3(V_2,\mathbf{U}_2)\rangle\,W_h(V_1-V_2)\,\mathcal{K}_h(\mathbf{U}_1-\mathbf{U}_2)\right]$ and $\mathcal{B}_{13,II}=E[\langle\pi_1$

$(V_1, \mathbf{U}_1), \pi_3(V_2, \mathbf{U}_2)\rangle W_h(V_3 - V_2)\mathcal{K}_h(\mathbf{U}_3 - \mathbf{U}_2)]$. It follows from Lemma 4.B.9 that

$$\begin{aligned}\frac{N-1}{N}E[\langle \pi_{11}, \zeta_{32}\rangle] = \frac{1}{N}\int \langle \widetilde{\pi}_1(v, \mathbf{u}), \pi_3(v, \mathbf{u})\rangle f^2_{V\mathbf{U}}(v, \mathbf{u})\, dv d\mathbf{u} + \left(1 - \frac{2}{N}\right)\langle \delta_1, \delta_3\rangle \\ + h^P \int \langle \delta_1, \pi_3(v, \mathbf{u})\rangle S_{W\mathcal{K}}(v, \mathbf{u})\, dv d\mathbf{u} \\ + o\left(\frac{h^P}{N} + \frac{1}{N}\right).\end{aligned}$$

Finally $\langle E[\widetilde{\pi}_{11}], E[\zeta_{31}]\rangle = \langle \delta_1, \delta_3\rangle + h^P \int \langle \delta_1, \pi_3(v, \mathbf{u})\rangle S_{W\mathcal{K}}(v, \mathbf{u})\, dv d\mathbf{u} + o(h^P)$. Therefore, we conclude

$$E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_3(h) - \delta_3\right\rangle\right] = O\left(\frac{1}{N}\right) + O\left(\frac{h^P}{N}\right) + o\left(\frac{h^P}{N} + \frac{1}{N}\right), \text{ and similarly,} \tag{4.A.35}$$

$$E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_2(h) - \delta_2\right\rangle\right] = O\left(\frac{1}{N}\right) + O\left(\frac{h^P}{N}\right) + o\left(\frac{h^P}{N} + \frac{1}{N}\right). \tag{4.A.36}$$

Terms: $E\left[\left\langle \widehat{\delta}_1(h) - \delta_1, \widehat{\delta}_4(h) - \delta_4\right\rangle\right]$ **and** $E\left[\left\langle \widehat{\delta}_1(h) - \delta_1, \widehat{\delta}_5(h) - \delta_5\right\rangle\right]$

As it was previously done, we have

$$\begin{aligned}E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_4(h) - \delta_4\right\rangle\right] &= E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{1i}, \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{4i}\widehat{f}_{V\mathbf{U}}(v_i, \mathbf{u}_i)\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)\right\rangle\right] \\ &+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}(\pi_{1i} - E[\pi_{1i}]), \frac{1}{N}\sum_{i=1}^{N}(\zeta_{4i} - E[\zeta_{4i}])\right\rangle\right],\end{aligned}$$

$$E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_5(h) - \delta_5\right\rangle\right] = E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\widetilde{\varepsilon}_{1i}, \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{5i}\widehat{f}^2_{V\mathbf{U}}(v_i, \mathbf{u}_i)\right\rangle\right] \tag{4.A.37}$$

$$+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\pi}_{1i} - E[\widetilde{\pi}_{1i}]), \frac{1}{N}\sum_{i=1}^{N}(\zeta_{5i} - E[\zeta_{5i}])\right\rangle\right]. \tag{4.A.38}$$

We show the working for $E\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_5(h) - \delta_5\rangle$ only. Firstly, Lemma 4.B.2 implies that (4.A.37) equals

$$\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N} E\left[\widetilde{\sigma}_{15}(v_i,\mathbf{u}_i)\,\widehat{f}^2_{V\mathbf{U}}(v_i,\mathbf{u}_i)\right] \\
&= \frac{1}{N(N-1)^2}\int \widetilde{\sigma}_{15}(v,\mathbf{u})\,E\left[\left\|\sum_{t=2}^{N} W_h(V_t-v)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right\|^2\right] f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} \\
&= \frac{C_{W\mathcal{K}}}{N^2h^d}\int \widetilde{\sigma}_{15}(v,\mathbf{u})\,f^2_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} + \frac{1}{N}\int \widetilde{\sigma}^2_{15}(v,\mathbf{u})\,f^3_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} \\
&+ O\left(h^P/N\right) + o\left(h^P\right) + o\left(N^{-2}h^{-d}\right), \text{ as } N\to\infty,
\end{aligned}$$

where $E\left[\widetilde{\varepsilon}^{\top}_{1i}\varepsilon_{5j}\middle|\,(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)\right] = 1(i=j)\,\widetilde{\sigma}_{15}(V_i,\mathbf{U}_i)$, from Assumption (A6). Also

$$\begin{aligned}
E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\pi}_{1i} - E[\widetilde{\pi}_{1i}]),\, \frac{1}{N}\sum_{i=1}^{N}(\zeta_{5i} - E[\zeta_{5i}])\right\rangle\right] &= \frac{1}{N}E[\langle\widetilde{\pi}_{11},\zeta_{51}\rangle] \\
&+ \frac{N-1}{N}E[\langle\widetilde{\pi}_{11},\zeta_{52}\rangle] - \langle E[\widetilde{\pi}_{11}], E[\zeta_{51}]\rangle.
\end{aligned}$$

The first term of the last equality, $N^{-1}E[\langle\widetilde{\pi}_{11},\zeta_{51}\rangle]$, is

$$\begin{aligned}
&= \frac{1}{N(N-1)^2}\int \langle\widetilde{\pi}_1(v,\mathbf{u}),\pi_5(v,\mathbf{u})\rangle\,E\left[\left\|\sum_{t=2}^{N} W_h(V_t-v)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{u})\right\|^2\right] f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} \\
&= \frac{C_{W\mathcal{K}}}{N^2h^d}\int \langle\widetilde{\pi}_1(v,\mathbf{u}),\pi_5(v,\mathbf{u})\rangle\,f^2_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} + \frac{1}{N}\int \langle\widetilde{\pi}_1(v,\mathbf{u}),\pi_5(v,\mathbf{u})\rangle\,f^3_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u} \\
&+ O\left(h^P/N\right) + o\left(h^P\right) + o\left(N^{-2}h^{-d}\right), \text{ as } N\to\infty, \text{ from Lemma 4.B.2}
\end{aligned}$$

The second term, $N^{-1}(N-1)\,E[\langle\widetilde{\pi}_{11},\zeta_{52}\rangle]$, equals

$$\begin{aligned}
&\frac{1}{N(N-1)}E\left[\langle\widetilde{\pi}_1(V_1,\mathbf{U}_1),\pi_5(V_2,\mathbf{U}_2)\rangle\left\|\sum_{\substack{t=1\\t\neq 2}}^{N} W_h(V_t-V_2)\,\mathcal{K}_h(\mathbf{U}_t-\mathbf{U}_2)\right\|^2\right] \\
&= \frac{1}{N(N-1)}\left[(N-2)\,\mathcal{B}_{15,I} + (N-2)(N-3)\,\mathcal{B}_{15,II} + 2(N-2)\,\mathcal{B}_{15,III} + \mathcal{B}_{15,IV}\right] \\
&= \frac{1}{N}\left[\mathcal{B}_{15,I} + 2\mathcal{B}_{15,III}\right] + \mathcal{B}_{15,II}\left[1-\frac{5}{N}\right] + \frac{1}{N^2}\mathcal{B}_{15,IV} \\
&+ \frac{1}{N^2}\left[6\mathcal{B}_{15,II} - 2\mathcal{B}_{15,I} - 4\mathcal{B}_{15,III}\right] + o\left(N^{-2}\right), \text{ where}
\end{aligned}$$

$$\mathcal{B}_{15,I} = E\left[\left\langle \widetilde{\pi}_1(V_1,\mathbf{U}_1), \pi_5(V_2,\mathbf{U}_2) W_h^2(V_3-V_2)\mathcal{K}_h^2(\mathbf{U}_3-\mathbf{U}_2)\right\rangle\right]$$
$$\mathcal{B}_{15,II} = E\left[\left\langle \widetilde{\pi}_1(V_1,\mathbf{U}_1), \pi_5(V_2,\mathbf{U}_2) W_h(V_4-V_2)\mathcal{K}_h(\mathbf{U}_4-\mathbf{U}_2) W_h(V_3-V_2)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_2)\right\rangle\right]$$
$$\mathcal{B}_{15,III} = E\left[\left\langle \widetilde{\pi}_1(V_1,\mathbf{U}_1) W_h(V_1-V_2)\mathcal{K}_h(\mathbf{U}_1-\mathbf{U}_2), \pi_5(V_2,\mathbf{U}_2) W_h(V_3-V_2)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_2)\right\rangle\right]$$
$$\mathcal{B}_{15,IV} = E\left[\left\langle \widetilde{\pi}_1(V_1,\mathbf{U}_1), \pi_5(V_2,\mathbf{U}_2) W_h^2(V_1-V_2)\mathcal{K}_h^2(\mathbf{U}_1-\mathbf{U}_2)\right\rangle\right]$$

Finally, $\langle E[\widetilde{\pi}_{11}], E[\zeta_{51}]\rangle = N^{-1}\mathcal{B}_{15,I} + \mathcal{B}_{15,II}$, and conclude from Lemma 4.B.10 that

$$E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_5(h) - \delta_5\right\rangle\right] = O\left(N^{-1}\right) + O\left(N^{-2}h^{-d}\right)$$
$$+O\left(\frac{h^P}{N} + \frac{1}{N^2}\right) + o\left(\frac{1}{N^2h^d} + \frac{h^P}{N} + \frac{1}{N^2} + h^{2P}\right), \text{ as } N \to \infty$$

Similarly, we infer from this result that

$$E\left[\left\langle \widehat{\delta}_1 - \delta_1, \widehat{\delta}_4(h) - \delta_4\right\rangle\right] = O\left(N^{-1}\right) + O\left(N^{-2}h^{-(d-1)}\right)$$
$$+O\left(\frac{h^P}{N} + \frac{1}{N^2}\right) + o\left(\frac{1}{N^2h^{(d-1)}} + \frac{h^P}{N} + \frac{1}{N^2} + h^{2P}\right), \text{ as } N \to \infty$$

and therefore of smaller order.

Terms: $E\left[\left\langle \widehat{\delta}_2(h) - \delta_2, \widehat{\delta}_3(h) - \delta_3\right\rangle\right]$

We have

$$E\left[\left\langle \widehat{\delta}_2(h) - \delta_2, \widehat{\delta}_3(h) - \delta_3\right\rangle\right] = E\left[\left\langle \frac{1}{N}\sum_{i=1}^N \widetilde{\varepsilon}_{2i}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i), \frac{1}{N}\sum_{j=1}^N \varepsilon_{3j}\widehat{f}_{V\mathbf{U}}(v_j,\mathbf{u}_j)\right\rangle\right] \quad (4.A.39)$$
$$+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^N\left(\widetilde{\zeta}_{2i} - E[\widetilde{\zeta}_{21}]\right), \frac{1}{N}\sum_{j=1}^N\left(\zeta_{3j} - E[\zeta_{31}]\right)\right\rangle\right] \quad (4.A.40)$$
$$+\left\langle E[\widetilde{\zeta}_{21}] - \delta_2, E[\zeta_{31}] - \delta_3\right\rangle, \quad (4.A.41)$$

where $\widetilde{\zeta}_{2i} \equiv \widetilde{\pi}_2(v_i,\mathbf{u}_i)\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)$, and by construction $E[\widetilde{\varepsilon}_{2i}|(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)] = 0$. As before

$$E\left[\widetilde{\varepsilon}_{2i}^{\top}\varepsilon_{3j}\middle|(V_1,\mathbf{U}_1),\ldots,(V_N,\mathbf{U}_N)\right] = \begin{cases} \widetilde{\sigma}_{23}(V_i,\mathbf{U}_i), & i=j, \\ 0, & i\neq j. \end{cases}$$

By Assumption (A6), the term (4.A.39) is equal to

$$\frac{1}{N(N-1)^2}\int \widetilde{\sigma}_{23}(v,\mathbf{u})\, E\left\langle \sum_{t=2}^N W_h(V_t-v)\mathcal{K}_h(\mathbf{U}_t-\mathbf{u}), \sum_{s=2}^N \mathcal{K}_h(\mathbf{U}_s-\mathbf{u})\right\rangle f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u},$$

where from Lemmas 4.B.1, 4.B.2 and 4.B.4, it follows that

$$\begin{aligned}
&E\left\langle \sum_{t=2}^{N} W_h(V_t - v)\,\mathcal{K}_h(\mathbf{U}_t - \mathbf{u}), \sum_{s=2}^{N} \mathcal{K}_h(\mathbf{U}_s - \mathbf{u}) \right\rangle \\
&= (N-1)\,E\left[W_h(V_1 - v)\,\|\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\|^2\right] \\
&+ (N-1)(N-2)\,E\left[W_h(V_1 - v)\,\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\right] E\left[\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\right] \\
&= (N-1)\left[C_{\mathcal{K}} f_{V\mathbf{U}}(v,\mathbf{u})\, h^{-(d-1)} + \psi^*_{W\mathcal{K}}(h,(v,\mathbf{u}))\right] \\
&+ (N-1)(N-2)\left[f_{V\mathbf{U}}(v,\mathbf{u}) + h^P S_{W\mathcal{K}}(v,\mathbf{u}) + \beta_{W\mathcal{K}}(h,(v,\mathbf{u}))\right] \times \\
&\qquad\qquad \left[f_{\mathbf{U}}(\mathbf{u}) + h^P S_{\mathcal{K}}(\mathbf{u}) + \beta_{\mathcal{K}}(h,\mathbf{u})\right].
\end{aligned}$$

Therefore, we have that the term (4.A.39) is equal to

$$\begin{aligned}
&\frac{1}{N}\int \widetilde{\sigma}_{23}(v,\mathbf{u}) f^2_{v\mathbf{u}}(v,\mathbf{u})\, f_{\mathbf{U}}(\mathbf{u})\, dv d\mathbf{u} + \frac{C_{\mathcal{K}}}{N^2 h^{d-1}}\int \widetilde{\sigma}_{23}(v,\mathbf{u})\, f^2_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u} \\
&\qquad + O\left(\frac{h^P}{N} + \frac{1}{N^2}\right) + o\left(\frac{1}{N^2 h^{d-1}}\right), \text{ as } N \to \infty.
\end{aligned}$$

Let us turn our attention to term (4.A.40),

$$\begin{aligned}
&E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\zeta}_{2i} - E[\widetilde{\zeta}_{21}]\right), \frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{3j} - E[\zeta_{31}]\right)\right\rangle\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left[\left\langle \widetilde{\zeta}_{2i}, \zeta_{3i}\right\rangle\right] + \frac{1}{N^2}\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} E\left[\left\langle \widetilde{\zeta}_{2i}, \zeta_{3j}\right\rangle\right] - \left\langle E[\widetilde{\zeta}_{21}], E[\zeta_{31}]\right\rangle \\
&= \frac{1}{N} E\left[\left\langle \widetilde{\zeta}_{21}, \zeta_{31}\right\rangle\right] + \frac{N-1}{N} E\left[\left\langle \widetilde{\zeta}_{21}, \zeta_{32}\right\rangle\right] - \langle q_2, q_3\rangle.
\end{aligned}$$

Also $E\left[\left\langle \widetilde{\zeta}_{21}, \zeta_{31}\right\rangle\right]$ equals

$$\frac{1}{N^2}\int \langle \widetilde{\pi}_2(v,\mathbf{u}), \pi_3(v,\mathbf{u})\rangle\, E\left[\sum_{t=2}^{N}\sum_{s=2}^{N} K_h(\mathbf{U}_t - \mathbf{u})\, W_h(V_s - v)\, \mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right] f_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u},$$

where $E\left[\sum_{t=2}^{N}\sum_{s=2}^{N} \mathcal{K}_h(\mathbf{U}_t - \mathbf{u})\, W_h(V_s - v)\, \mathcal{K}_h(\mathbf{U}_s - \mathbf{u})\right] = (N-1)\,E[W_h(V_1 - v) \times \|\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})\|^2\Big] + (N-1)(N-2)\,E[\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})]\, E[W_h(V_1 - v)\,\mathcal{K}_h(\mathbf{U}_1 - \mathbf{u})]$. It follows from Lemmas 4.B.1, 4.B.2 and 4.B.4 that

$$
\begin{aligned}
& E\left[\sum_{t=2}^{N}\sum_{s=2}^{N}\mathcal{K}_h\left(\mathbf{U}_t-\mathbf{u}\right)W_h\left(V_s-v\right)\mathcal{K}_h\left(\mathbf{U}_s-\mathbf{u}\right)\right] \\
&= (N-1)\left(C_{\mathcal{K}}f_{V\mathbf{U}}\left(v,\mathbf{u}\right)h^{-(d-1)}+\psi^*_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right) \\
&+ (N-1)(N-2)\left[f_{\mathbf{U}}\left(\mathbf{u}\right)+h^P S_{\mathcal{K}}\left(\mathbf{u}\right)+\beta_{\mathcal{K}}\left(h,\mathbf{u}\right)\right]\times \\
&\qquad\qquad \left[f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+h^P S_{W\mathcal{K}}\left(v,\mathbf{u}\right)+\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{31}\right\rangle\right] &= \frac{1}{N}\int\left\langle\widetilde{\pi}_2\left(v,\mathbf{u}\right),\pi_3\left(v,\mathbf{u}\right)\right\rangle f^2_{V\mathbf{U}}\left(v,\mathbf{u}\right)f_{\mathbf{u}}\left(\mathbf{u}\right)dvd\mathbf{u} \\
&+\frac{1}{N^2h^{d-1}}C_{\mathcal{K}}\int\left\langle\widetilde{\pi}_2\left(v,\mathbf{u}\right),\pi_3\left(v,\mathbf{u}\right)\right\rangle f^2_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u} \\
&+O\left(N^{-1}h^P+N^{-2}\right)+o\left(N^{-2}h^{-(d-1)}\right),\ \text{as } N\to\infty.
\end{aligned}
$$

We now turn our attention to

$$
\frac{N-1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{32}\right\rangle\right]=N^{-3}\left(N-1\right)\sum_{\substack{t=1\\ t\neq 1,s\neq 2}}^{N}\sum_{s=1}^{N}\Delta_{23,ts},
$$

where $\Delta_{23,ts}=E[\langle\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_t-\mathbf{U}_1\right),\pi_3\left(V_2,\mathbf{U}_2\right)W_h\left(V_s-V_2\right)\mathcal{K}_h\left(\mathbf{U}_s-\mathbf{U}_2\right)\rangle]$. Furthermore, we write

$$
\Delta_{23,ts}=\begin{cases}
\mathcal{B}_{23,I} & ;\quad s=t, \\
\mathcal{B}_{23,II} & ;\quad s\neq t,\ t\neq 2,\ s\neq 1, \\
\mathcal{B}_{23,III} & ;\quad s=t\ \&\ t=2,\ s\neq 1 \text{ or } t\neq 2,\ s=1, \\
\mathcal{B}_{23,IV} & ;\quad s\neq t,\ t=2,\ s=1.
\end{cases}
$$

Here we make the following definitions:

$$
\begin{aligned}
\mathcal{B}_{23,I} &= E\left[\left\langle E\left[\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\middle|\,\mathbf{U}_3\right],\right.\right. \\
&\qquad \left.\left. E\left[\pi_3\left(V_1,\mathbf{U}_1\right)W_h\left(V_3-V_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\middle|\,V_3,\mathbf{U}_3\right]\right\rangle\right] & ; \\
\mathcal{B}_{23,II} &= \left\langle E\left[\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\right],\right. \\
&\qquad \left. E\left[\pi_3\left(V_1,\mathbf{U}_1\right)W_h\left(V_3-V_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\right]\right\rangle\equiv\left\langle q_2,q_3\right\rangle & ; \\
\mathcal{B}_{23,III} &= E\left[\left\langle\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right),\pi_3\left(V_2,\mathbf{U}_2\right)\right\rangle\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)W_h\left(V_3-V_2\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_2\right)\right] & ; \\
\mathcal{B}_{23,IV} &= E\left[\left\langle\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right),\pi_3\left(V_2,\mathbf{U}_2\right)\right\rangle W_h\left(V_1-V_2\right)\left\|\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{U}_2\right)\right\|^2\right].
\end{aligned}
$$

Therefore, we are able to write

$$\frac{N-1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{32}\right\rangle\right]=\frac{1}{N}\left[\mathcal{B}_{23,I}+2\mathcal{B}_{23,III}\right]+\langle q_2,q_3\rangle\left[1-\frac{5}{N}\right]+\frac{1}{N^2}\mathcal{B}_{23,IV}$$
$$+\frac{1}{N^2}\left[6\left\langle\widetilde{q}_2,q_3\right\rangle-2\mathcal{B}_{23,I}-4\mathcal{B}_{23,III}\right]+o\left(N^{-2}\right).$$

Finally, term (4.A.41) is such that

$$\left\langle E[\widetilde{\zeta}_{21}]-\delta_2\,,\,E\left[\zeta_{31}\right]-\delta_3\right\rangle=\langle q_2-\delta_2,q_3-\delta_3\rangle$$
$$=h^{2P}\left\langle\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)d\mathbf{u},\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\rangle$$
$$+\,o\left(h^{2P}\right),\text{ as } N\to\infty,$$

which follows from Lemma 4.B.5.

Combining these pieces together, we obtain

$$E\left[\left\langle\widehat{\delta}_2\left(h\right)-\delta_2,\widehat{\delta}_3\left(h\right)-\delta_3\right\rangle\right]$$
$$=h^{2P}\left\langle\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)d\mathbf{u},\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right\rangle$$
$$+\,O\left(N^{-1}\right)+O\left(N^2h^{d-1}\right)+O\left(\frac{h^P}{N}+\frac{1}{N^2}\right)$$
$$+\,o\left(\frac{h^P}{N}+\frac{1}{N^2}+h^{2P}\right),\text{ as } N\to\infty, \tag{4.A.42}$$

Terms: $E\left[\left\langle\widehat{\delta}_2\left(h\right)-\delta_2,\widehat{\delta}_4\left(h\right)-\delta_4\right\rangle\right]$ **and** $E\left[\left\langle\widehat{\delta}_2\left(h\right)-\delta_2,\widehat{\delta}_5\left(h\right)-\delta_5\right\rangle\right]$

As before,

$$E\left[\left\langle\widehat{\delta}_2\left(h\right)-\delta_2,\widehat{\delta}_4\left(h\right)-\delta_4\right\rangle\right]=E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\widetilde{\varepsilon}_{2i}\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i\right),\frac{1}{N}\sum_{j=1}^{N}\varepsilon_{4j}\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i\right)\widehat{f}_{V\mathbf{U}}\left(v_j,\mathbf{u}_j\right)\right\rangle\right] \tag{4.A.43}$$

$$+\,E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\zeta}_{2i}-E[\widetilde{\zeta}_{21}]\right),\frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{4j}-E\left[\zeta_{41}\right]\right)\right\rangle\right] \tag{4.A.44}$$

$$+\left\langle E[\widetilde{\zeta}_{21}]-\delta_2,E\left[\zeta_{41}\right]-\delta_4\right\rangle,\text{ similarly} \tag{4.A.45}$$

$$E\left[\left\langle \widehat{\delta}_2(h) - \delta_2, \widehat{\delta}_5(h) - \delta_5 \right\rangle\right] = E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N} \widetilde{\varepsilon}_{2i} \widehat{f}_{\mathbf{U}}(\mathbf{u}_i), \frac{1}{N}\sum_{j=1}^{N} \varepsilon_{5j} \widehat{f}^2_{V\mathbf{U}}(v_j, \mathbf{u}_j) \right\rangle\right] \tag{4.A.46}$$

$$+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N} \left(\widetilde{\zeta}_{2i} - E[\widetilde{\zeta}_{21}]\right), \frac{1}{N}\sum_{j=1}^{N} \left(\zeta_{5j} - E[\zeta_{51}]\right) \right\rangle\right] \tag{4.A.47}$$

$$+ \left\langle E[\widetilde{\zeta}_{21}] - \delta_2, E[\zeta_{51}] - \delta_5 \right\rangle. \tag{4.A.48}$$

We also notice that

$$E\left[\widetilde{\varepsilon}_{2i}^{\top} \varepsilon_{4j} \middle| (V_1, \mathbf{U}_1), \ldots, (V_N, \mathbf{U}_N)\right] = \begin{cases} \widetilde{\sigma}_{24}(V_i, \mathbf{U}_i), & i = j, \\ 0, & i \neq j. \end{cases}$$

$$E\left[\widetilde{\varepsilon}_{2i}^{\top} \varepsilon_{5j} \middle| (V_1, \mathbf{U}_1), \ldots, (V_N, \mathbf{U}_N)\right] = \begin{cases} \widetilde{\sigma}_{25}(V_i, \mathbf{U}_i), & i = j, \\ 0, & i \neq j. \end{cases}$$

In what follows, we also use the following quantities:

$$E\left\langle \sum_{i=2}^{N} \mathcal{K}_h(\mathbf{U}_i - \mathbf{u}), \sum_{j=2}^{N}\sum_{k=2}^{N} \mathcal{K}_h(\mathbf{U}_j - \mathbf{u}) W_h(V_k - v) \mathcal{K}_h(\mathbf{U}_k - \mathbf{u}) \right\rangle \tag{4.A.49}$$
$$= (N-1) E\left[W_h(V_2 - v) \mathcal{K}_h^3(\mathbf{U}_2 - \mathbf{u})\right]$$
$$+ (N-1)(N-2) E\left[W_h(V_3 - v) \mathcal{K}_h(\mathbf{U}_3 - \mathbf{u})\right] E\left[\mathcal{K}_h^2(\mathbf{U}_2 - \mathbf{u})\right]$$
$$+ 2(N-1)(N-2) E\left[W_h(V_2 - v) \mathcal{K}_h^2(\mathbf{U}_2 - \mathbf{u})\right] E\left[\mathcal{K}_h(\mathbf{U}_3 - \mathbf{u})\right]$$
$$+ (N-1)(N-2)(N-3) \left\|E\left[\mathcal{K}_h(\mathbf{U}_4 - \mathbf{u})\right]\right\|^2 E\left[W_h(V_3 - v) \mathcal{K}_h(\mathbf{U}_3 - \mathbf{u})\right]$$

$$E\left\langle \sum_{i=2}^{N} \mathcal{K}_h(\mathbf{U}_i - \mathbf{u}), \sum_{j=2}^{N}\sum_{k=2}^{N} W_h(V_j - v) \mathcal{K}_h(\mathbf{U}_j - \mathbf{u}) W_h(V_k - v) \mathcal{K}_h(\mathbf{U}_k - \mathbf{u}) \right\rangle \tag{4.A.50}$$
$$= (N-1) E\left[W_h^2(V_2 - v) \mathcal{K}_h^3(\mathbf{U}_2 - \mathbf{u})\right]$$
$$+ 2(N-1)(N-2) E\left[W_h(V_3 - v) \mathcal{K}_h^2(\mathbf{U}_3 - \mathbf{u})\right] E\left[W_h(V_2 - v) \mathcal{K}_h(\mathbf{U}_2 - \mathbf{u})\right]$$
$$+ (N-1)(N-2) E\left[\mathcal{K}_h(\mathbf{U}_3 - \mathbf{u})\right] E\left[W_h^2(V_3 - v) \mathcal{K}_h^2(\mathbf{U}_3 - \mathbf{u})\right]$$
$$+ (N-1)(N-2)(N-3) E\left[\mathcal{K}_h(\mathbf{U}_4 - \mathbf{u})\right] \left\|E\left[W_h(V_3 - v) \mathcal{K}_h(\mathbf{U}_3 - \mathbf{u})\right]\right\|^2.$$

By Assumption (A6), the terms (4.A.43) and (4.A.46) are

$$\frac{1}{N(N-1)^3}\int \widetilde{\sigma}_{24}(v,\mathbf{u}) \times E\left\langle \sum_{i=2}^{N}\mathcal{K}_h(\mathbf{U}_i-\mathbf{u}), \sum_{j=2}^{N}\sum_{k=2}^{N}\mathcal{K}_h(\mathbf{U}_j-\mathbf{u})W_h(V_k-v)\mathcal{K}_h(\mathbf{U}_k-\mathbf{u})\right\rangle f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}, \text{ and} \tag{4.A.51}$$

$$\frac{1}{N(N-1)^3}\int \widetilde{\sigma}_{25}(v,\mathbf{u}) \times E\left\langle \sum_{i=2}^{N}\mathcal{K}_h(\mathbf{U}_i-\mathbf{u}), \sum_{j=2}^{N}\sum_{k=2}^{N}W_h(V_j-v)\mathcal{K}_h(\mathbf{U}_j-\mathbf{u})W_h(V_k-v)\mathcal{K}_h(\mathbf{U}_k-\mathbf{u})\right\rangle \tag{4.A.52}$$

$$\times f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}.$$

After plugging (4.A.49) and (4.A.50) in (4.A.51) and (4.A.52) respectively, and using Lemmas 4.B.1, 4.B.2, and 4.B.4, it follows that

$$\begin{aligned}
&E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\widetilde{\varepsilon}_{2i}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i), \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{4j}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)\widehat{f}_{V\mathbf{U}}(v_j,\mathbf{u}_j)\right\rangle\right]\\
&= O\left(N^{-1}\right)+O\left(N^{-3}h^{-2(d-1)}\right)+O\left(N^{-2}h^{-(d-1)}\right)+O\left(N^{-1}h^{P}\right), \text{ and}\\
&E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\widetilde{\varepsilon}_{2i}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i), \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{5j}\widehat{f}^{2}_{V\mathbf{U}}(v_j,\mathbf{u}_j)\right\rangle\right]\\
&= O\left(N^{-1}\right)+O\left(N^{-3}h^{-(2d-1)}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-2}h^{-(d-1)}\right)+O\left(N^{-1}h^{P}\right).
\end{aligned}$$

Now, terms (4.A.44) and (4.A.47) can be written as

$$\begin{aligned}
&E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\zeta}_{2i}-E[\widetilde{\zeta}_{21}]\right), \frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{4j}-E\left[\zeta_{41}\right]\right)\right\rangle\right]\\
&= \frac{1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{41}\right\rangle\right]+\frac{N-1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{42}\right\rangle\right]-\left\langle E[\widetilde{\zeta}_{21}],E\left[\zeta_{41}\right]\right\rangle, \text{ and}\\
&E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\zeta}_{2i}-E[\widetilde{\zeta}_{21}]\right), \frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{5j}-E\left[\zeta_{51}\right]\right)\right\rangle\right]\\
&= \frac{1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{51}\right\rangle\right]+\frac{N-1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{52}\right\rangle\right]-\left\langle E[\widetilde{\zeta}_{21}],E\left[\zeta_{51}\right]\right\rangle \text{ respectively.}
\end{aligned}$$

The terms $N^{-1}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{41}\right\rangle\right]$ and $N^{-1}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{51}\right\rangle\right]$ are like (4.A.51) and (4.A.52) after replacing $\widetilde{\sigma}_{24}(v,\mathbf{u})$ and $\widetilde{\sigma}_{25}(v,\mathbf{u})$ by $\langle\widetilde{\pi}_2(v,\mathbf{u}),\pi_4(v,\mathbf{u})\rangle$, and $\langle\widetilde{\pi}_2(v,\mathbf{u}),\pi_5(v,\mathbf{u})\rangle$ respectively. Therefore, they are of the same order. In particular,

$$\begin{aligned}
&\frac{1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{41}\right\rangle\right]\\
&=O\left(N^{-1}\right)+O\left(N^{-3}h^{-2(d-1)}\right)+O\left(N^{-2}h^{-(d-1)}\right)+O\left(N^{-1}h^{P}\right),\text{ and}\\
&\frac{1}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{51}\right\rangle\right]\\
&=O\left(N^{-1}\right)+O\left(N^{-3}h^{-(2d-1)}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-2}h^{-(d-1)}\right)+O\left(N^{-1}h^{P}\right).
\end{aligned}$$

We only show the working for $N^{-1}(N-1)E[\langle\widetilde{\zeta}_{21},\zeta_{52}\rangle]$, which is the leading term, in any case:

$$\begin{aligned}
&\frac{(N-1)}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{52}\right\rangle\right]\\
&=\frac{1}{N(N-1)^2}\left[\mathcal{B}_{25,I}+\mathcal{B}_{25,II}+\mathcal{B}_{25,III}+\mathcal{B}_{25,IV}+\mathcal{B}_{25,V}+\mathcal{B}_{25,VI}\right]+\mathcal{B}_{25,IV},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{B}_{25,I}&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}^{3}\right],\\
\mathcal{B}_{25,II}&=(N-2)E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle \mathcal{K}_{h;12}^{2}W_{h;32}\mathcal{K}_{h;32}\right],\\
\mathcal{B}_{25,III}&=2(N-2)E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}^{2}\mathcal{K}_{h;32}\right],\\
\mathcal{B}_{25,IV}&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle \mathcal{K}_{h;12}\left\{(N-2)W_{h;32}\mathcal{K}_{h;32}^{3}+(N-2)(N-3)\mathcal{K}_{h;32}W_{h;42}\mathcal{K}_{h;42}\right\}\right],\\
\mathcal{B}_{25,V}&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle \mathcal{K}_{h;12}\left\{(N-2)\mathcal{K}_{h;31}W_{h;32}\mathcal{K}_{h;32}+(N-2)(N-3)\mathcal{K}_{h;31}W_{h;42}\mathcal{K}_{h;42}\right\}\right],\\
\mathcal{B}_{25,VI}&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}\left\{(N-2)\mathcal{K}_{h;31}\mathcal{K}_{h;32}+(N-2)(N-3)\mathcal{K}_{h;31}\mathcal{K}_{h;42}\right\}\right],\\
\mathcal{B}_{25,VII}&=\frac{1}{N(N-1)^2}E\left[\left\langle\widetilde{\pi}_{2;1}\textstyle\sum_{t=3}^{N}\mathcal{K}_{h;t1},\pi_{5;2}\left\|\sum_{t=3}^{N}W_{h;t2}\mathcal{K}_{h;t2}\right\|^{2}\right\rangle\right].
\end{aligned}$$

It then follows from Lemmas 4.B.1, 4.B.12 and 4.B.13, that

$$\begin{aligned}
\frac{(N-1)}{N}E\left[\left\langle\widetilde{\zeta}_{21},\zeta_{52}\right\rangle\right]&=\left\langle E[\widetilde{\zeta}_{21}],E\left[\zeta_{51}\right]\right\rangle+O\left(N^{-1}\right)\\
&\quad+O\left(N^{-2}h^{-2(d-1)}\right)+O\left(N^{-1}h^{P}\right)
\end{aligned}$$

and conclude

$$\begin{aligned}
&E\left[\left\langle\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\zeta}_{2i}-E[\widetilde{\zeta}_{21}]\right),\frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{4j}-E\left[\zeta_{41}\right]\right)\right\rangle\right]\\
&=O\left(N^{-1}\right)+O\left(N^{-2}h^{-2(d-1)}\right)+O\left(N^{-2}h^{-d}\right)+O\left(N^{-2}h^{-(d-1)}\right)+O\left(N^{-1}h^{P}\right).
\end{aligned}$$

Likewise, (4.A.44) will be of smaller order than this term and therefore, it will not contribute towards the leading terms in the expansion. The only contributions will be from terms

(4.A.45), and (4.A.48). In particular,

$$\begin{aligned}
&\left\langle E[\widetilde{\zeta}_{21}]-\delta_2, E\left[\zeta_{41}\right]-\delta_4\right\rangle \\
&= h^{2P}\left\langle \int \pi_2\left(\mathbf{u}\right) S_{\mathcal{K}}\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) d\mathbf{u} \right. \\
&\left. , \int \pi_2\left(\mathbf{u}\right) S_{\mathcal{K}}\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) d\mathbf{u} + \int \pi_3\left(v,\mathbf{u}\right) S_{W\mathcal{K}}\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) dv d\mathbf{u}\right\rangle + o\left(N^{-1}h^{P-d}\right), \\
&\left\langle E[\widetilde{\zeta}_{21}]-\delta_2, E\left[\zeta_{51}\right]-\delta_5\right\rangle \\
&= 2h^{2P}\left\langle \int \pi_2\left(\mathbf{u}\right) S_{\mathcal{K}}\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) d\mathbf{u}, \int \pi_3\left(v,\mathbf{u}\right) S_{W\mathcal{K}}\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) dv d\mathbf{u}\right\rangle \\
&+ N^{-1}h^{P-d}\left\langle \int \pi_2\left(\mathbf{u}\right) S_{\mathcal{K}}\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) d\mathbf{u}, C_{W\mathcal{K}} \int \pi_3\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) dv d\mathbf{u}\right\rangle.
\end{aligned}$$

by Lemma 4.B.1.

Terms: $E\left[\left\langle \widehat{\delta}_3\left(h\right)-\delta_3, \widehat{\delta}_4\left(h\right)-\delta_4\right\rangle\right]$**,** $E\left[\left\langle \widehat{\delta}_3\left(h\right)-\delta_3, \widehat{\delta}_5\left(h\right)-\delta_5\right\rangle\right]$

and $E\left[\left\langle \widehat{\delta}_4\left(h\right)-\delta_4, \widehat{\delta}_5\left(h\right)-\delta_5\right\rangle\right]$

As before,

$$E\left[\left\langle \widehat{\delta}_3\left(h\right)-\delta_3, \widehat{\delta}_4\left(h\right)-\delta_4\right\rangle\right] = E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}\left(v,\mathbf{u}_i\right), \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{4j}\widehat{f}_{\mathbf{U}}\left(\mathbf{u}_i\right)\widehat{f}_{V\mathbf{U}}\left(v_j,\mathbf{u}_j\right)\right\rangle\right] \tag{4.A.53}$$

$$+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{3i}-E[\zeta_{31}]\right), \frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{4j}-E\left[\zeta_{41}\right]\right)\right\rangle\right] \tag{4.A.54}$$

$$+ \left\langle E[\zeta_{31}]-\delta_3, E\left[\zeta_{41}\right]-\delta_4\right\rangle, \tag{4.A.55}$$

$$E\left[\left\langle \widehat{\delta}_3\left(h\right)-\delta_3, \widehat{\delta}_5\left(h\right)-\delta_5\right\rangle\right] = E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{3i}\widehat{f}_{V\mathbf{U}}\left(v,\mathbf{u}_i\right), \frac{1}{N}\sum_{j=1}^{N}\varepsilon_{5j}\widehat{f}^{2}_{V\mathbf{U}}\left(v_j,\mathbf{u}_j\right)\right\rangle\right] \tag{4.A.56}$$

$$+ E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\left(\zeta_{3i}-E[\zeta_{31}]\right), \frac{1}{N}\sum_{j=1}^{N}\left(\zeta_{5j}-E\left[\zeta_{51}\right]\right)\right\rangle\right] \tag{4.A.57}$$

$$+ \left\langle E[\zeta_{31}]-\delta_3, E\left[\zeta_{51}\right]-\delta_5\right\rangle, \text{ similarly} \tag{4.A.58}$$

$$E\left[\left\langle \widehat{\delta}_4(h)-\delta_4,\widehat{\delta}_5(h)-\delta_5\right\rangle\right]=E\left[\left\langle \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{4j}\widehat{f}_{\mathbf{U}}(\mathbf{u}_i)\,\widehat{f}_{V\mathbf{U}}(v_j,\mathbf{u}_j),\frac{1}{N}\sum_{j=1}^{N}\varepsilon_{5j}\widehat{f}^2_{V\mathbf{U}}(v_j,\mathbf{u}_j)\right\rangle\right] \tag{4.A.59}$$

$$+E\left[\left\langle \frac{1}{N}\sum_{j=1}^{N}(\zeta_{4j}-E[\zeta_{41}]),\frac{1}{N}\sum_{j=1}^{N}(\zeta_{5j}-E[\zeta_{51}])\right\rangle\right] \tag{4.A.60}$$

$$+\langle E[\zeta_{41}]-\delta_4,E[\zeta_{51}]-\delta_5\rangle \tag{4.A.61}$$

As it was proven above, terms such as (4.A.53), (4.A.54), (4.A.56), (4.A.57), (4.A.59), and (4.A.60) will not contribute towards the leading terms in the expansion. However, terms (4.A.55), (4.A.58) and (4.A.61) will. In particular, in view of Lemma 4.B.1, it follows

$$\begin{aligned}
&\langle E[\zeta_{31}]-\delta_3,E[\zeta_{41}]-\delta_4\rangle\\
&=h^{2P}\left\langle\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right.\\
&\left.\int\pi_2(\mathbf{u})\,S_{\mathcal{K}}(\mathbf{u})\,f_{\mathbf{U}}(\mathbf{u})\,d\mathbf{u}+\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right\rangle+o\left(N^{-1}h^{P-d}\right),
\end{aligned}$$

$$\begin{aligned}
&\langle E[\zeta_{31}]-\delta_3,E[\zeta_{51}]-\delta_5\rangle\\
&=2h^{2P}\left\|\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right\|^2\\
&+N^{-1}h^{P-d}\left\langle\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u},C_{W\mathcal{K}}\int\pi_3(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right\rangle
\end{aligned}$$

$$\begin{aligned}
&\langle E[\zeta_{41}]-\delta_4,E[\zeta_{51}]-\delta_5\rangle\\
&=2h^{2P}\left\langle\int\pi_2(\mathbf{u})\,S_{\mathcal{K}}(\mathbf{u})\,f_{\mathbf{U}}(\mathbf{u})\,d\mathbf{u}+\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u},\right.\\
&\left.\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right\rangle\\
&+N^{-1}h^{P-d}\left\langle\int\pi_2(\mathbf{u})\,S_{\mathcal{K}}(\mathbf{u})\,f_{\mathbf{U}}(\mathbf{u})\,d\mathbf{u}+\int\pi_3(v,\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u},\right.\\
&\left.C_{W\mathcal{K}}\int\pi_3(v,\mathbf{u})\,f_{V\mathbf{U}}(v,\mathbf{u})\,dvd\mathbf{u}\right\rangle.
\end{aligned}$$

Summary

Summarizing, let us define the following quantities:

$$\mathfrak{B}_{1,1} = \int \pi_2(\mathbf{u})\, S_{\mathcal{K}}(\mathbf{u})\, f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u},$$
$$\mathfrak{B}_{1,2} = \int \pi_3(v,\mathbf{u})\, S_{W\mathcal{K}}(v,\mathbf{u})\, f_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u}, \text{ and}$$
$$\mathfrak{B}_2 = C_{W\mathcal{K}} \int \pi_3(v,\mathbf{u})\, f_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u},$$

then the contributions of each of the terms analyzed above will be:

Term:	Contribution: h^{2P}	Contribution: $N^{-1}h^{P-d}$	Contribution: $N^{-2}h^{-2d}$
(4.A.4)	–	–	–
(4.A.5)	$+4\,\Vert\mathfrak{B}_{1,1}\Vert^2$	–	–
(4.A.6)	$+4\,\Vert\mathfrak{B}_{1,2}\Vert^2$	–	–
(4.A.7)	$+\,\Vert\mathfrak{B}_{1,1}+\mathfrak{B}_{1,2}\Vert^2$	–	–
(4.A.8)	$+4\,\Vert\mathfrak{B}_{1,2}\Vert^2$	$+4\,\langle\mathfrak{B}_2,\mathfrak{B}_{1,2}\rangle$	$\Vert\mathfrak{B}_2\Vert^2$
(4.A.9)	–	–	–
(4.A.10)	–	–	–
(4.A.11)	–	–	–
(4.A.12)	–	–	–
(4.A.13)	$-8\,\langle\mathfrak{B}_{1,1},\mathfrak{B}_{1,2}\rangle$	–	–
(4.A.14)	$-4\,\langle\mathfrak{B}_{1,1},\mathfrak{B}_{1,1}+\mathfrak{B}_{1,2}\rangle$	–	–
(4.A.15)	$+8\,\langle\mathfrak{B}_{1,1},\mathfrak{B}_{1,2}\rangle$	$+4\,\langle\mathfrak{B}_2,\mathfrak{B}_{1,1}\rangle$	–
(4.A.16)	$+4\,\langle\mathfrak{B}_{1,2},\mathfrak{B}_{1,1}+\mathfrak{B}_{1,2}\rangle$	–	–
(4.A.17)	$-8\,\Vert\mathfrak{B}_{1,2}\Vert^2$	$-4\,\langle\mathfrak{B}_2,\mathfrak{B}_{1,2}\rangle$	–
(4.A.18)	$-4\,\langle\mathfrak{B}_{1,1}+\mathfrak{B}_{1,2},\mathfrak{B}_{1,2}\rangle$	$-2\,\langle\mathfrak{B}_2,\mathfrak{B}_{1,1}+\mathfrak{B}_{1,2}\rangle$	–
Net:	$\Vert\mathfrak{B}_{1,1}-\mathfrak{B}_{1,2}\Vert^2$	$2\,\langle\mathfrak{B}_{1,1}-\mathfrak{B}_{1,2},\mathfrak{B}_2\rangle$	$\Vert\mathfrak{B}_2\Vert^2$

Therefore, we conclude that the leading terms are:

$$
\begin{aligned}
& h^{2P}\left\|\int \pi_2(\mathbf{u}) S_{\mathcal{K}}(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} - \int \pi_3(v,\mathbf{u}) S_{W\mathcal{K}}(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) dv d\mathbf{u}\right\|^2 \\
& + \frac{2C_{W\mathcal{K}}}{Nh^d} h^P \left\langle \int \pi_3(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) dv d\mathbf{u}\right. \\
& \left., \int \pi_2(\mathbf{u}) S_{\mathcal{K}}(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} - \int \pi_3(v,\mathbf{u}) S_{W\mathcal{K}}(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) dv d\mathbf{u}\right\rangle \\
& + \frac{C_{W\mathcal{K}}^2}{N^2h^{2d}}\left\|\int \pi_3(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) dv d\mathbf{u}\right\|^2 \\
& = h^{2P}\left\|\mathfrak{B}_{1,1} - \mathfrak{B}_{1,2}\right\|^2 + 2\frac{h^P}{Nh^d}\left\langle \mathfrak{B}_2, \mathfrak{B}_{1,1} - \mathfrak{B}_{1,2}\right\rangle + \frac{1}{N^2h^{2d}}\left\|\mathfrak{B}_2\right\|^2 \\
& = \left\|h^P\mathfrak{B}_1 + \mathfrak{B}_2 N^{-1} h^{-d}\right\|^2,
\end{aligned}
$$

where $\mathfrak{B}_1 = \mathfrak{B}_{1,1} - \mathfrak{B}_{1,2}$, as required.

Proof of Proposition 4.3.1

In what follows, we make use of the following identities:

$$
\begin{aligned}
\frac{\widehat{f}_{\mathbf{U}i}}{\widehat{f}_{V\mathbf{U}i}} - \frac{f_{\mathbf{U}i}}{f_{V\mathbf{U}i}} &= \frac{\widehat{f}_{\mathbf{U}i} - f_{\mathbf{U}i}}{f_{V\mathbf{U}i}} - \frac{f_{\mathbf{U}i}\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)}{f_{V\mathbf{U}i}^2} \\
&+ \frac{f_{\mathbf{U}i}\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)^2}{f_{V\mathbf{U}i}^2\widehat{f}_{V\mathbf{U}i}} - \frac{\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)\left(\widehat{f}_{\mathbf{U}i} - f_{\mathbf{U}i}\right)}{f_{V\mathbf{U}i}\widehat{f}_{V\mathbf{U}i}}, \text{ and}
\end{aligned} \tag{4.A.62}
$$

$$
\begin{aligned}
\frac{\widehat{f}_{\mathbf{U}i}}{\widehat{f}_{V\mathbf{U}i}^2} - \frac{f_{\mathbf{U}i}}{f_{V\mathbf{U}i}^2} &= \frac{\widehat{f}_{\mathbf{U}i} - f_{\mathbf{U}i}}{f_{V\mathbf{U}i}^2} - \frac{f_{\mathbf{U}i}\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)\left(\widehat{f}_{V\mathbf{U}i} + f_{V\mathbf{U}i}\right)}{f_{V\mathbf{U}i}^4} \\
&+ \frac{f_{\mathbf{U}i}\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)^2\left(\widehat{f}_{V\mathbf{U}i} + f_{V\mathbf{U}i}\right)^2}{f_{V\mathbf{U}i}^4\widehat{f}_{V\mathbf{U}i}^2} \\
&- \frac{\left(\widehat{f}_{V\mathbf{U}i} - f_{V\mathbf{U}i}\right)\left(\widehat{f}_{\mathbf{U}i} - f_{\mathbf{U}i}\right)\left(\widehat{f}_{V\mathbf{U}i} + f_{V\mathbf{U}i}\right)}{f_{V\mathbf{U}i}^2\widehat{f}_{V\mathbf{U}i}^2}.
\end{aligned} \tag{4.A.63}
$$

Term: $\widehat{\mathfrak{B}}_1(h_0)$

Firstly, it follows from (4.A.62) that

$$\begin{aligned}\widetilde{\eta}(\Delta h_0)-\widetilde{\eta}(h_0)&=\widehat{\delta}_2(\Delta h_0)-\widehat{\delta}_2(h_0)-\left[\widehat{\delta}_3(\Delta h_0)-\widehat{\delta}_3(h_0)\right]\\&+N^{-1}\sum_{i=1}^{N}(\widehat{\vartheta}_{1i}(\Delta h_0)-\widehat{\vartheta}_{2i}(\Delta h_0))\omega_i a_\tau(v_i,\mathbf{u}_i) \qquad (4.A.64)\\&-N^{-1}\sum_{i=1}^{N}(\widehat{\vartheta}_{1i}(h_0)-\widehat{\vartheta}_{2i}(h_0))\omega_i a_\tau(v_i,\mathbf{u}_i), \qquad (4.A.65)\end{aligned}$$

where (4.A.64) and (4.A.65) are $O_p(N^{-1}h_0^{-d}\log N+h_0^{2P})$ because of Assumptions (A1), (A2), (A3), (A4), and results (4.A.63), (4.A.1), (4.A.2). That is,

$$\begin{aligned}\frac{\widetilde{\eta}(\Delta h_0)-\widetilde{\eta}(h_0)}{h_0^P(\Delta^P-1)}&=\begin{pmatrix}N\\2\end{pmatrix}^{-1}\sum_{i<j}\frac{p_2(\mathbf{t}_{2\tau i},\mathbf{t}_{2\tau j};\Delta h_0)-p_2(\mathbf{t}_{2\tau i},\mathbf{t}_{2\tau j};h_0)}{h_0^P(\Delta^P-1)}\\&-\begin{pmatrix}N\\2\end{pmatrix}^{-1}\sum_{i<j}\frac{p_3(\mathbf{t}_{3\tau i},\mathbf{t}_{3\tau j};\Delta h_0)-p_3(\mathbf{t}_{3\tau i},\mathbf{t}_{3\tau j};h_0)}{h_0^P(\Delta^P-1)}\\&+O_p((Nh_0^{P+d})^{-1}\log N+h_0^P),\end{aligned}$$

which means that $\widehat{\mathfrak{B}}_1(h_0)$ is the sum of two U-statistics plus a reminder that is $o_p(1)$, because under the conditions of the proposition, $h_0\to 0$ and $Nh_0^{P+d}\to\infty$ as $N\to\infty$. Given Lemma 4.B.14, it then follows from Lemma 3.1 (page 1410) in Powell, Stock, and Stoker (1989), and Theorem A (page 4) in Lewbel (2000a), that $(\widetilde{\eta}(\Delta h_0)-\widetilde{\eta}(h_0))/(h_0^P(\Delta^P-1))$ is consistent for

$$E\left[\omega\left(\frac{S_{\mathcal{K}}(\mathbf{U})}{f_{V\mathbf{U}}(V,\mathbf{U})}-\frac{f_{\mathbf{U}}(\mathbf{U})\,S_{W\mathcal{K}}(V,\mathbf{U})}{f_{V\mathbf{U}}^2(V,\mathbf{U})}\right)a_\tau(V,\mathbf{U})\right],$$

This is true because,

$$\begin{aligned}&\left\|E\left[\omega\left(\frac{S_{\mathcal{K}}(\mathbf{U})}{f_{V\mathbf{U}}(V,\mathbf{U})}-\frac{f_{\mathbf{U}}(\mathbf{U})\,S_{W\mathcal{K}}(V,\mathbf{U})}{f_{V\mathbf{U}}^2(V,\mathbf{U})}\right)(1-a_\tau(V,\mathbf{U}))\right]\right\|^2\\&\leq\left[\sup_{\Omega_{V\mathbf{U}}}\|\omega\|\sup_{\Omega_{V\mathbf{U}}}\left|\frac{S_{\mathcal{K}}(\mathbf{u})}{f_{V\mathbf{U}}(v,\mathbf{u})}-\frac{f_{\mathbf{U}}(\mathbf{u})\,S_{W\mathcal{K}}(v,\mathbf{u})}{f_{V\mathbf{U}}^2(v,\mathbf{u})}\right|E\left[1-a_\tau(V,\mathbf{U})\right]\right]^2. \qquad (4.A.66)\end{aligned}$$

Now $E[1-a_\tau(V,\mathbf{U})]$ equals the probability that $(v,\mathbf{u})$ is within a distance τ of the boundary of $\Omega_{V\mathbf{U}}$, which is less or equal to $\sup_{\Omega_{V\mathbf{U}}}f_{V\mathbf{U}}(v,\mathbf{u})$ times the volume of the space within a distance τ of the boundary or $\Omega_{V\mathbf{U}}$. This volume is $O(\tau)$, so from Assumptions (A3) and (A6), we have that (4.A.66) is $O(\tau)=O(N^{-1/2}(N^{1/2}\tau))=o(N^{-1/2})$, where the last equality follows from Assumption (A7). Therefore, under the conditions of the proposition, we conclude that $\widehat{\mathfrak{B}}_1(h_0) \xrightarrow{p} \mathfrak{B}_1$ as $N \to \infty$.

Term: $\widehat{\mathfrak{B}}_2(h_*)$

Notice that,

$$\widehat{\mathfrak{B}}_2(h_*) = \frac{C_{W\mathcal{K}}}{N}\sum_{i=1}^{N}\varpi_{3\tau i} + \frac{C_{W\mathcal{K}}}{N}\sum_{i=1}^{N}\widehat{\varpi}_{*3\tau i} - \varpi_{3\tau i}, \tag{4.A.67}$$

where the second term on the right–hand side of (4.A.67) is bounded above by

$$C_{W\mathcal{K}}\left(\frac{1}{N}\sum_{i=1}^{N}\|\omega_i\|^2\right)\max_{i=1,\dots,n}\left|\frac{\widehat{f}_{\mathbf{U}i}}{\widehat{f}^2_{V\mathbf{U}i}} - \frac{f_{\mathbf{U}i}}{f^2_{V\mathbf{U}i}}\right| = O_p\left(\sqrt{\frac{\log N}{Nh_*^d}} + h_*^P\right),$$

which is $o_p(1)$ by Assumption (A4), representation (4.A.63), and the assumptions of the proposition ($h_* \to 0$ and $Nh_*^d \to \infty$ as $N \to \infty$). The result follows from Kolmogorov's Law of Large Numbers when applied to the first term in the right–hand side of (4.A.67), and conclude that

$$\widehat{\mathfrak{B}}_2(h_*) \xrightarrow{p} \mathfrak{B}_2, \text{ as } N \to \infty.$$

4.B Technical Lemmas

Lemma 4.B.1 *Let Assumptions* (A1)-(A3) *hold. Then*

$$E\left(W_h\left(V_1-v\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right)=f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+h^P S_{W\mathcal{K}}\left(v,\mathbf{u}\right)+\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right), \tag{4.B.1}$$

$$E\left(\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right)=f_{\mathbf{U}}\left(\mathbf{u}\right)+h^P S_{\mathcal{K}}\left(\mathbf{u}\right)+\beta_{\mathcal{K}}\left(h,\mathbf{u}\right), \quad \forall\left(v,\mathbf{u}\right)\in\Re^d \tag{4.B.2}$$

where $\sup_{(v,\mathbf{u})}\left|\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right|=o\left(h^P\right)$, $\sup_{\mathbf{U}}\left|\beta_{\mathcal{K}}\left(h,\mathbf{u}\right)\right|=o\left(h^P\right)$ *as* $h\to0$, *and*

$$q_3=\delta_3+h^P\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}+\gamma_{W\mathcal{K}}\left(h\right), \tag{4.B.3}$$

$$q_2=\delta_2+h^P\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)d\mathbf{u}+\gamma_{\mathcal{K}}\left(h\right), \tag{4.B.4}$$

$$q_5=\delta_5+2h^P\left[\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right]+\gamma_{W\mathcal{K}}\left(h\right), \tag{4.B.5}$$

$$q_4=\delta_4+h^P\left[\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)d\mathbf{u}+\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right]+\gamma_{W\mathcal{K}}\left(h\right), \tag{4.B.6}$$

where $\left|\gamma_{W\mathcal{K}}\left(h\right)\right|=o\left(h^P\right)$, *and* $\left|\gamma_{\mathcal{K}}\left(h\right)\right|=o\left(h^P\right)$ *as* $h\to0$.

Proof. We prove (4.B.1) and (4.B.3) only, as (4.B.2) and (4.B.4) follow the exact same arguments. By a simple change of argument,

$$E\left(W_h\left(V_1-v\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right)=\frac{1}{h^d}\int f_{V\mathbf{U}}\left(v+ch,\mathbf{u}+\mathbf{c}h\right)W\left(c\right)\mathcal{K}\left(\mathbf{c}\right)dcd\mathbf{c}.$$

Assumption (A3) ensures that a Taylor series expansion is valid, and, uniformly in $(c,\mathbf{c})\in[-1,1]^d$,

$$\left|f_{V\mathbf{U}}\left(v+ch,\mathbf{u}+\mathbf{c}h\right)-\sum_{0\le|\alpha|\le P}\frac{h^{|\alpha|}}{\alpha!}D^\alpha f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\left(c,\mathbf{c}\right)^\alpha\right|\le\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right),$$

where $\sup_{(v,\mathbf{u})}\left|\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right|=o\left(h^P\right)$, with $h\to0$. We use the notation $\alpha=(\alpha_1,\ldots,\alpha_d)$, $\alpha!=\alpha_1!\times\ldots\times\alpha_d!$, $|\alpha|=\sum_{j=1}^d\alpha_j$, $(c,\mathbf{c})^\alpha=c^{\alpha_1}\times c_1^{\alpha_2}\times\ldots\times c_{d-1}^{\alpha_d}$, $\sum_{0\le|\alpha|\le P}=\sum_{j=0}^P\sum_{\substack{\alpha_1=0\\ \alpha_1+\ldots+\alpha_d=j}}^j\cdots\sum_{\alpha_d=0}^j$, and

$$D^\alpha f_{V\mathbf{U}}\left(v,\mathbf{u}\right)=\frac{\partial^\alpha f_{V\mathbf{U}}\left(v,\mathbf{u}\right)}{\partial v^{\alpha_1}\partial u_1^{\alpha_2}\ldots\partial u_{d-1}^{\alpha_d}}.$$

It follows from Assumption (A2),

$$
\begin{aligned}
E\left(W_h\left(V_1-v\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right) &= f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+\sum_{|\alpha|=P}\frac{h^{|\alpha|}}{\alpha!}D^{\alpha}f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\\
&\quad\times\int\left(c,\mathbf{c}\right)^{\alpha}W\left(c\right)\mathcal{K}\left(\mathbf{c}\right)dcd\mathbf{c}+\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\\
&= f_{V\mathbf{U}}\left(v,\mathbf{u}\right)+h^{P}S_{W\mathcal{K}}\left(v,\mathbf{u}\right)+\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right).
\end{aligned}
$$

Given this expression, It also follows

$$
\begin{aligned}
q_3 &= E\left[\pi_3\left(V_1,\mathbf{U}_1\right)W_h\left(V_2-V_1\right)\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)\right]\\
&= E\left[\pi_3\left(V_1,\mathbf{U}_1\right)E\left[W_h\left(V_2-V_1\right)\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)\middle| V=V_1,\mathbf{U}=\mathbf{U}_1\right]\right]\\
&= E\left[\pi_3\left(V_1,\mathbf{U}_1\right)\left(f_{V\mathbf{U}}\left(V_1,\mathbf{U}_1\right)+h^{P}S_{W\mathcal{K}}\left(V_1,\mathbf{U}_1\right)+\beta_{W\mathcal{K}}\left(h,\left(V_1,\mathbf{U}_1\right)\right)\right)\right]\\
&= \delta_3+h^{P}\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}+\gamma_{W\mathcal{K}}\left(h\right),
\end{aligned}
$$

where the last equality follows from $\sup_{(v,\mathbf{u})}\left|\beta_{W\mathcal{K}}\left(h,\left(v,\mathbf{u}\right)\right)\right| = o\left(h^{P}\right)$, and $\int\pi_3\left(v,\mathbf{u}\right)\times f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}<\infty$ since π_3 and $f_{V\mathbf{U}}$ are bounded on the compact support $\Omega_{V\mathbf{U}}$. Similarly, we have

$$
\begin{aligned}
q_4 &= E\left[\pi_4\left(V_1,\mathbf{U}_1\right)W_h\left(V_2-V_1\right)\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\right]\\
&= E\left[\pi_4\left(V_1,\mathbf{U}_1\right)E\left[W_h\left(V_2-V_1\right)\mathcal{K}_h\left(\mathbf{U}_2-\mathbf{U}_1\right)\middle| V=V_1,\mathbf{U}=\mathbf{U}_1\right]\right.\\
&\quad\left.\times E\left[\mathcal{K}_h\left(\mathbf{U}_3-\mathbf{U}_1\right)\middle| V=V_1,\mathbf{U}=\mathbf{U}_1\right]\right]\\
&= E\left[\pi_4\left(V_1,\mathbf{U}_1\right)\left(f_{V\mathbf{U}}\left(V_1,\mathbf{U}_1\right)+h^{P}S_{W\mathcal{K}}\left(V_1,\mathbf{U}_1\right)+\beta_{W\mathcal{K}}\left(h,\left(V_1,\mathbf{U}_1\right)\right)\right)\right.\\
&\quad\left.\times\left(f_{\mathbf{U}}\left(\mathbf{U}_1\right)+h^{P}S_{\mathcal{K}}\left(\mathbf{U}_1\right)+\beta_{\mathcal{K}}\left(h,\mathbf{U}_1\right)\right)\right]\\
&= \int\pi_4\left(v,\mathbf{u}\right)f_{V\mathbf{U}}^{2}\left(v,\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)dvd\mathbf{u}+h^{P}\int\pi_4\left(v,\mathbf{u}\right)f_{V\mathbf{U}}^{2}\left(v,\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)dvd\mathbf{u}\\
&\quad+h^{P}\int\pi_4\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)f_{\mathbf{U}}\left(\mathbf{u}\right)dvd\mathbf{u}+\gamma_{W\mathcal{K}}\left(h\right)\\
&= \delta_4+h^{P}\left[\int\pi_2\left(\mathbf{u}\right)S_{\mathcal{K}}\left(\mathbf{u}\right)d\mathbf{u}+\int\pi_3\left(v,\mathbf{u}\right)S_{W\mathcal{K}}\left(v,\mathbf{u}\right)f_{V\mathbf{U}}\left(v,\mathbf{u}\right)dvd\mathbf{u}\right]+\gamma_{W\mathcal{K}}\left(h\right),
\end{aligned}
$$

where the last equality follows from observing that

$$\begin{aligned}
\pi_4(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) &= E\left[\left.\frac{\omega}{f_{V\mathbf{U}}^2(V,\mathbf{U})}\right| V=v, \mathbf{U}=\mathbf{u}\right] f_{V\mathbf{U}}(v,\mathbf{u}) \\
&= E\left[\left.\frac{\omega}{f_{V\mathbf{U}}(V,\mathbf{U})}\right| V=v, \mathbf{U}=\mathbf{u}\right] \\
&= \widetilde{\pi}_2(v,\mathbf{u}), \text{ such that} \\
E\left[\widetilde{\pi}_2(V,\mathbf{U})\right| \mathbf{U}=\mathbf{u}] &= \int \widetilde{\pi}_2(v,\mathbf{u}) f_{V|\mathbf{U}}(v|\,\mathbf{u})\, dv = \pi_2(\mathbf{u}), \text{ and} \\
\pi_4(v,\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) &= E\left[\left.\frac{\omega}{f_{V\mathbf{U}}^2(V,\mathbf{U})}\right| V=v, \mathbf{U}=\mathbf{u}\right] f_{\mathbf{U}}(\mathbf{u}) \\
&= E\left[\left.\frac{\omega f_{\mathbf{U}}(\mathbf{U})}{f_{V\mathbf{U}}^2(V,\mathbf{U})}\right| V=v, \mathbf{U}=\mathbf{u}\right] \\
&= \pi_3(v,\mathbf{u}).
\end{aligned}$$

By the exact same arguments, we have

$$\begin{aligned}
q_5 &= E\left[\pi_5(V_1,\mathbf{U}_1) \left\|E\left[W_h(V_2-V_1)\mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1)\right| V=V_1, \mathbf{U}=\mathbf{U}_1]\right\|^2\right] \\
&= E\left[\pi_5(V_1,\mathbf{U}_1)\left(f_{V\mathbf{U}}(V_1,\mathbf{U}_1) + h^P S_{W\mathcal{K}}(V_1,\mathbf{U}_1) + \beta_{W\mathcal{K}}(h,(V_1,\mathbf{U}_1))\right)^2\right] \\
&= \int \pi_5(v,\mathbf{u}) f_{V\mathbf{U}}^3(v,\mathbf{u})\, dv d\mathbf{u} + 2h^P\left[\int \pi_5(v,\mathbf{u}) S_{W\mathcal{K}}(V_1,\mathbf{U}_1) f_{V\mathbf{U}}^2(v,\mathbf{u})\, dv d\mathbf{u}\right] + \gamma_{W\mathcal{K}}(h) \\
&= \delta_5 + 2h^P\left[\int \pi_3(v,\mathbf{u}) S_{W\mathcal{K}}(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u})\, dv d\mathbf{u}\right] + \gamma_{W\mathcal{K}}(h).
\end{aligned}$$

as needed. ■

Lemma 4.B.2 *Let Assumptions* (A1)-(A3) *hold. Then*

$$\begin{aligned}
E\left[W_h(V_1-v)\mathcal{K}_h^2(\mathbf{U}_1-\mathbf{u})\right] &= C_{\mathcal{K}} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-(d-1)} + \psi_{W\mathcal{K},12}^*(h,(v,\mathbf{u})), \\
E\left[W_h(V_1-v)\mathcal{K}_h^3(\mathbf{U}_1-\mathbf{u})\right] &= C_{\mathcal{K},3} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-2(d-1)} + \psi_{W\mathcal{K},13}^*(h,(v,\mathbf{u})), \\
E\left[W_h^2(V_1-v)\mathcal{K}_h^2(\mathbf{U}_1-\mathbf{u})\right] &= C_{W\mathcal{K}} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-d} + \psi_{W\mathcal{K}}(h,(v,\mathbf{u})), \\
E\left[W_h^2(V_1-v)\mathcal{K}_h^3(\mathbf{U}_1-\mathbf{u})\right] &= C_{W\mathcal{K},23} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-(2d-1)} + \psi_{W\mathcal{K},23}(h,(v,\mathbf{u})), \\
E\left[W_h^3(V_1-v)\mathcal{K}_h^3(\mathbf{U}_1-\mathbf{u})\right] &= C_{W\mathcal{K},33} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-2d} + \psi_{W\mathcal{K},33}(h,(v,\mathbf{u})), \\
E\left[W_h^3(V_1-v)\mathcal{K}_h^4(\mathbf{U}_1-\mathbf{u})\right] &= C_{W\mathcal{K},34} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-(3d-1)} + \psi_{W\mathcal{K},34}(h,(v,\mathbf{u})), \\
E\left[W_h^4(V_1-v)\mathcal{K}_h^4(\mathbf{U}_1-\mathbf{u})\right] &= C_{W\mathcal{K},44} f_{V\mathbf{U}}(v,\mathbf{u}) h^{-3d} + \psi_{W\mathcal{K},44}(h,(v,\mathbf{u})),
\end{aligned}$$

where $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},12}^*(h,(v,\mathbf{u}))\right| = o\left(h^{-(d-1)}\right)$, $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},13}^*(h,(v,\mathbf{u}))\right| = o\left(h^{-2(d-1)}\right)$, $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K}}(h,(v,\mathbf{u}))\right| = o\left(h^{-d}\right)$, $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},23}(h,(v,\mathbf{u}))\right| = o\left(h^{-(2d-1)}\right)$, $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},33}(h,(v,\mathbf{u}))\right| = o\left(h^{-2d}\right)$, $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},34}(h,(v,\mathbf{u}))\right| = o\left(h^{-(3d-1)}\right)$, *and* $\sup_{(v,\mathbf{u})}\left|\psi_{W\mathcal{K},44}(h,(v,\mathbf{u}))\right| = o\left(h^{-3d}\right)$.

Proof. Firstly,

$$\begin{aligned} E\left[W_h^2(V_1 - v)\mathcal{K}_h^2(\mathbf{U}_1 - \mathbf{u})\right] &= \frac{1}{h^{2d}}\int W^2\left(\frac{t-v}{h}\right)\mathcal{K}^2\left(\frac{\mathbf{t}-\mathbf{u}}{h}\right) f_{V\mathbf{U}}(t,\mathbf{t})\, dt d\mathbf{t} \\ &= \frac{1}{h^d}\int W^2(c)\mathcal{K}^2(\mathbf{c}) f_{V\mathbf{U}}(v+ch, \mathbf{u}+\mathbf{c}h)\, dc d\mathbf{c}. \end{aligned}$$

Similarly, we have that

$$\begin{aligned} E\left[W_h(V_1 - v)\mathcal{K}_h^2(\mathbf{U}_1 - \mathbf{u})\right] &= \frac{1}{h^{2d-1}}\int W\left(\frac{t-v}{h}\right)\mathcal{K}^2\left(\frac{\mathbf{t}-\mathbf{u}}{h}\right) f_{V\mathbf{U}}(t,\mathbf{t})\, dt d\mathbf{t} \\ &= \frac{1}{h^{d-1}}\int W(c)\mathcal{K}^2(\mathbf{c}) f_{V\mathbf{U}}(v+ch, \mathbf{u}+\mathbf{c}h)\, dc d\mathbf{c}. \end{aligned}$$

It follows from Assumption (A3), that $f_{V\mathbf{U}}$, and $f_{\mathbf{U}}$ are Lipschitz continuous on $\Omega_{V\mathbf{U}}$ and $\Omega_{\mathbf{U}}$ respectively, with some Lipschitz constants $L_{f_{V\mathbf{U}}}$ and $L_{f_{\mathbf{U}}}$. Thus,

$$\begin{aligned} &\left|\frac{1}{h^d}\int W^2(c)\mathcal{K}^2(\mathbf{c}) f_{V\mathbf{U}}(v+ch, \mathbf{u}+\mathbf{c}h)\, dc d\mathbf{c} - \frac{1}{h^d} C_{W\mathcal{K}} f_{V\mathbf{U}}(v, \mathbf{u})\right| \\ &\le \frac{L_{f_{V\mathbf{U}}}}{h^{d-1}}\int \|W(c)\mathcal{K}(\mathbf{c})\|^2 \|(c, \mathbf{c})\|\, dc d\mathbf{c}, \\ &\left|\frac{1}{h^{d-1}}\int W(c)\mathcal{K}^2(\mathbf{c}) f_{V\mathbf{U}}(v+ch, \mathbf{u}+\mathbf{c}h)\, dc d\mathbf{c} - \frac{1}{h^{d-1}} C_{\mathcal{K}} f_{V\mathbf{U}}(v, \mathbf{u})\right| \\ &\le \frac{L_{f_{V\mathbf{U}}}}{h^{d-2}}\int \|W(c)\| \|\mathcal{K}(\mathbf{c})\|^2 \|(c, \mathbf{c})\|\, dc d\mathbf{c}, \end{aligned}$$

which proves this Lemma. The same argument applies to each of the other components. ∎

Lemma 4.B.3 *Let Assumptions* (A1)-(A3) *hold. Then*

$$\begin{aligned} &E\left[W_h(V - v_1)\mathcal{K}_h(\mathbf{U} - \mathbf{u}_1) W_h(V - v_2)\mathcal{K}_h(\mathbf{U} - \mathbf{u}_2)\right] \qquad (4.B.7) \\ &= h^{-d}\langle W\mathcal{K}, W\mathcal{K}\rangle\left(\frac{v_1 - v_2}{h}, \frac{\mathbf{u}_1 - \mathbf{u}_2}{h}\right) f_{V\mathbf{U}}(v_1, \mathbf{u}_1) + \psi_{\langle W\mathcal{K}, W\mathcal{K}\rangle}(h, (v, \mathbf{u})), \\ &E\left[\mathcal{K}_h(\mathbf{U} - \mathbf{u}_1)\mathcal{K}_h(\mathbf{U} - \mathbf{u}_2)\right] \\ &= h^{-(d-1)}\langle \mathcal{K}, \mathcal{K}\rangle\left(\frac{\mathbf{u}_1 - \mathbf{u}_2}{h}\right) f_{\mathbf{U}}(\mathbf{u}_1) + \psi_{\langle \mathcal{K}, \mathcal{K}\rangle}(h, \mathbf{u}) \end{aligned}$$

$$
\begin{aligned}
&E\left[\mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_1\right) W_h\left(V-v_2\right) \mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_2\right)\right] \\
&= h^{-(d-1)}\langle\mathcal{K}, W\mathcal{K}\rangle\left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1, \mathbf{u}_1\right)+\psi_{\langle\mathcal{K}, W\mathcal{K}\rangle}(h,(v, \mathbf{u})), \\
&E\left[\mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_1\right) W_h^2\left(V-v_2\right) \mathcal{K}_h^2\left(\mathbf{U}-\mathbf{u}_2\right)\right] \\
&= h^{-(2d-1)}\left\langle\mathcal{K}, W^2\mathcal{K}^2\right\rangle\left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1, \mathbf{u}_1\right)+\psi_{\langle\mathcal{K}, W^2\mathcal{K}^2\rangle}(h,(v, \mathbf{u})), \\
&E\left[W_h\left(V-v_1\right) \mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_1\right) W_h^2\left(V-v_2\right) \mathcal{K}_h^2\left(\mathbf{U}-\mathbf{u}_2\right)\right] \\
&= h^{-2d}\left\langle W\mathcal{K}, W^2\mathcal{K}^2\right\rangle\left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1, \mathbf{u}_1\right)+\psi_{\langle W\mathcal{K}, W^2\mathcal{K}^2\rangle}(h,(v, \mathbf{u})),
\end{aligned}
$$

where $\langle f,g\rangle(v,u)=\int f(c,\mathbf{c})\,g(c-v,\mathbf{c}-\mathbf{u})\,dcd\mathbf{c}$, *and*

$$
\begin{aligned}
\sup_{\Omega_{V\mathbf{U}}} \psi_{\langle W\mathcal{K}, W\mathcal{K}\rangle}(h,(v, \mathbf{u})) &= o\left(h^{-d}\right), \\
\sup_{\Omega_{\mathbf{U}}} \psi_{\langle\mathcal{K}, \mathcal{K}\rangle}(h, \mathbf{u}) &= o\left(h^{-(d-1)}\right), \\
\sup_{\Omega_{V\mathbf{U}}} \psi_{\langle\mathcal{K}, W\mathcal{K}\rangle}(h,(v, \mathbf{u})) &= o\left(h^{-(d-1)}\right), \\
\sup_{\Omega_{V\mathbf{U}}} \psi_{\langle\mathcal{K}, W^2\mathcal{K}^2\rangle}(h,(v, \mathbf{u})) &= o\left(h^{-(2d-1)}\right), \\
\sup_{\Omega_{V\mathbf{U}}} \psi_{\langle W\mathcal{K}, W^2\mathcal{K}^2\rangle}(h,(v, \mathbf{u})) &= o\left(h^{-2d}\right),
\end{aligned}
$$

Proof. We only show the working for (4.B.7), as all other terms follow the exact same argument. Firstly,

$$
\begin{aligned}
&E\left[W_h\left(V-v_1\right) \mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_1\right) W_h\left(V-v_2\right) \mathcal{K}_h\left(\mathbf{U}-\mathbf{u}_2\right)\right] \\
&= \frac{1}{h^{2d}} \int W\left(\frac{t-v_1}{h}\right) \mathcal{K}\left(\frac{\mathbf{t}-\mathbf{u}_1}{h}\right) W\left(\frac{t-v_2}{h}\right) \mathcal{K}\left(\frac{\mathbf{t}-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}(t, \mathbf{t})\, dtd\mathbf{t} \\
&= \frac{1}{h^{d}} \int W(c) \mathcal{K}(\mathbf{c}) W\left(c+\frac{v_1-v_2}{h}\right) \mathcal{K}\left(\mathbf{c}+\frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1+ch, \mathbf{u}_1+\mathbf{c}h\right) dcd\mathbf{c},
\end{aligned}
$$

It follows from Assumption (A3), that $f_{V\mathbf{U}}$, and $f_{\mathbf{U}}$ are Lipschitz continuous on $\Omega_{V\mathbf{U}}$ and $\Omega_{\mathbf{U}}$ respectively, with some Lipschitz constants $L_{f_{V\mathbf{U}}}$ and $L_{f_{\mathbf{U}}}$. Similarly, from Assumption (A2), $W\mathcal{K}$ is also Lipschitz continuous on $[0,1]^d$, with some Lipschitz constant $L_{W\mathcal{K}}$. Thus,

$$
\begin{aligned}
&\left|\frac{1}{h^{d}} \int W(c) \mathcal{K}(\mathbf{c}) W\left(c+\frac{v_1-v_2}{h}\right) \mathcal{K}\left(\mathbf{c}+\frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1+ch, \mathbf{u}_1+\mathbf{c}h\right) dcd\mathbf{c}\right. \\
&\left.-\frac{1}{h^{d}}\langle W\mathcal{K}, W\mathcal{K}\rangle\left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}\left(v_1, \mathbf{u}_1\right)\right| \\
&\leq \frac{L_{W\mathcal{K}} L_{f_{V\mathbf{U}}}}{h^{d-1}} \int \|W(c) \mathcal{K}(\mathbf{c})\| \|(c, \mathbf{c})\|\, dcd\mathbf{c}.
\end{aligned}
$$

■

Lemma 4.B.4 *Let Assumptions* (A1)-(A3) *hold. Then*

$$E\left[\left\|\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right\|^2\right] = C_{\mathcal{K}} f_{\mathbf{U}}\left(\mathbf{u}\right) h^{-(d-1)} + \psi_{\mathcal{K}}\left(h,\mathbf{u}\right),$$
$$E\left[\mathcal{K}_h^3\left(\mathbf{U}_1-\mathbf{u}\right)\right] = C_{\mathcal{K},3} f_{\mathbf{U}}\left(\mathbf{u}\right) h^{-2(d-1)} + \psi_{\mathcal{K},3}\left(h,\mathbf{u}\right),$$
$$E\left[\left\|\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right\|^4\right] = C_{\mathcal{K},4} f_{\mathbf{U}}\left(\mathbf{u}\right) h^{-3(d-1)} + \psi_{\mathcal{K},4}\left(h,\mathbf{u}\right),$$

where $\sup_{\mathbf{u}}\left|\psi_{\mathcal{K}}\left(h,\mathbf{u}\right)\right| = o\left(h^{-(d-1)}\right)$, $\sup_{\mathbf{u}}\left|\psi_{\mathcal{K},3}\left(h,\mathbf{u}\right)\right| = o\left(h^{-2(d-1)}\right)$, *and* $\sup_{\mathbf{u}}\left|\psi_{\mathcal{K},4}\left(h,\mathbf{u}\right)\right| = o\left(h^{-3(d-1)}\right)$.

Proof. These are special cases of those in Lemma 4.B.2, with $W_h(\cdot)$ empty. The result follows by the same arguments. ■

Lemma 4.B.5 *Let Assumptions* (A1)-(A3) *hold. Then*

$$\begin{aligned} E\left[\pi_3\left(V_1,\mathbf{U}_1\right) W_h\left(V_1-v\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right] &= \pi_3\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) \\ &\quad + h^P \widetilde{S}_{W\mathcal{K},3}\left(v,\mathbf{u}\right) + \widetilde{\beta}_{W\mathcal{K},3}\left(h,\left(v,\mathbf{u}\right)\right), \end{aligned} \tag{4.B.8}$$

$$\begin{aligned} E\left[\pi_5\left(V_1,\mathbf{U}_1\right) W_h\left(V_1-v\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right] &= \pi_5\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) \\ &\quad + h^P \widetilde{S}_{W\mathcal{K},5}\left(v,\mathbf{u}\right) + \widetilde{\beta}_{W\mathcal{K},5}\left(h,\left(v,\mathbf{u}\right)\right), \end{aligned}$$

$$E\left[\pi_2\left(\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right] = \pi_2\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) + h^P \widetilde{S}_{\mathcal{K},2}\left(\mathbf{u}\right) + \widetilde{\beta}_{\mathcal{K},2}\left(h,\mathbf{u}\right), \tag{4.B.9}$$

$$E\left[\widetilde{\pi}_2\left(V_1,\mathbf{U}_1\right)\mathcal{K}_h\left(\mathbf{U}_1-\mathbf{u}\right)\right] = \pi_2\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right) + h^P \widetilde{S}^*_{\mathcal{K},2}\left(\mathbf{u}\right) + \widetilde{\beta}^*_{\mathcal{K},2}\left(h,\mathbf{u}\right), \tag{4.B.10}$$

$$E\left[\pi_5\left(V_2,\mathbf{U}_2\right) W_h^2\left(V_2-v\right)\mathcal{K}_h^2\left(\mathbf{U}_2-\mathbf{u}\right)\right] = h^{-d} C_{W\mathcal{K}} \pi_5\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right) + \beta_{W\mathcal{K},5}\left(h,\left(v,\mathbf{u}\right)\right) \tag{4.B.11}$$

where

$$\widetilde{S}_{W\mathcal{K},l}\left(v,\mathbf{u}\right) = \frac{1}{P}\left[d_W \frac{\partial^P}{\partial v^P}\left[\pi_l\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\right] + d_K \sum_{j=1}^{d-1} \frac{\partial^P}{\partial u_j^P}\left[\pi_l\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\right]\right],$$

$$\widetilde{S}_{\mathcal{K},2}\left(\mathbf{u}\right) = \frac{d_K}{P}\sum_{j=1}^{d-1}\frac{\partial^P}{\partial u_j^P}\left[\pi_2\left(\mathbf{u}\right) f_{\mathbf{U}}\left(\mathbf{u}\right)\right],$$

$$\widetilde{S}^*_{\mathcal{K},2}\left(\mathbf{u}\right) = \frac{d_K}{P}\sum_{j=1}^{d-1}\int \frac{\partial^P}{\partial u_j^P}\left[\widetilde{\pi}_2\left(v,\mathbf{u}\right) f_{V\mathbf{U}}\left(v,\mathbf{u}\right)\right] dv,$$

and $\sup_{(v,\mathbf{u})}\left|\widetilde{\beta}_{W\mathcal{K},l}\left(h,\left(v,\mathbf{u}\right)\right)\right| = o\left(h^{-d}\right)$, *for* $l = 3$ *and* 5, $\sup_{\mathbf{U}}\left|\widetilde{\beta}_{\mathcal{K}}\left(h,\mathbf{u}\right)\right| = o\left(h^{-(d-1)}\right)$,

$\sup_{(v,\mathbf{u})}\left|\widetilde{\beta}_{\mathcal{K}}^{*}(h,\mathbf{u})\right| = o\left(h^{-(d-1)}\right)$, *and* $\sup_{(v,\mathbf{u})}\left|\beta_{W\mathcal{K},5}(h,(v,\mathbf{u}))\right| = o\left(h^{-d}\right)$ *as* $h \to 0$.

Proof. As before, we only show the results for (4.B.8), (4.B.10), and (4.B.11) as the others follow the exact same arguments. By a simple change of argument

$$\begin{aligned}
&E\left[\pi_3(V_1,\mathbf{U}_1)W_h(V_1-v)\mathcal{K}_h(\mathbf{U}_1-\mathbf{u})\right] \\
&= \int \pi_3(v+ch,\mathbf{u}+\mathbf{c}h) f_{V\mathbf{U}}(v+ch,\mathbf{u}+\mathbf{c}h) W(c)\mathcal{K}(\mathbf{c})\,dcd\mathbf{c} \\
&= \pi_3(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) + h^P \sum_{|\alpha|=P}\frac{1}{\alpha!}D^{\alpha}\left[\pi_3(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u})\right] \\
&\times \int (c,\mathbf{c})^{\alpha} W(c)\mathcal{K}(\mathbf{c})\,dcd\mathbf{c} + \widetilde{\beta}_{W\mathcal{K}}(h,(v,\mathbf{u})),
\end{aligned}$$

where the last equality follows from Assumptions (A3) and (A2). Similarly, we can write (4.B.10) as

$$\begin{aligned}
E\left[\widetilde{\pi}_2(V_1,\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_1-\mathbf{u})\right] &= \int \widetilde{\pi}_2(v,\mathbf{u}+\mathbf{c}h) f_{V\mathbf{U}}(v,\mathbf{u}+\mathbf{c}h)\mathcal{K}(\mathbf{c})\,dvd\mathbf{c} \\
&= \pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) + h^P \sum_{|\alpha|=P}\frac{1}{\alpha!}\int D_{\mathbf{U}}^{\alpha}\left[\widetilde{\pi}_2(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u})\right]dv \\
&\times \int \mathbf{c}^{\alpha}\mathcal{K}(\mathbf{c})\,d\mathbf{c} + \widetilde{\beta}_{\mathcal{K}}^{*}(h,(v,\mathbf{u})).
\end{aligned}$$

Also,

$$\begin{aligned}
&E\left[\pi_5(V_2,\mathbf{U}_2)W_h^2(V_2-v)\mathcal{K}_h^2(\mathbf{U}_2-\mathbf{u})\right] \\
&= h^{-d}\int \pi_5(v+ch,\mathbf{u}+\mathbf{c}h) f_{V\mathbf{U}}(v+ch,\mathbf{u}+\mathbf{c}h) W^2(c)\left\|\mathcal{K}(\mathbf{c})\right\|^2 dcd\mathbf{c} \\
&= h^{-d}C_{W\mathcal{K}}\pi_5(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u}) + \widetilde{\beta}_{W\mathcal{K},5}(h,(v,\mathbf{u}))
\end{aligned}$$

as required. ■

Lemma 4.B.6 *Let Assumptions* (A1)-(A3) *hold. Then as* $h \to 0$,

$$\begin{aligned}
\mathcal{B}_{3,I} &= \int \|\pi_3(v,\mathbf{u})\|^2 f_{V\mathbf{U}}^3(v,\mathbf{u})\, d\mathbf{u} + O\left(h^P\right), \\
\mathcal{B}_{3,III} &= \int \|\pi_3(v,\mathbf{u})\|^2 f_{V\mathbf{U}}^3(v,\mathbf{u})\, d\mathbf{u} + O\left(h^P\right), \\
\mathcal{B}_{3,IV} &= \frac{1}{h^d}\left[C_{W\mathcal{K}} \int \|\pi_3(v,\mathbf{u}) f_{V\mathbf{U}}(v,\mathbf{u})\|^2 d\mathbf{u}\,\{1+o(1)\}\right], \\
\mathcal{B}_{2,I} &= \int \|\pi_2(\mathbf{u})\|^2 f_{\mathbf{U}}^3(\mathbf{u})\, d\mathbf{u} + O\left(h^P\right), \\
\mathcal{B}_{2,III} &= \int \|\pi_2(\mathbf{u})\|^2 f_{\mathbf{U}}^3(\mathbf{u})\, d\mathbf{u} + O\left(h^P\right), \\
\mathcal{B}_{2,IV} &= \frac{1}{h^{d-1}}\left[C_{\mathcal{K}} \int \|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\|^2 d\mathbf{u}\,\{1+o(1)\}\right].
\end{aligned}$$

Proof. For this Lemma, we show the results for $\mathcal{B}_{2,l}$, for $l = I, III, IV$, for notational convenience. The proof of $\mathcal{B}_{3,l}$, for $l = I, III, IV$ follows the exact same arguments, and therefore is omitted.

By Lemma 4.B.5, it follows

$$\begin{aligned}
\mathcal{B}_{2,I} &= E\left[\|E[\pi_2(\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_1)|\,\mathbf{U}_3]\|^2\right] \\
&= \int f_{\mathbf{U}}(\mathbf{u})\, \|E[\pi_2(\mathbf{U}_1)\mathcal{K}_h(\mathbf{u}-\mathbf{U}_1)]\|^2 d\mathbf{u} \\
&= \int f_{\mathbf{U}}(\mathbf{u}) \left\|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u}) + h^P \widetilde{S}_{\mathcal{K}}(\mathbf{u}) + \widetilde{\beta}_{\mathcal{K}}(h,\mathbf{u})\right\|^2 d\mathbf{u} \\
&= \int \|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\|^2 f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u} + O\left(h^P\right).
\end{aligned}$$

Using Lemmas 4.B.1 and 4.B.5, we obtain

$$\begin{aligned}
\mathcal{B}_{2,III} &= E\left[\langle \pi_2(\mathbf{U}_1), \pi_2(\mathbf{U}_2)\rangle \mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_2)\right] \\
&= \int \langle \pi_2(\mathbf{x}), \pi_2(\mathbf{y})\rangle \mathcal{K}_h(\mathbf{y}-\mathbf{x})\mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{\mathbf{U}}(\mathbf{x}) f_{\mathbf{U}}(\mathbf{y}) f_{\mathbf{U}}(\mathbf{z})\, d\mathbf{x}d\mathbf{y}d\mathbf{z} \\
&= \int \left\langle \int \pi_2(\mathbf{x})\mathcal{K}_h(\mathbf{y}-\mathbf{x}) f_{\mathbf{U}}(\mathbf{x})\, d\mathbf{x}, \int \pi_2(\mathbf{y})\mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{\mathbf{U}}(\mathbf{z})\, d\mathbf{z}\right\rangle f_{\mathbf{U}}(\mathbf{y})\, d\mathbf{y} \\
&= \int \Big\langle \pi_2(\mathbf{y}) f_{\mathbf{U}}(\mathbf{y}) + h^P \widetilde{S}_{\mathcal{K}}(\mathbf{y}) + \widetilde{\beta}_{\mathcal{K}}(h,\mathbf{y}), \\
&\quad \pi_2(\mathbf{y})\left[f_{\mathbf{U}}(\mathbf{y}) + h^P S_{\mathcal{K}}(\mathbf{y}) + \beta_{\mathcal{K}}(h,\mathbf{y})\right]\Big\rangle f_{\mathbf{U}}(\mathbf{y})\, d\mathbf{y} \\
&= \int \|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\|^2 f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u} + O\left(h^P\right).
\end{aligned}$$

Finally, by the change of variables: $\mathbf{c} = (\mathbf{x} - \mathbf{y})\, h^{-1}$, we have

$$\begin{aligned}
\mathcal{B}_{2,IV} &= E\left[\langle \pi_2(\mathbf{U}_1), \pi_2(\mathbf{U}_2)\rangle \|\mathcal{K}_h(\mathbf{U}_1 - \mathbf{U}_2)\|^2\right] \\
&= \int \langle \pi_2(\mathbf{x}), \pi_2(\mathbf{y})\rangle \|\mathcal{K}_h(\mathbf{x} - \mathbf{y})\|^2 f_{\mathbf{U}}(\mathbf{x}) f_{\mathbf{U}}(\mathbf{y})\, d\mathbf{x}d\mathbf{y} \\
&= \frac{1}{h^{d-1}} \int \langle \pi_2(\mathbf{y} + \mathbf{c}h), \pi_2(\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y} + \mathbf{c}h) f_{\mathbf{U}}(\mathbf{y}) \|\mathcal{K}(\mathbf{c})\|^2 d\mathbf{y}d\mathbf{c}.
\end{aligned}$$

From Assumption (A3), it follows that $f_{\mathbf{U}}$ and π_2 are Lipschitz continuous on $\Omega_{\mathbf{U}}$. Then

$$\begin{aligned}
\mathcal{B}_{2,IV} &= \frac{1}{h^{d-1}} \left[\int \|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\|^2 d\mathbf{u} \int \|\mathcal{K}(\mathbf{c})\|^2 d\mathbf{c}\, \{1 + O(h)\}\right] \\
&= \frac{1}{h^{d-1}} \left[\int \|\pi_2(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\|^2 d\mathbf{u} \int \|\mathcal{K}(\mathbf{c})\|^2 d\mathbf{c}\, \{1 + o(1)\}\right],
\end{aligned}$$

which concludes the proof. ∎

Lemma 4.B.7 *Let Assumptions* (A1)-(A3) *hold. Then as* $h \to 0$,

$$\begin{aligned}
\mathcal{B}_{5,I} &= O\left(h^{-3d}\right), \\
\mathcal{B}_{5,II} &= O\left(Nh^{-2d}\right), \\
\mathcal{B}_{5,III} &= O\left(Nh^{-2d}\right), \\
\mathcal{B}_{5,IV} &= O\left(N^2h^{-d}\right), \\
\mathcal{B}_{5,V} &= O\left(Nh^{-2d}\right) + O\left(N^2h^{-d}\right) + O\left(N^3\right) + O\left(N^3h^P\right), \\
\mathcal{B}_{5,VI} &= O\left(Nh^{-2d}\right) + O\left(N^2h^{-d}\right).
\end{aligned}$$

Proof. As before, it follows from Lemma 4.B.2,

$$\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W^4_{h;12} \mathcal{K}^4_{h;12}\right] \\
&= E\left[\pi_5(V_1, \mathbf{U}_1) E\left[\pi_5(V_2, \mathbf{U}_2) W^4_h(V_1 - V_2) \mathcal{K}^4_h(\mathbf{U}_1 - \mathbf{U}_2) \middle| \mathbf{U}_1\right]\right] \\
&= E\left[\pi_5(V_1, \mathbf{U}_1)\left(h^{-3d} C_{W\mathcal{K},44} \pi_5(V_1, \mathbf{U}_1) f_{V\mathbf{U}}(V_1, \mathbf{U}_1) + \widetilde{\beta}_{W\mathcal{K},44}(h, (V_1, \mathbf{U}_1))\right)\right] \\
&= \frac{C_{W\mathcal{K}}}{h^{-3d}} \int \|\pi_5(v, \mathbf{u})\|^2 f^2_{V\mathbf{U}}(v, \mathbf{u})\, dv d\mathbf{u}\, \{1 + o(1)\},
\end{aligned}$$

Similarly, we have

$$
\begin{aligned}
& E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^3 \mathcal{K}_{h;12}^3 W_{h;32}\mathcal{K}_{h;32}\right] \\
&= \left\langle \int \pi_5(x,\mathbf{x}) W_h^3(y-x)\mathcal{K}_h^3(\mathbf{y}-\mathbf{x}) f_{V\mathbf{U}}(x,\mathbf{x})\,dxd\mathbf{x} \right. \\
&\left. , \pi_5(y,\mathbf{y}) \int W_h(z-y)\mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{V\mathbf{U}}(z,\mathbf{z})\,dzd\mathbf{z} \right\rangle f_{V\mathbf{U}}(y,\mathbf{y})\,dyd\mathbf{y} \\
&= \int \Big\langle \pi_5(y,\mathbf{y}) \left(C_{W\mathcal{K},33} f_{V\mathbf{U}}(y,\mathbf{y}) h^{-2d} + \psi_{W\mathcal{K},33}(h,(y,\mathbf{y}))\right) \\
& \pi_5(y,\mathbf{y}) \left[f_{V\mathbf{U}}(y,\mathbf{y}) + h^P S_{W\mathcal{K},5}(y,\mathbf{y}) + \beta_{W\mathcal{K},5}(h,(y,\mathbf{y}))\right]\Big\rangle f_{V\mathbf{U}}(y,\mathbf{y})\,dyd\mathbf{y} \\
&= \frac{C_{W\mathcal{K},33}}{h^{-2d}} \int \|\pi_5(y,\mathbf{y})\|^2 f_{V\mathbf{U}}^3(y,\mathbf{y})\,dyd\mathbf{y}\,\{1+o(1)\}, \text{ and}
\end{aligned}
$$

from Lemma 4.B.3, it also follows

$$
\begin{aligned}
& E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2 \mathcal{K}_{h;12}^2 W_{h;31}\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32}\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2 \mathcal{K}_{h;12}^2 \times E\left[W_{h;31}\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} \middle| (V_1,\mathbf{U}_1),(V_2,\mathbf{U}_2)\right]\right] \\
&= \frac{1}{h^d} E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2 \mathcal{K}_{h;12}^2 \langle W\mathcal{K}, W\mathcal{K}\rangle \left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}(v_1,\mathbf{u}_1)\{1+o(1)\}\right] \\
&= O\left(h^{-2d}\right).
\end{aligned}
$$

Also, from Lemma 4.B.1 it follows

$$
\begin{aligned}
& E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2 \mathcal{K}_{h;12}^2 W_{h;41}\mathcal{K}_{h;41} W_{h;32}\mathcal{K}_{h;32}\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2 \mathcal{K}_{h;12}^2 \times E\left[W_{h;41}\mathcal{K}_{h;41} \middle| (V_1,\mathbf{U}_1)\right] E\left[W_{h;32}\mathcal{K}_{h;32} \middle| (V_2,\mathbf{U}_2)\right]\right] \\
&= O\left(h^{-d}\right).
\end{aligned}
$$

By Lemmas 4.B.1 and 4.B.3, all other terms are

$$
\begin{aligned}
& E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12} W_{h;31}\mathcal{K}_{h;31} W_{h;32}^2\mathcal{K}_{h;32}^2\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12} E\left[W_{h;31}\mathcal{K}_{h;31} W_{h;32}^2\mathcal{K}_{h;32}^2 \middle| (V_1,\mathbf{U}_1),(V_2,\mathbf{U}_2)\right]\right] \\
&= \frac{1}{h^{2d}} E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12} \langle W\mathcal{K}, W^2\mathcal{K}^2\rangle \left(\frac{v_1-v_2}{h}, \frac{\mathbf{u}_1-\mathbf{u}_2}{h}\right) f_{V\mathbf{U}}(v_1,\mathbf{u}_1)\{1+o(1)\}\right] \\
&= O\left(h^{-2d}\right).
\end{aligned}
$$

$$
\begin{aligned}
& E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12} W_{h;31}\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} W_{h;42}\mathcal{K}_{h;42}\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}\right. \\
&\left. \times E\left[W_{h;31}\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} \middle| (V_1,\mathbf{U}_1),(V_2,\mathbf{U}_2)\right] E\left[W_{h;42}\mathcal{K}_{h;42} \middle| V_2,\mathbf{U}_2\right]\right] \\
&= O\left(h^{-d}\right),
\end{aligned}
$$

$$\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}W_{h;41}\mathcal{K}_{h;41}W_{h;32}^2\mathcal{K}_{h;32}^2\right]\\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}\right.\\
&\left.\times E\left[W_{h;41}\mathcal{K}_{h;41}\right| V_1, \mathbf{U}_1\right] E\left[W_{h;32}^2\mathcal{K}_{h;32}^2\right| V_2, \mathbf{U}_2\left.\right]\Big]\\
&= O\left(h^{-d}\right),
\end{aligned}$$

$$\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}W_{h;31}\mathcal{K}_{h;31}W_{h;42}\mathcal{K}_{h;42}W_{h;52}\mathcal{K}_{h;52}\right]\\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}\mathcal{K}_{h;12}\right.\\
&\left.\times E\left[W_{h;31}\mathcal{K}_{h;31}\right| V_1, \mathbf{U}_1\right]\left(E\left[W_{h;42}\mathcal{K}_{h;42}\right| V_2, \mathbf{U}_2\right])^3\Big]\\
&= O(1) + O\left(h^P\right),
\end{aligned}$$

$$\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2\mathcal{K}_{h;12}^2W_{h;32}^2\mathcal{K}_{h;32}^2\right]\\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2\mathcal{K}_{h;12}^2 E\left[W_{h;32}^2\mathcal{K}_{h;32}^2\right| V_2, \mathbf{U}_2\right]\Big]\\
&= O\left(h^{-2d}\right),
\end{aligned}$$

$$\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2\mathcal{K}_{h;12}^2W_{h;32}\mathcal{K}_{h;32}W_{h;42}\mathcal{K}_{h;42}\right]\\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;12}^2\mathcal{K}_{h;12}^2\left(E\left[W_{h;32}\mathcal{K}_{h;32}\right| V_2, \mathbf{U}_2\right])^2\right]\\
&= O\left(h^{-d}\right),
\end{aligned}$$

as required. ■

Lemma 4.B.8 *Let Assumptions* (A1)-(A3) *hold. Then*

$$\mathcal{B}_{5,VI} = \|E[\zeta_{51}]\|^2 + O\left(N^{-2}h^{-d}\right)$$

Proof. Firstly,

$$
\begin{aligned}
&\mathcal{B}_{5,VI} = \\
&\frac{1}{N(N-1)^3} E\left[\left\langle \pi_{5;1}\left\|\sum_{t=3}^{N} W_{h;t1}\mathcal{K}_{h;t1}\right\|^2, \pi_{5;2}\left\|\sum_{s=3}^{N} W_{h;s2}\mathcal{K}_{h;s2}\right\|^2\right\rangle\right] \\
&= \frac{1}{N(N-1)^2} E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;31}^2 \mathcal{K}_{h;31}^2 W_{h;32}^2 \mathcal{K}_{h;32}^2\right] \\
&+ \frac{1}{N(N-1)} E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;41}\mathcal{K}_{h;41} W_{h;42}\mathcal{K}_{h;42} W_{h;51}\mathcal{K}_{h;51} W_{h;52}\mathcal{K}_{h;52}\right] \\
&+ \frac{1}{N(N-1)} \left\langle E\left[\pi_{5;1} W_{h;31}^2 \mathcal{K}_{h;31}^2\right], E\left[\pi_{5;2} W_{h;32}^2 \mathcal{K}_{h;32}^2\right]\right\rangle \qquad (4.B.12) \\
&+ \frac{2}{N} \left\langle E\left[\pi_{5;1} W_{h;31}^2 \mathcal{K}_{h;31}^2\right], E\left[\pi_{5;2} W_{h;42}\mathcal{K}_{h;42} W_{h;52}\mathcal{K}_{h;52}\right]\right\rangle \qquad (4.B.13) \\
&+ \left\langle E\left[\pi_{5;1} W_{h;41}\mathcal{K}_{h;41} W_{h;51}\mathcal{K}_{h;51}\right], E\left[\pi_{5;2} W_{h;42}\mathcal{K}_{h;42} W_{h;52}\mathcal{K}_{h;52}\right]\right\rangle. \qquad (4.B.14)
\end{aligned}
$$

Now, it follows from Lemma 4.B.3, that

$$
\begin{aligned}
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;31}^2 \mathcal{K}_{h;31}^2 W_{h;32}^2 \mathcal{K}_{h;32}^2\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle E\left[W_{h;31}^2 \mathcal{K}_{h;31}^2 \middle| V_1, \mathbf{U}_1\right] E\left[W_{h;32}^2 \mathcal{K}_{h;32}^2 \middle| V_2, \mathbf{U}_2\right]\right] \\
&= O\left(h^{-d}\right), \text{ and} \\
&E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle W_{h;41}\mathcal{K}_{h;41} W_{h;42}\mathcal{K}_{h;42} W_{h;51}\mathcal{K}_{h;51} W_{h;52}\mathcal{K}_{h;52}\right] \\
&= E\left[\langle \pi_{5;1}, \pi_{5;2}\rangle E\left[W_{h;41}\mathcal{K}_{h;41} W_{h;42}\mathcal{K}_{h;42} \middle| (V_1, \mathbf{U}_1), (V_2, \mathbf{U}_2)\right]\right. \\
&\times E\left[W_{h;51}\mathcal{K}_{h;51} W_{h;52}\mathcal{K}_{h;52} \middle| (V_1, \mathbf{U}_1), (V_2, \mathbf{U}_2)\right]\left.\right] \\
&= O\left(h^{-d}\right).
\end{aligned}
$$

The result follows after noticing that $\|E[\zeta_{51}]\|^2 = (4.B.12) + (4.B.13) + (4.B.14)$. ■

Lemma 4.B.9 *Let Assumptions* (A1)-(A3) *hold. Then as* $h \to 0$,

$$
\begin{aligned}
\mathcal{B}_{13,I} &= \int \langle \widetilde{\pi}_1(v, \mathbf{u}), \pi_3(v, \mathbf{u})\rangle f_{V\mathbf{U}}^2(v, \mathbf{u})\, dv d\mathbf{u}\, \{1 + o(1)\}, \\
\mathcal{B}_{13,II} &= \langle \delta_1, \delta_3\rangle + h^P \int \langle \delta_1, \pi_3(v, \mathbf{u})\rangle S_{W\mathcal{K}}(v, \mathbf{u})\, dv d\mathbf{u} + \mathbf{o}\left(h^P\right), \\
\mathcal{B}_{12,I} &= \int \langle \pi_1(\mathbf{u}), \pi_2(\mathbf{u})\rangle f_{\mathbf{U}}^2(\mathbf{u})\, d\mathbf{u}\, \{1 + o(1)\}, \\
\mathcal{B}_{12,II} &= \langle \delta_1, \delta_2\rangle + h^P \int \langle \delta_1, \pi_2(\mathbf{u})\rangle S_{\mathcal{K}}(\mathbf{u})\, d\mathbf{u} + \mathbf{o}\left(h^P\right).
\end{aligned}
$$

Proof. As before, we show the results for $\mathcal{B}_{12,l}$, for $l = I, II$. From Assumption (A2) and (A3), we have

$$\begin{aligned}
\mathcal{B}_{12,I} &= E\left[\langle \pi_1(\mathbf{U}_1), \pi_2(\mathbf{U}_2)\rangle \mathcal{K}_h(\mathbf{U}_1 - \mathbf{U}_2)\right] \\
&= \int \langle \pi_1(\mathbf{x}), \pi_2(\mathbf{y})\rangle \mathcal{K}_h(\mathbf{x}-\mathbf{y}) f_{\mathbf{U}}(\mathbf{x}) f_{\mathbf{U}}(\mathbf{y})\, d\mathbf{x}d\mathbf{y} \\
&= \int \langle \pi_1(\mathbf{y}+\mathbf{c}h), \pi_2(\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y}+\mathbf{c}h) f_{\mathbf{U}}(\mathbf{y}) \mathcal{K}(\mathbf{c})\, d\mathbf{y}d\mathbf{c} \\
&= \int \langle \pi_1(\mathbf{y}), \pi_2(\mathbf{y})\rangle f_{\mathbf{U}}^2(\mathbf{y})\, d\mathbf{y}\left\{1 + o(1)\right\},
\end{aligned}$$

where the last equality comes from the change of variables: $\mathbf{c} = h^{-1}(\mathbf{x}-\mathbf{y})$. Finally, from Lemma 4.B.1, we obtain

$$\begin{aligned}
\mathcal{B}_{12,II} &= E[\langle \pi_1(\mathbf{U}_1), \pi_2(\mathbf{U}_2)\rangle \mathcal{K}_h(\mathbf{U}_3 - \mathbf{U}_2)] \\
&= \int \langle \pi_1(\mathbf{x}), \pi_2(\mathbf{y})\rangle \mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{\mathbf{U}}(\mathbf{x}) f_{\mathbf{U}}(\mathbf{y}) f_{\mathbf{U}}(\mathbf{z})\, d\mathbf{x}d\mathbf{y}d\mathbf{z} \\
&= \int \left\langle \int \pi_1(\mathbf{x}) f_{\mathbf{U}}(\mathbf{x})\, d\mathbf{x}, \int \pi_2(\mathbf{y}) \mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{\mathbf{U}}(\mathbf{z})\, d\mathbf{z}\right\rangle f_{\mathbf{U}}(\mathbf{y})\, d\mathbf{y} \\
&= \int \langle \delta_1, \pi_2(\mathbf{y})\rangle f_{\mathbf{U}}^2(\mathbf{y})\, d\mathbf{y} + h^P \int \langle \delta_1, \pi_2(\mathbf{y})\rangle S_{\mathcal{K}}(\mathbf{y})\, d\mathbf{y} + o\left(h^P\right),
\end{aligned}$$

the result follows after noticing that $E\left[\pi_2(\mathbf{y}) f_{\mathbf{U}}(\mathbf{y})\right] = \delta_2$. ■

Lemma 4.B.10 *Let Assumptions* (A1)-(A3) *hold. Then as* $h \to 0$,

$$\begin{aligned}
\mathcal{B}_{15,I} &= \frac{1}{h^d}\left[C_{W\mathcal{K}} \int \langle \delta_1^*, \pi_5(y,\mathbf{y})\rangle f_{V\mathbf{U}}^2(y,\mathbf{y})\, dy d\mathbf{y}\left\{1 + o(1)\right\}\right], \\
\mathcal{B}_{15,II} &= \left\langle \delta_1^*, \int \pi_5(y,\mathbf{y}) f_{V\mathbf{U}}^3(y,\mathbf{y})\, dy d\mathbf{y}\right\rangle + O\left(h^P\right), \\
\mathcal{B}_{15,III} &= \int \langle \widetilde{\pi}_1(y,\mathbf{y}), \pi_5(y,\mathbf{y}) f_{V\mathbf{U}}(y,\mathbf{y})\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dy d\mathbf{y} + O\left(h^P\right), \text{ and} \\
\mathcal{B}_{15,IV} &= \frac{1}{h^d}\left[C_{W\mathcal{K}} \int \langle \widetilde{\pi}_1(x,\mathbf{x}), \pi_5(x,\mathbf{x}) f_{V\mathbf{U}}(x,\mathbf{x})\rangle f_{V\mathbf{U}}(x,\mathbf{x})\left\{1 + o(1)\right\}\right].
\end{aligned}$$

Proof. Firstly, by Lemma 4.B.2, it follows

$$
\begin{aligned}
\mathcal{B}_{15,I} = \Big\langle \int & \pi_1(x,\mathbf{x}) f_{V\mathbf{U}}(x,\mathbf{x})\, dxd\mathbf{x}, \\
& \int \pi_5(y,\mathbf{y}) W_h^2(z-y) \mathcal{K}_h^2(\mathbf{z}-\mathbf{y}) f_{V\mathbf{U}}(z,\mathbf{z})\, dzd\mathbf{z} f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \Big\rangle \\
&= \left\langle \delta_1^*, \int \pi_5(y,\mathbf{y}) \left[C_{W\mathcal{K}} f_{V\mathbf{U}}(y,\mathbf{y}) h^{-d} + \psi_{W\mathcal{K}}(h,(y,\mathbf{y}))\right] f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \right\rangle \\
&= \frac{1}{h^d}\left[C_{W\mathcal{K}} \int \langle \delta_1^*, \pi_5(y,\mathbf{y})\rangle f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y}\{1+o(1)\}\right].
\end{aligned}
$$

Also, from Lemma 4.B.1, it follows

$$
\begin{aligned}
\mathcal{B}_{15,II} &= \left\langle \delta_1^*, \int \pi_5(y,\mathbf{y}) \left[f_{V\mathbf{U}}(y,\mathbf{y}) + h^P S_{W\mathcal{K}}(y,\mathbf{y}) + \beta_{W\mathcal{K}}(h,(y,\mathbf{y}))\right]^2 f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \right\rangle \\
&= \left\langle \delta_1^*, \int \pi_5(y,\mathbf{y}) f_{V\mathbf{U}}^3(y,\mathbf{y})\, dyd\mathbf{y} \right\rangle \\
&+ 2h^P \left\langle \delta_1^*, \pi_5(y,\mathbf{y}) S_{W\mathcal{K}}(y,\mathbf{y}) f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y} \right\rangle + o\left(h^P\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathcal{B}_{15,III} &= \int \langle \widetilde{\pi}_1(x,\mathbf{x}) W_h(x-y) \mathcal{K}_h(\mathbf{x}-\mathbf{y}) f_{V\mathbf{U}}(x,\mathbf{x})\, dxd\mathbf{x}, \\
& \pi_5(y,\mathbf{y}) W_h(z-y) \mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{V\mathbf{U}}(z,\mathbf{z})\, dzd\mathbf{z}\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \\
&= \int \langle \widetilde{\pi}_1(y+ch,\mathbf{y}+\mathbf{c}h) W(c) \mathcal{K}(\mathbf{c}) f_{V\mathbf{U}}(y+ch,\mathbf{y}+\mathbf{c}h)\, dcd\mathbf{c}, \\
& \pi_5(y,\mathbf{y}) \left[f_{V\mathbf{U}}(y,\mathbf{y}) + h^P S_{W\mathcal{K}}(y,\mathbf{y}) + \beta_{W\mathcal{K}}(h,(y,\mathbf{y}))\right]\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \\
&= \int \langle \widetilde{\pi}_1(y,\mathbf{y}), \pi_5(y,\mathbf{y}) f_{V\mathbf{U}}(y,\mathbf{y})\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} \\
&+ h^P \langle \widetilde{\pi}_1(y,\mathbf{y}), \pi_5(y,\mathbf{y}) S_{W\mathcal{K}}(y,\mathbf{y})\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y} + o\left(h^P\right).
\end{aligned}
$$

Finally, from Assumption (A3), it follows that $f_{V\mathbf{U}}$ and π_5 are Lipschitz continuous on $\Omega_{V\mathbf{U}}$. Then

$$
\begin{aligned}
\mathcal{B}_{15,IV} &= E\left[\langle \widetilde{\pi}_1(V_1,\mathbf{U}_1), \pi_5(V_2,\mathbf{U}_2) W_h^2(V_1-V_2) \mathcal{K}_h^2(\mathbf{U}_1-\mathbf{U}_2)\rangle\right] \\
&= \int \langle \widetilde{\pi}_1(x,\mathbf{x}), \pi_5(y,\mathbf{y})\rangle \\
&\times W_h^2(x-y) \mathcal{K}_h^2(\mathbf{x}-\mathbf{y}) f_{V\mathbf{U}}(x,\mathbf{x}) f_{V\mathbf{U}}(y,\mathbf{y})\, dxd\mathbf{x}dyd\mathbf{y} \\
&= \frac{1}{h^d}\int \left\langle \widetilde{\pi}_1(x,\mathbf{x}), \int \pi_5(x+ch,\mathbf{x}+\mathbf{c}h) W^2(c) \|\mathcal{K}(\mathbf{c})\| f_{V\mathbf{U}}(x+ch,\mathbf{x}+\mathbf{c}h)\, dcd\mathbf{c} \right\rangle \\
&\times f_{V\mathbf{U}}(x,\mathbf{x})\, dxd\mathbf{x} \\
&= \frac{C_{W\mathcal{K}}}{h^d}\int \langle \widetilde{\pi}_1(x,\mathbf{x}), \pi_5(x,\mathbf{x}) f_{V\mathbf{U}}(x,\mathbf{x})\rangle f_{V\mathbf{U}}(x,\mathbf{x})\, dxd\mathbf{x}\{1+o(1)\},
\end{aligned}
$$

as required. ∎

Lemma 4.B.11 *Let Assumptions* (A1)-(A3) *hold. Then as* $h \to 0$,

$$\begin{aligned}
\mathcal{B}_{23,I} &= \int \langle \pi_2(\mathbf{y}), \pi_3(y,\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y}) f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y} + O\left(h^P\right),\\
\mathcal{B}_{23,III} &= \int \langle \pi_2(\mathbf{y}), \pi_3(y,\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y}) f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y} + O\left(h^P\right),\\
\mathcal{B}_{23,IV} &= \frac{1}{h^{d-1}}\left[C_{\mathcal{K}} \int \langle \widetilde{\pi}_2(v,\mathbf{u}), \pi_3(v,\mathbf{u})\rangle f_{V\mathbf{U}}^2(v,\mathbf{u})\, dvd\mathbf{u}\,\{1+o(1)\}\right].
\end{aligned}$$

Proof. By Lemma 4.B.5, it follows

$$\begin{aligned}
\mathcal{B}_{23,I} &= E\left[\langle E\left[\widetilde{\pi}_2(V_1,\mathbf{U}_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_1)\right|\mathbf{U}_3\right],\\
&\quad E\left[\pi_3(V_1,\mathbf{U}_1) W_h(V_3-V_1)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_1)\right| V_3,\mathbf{U}_3]\rangle]\\
&= \int \Big\langle \pi_2(\mathbf{y}) f_{\mathbf{U}}(\mathbf{y}) + h^P \widetilde{S}_{\mathcal{K}}^*(\mathbf{y}) + \widetilde{\beta}_{\mathcal{K}}^*(h,\mathbf{y}),\\
&\quad \pi_3(y,\mathbf{y}) f_{V\mathbf{U}}(y,\mathbf{y}) + h^P \widetilde{S}_{W\mathcal{K}}(y,\mathbf{y}) + \widetilde{\beta}_{W\mathcal{K}}(h,(y,\mathbf{y}))\Big\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y}\\
&= \int \langle \pi_2(\mathbf{y}), \pi_3(y,\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y}) f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y} + O\left(h^P\right).
\end{aligned}$$

Also, by Lemmas (4.B.1) and (4.B.5), we have

$$\begin{aligned}
\mathcal{B}_{23,III} &= E\left[\langle \widetilde{\pi}_2(V_1,\mathbf{U}_1), \pi_3(V_2,\mathbf{U}_2)\rangle \mathcal{K}_h(\mathbf{U}_2-\mathbf{U}_1) W_h(V_3-V_2)\mathcal{K}_h(\mathbf{U}_3-\mathbf{U}_2)\right]\\
&= \int \langle \widetilde{\pi}_2(x,\mathbf{x}), \pi_3(y,\mathbf{y})\rangle \mathcal{K}_h(\mathbf{y}-\mathbf{x}) \times\\
&\quad W_h(z-y)\mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{V\mathbf{U}}(x,\mathbf{x}) f_{V\mathbf{U}}(y,\mathbf{y}) f_{V\mathbf{U}}(z,\mathbf{z})\, dxd\mathbf{x}dyd\mathbf{y}dzd\mathbf{z}\\
&= \int \Big\langle \int \widetilde{\pi}_2(x,\mathbf{x})\mathcal{K}_h(\mathbf{y}-\mathbf{x}) f_{V\mathbf{U}}(x,\mathbf{x})\, dxd\mathbf{x},\\
&\quad \int \pi_3(y,\mathbf{y}) W_h(z-y)\mathcal{K}_h(\mathbf{z}-\mathbf{y}) f_{V\mathbf{U}}(z,\mathbf{z})\, dzd\mathbf{z}\Big\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y}\\
&= \int \Big\langle \pi_2(\mathbf{y}) f_{\mathbf{U}}(\mathbf{y}) + h^P \widetilde{S}_{\mathcal{K}}^*(\mathbf{y}) + \widetilde{\beta}_{\mathcal{K}}^*(h,\mathbf{y}),\\
&\quad \pi_3(y,\mathbf{y})\left[f_{V\mathbf{U}}(y,\mathbf{y}) + h^P \widetilde{S}_{W\mathcal{K}}(y,\mathbf{y}) + \beta_{W\mathcal{K}}(h,(y,\mathbf{y}))\right]\Big\rangle f_{V\mathbf{U}}(y,\mathbf{y})\, dyd\mathbf{y}\\
&= \int \langle \pi_2(\mathbf{y}), \pi_3(y,\mathbf{y})\rangle f_{\mathbf{U}}(\mathbf{y}) f_{V\mathbf{U}}^2(y,\mathbf{y})\, dyd\mathbf{y} + O\left(h^P\right).
\end{aligned}$$

Finally, after a change of variables $c = (z - x) h^{-1}$ and $\mathbf{c} = (\mathbf{z} - \mathbf{x}) h^{-1}$, we have

$$\begin{aligned}
\mathcal{B}_{23,IV} &= E\left[\langle \widetilde{\pi}_2 (V_1, \mathbf{U}_1), \pi_3 (V_2, \mathbf{U}_2)\rangle W_h (V_1 - V_2) \left\|\mathcal{K}_h (\mathbf{U}_1 - \mathbf{U}_2)\right\|^2\right] \\
&= \int \langle \widetilde{\pi}_2 (x, \mathbf{x}), \pi_3 (z, \mathbf{z})\rangle W_h (z - x) \left\|\mathcal{K}_h (\mathbf{z} - \mathbf{x})\right\|^2 f_{V\mathbf{U}} (x, \mathbf{x}) f_{V\mathbf{U}} (z, \mathbf{z})\, dx d\mathbf{x} dz d\mathbf{z} \\
&= \frac{1}{h^{d-1}} \int \langle \widetilde{\pi}_2 (x, \mathbf{x}), \pi_3 (x + ch, \mathbf{x} + \mathbf{c}h)\rangle \times \\
&\qquad f_{V\mathbf{U}} (x + ch, \mathbf{x} + \mathbf{c}h) f_{V\mathbf{U}} (x, \mathbf{x}) W (c) \left\|\mathcal{K} (\mathbf{c})\right\|^2 dx d\mathbf{x} dc d\mathbf{c} \\
&= \frac{1}{h^{d-1}} \int \langle \widetilde{\pi}_2 (x, \mathbf{x}), \pi_3 (x, \mathbf{x})\rangle f_{V\mathbf{U}}^2 (x, \mathbf{x})\, dx d\mathbf{x} \int \left\|\mathcal{K} (\mathbf{c})\right\|^2 d\mathbf{c} \left\{1 + O\left(h^P\right)\right\},
\end{aligned}$$

as required. ∎

Lemma 4.B.12 *Let Assumptions* (A1)–(A3) *hold. Then*

$$\begin{aligned}
\mathcal{B}_{25,I} &= O\left(h^{-2(d-1)}\right), \\
\mathcal{B}_{25,II} &= O\left(h^{-(d-1)}\right), \\
\mathcal{B}_{25,III} &= O\left(N h^{-2(d-1)}\right) + O\left(N^2\right) + O\left(N^2 h^P\right), \\
\mathcal{B}_{25,IV} &= O\left(N h^{-(d-1)}\right) + O\left(N^2\right) + O\left(N^2 h^P\right), \\
\mathcal{B}_{25,V} &= O\left(N h^{-(d-1)}\right) + O\left(N^2\right) + O\left(N^2 h^P\right), \\
\mathcal{B}_{25,VI} &= O\left(N h^{-(d-1)}\right) + O\left(N^2\right) + O\left(N^2 h^P\right).
\end{aligned}$$

Proof. Firstly,

$$\begin{aligned}
\mathcal{B}_{25,I} &= E\left[\langle \widetilde{\pi}_2 (V_1, \mathbf{U}_1), \pi_5 (V_2, \mathbf{U}_2)\rangle W_h (V_1 - V_2) \mathcal{K}_h^3 (\mathbf{U}_2 - \mathbf{U}_1)\right] \\
&= \int \langle \widetilde{\pi}_2 (x, \mathbf{x}), \pi_5 (y, \mathbf{y})\rangle W_h (y - x) \mathcal{K}_h^3 (\mathbf{y} - \mathbf{x}) f_{V\mathbf{U}} (x, \mathbf{x}) f_{V\mathbf{U}} (y, \mathbf{y})\, dx d\mathbf{x} dy d\mathbf{y} \\
&= \frac{1}{h^{2(d-1)}} \int \langle \widetilde{\pi}_2 (x, \mathbf{x}), \pi_5 (x + ch, \mathbf{x} + \mathbf{c}h)\rangle \\
&\qquad \times W (c) K^3 (\mathbf{c}) f_{V\mathbf{U}} (x, \mathbf{x}) f_{V\mathbf{U}} (x + ch, \mathbf{x} + \mathbf{c}h)\, dc d\mathbf{c} dx d\mathbf{x} \\
&= O\left(h^{-2(d-1)}\right), \text{ by a further Taylor–series expansion.}
\end{aligned}$$

It follows from Lemmas 4.B.1, 4.B.2, 4.B.3, and 4.B.4 that all other terms have the following orders of magnitude:

$$
\begin{aligned}
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}^{2}W_{h;32}\mathcal{K}_{h;32}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}^{2}E\left[W_{h;32}\mathcal{K}_{h;32}\right|V_2,\mathbf{U}_2]\right]\\
&=O\left(h^{-(d-1)}\right),\\
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}^{2}\mathcal{K}_{h;32}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}^{2}E\left[\mathcal{K}_{h;32}\right|V_2,\mathbf{U}_2]\right]\\
&=O\left(h^{-(d-1)}\right),\\
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}W_{h;32}\mathcal{K}_{h;32}^{3}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}E\left[W_{h;32}\mathcal{K}_{h;32}^{3}\right|V_2,\mathbf{U}_2]\right]\\
&=O\left(h^{-2(d-1)}\right),\\
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}\mathcal{K}_{h;32}W_{h;42}\mathcal{K}_{h;42}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}E\left[\mathcal{K}_{h;32}\right|V_2,\mathbf{U}_2]\,E\left[W_{h;42}\mathcal{K}_{h;42}\right|V_2,\mathbf{U}_2]\right]\\
&=O(1)+O\left(h^{P}\right),\\
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}\mathcal{K}_{h;31}W_{h;32}\mathcal{K}_{h;32}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}E\left[\mathcal{K}_{h;31}W_{h;32}\mathcal{K}_{h;32}\right|(V_1,\mathbf{U}_1),(V_2,\mathbf{U}_2)]\right]\\
&=O\left(h^{-(d-1)}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}\mathcal{K}_{h;31}W_{h;42}\mathcal{K}_{h;42}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,\mathcal{K}_{h;12}E\left[\mathcal{K}_{h;31}\right|V_1,\mathbf{U}_1]\,E\left[W_{h;42}\mathcal{K}_{h;42}\right|V_2,\mathbf{U}_2]\right]\\
&=O(1)+O\left(h^{P}\right),
\end{aligned}
$$

$$
\begin{aligned}
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}\mathcal{K}_{h;31}\mathcal{K}_{h;32}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}E\left[\mathcal{K}_{h;31}\mathcal{K}_{h;32}\right|(V_1,\mathbf{U}_1),(V_2,\mathbf{U}_2)]\right]\\
&=O\left(h^{-(d-1)}\right),\text{ and}
\end{aligned}
$$

$$
\begin{aligned}
&E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}\mathcal{K}_{h;31}\mathcal{K}_{h;32}\right]\\
&=E\left[\langle\widetilde{\pi}_{2;1},\pi_{5;2}\rangle\,W_{h;12}\mathcal{K}_{h;12}E\left[\mathcal{K}_{h;31}\right|V_1,\mathbf{U}_1]\,E\left[\mathcal{K}_{h;42}\right|V_2,\mathbf{U}_2]\right]\\
&=O(1)+O\left(h^{P}\right),
\end{aligned}
$$

as needed. ■

Lemma 4.B.13 *Let Assumptions* (A1)-(A3) *hold. Then*

$$\mathcal{B}_{25,VII} = \left\langle E[\widetilde{\zeta}_{21}], E[\zeta_{51}]\right\rangle + O\left(N^{-2}h^{-(d-1)}\right) + o\left(N^{-1}\right).$$

Proof. Firstly,

$$\begin{aligned}
&\mathcal{B}_{25,VII} \\
&= \frac{1}{N(N-1)^2} E\left[\left\langle \widetilde{\pi}_{2;1}\textstyle\sum_{t=3}^N \mathcal{K}_{h;t1}, \pi_{5;2}\left\|\textstyle\sum_{t=3}^N W_{h;t2}\mathcal{K}_{h;t2}\right\|^2\right\rangle\right] \\
&= \frac{1}{N^2} E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle \mathcal{K}_{h;31} W_{h;32}^2 \mathcal{K}_{h;32}^2\right] \\
&+ \frac{2}{N} E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle \mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} W_{h;42}\mathcal{K}_{h;42}\right] \\
&+ \frac{1}{N}\left\langle E\left[\widetilde{\pi}_{2;1}\mathcal{K}_{h;31}\right], E\left[\pi_{5;2} W_{h;42}^2\mathcal{K}_{h;42}^2\right]\right\rangle && (4.B.15) \\
&+ \left\langle E\left[\widetilde{\pi}_{2;1}\mathcal{K}_{h;31}\right], E\left[\pi_{5;2} W_{h;42}\mathcal{K}_{h;42} W_{h;52}\mathcal{K}_{h;52}\right]\right\rangle. && (4.B.16)
\end{aligned}$$

Now, it follows from Lemma 4.B.3, that

$$\begin{aligned}
&E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle \mathcal{K}_{h;31} W_{h;32}^2\mathcal{K}_{h;32}^2\right] \\
&= E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle E\left[\mathcal{K}_{h;31} W_{h;32}^2\mathcal{K}_{h;32}^2 \middle| (V_1, \mathbf{U}_1), (V_2, \mathbf{U}_2)\right]\right] \\
&= O\left(h^{-(d-1)}\right), \text{ and} \\
&E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle \mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} W_{h;42}\mathcal{K}_{h;42}\right] \\
&= E\left[\langle \widetilde{\pi}_{2;1}, \pi_{5;2}\rangle E\left[\mathcal{K}_{h;31} W_{h;32}\mathcal{K}_{h;32} \middle| (V_1, \mathbf{U}_1), (V_2, \mathbf{U}_2)\right]\right. \\
&\left.\times E\left[W_{h;42}\mathcal{K}_{h;42} \middle| (V_2, \mathbf{U}_2)\right]\right] \\
&= O(h).
\end{aligned}$$

The result follows after noticing that $\langle\, E[\widetilde{\zeta}_{21}], E[\zeta_{51}]\,\rangle = (4.B.15) + (4.B.16)$. ∎

Lemma 4.B.14 *Let Assumptions* (A1)-(A3), (A4), (A6) *and* (A7) *hold, then*

$$E\left[\|p_3(\mathbf{t}_{3\tau1}, \mathbf{t}_{3\tau2}; \Delta h_0) - p_3(\mathbf{t}_{3\tau1}, \mathbf{t}_{3\tau2}; h_0)\|^2 / \left[h_0^P\left(1-\Delta^P\right)\right]^2\right] = o(N), \qquad (4.B.17)$$

$$E\left[\|p_2(\mathbf{t}_{2\tau1}, \mathbf{t}_{2\tau2}; \Delta h_0) - p_2(\mathbf{t}_{2\tau1}, \mathbf{t}_{2\tau2}; h_0)\|^2 / \left[h_0^P\left(1-\Delta^P\right)\right]^2\right] = o(N). \qquad (4.B.18)$$

Proof. Recall $\mathbf{t}_{3\tau 1}^{\top} = \left(\varpi_{3\tau 1}^{\top}, V_1, \mathbf{U}_1^{\top}\right)$, where $\varpi_{3\tau 1} = \varpi_{31} a_\tau \left(V_1, \mathbf{U}_1\right)$, and define $\varrho_{\epsilon\tau}\left(v, \mathbf{u}\right) \equiv E\left[\left\|\varpi_{3\tau 1}\right\|^{\epsilon} \middle| V_1 = v, \mathbf{U}_1 = \mathbf{u}\right]$ for $\epsilon = 1, 2, 3, 4$. Then

$$
\begin{aligned}
& E\left[\left\|p_3\left(\mathbf{t}_{3\tau 1}, \mathbf{t}_{3\tau 2}; \Delta h_0\right) - p_3\left(\mathbf{t}_{3\tau 1}, \mathbf{t}_{3\tau 2}; h_0\right)\right\|^2 / \left[h_0^P\left(1 - \Delta^P\right)\right]^2\right] \\
& = E\left[\left(\frac{c_\Delta}{h_0^{2P}}\right)\left\|\frac{\varpi_{3\tau 1} + \varpi_{3\tau 2}}{2h_0^d}\right\|^2 W^2\left(\frac{V_1 - V_2}{h_0}\right)\mathcal{K}^2\left(\frac{\mathbf{U}_1 - \mathbf{U}_2}{h_0}\right)\right] \\
& = \int\left(\frac{c_\Delta}{4h_0^{2P+d}}\right) f_{V\mathbf{U}}\left(z + ch_0, \mathbf{z} + \mathbf{c}h_0\right) f_{V\mathbf{U}}\left(z, \mathbf{z}\right) \times \\
& \left[\varrho_{2\tau}\left(z + ch_0, \mathbf{z} + \mathbf{c}h_0\right) + \varrho_{2\tau}\left(z, \mathbf{z}\right) + 2\left\langle\varrho_{0\tau}\left(z + ch_0, \mathbf{z} + \mathbf{c}h_0\right), \varrho_{0\tau}\left(z, \mathbf{z}\right)\right\rangle\right] \times \\
& W^2\left(c\right)\mathcal{K}^2\left(\mathbf{c}\right) dz dc d\mathbf{z} d\mathbf{c} \\
& = O(h_0^{-(2P+d)}) = O(N(Nh_0^{2P+d})^{-1}) = o\left(N\right),
\end{aligned}
$$

where $c_\Delta = \left(1 - \Delta^2\right) / \left(\Delta\left(1 - \Delta^P\right)\right)^2$, and $\varrho_{0\tau}\left(v, \mathbf{u}\right) \equiv E\left[\varpi_{3\tau 1} \middle| V_1 = v, \mathbf{U}_1 = \mathbf{u}\right]$. The second equality uses the change of variables from $(y, \mathbf{y}^{\top}, z, \mathbf{z}^{\top})$ to $(c = h_0^{-1}\left(y - z\right), \mathbf{c}^{\top} = h_0^{-1}(\mathbf{y} - \mathbf{z})^{\top}, z, \mathbf{z}^{\top})$ with jacobian h_0^d. This change of variables is not affected by boundary effects because of Assumptions (A1) and (A2), and the fact that $a_\tau\left(z, \mathbf{z}\right) = 0$ for all $\left(z, \mathbf{z}\right)$ within a distance τ of the boundary of $\Omega_{V\mathbf{U}}$, with $h_0/\tau \to 0$. The last equality uses the continuity of the $\varrho_{\epsilon\tau}$'s and Assumption (A7). (4.B.18) follows the exact same arguments and therefore it is omitted. ■

4.C Tables & Figures

Table 4.1: Monte Carlo results for Design 1: Bandwidth Estimation

		$N = 200$		$N = 400$		$N = 600$	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
$\rho = 0$	$h_{\mathbf{opt}}$	0.6530	–	0.5491	–	0.4962	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7006	0.0583	0.5654	0.0228	0.5024	0.0151
	$\widehat{h}_{\mathbf{opt};R2}$	0.6994	0.0628	0.5651	0.0227	0.5024	0.0144
	$\widehat{h}_{\mathbf{opt};R3}$	0.6877	0.0311	0.5626	0.0146	0.5019	0.0104
	$\widehat{h}_{R1}$	0.5955	0.0220	0.5317	0.0135	0.4968	0.0102
	$\widehat{h}_{R2}$	0.6055	0.0395	0.5450	0.0237	0.5100	0.0178
	$\widehat{h}_{R3}$	0.6846	0.0252	0.6110	0.0155	0.5706	0.0117
$\rho = 1/4$	$h_{\mathbf{opt}}$	0.6664	–	0.5603	–	0.5063	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7024	0.0802	0.5698	0.0354	0.5059	0.0192
	$\widehat{h}_{\mathbf{opt};R2}$	0.7029	0.1012	0.5693	0.0399	0.5058	0.0195
	$\widehat{h}_{\mathbf{opt};R3}$	0.6872	0.0399	0.5636	0.0190	0.5033	0.0124
	$\widehat{h}_{R1}$	0.5918	0.0210	0.5283	0.0133	0.4938	0.0102
	$\widehat{h}_{R2}$	0.5931	0.0385	0.5336	0.0237	0.5003	0.0181
	$\widehat{h}_{R3}$	0.6823	0.0243	0.6083	0.0153	0.5686	0.0118

[a] Means and standard deviations (SD) are based on 2000 replications.

[b] Estimated bandwidths: $\widehat{h}_{\mathbf{opt};R1}$, $\widehat{h}_{\mathbf{opt};R2}$, and $\widehat{h}_{\mathbf{opt};R3}$ were calculated by setting $h_0 = \widehat{h}_{R1}N^{1/12}$, $h_0 = \widehat{h}_{R2}N^{1/12}$, and $h_0 = \widehat{h}_{R3}N^{1/12}$ in Section 4.3, respectively. Similarly, auxiliary bandwidths were set $h_* = \widehat{h}_{R1}$, $h_* = \widehat{h}_{R2}$, and $h_* = \widehat{h}_{R3}$ respectively.

Table 4.2: Monte Carlo results for Design 1: Parameter η

		$N=200$			$N=400$			$N=600$		
		Bias	*SD*	*MSE*	*Bias*	*SD*	*MSE*	*Bias*	*SD*	*MSE*
$\rho=0$	$\widetilde{\eta}(h_{\mathbf{opt}})$	0.8092	0.1598	0.6548	0.6593	0.0953	0.4347	0.5882	0.0731	0.3460
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7932	0.1594	0.6291	0.6583	0.0961	0.4333	0.5878	0.0741	0.3455
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7924	0.1595	0.6279	0.6585	0.0962	0.4336	0.5880	0.0741	0.3458
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7980	0.1576	0.6368	0.6596	0.0959	0.4350	0.5884	0.0740	0.3463
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8378	0.1658	0.7019	0.6616	0.0955	0.4377	0.5896	0.0737	0.3477
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8387	0.1641	0.7034	0.6637	0.0960	0.4405	0.5924	0.0747	0.3510
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.8124	0.1575	0.6600	0.6700	0.0978	0.4489	0.6062	0.0764	0.3675
$\rho=1/4$	$\widetilde{\eta}(h_{\mathbf{opt}})$	0.7816	0.1650	0.6108	0.6395	0.1016	0.4090	0.5751	0.0773	0.3307
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7688	0.1641	0.5911	0.6357	0.1014	0.4041	0.5738	0.0780	0.3292
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7688	0.1725	0.5911	0.6355	0.1018	0.4039	0.5737	0.0780	0.3291
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7720	0.1591	0.5959	0.6371	0.1010	0.4059	0.5740	0.0778	0.3295
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8166	0.1855	0.6669	0.6453	0.1037	0.4164	0.5759	0.0776	0.3317
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8230	0.1989	0.6774	0.6468	0.1047	0.4183	0.5783	0.0780	0.3344
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.7851	0.1628	0.6164	0.6436	0.1018	0.4142	0.5868	0.0794	0.3444

[a] Simulated biases, standard deviations, and average Mean Squared Error (MSE) are based on 2000 replications.

Table 4.3: Monte Carlo results for Design 2: Bandwidth Estimation

		$N = 200$		$N = 400$		$N = 600$	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
$\rho = 0$	$h_{\mathbf{opt}}$	0.6595	–	0.5545	–	0.5011	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7027	0.0826	0.5671	0.0331	0.5032	0.0175
	$\widehat{h}_{\mathbf{opt};R2}$	0.7011	0.0951	0.5665	0.0343	0.5031	0.0166
	$\widehat{h}_{\mathbf{opt};R3}$	0.6883	0.0408	0.5630	0.0186	0.5021	0.0112
	$\widehat{h}_{R1}$	0.5961	0.0214	0.5317	0.0134	0.4971	0.0103
	$\widehat{h}_{R2}$	0.6053	0.0398	0.5445	0.0244	0.5107	0.0185
	$\widehat{h}_{R3}$	0.6853	0.0246	0.6109	0.0154	0.5710	0.0118
$\rho = 1/4$	$h_{\mathbf{opt}}$	0.6755	–	0.5680	–	0.5133	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7090	0.1729	0.5722	0.0652	0.5084	0.0299
	$\widehat{h}_{\mathbf{opt};R2}$	0.7100	0.2604	0.5719	0.0724	0.5082	0.0322
	$\widehat{h}_{\mathbf{opt};R3}$	0.6906	0.0712	0.5655	0.0324	0.5050	0.0164
	$\widehat{h}_{R1}$	0.5917	0.0216	0.5276	0.0132	0.4934	0.0101
	$\widehat{h}_{R2}$	0.5924	0.0407	0.5332	0.0246	0.4996	0.0184
	$\widehat{h}_{R3}$	0.6820	0.0249	0.6071	0.0153	0.5681	0.0116

[a] Means and standard deviations (SD) are based on 2000 replications.

[b] Estimated bandwidths: $\widehat{h}_{\mathbf{opt};R1}$, $\widehat{h}_{\mathbf{opt};R2}$, and $\widehat{h}_{\mathbf{opt};R3}$ were calculated by setting $h_0 = \widehat{h}_{R1}N^{1/12}$, $h_0 = \widehat{h}_{R2}N^{1/12}$, and $h_0 = \widehat{h}_{R3}N^{1/12}$ in Section 4.3, respectively. Similarly, auxiliary bandwidths were set $h_* = \widehat{h}_{R1}$, $h_* = \widehat{h}_{R2}$, and $h_* = \widehat{h}_{R3}$ respectively.

Table 4.4: Monte Carlo results for Design 2: Parameter η

		$N=200$			$N=400$			$N=600$		
		Bias	*SD*	*MSE*	*Bias*	*SD*	*MSE*	*Bias*	*SD*	*MSE*
$\rho=0$	$\widetilde{\eta}(h_{\mathbf{opt}})$	0.8048	0.1655	0.6477	0.6640	0.0977	0.4409	0.5925	0.0729	0.3511
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7952	0.1638	0.6323	0.6621	0.0984	0.4383	0.5920	0.0737	0.3504
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7955	0.1643	0.6329	0.6622	0.0984	0.4385	0.5918	0.0737	0.3503
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7978	0.1601	0.6365	0.6634	0.0976	0.4401	0.5919	0.0736	0.3503
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8336	0.1814	0.6950	0.6667	0.0982	0.4445	0.5932	0.0734	0.3519
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8362	0.1852	0.6993	0.6682	0.0988	0.4465	0.5965	0.0744	0.3559
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.8098	0.1625	0.6558	0.6732	0.0991	0.4532	0.6100	0.0761	0.3721
$\rho=1/4$	$\widetilde{\eta}(h_{\mathbf{opt}})$	0.7750	0.1804	0.6006	0.6410	0.1093	0.4109	0.5749	0.0779	0.3305
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7630	0.2567	0.5822	0.6363	0.1113	0.4049	0.5719	0.0778	0.3271
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7646	0.4242	0.5846	0.6365	0.1129	0.4052	0.5721	0.0779	0.3273
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7644	0.1704	0.5843	0.6375	0.1062	0.4064	0.5724	0.0775	0.3276
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8187	0.2796	0.6702	0.6470	0.1168	0.4186	0.5748	0.0787	0.3304
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8274	0.4854	0.6847	0.6486	0.1192	0.4207	0.5762	0.0795	0.3320
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.7811	0.1797	0.6101	0.6461	0.1082	0.4174	0.5844	0.0799	0.3416

[a] Simulated biases, standard deviations, and average Mean Squared Error (MSE) are based on 2000 replications.

Table 4.5: Monte Carlo results for Design 3: Bandwidth Estimation

		$N=200$		$N=400$		$N=600$	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
$\rho=0$	$h_{\mathbf{opt}}$	0.6834	–	0.5747	–	0.5193	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7090	0.1552	0.5706	0.0672	0.5065	0.0395
	$\widehat{h}_{\mathbf{opt};R2}$	0.7070	0.1991	0.5694	0.0692	0.5061	0.0376
	$\widehat{h}_{\mathbf{opt};R3}$	0.6904	0.0694	0.5648	0.0311	0.5037	0.0200
	$\widehat{h}_{R1}$	0.5947	0.0213	0.5309	0.0133	0.4961	0.0100
	$\widehat{h}_{R2}$	0.6040	0.0444	0.5439	0.0264	0.5096	0.0197
	$\widehat{h}_{R3}$	0.6837	0.0245	0.6102	0.0153	0.5699	0.0115
$\rho=1/4$	$h_{\mathbf{opt}}$	0.7101	–	0.5971	–	0.5396	–
	$\widehat{h}_{\mathbf{opt};R1}$	0.7169	0.4016	0.5797	0.1363	0.5138	0.0955
	$\widehat{h}_{\mathbf{opt};R2}$	0.7183	0.8889	0.5797	0.1708	0.5134	0.1141
	$\widehat{h}_{\mathbf{opt};R3}$	0.6937	0.1440	0.5699	0.0624	0.5085	0.0433
	$\widehat{h}_{R1}$	0.5919	0.0208	0.5273	0.0134	0.4931	0.0102
	$\widehat{h}_{R2}$	0.5910	0.0451	0.5305	0.0283	0.4972	0.0212
	$\widehat{h}_{R3}$	0.6816	0.0239	0.6068	0.0154	0.5675	0.0118

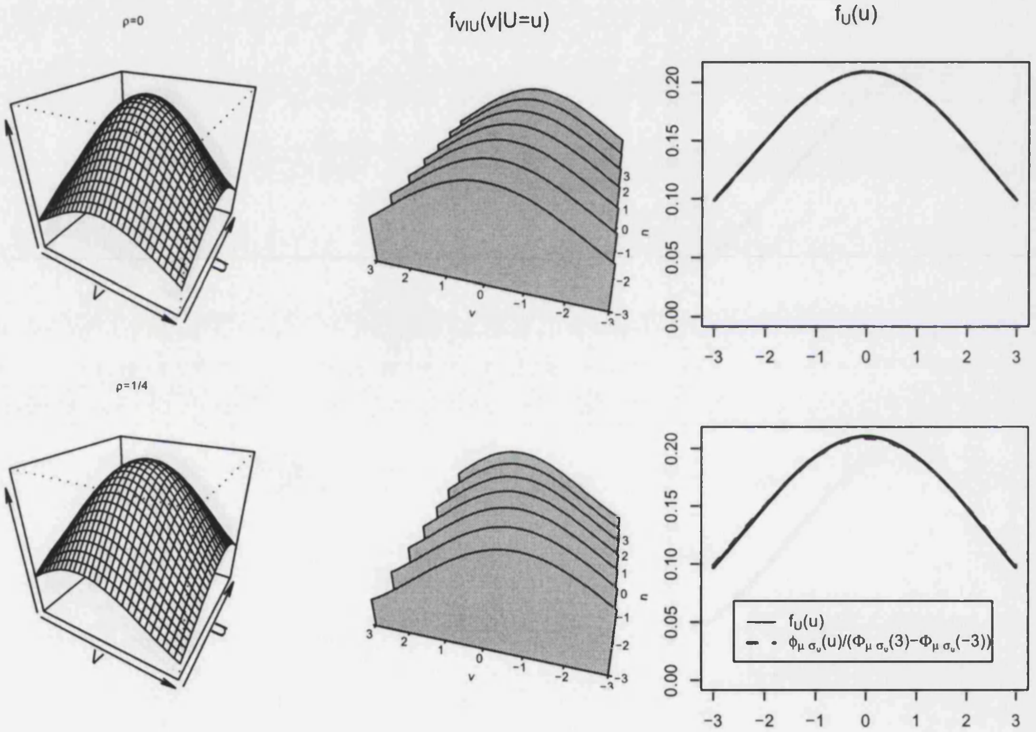
[a] Means and standard deviations (SD) are based on 2000 replications.
[b] Estimated bandwidths: $\widehat{h}_{\mathbf{opt};R1}$, $\widehat{h}_{\mathbf{opt};R2}$, and $\widehat{h}_{\mathbf{opt};R3}$ were calculated by setting $h_0 = \widehat{h}_{R1}N^{1/12}$, $h_0 = \widehat{h}_{R2}N^{1/12}$, and $h_0 = \widehat{h}_{R3}N^{1/12}$ in Section 4.3, respectively. Similarly, auxiliary bandwidths were set $h_* = \widehat{h}_{R1}$, $h_* = \widehat{h}_{R2}$, and $h_* = \widehat{h}_{R3}$ respectively.

Table 4.6: Monte Carlo results for Design 3: Parameter η

		$N = 200$			$N = 400$			$N = 600$		
		Bias	*SD*	*MSE*	*Bias*	*SD*	*MSE*	*Bias*	*SD*	*MSE*
$\rho = 0$	$\widetilde{\eta}\,(h_{\mathbf{opt}})$	0.7943	0.1692	0.6310	0.6539	0.1008	0.4276	0.5891	0.0756	0.3470
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7863	0.2013	0.6183	0.6515	0.1048	0.4244	0.5855	0.0779	0.3428
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7879	0.2506	0.6208	0.6515	0.1038	0.4244	0.5855	0.0771	0.3428
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7858	0.1639	0.6175	0.6510	0.1001	0.4238	0.5857	0.0759	0.3430
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8282	0.2195	0.6859	0.6600	0.1051	0.4357	0.5871	0.0764	0.3447
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8334	0.2643	0.6946	0.6602	0.1062	0.4359	0.5902	0.0771	0.3483
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.8020	0.1688	0.6433	0.6612	0.1014	0.4371	0.6027	0.0778	0.3632
$\rho = 1/4$	$\widetilde{\eta}\,(h_{\mathbf{opt}})$	0.7791	0.1957	0.6069	0.6487	0.1101	0.4208	0.5801	0.0818	0.3365
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R1})$	0.7721	0.7231	0.5961	0.6427	0.1535	0.4130	0.5741	0.1041	0.3296
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R2})$	0.7758	1.9588	0.6019	0.6431	0.2037	0.4136	0.5745	0.1269	0.3301
	$\widetilde{\eta}(\widehat{h}_{\mathbf{opt};R3})$	0.7718	0.2156	0.5957	0.6429	0.1079	0.4133	0.5738	0.0801	0.3293
	$\widetilde{\eta}(\widehat{h}_{R1})$	0.8268	0.8472	0.6835	0.6556	0.1488	0.4298	0.5787	0.0963	0.3349
	$\widetilde{\eta}(\widehat{h}_{R2})$	0.8383	0.8563	0.7027	0.6575	0.1869	0.4323	0.5810	0.1095	0.3375
	$\widetilde{\eta}(\widehat{h}_{R3})$	0.7903	0.2415	0.6246	0.6526	0.1115	0.4259	0.5878	0.0815	0.3455

[a] Simulated biases, standard deviations, and average Mean Squared Error (MSE) are based on 2000 replications.

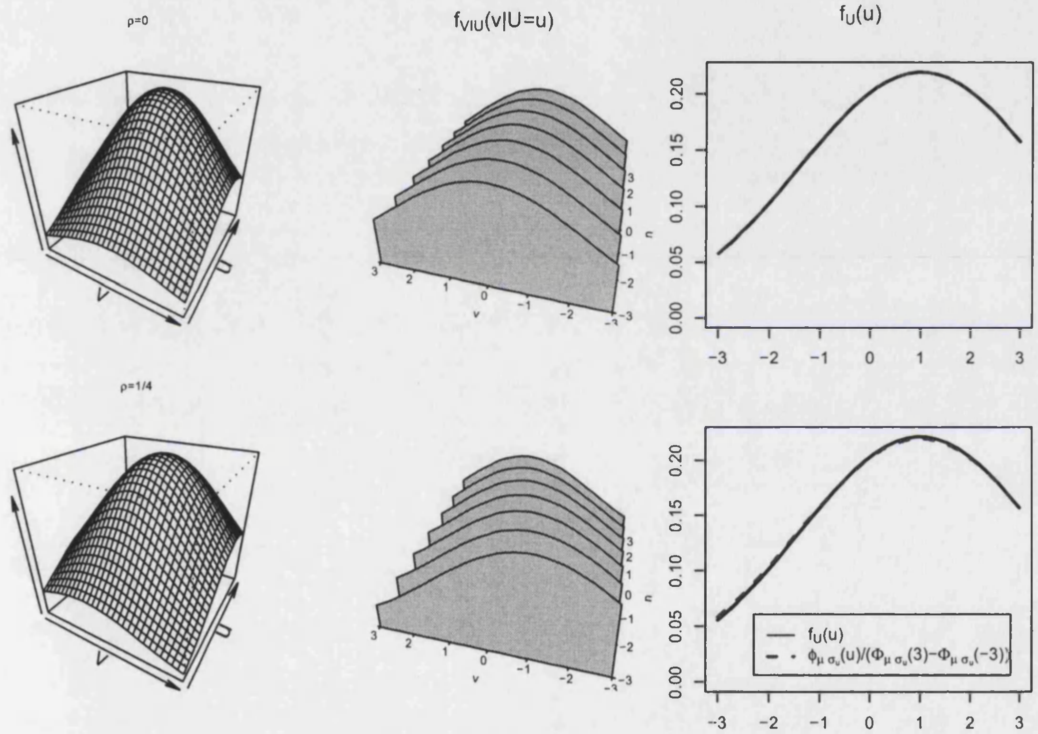
Figure 4.1: Visualization of Design 1



[a] Each row represents a variation of Design 1: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] First column from the left shows their joint densities, $f_{VU}(v, u)$. Middle column shows their associated conditional densities, $f_{V|U}(v|\, U = u)$, and last column shows their marginal distribution, $f_U(u)$, with respect to U, as well as that of a univariate truncated, $[-3, 3]$, normal with parameters: $\mu_u = 0$, and $\sigma_u^2 = 6$.

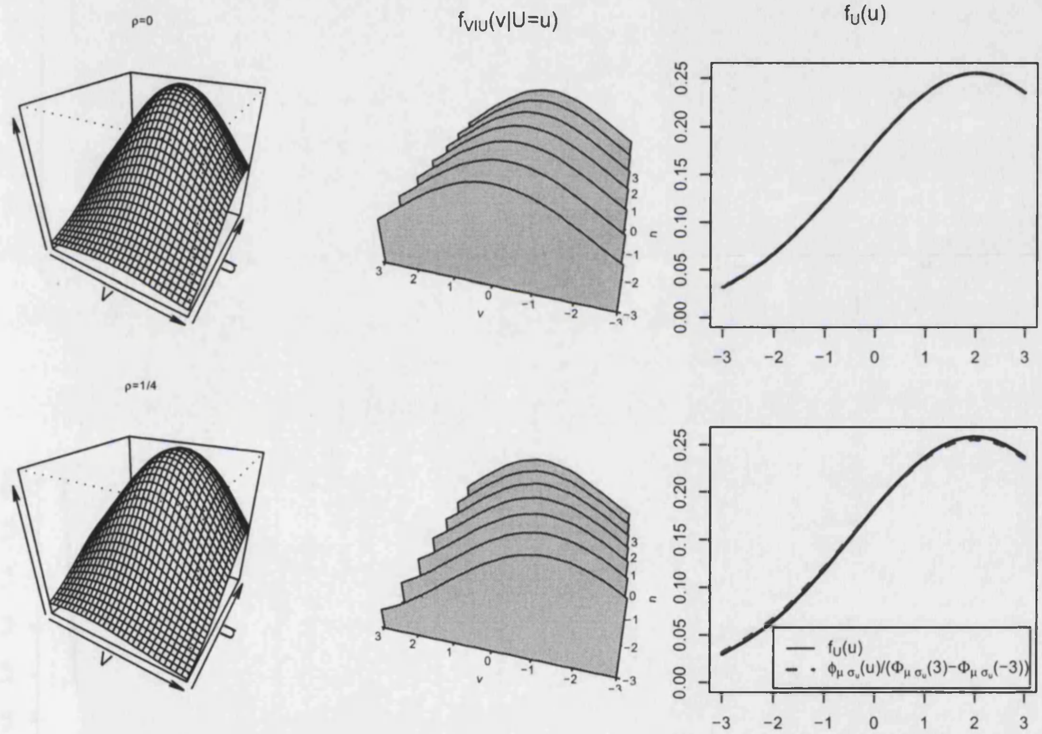
Figure 4.2: Visualization of Design 2



[a] Each row represents a variation of Design 2: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] First column from the left shows their joint densities, $f_{VU}(v, u)$. Middle column shows their associated conditional densities, $f_{V|U}(v|\,U = u)$, and last column shows their marginal distribution, $f_U(u)$, with respect to U, as well as that of a univariate truncated, $[-3, 3]$, normal with parameters: $\mu_u = 1$, and $\sigma_u^2 = 6$.

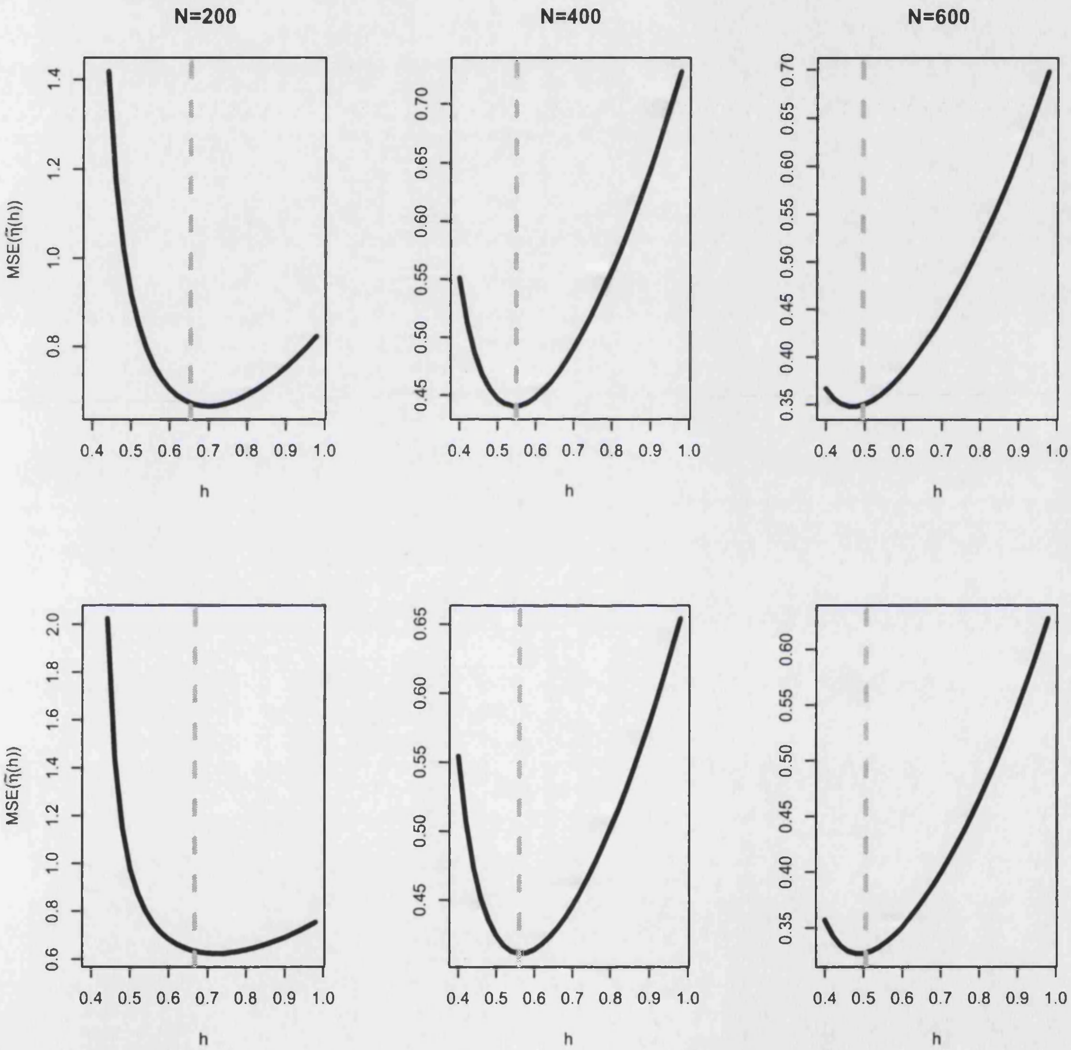
Figure 4.3: Visualization of Design 3



[a] Each row represents a variation of Design 3: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] First column from the left shows their joint densities, $f_{VU}(v, u)$. Middle column shows their associated conditional densities, $f_{V|U}(v|\, U = u)$, and last column shows their marginal distribution, $f_U(u)$, with respect to U, as well as that of a univariate truncated, $[-3, 3]$, normal with parameters: $\mu_u = 2$, and $\sigma_u^2 = 6$.

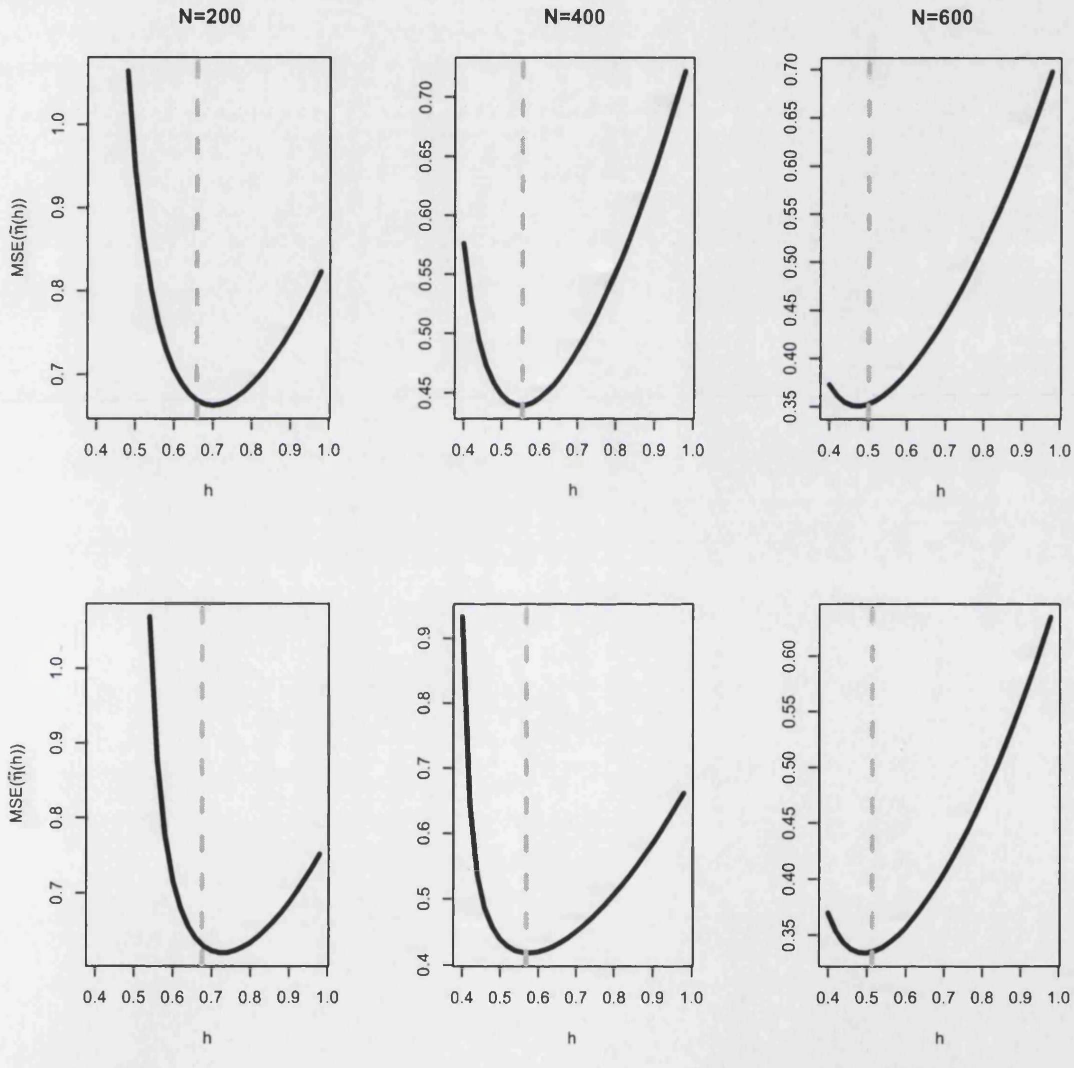
Figure 4.4: Simulated MSE of Design 1



[a] Each row represents a variation of Design 1: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] Simulation based on 1000 replications. Dashed gray lines represent the optimal bandwidth predicted by our results in each case.

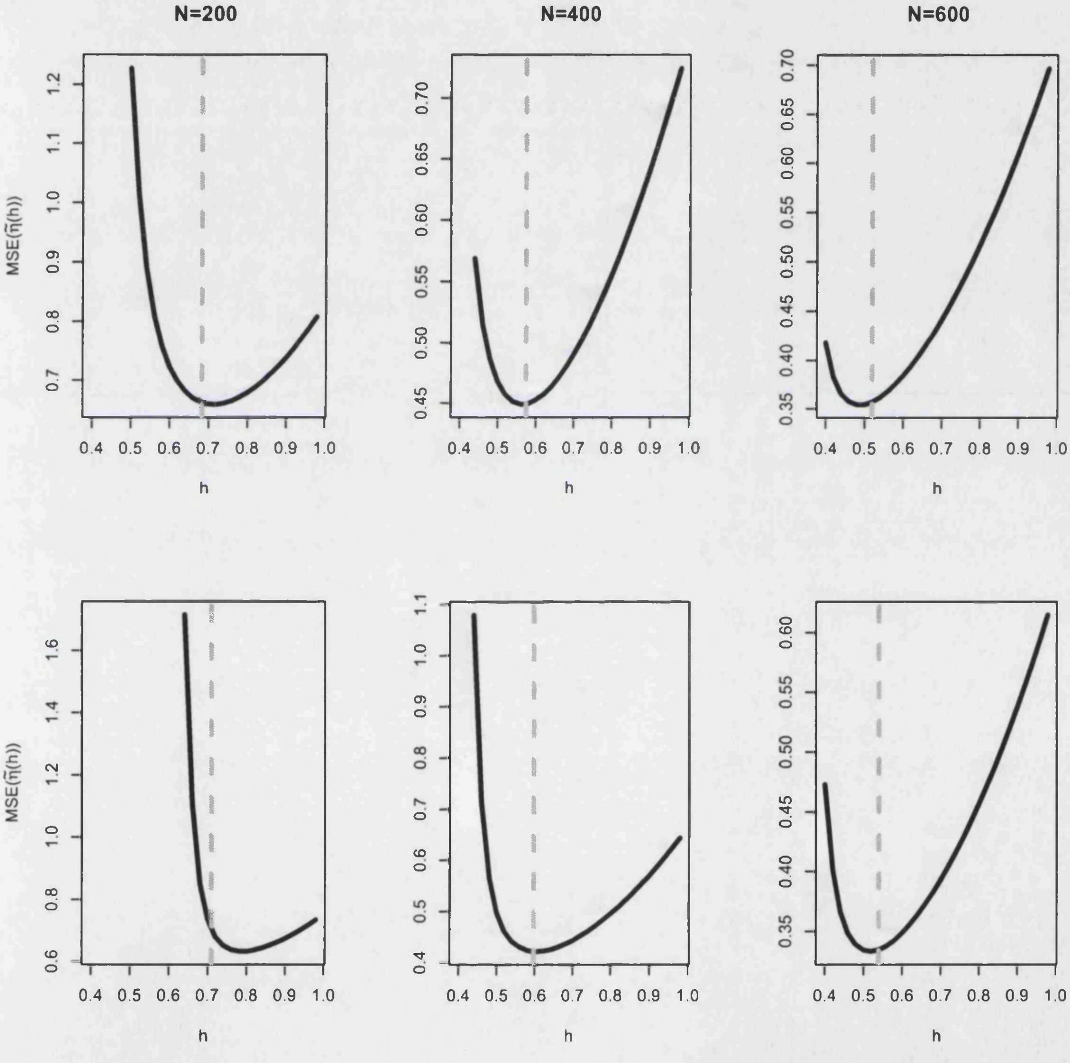
Figure 4.5: Simulated MSE of Design 2



[a] Each row represents a variation of Design 2: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] Simulation based on 1000 replications. Dashed gray lines represent the optimal bandwidth predicted by our results in each case.

Figure 4.6: Simulated MSE of Design 3



[a] Each row represents a variation of Design 3: (a) $\rho = 0$, and (b) $\rho = 1/4$ in descending order.

[b] Simulation based on 1000 replications. Dashed gray lines represent the optimal bandwidth predicted by our results in each case.

Bibliography

AHN, H. (1995): "Nonparametric Two-stage Estimation of Conditional Choice Probabilities in a Binary Choice Model under Uncertainty," *Journal of Econometrics*, 67, 337–378.

——— (1997): "Semiparametric Estimation of a Single-index Model with Nonparametric Generated Regressors," *Econometric Theory*, 13(1), 3–31.

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

ANDREWS, D. W. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

BAIRAM, E. I. (1994): *Homogeneous and Nonhomogeneous Production Functions*. Avebury, Vermont, 1 edn.

BASHTANNYK, D. M., AND R. J. HYNDMAN (2001): "Bandwidth Selection for Kernel Conditional Density Estimation," *Computational Statistics & Data Analysis*, 36, 279–298.

BIERENS, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58(6), 1443–1458.

BILLINGSLEY, P. (1986): *Probability and Measure*, Wiley series in probability and mathematical statistics. Wiley, 2 edn.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econonometrics: Theory and Applications*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, vol. II of *Eighth World Congress*. Cambridge University Press, Cambridge.

BREIMAN, L., AND J. H. FRIEDMAN (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580–598.

BUCKLEY, J., AND I. JAMES (1979): "Linear Regression with Censored Data," *Biometrika*, 66(3), 429–436.

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Semi-parametric Models with Censoring," *Journal of Econometrics*, 32, 189–218.

——— (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60(3), 567–596.

CHEN, X., O. B. LINTON, AND P. M. ROBINSON (2001): "The Estimation of Conditional Densities," in *In: Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas*, ed. by M. L. Puri, pp. 71–84. VSP International Science Publishers, The Netherlands, 1 edn.

CHESHER, A. (2001): "Quantile Driven Identification of Structural Derivatives," Working Paper CWP08/01, Cemmap.

CHIANG, A. (1984): *Fundamental Methods of Mathematical Economics.* McGraw-Hill, New York, 3 edn.

CHRISTENSEN, L. R., D. W. JORGENSON, AND L. J. LAU (1973): "Transcendental Logarithmic Production Frontiers," *The Review of Economics and Statistics*, 55, 28–45.

CHUNG, J. W. (1994): *Utility and Production Functions.* Blackwell Publishers, Massachusetts, 1 edn.

COBB, C. W., AND P. H. DOUGLAS (1928): "Theory of Production," *American Economic Review*, 18, 139–165.

COLLOMB, G., AND W. HÄRDLE (1986): "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: kernel Regression Estimation from Dependent Observations," *Stochastic Processes and their Application*, 23(1), 77–89.

COSSLETT, S. R. (1983): "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765–782.

——— (1987): "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and the Censored Regression Models," *Econometrica*, 55, 559–585.

——— (2004): "Efficient Semiparametric Estimation of Censored and Truncated Regressions via a Smoothed Self-Consistency Equation," *Econometrica*, 72(4), 1277–1293.

CRÉPON, B., F. KRAMARZ, AND A. TROGNON (1998): "Parameter of Interest, Nuisance parameter and Orthogonality Conditions: An Application to Autoregressive Error Component Models," *Journal of Econometrics*, 82(1), 135–156.

DE GOOIJER, J. G., AND D. ZEROM (2003): "On Conditional Density Estimation," *Statistica Neerlandica*, 57(2), 159–176.

DELGADO, M. A., AND J. MORA (1995): "On Asymptotic Inferences in Non-parametric and Semiparametric Models with Discrete and Mixed Regressors," *Investigaciones Económicas*, 19(3), 435–467.

DOUGLAS, P. H. (1967): "Comments on the Cobb-Douglas Production Function," in *The Theory and Empirical Analysis of Production*, ed. by M. Brown, pp. 15–22. Columbia University Press, New York.

EKELAND, I., J. J. HECKMAN, AND L. NESHEIM (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–S109.

FAN, J. (1992): "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87(420), 998–1004.

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, 1 edn.

FAN, J., Q. YAO, AND H. TONG (1996): "Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems," *Biometrika*, 83(1), 189–206.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64(4), 865 – 890.

FERNANDES, M., AND P. K. MONTEIRO (2005): "Central Limit Theorem for Asymmetric Kernel Functionals," *Annals of the Institute of Statistical Mathematics*, 57(3), 425–442.

FRIEDMAN, J. H., AND W. STUTZLE (1981): "Projection Pursuit Regression," *Journal of The American Statistical Association*, 76, 817–823.

GABLER, S., F. LAISNEY, AND M. LECHNER (1993): "Seminonparametric Estimation of Binary Choice Models with an Application to Labor Force Participation," *Journal of Business and Economic Statistics*, 11(1), 61–80.

GASSER, T., H.-G. MÜLLER, AND V. MAMMITZSCH (1985): "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 47, 238–252.

GOH, C. (2004): "Smoothing Choices and Distributional Approximations for Econometric Inference," Ph.D. thesis, University of California, Berkeley.

GOLDMAN, S. M., AND H. UZAWA (1964): "A Note on Separability in Demand Analysis," *Econometrica*, 32(3), 387–398.

GOLDSTEIN, L., AND K. MESSER (1992): "Optimal Plug-in Estimators for Nonparametric Functional Estimation," *The Annals of Statistics*, 20(3), 1306–1328.

GOZALO, P., AND O. B. LINTON (2000): "Local nonlinear Least Squares: Using Parametric Information in Nonparametric Regression," *Journal of Econometrics*, 99(1), 63–106.

GOZALO, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9(3), 451–477.

GOZALO, P. L., AND O. B. LINTON (2001): "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," *Journal of Econometrics*, 104(1), 1–48.

HAHN, J. (1998): "On the Role of The Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66(2), 315–331.

HALL, P., AND J. S. MARRON (1987): "Estimation of Integrated Squared Density Derivatives," *Statistica and Probability Letters*, 6(2), 109–115.

HAMPEL, F. R. (1974): "The Influence Function and its Role in Robust Estimation," *Journal of the American Statistical Association*, 62, 1179–1186.

HANOCH, G., AND M. ROTHSCHILD (1972): "Testing the Assumptions of Production Theory: A Nonparametric Approach," *Journal of Political Economy*, 80, 256–275.

HANSEN, B. E. (2004): "Nonparametric Conditional Density Estimation," Unpublished Manuscript.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029–1054.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*, vol. 9 of *Econometric Society Monographs*. Cambridge University Press, 1 edn.

HÄRDLE, W., J. D. HART, J. S. MARRON, AND A. B. TSYBAKOV (1992): "Bandwidth Choice for Average Derivative Estimation," *Journal of The American Statistical Association*, 87(417), 218–226.

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21(4), 1926–1947.

HÄRDLE, W., AND T. M. STOKER (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

HÄRDLE, W., AND A. B. TSYBAKOV (1993): "How Sensitive are Average Derivatives?," *Journal of Econometrics*, 58, 31–48.

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

HONORÉ, B. E., AND A. LEWBEL (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica*, 70(5), 2053–2063.

HOROWITZ, J. L. (1986): "A Distribution Free Least Squares Estimator for Censored Linear Regression Models," *Journal of Econometrics*, 32, 59–84.

——— (1988): "Semiparametric M-Estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45–83.

——— (1998): *Semiparametric Methods in Econometrics*, vol. 131 of *Lectures Notes in Statistics*. Springer-Verlag New York, Inc, New York, 1 edn.

——— (2001): "Nonparametric Estimation of a Generalized Additive Model with an Unknown Link Function," *Econometrica*, 69, 499–513.

HOROWITZ, J. L., AND W. HÄRDLE (1996): "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates," *Journal of the American Statistical Association*, 91, 1632–1640.

HOROWITZ, J. L., AND E. MAMMEN (2004): "Nonparametric Estimation of an additive Model with a Link Function," *Annals of Statistics*, 36(2), 2412–2443.

HOROWITZ, J. L., AND V. G. SPOKOINY (2001): "An Adaptive, Rate-Optimal Test of a Parametric Model Against a Nonparametric Alternative," *Econometrica*, 69(3), 599–631.

HYNDMAN, R. J., D. M. BASHTANNYK, AND G. K. GRUNWALD (1996): "Estimating and Visualizing Conditional Densities," *Journal of Computational and Graphical Statistics*, 5(4), 315–336.

HYNDMAN, R. J., AND Q. YAO (2002): "Nonparametric Estimation and Symmetry Tests for Conditional Density Functions," *Journal of Nonparametric Statistics*, 14(3), 259–278.

ICHIMURA, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71–120.

ICHIMURA, H., AND O. B. LINTON (2005): "Asymptotic Expansions for some Semiparametric Program Evaluation Estimators," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,*, ed. by D. W. K. Andrews, and J. H. Stock, chap. 8, pp. 149–170. Cambridge University Press, Cambridge, 1 edn.

IHAKA, R., AND R. GENTLEMAN (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5(3), 299–314.

IMBENS, G. W., AND W. K. NEWEY (2002): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper 285, NBER.

JEFFERSON, G., A. G. Z. HU, X. GUAN, AND X. YU (2003): "Ownership, Performance, and Innovation in China's Large- and Medium-size Industrial Enterprise Sector," *China Economic Review*, 14(1), 89–113.

JONES, M. C., AND S. J. SHEATHER (1991): "Using Non-Stochastic Terms to Advantage in Kernel-Based Estimation of Integrated Squared density derivatives," *Statistica and Probability Letters*, 11(6), 511–514.

KHAN, S., AND A. LEWBEL (2006): "Weighted and Two Stage Least Squares Estimation of Semiparametric Truncated Regression Models," Forthcoming in Econometric Theory.

KIM, C.-K., AND T. L. LAI (2000): "Efficient Score Estimation and Adaptive M-Estimators in Censored and Truncated Regression Models," *Statistica Sinica*, 10, 731–749.

KLEIN, R. W., AND R. H. SPADY (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61(2), 387–421.

KONISHI, Y., AND Y. NISHIYAMA (2002): "Nonparametric Test for Translog Specification of Production Function in Japanese Manufacturing Industry," in *Integrated Assessment and Decision Support, Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society*, ed. by A. E. Rizzoli, and A. J. Jakeman, vol. 2, pp. 597–602. iEMSs.

LAI, T. L., AND Z. YING (1992): "Asymptotically Efficient Estimation in Censored and Truncated Regression Models," *Statistica Sinica*, 2, 17–46.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–341.

LEWBEL, A. (1998): "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 66(1), 105–121.

——— (2000a): "Asymptotic Trimming for Bounded Density Plug-in Estimators," Boston College. Unpublished Manuscript.

——— (2000b): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97(1), 145–177.

——— (2006): "Endogenous Selection or Treatment Model Estimation," Forthcoming in Journal of Econometrics.

LEWBEL, A., AND O. B. LINTON (1999): "Nonparametric Censored and Truncated Regression," Sticerd Working Paper EM/00/389, STICERD.

——— (2002): "Nonparametric Censored and Truncated Regression," *Econometrica*, 70(2), 765–779.

——— (2006): "Matching Estimation and Efficient Estimators of Nonparametric Homothetically Separable Functions," Unpublished Manuscript.

LEWBEL, A., AND S. M. SCHENNACH (2005): "A Simple Ordered Data Estimator for Inverse Density Weighted Expectations," Forthcoming in Journal of Econometrics.

LINTON, O. B. (1991): "Edgeworth Approximation in Semiparametric Regression Models," Ph.D. thesis, University of California, Berkeley.

——— (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63(3), 1079–1112.

——— (2000): "Efficient Estimation of Generalized Additive Nonparametric Regression Models," *Econometric Theory*, 16(4), 502–523.

LINTON, O. B., AND W. HÄRDLE (1996): "Estimating Additive Regression with Known Links," *Biometrika*, 83, 529–540.

LINTON, O. B., AND J. P. NIELSEN (1995): "A Kernel Model of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

MAGNAC, T., AND E. MAURIN (2004): "Identification and Information in Monotone Binary Models," Forthcoming in Journal of Econometrics.

MASRY, E. (1996a): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17(6), 571–599.

——— (1996b): "Multivariate Regression Estimation Local Polynomial Fitting for Time Series," *Stochastic Processes and their Application*, 65, 81–101.

MATZKIN, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

MCCULLAGH, P., AND J. A. NELDER (1989): *Generalized Linear Models*, vol. 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 2 edn.

MOON, C.-G. (1989): "A Monte Carlo Comparison of Semiparametric Tobit Estimators," *Journal of Applied Econometrics*, 4, 361–382.

NEWEY, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5(2), 99–135.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by D. McFadden, and R. F. Engle, vol. IV, pp. 2111–2245. Elsevier, North-Holland, Amsterdam.

NEWEY, W. K., AND J. L. POWELL (1990): "Efficient Estimation of Linear and Type 1 Censored Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equation Models," *Econometrica*, 67, 565–603.

OLLEY, S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunication Equipment Industry," *Econometrica*, 64(6), 1263–1298.

PINKSE, J. (2001): "Nonparametric Regression Estimation using Weak Separability," Unpublished manuscript.

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

POWELL, J. L., AND T. M. STOKER (1996): "Optimal Bandwidth Choice for Density-weighted Averages," *Journal of Econometrics*, 75(2), 291–316.

PRIMONT, D., AND D. PRIMONT (1994): "Homothetic Non-parametric Production Models," *Economics Letters*, 45, 191–195.

ROBINSON, P. M. (1988): "Root n-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

ROSENBLATT, M. (1969): "Conditional Probability Density and Regression Estimators," in *Multivariate Analysis II*, ed. by P. R. Krishnaiah, pp. 25–31. Academic Press, New York.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

SEVERINI, T. A., AND G. TRIPATHI (2001): "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models," *Journal of Econometrics*, 102, 23–66.

SHEPHARD, R. W. (1953): *Cost and Production Functions.* Princeton University Press, New Jersey.

——— (1970): *Theory of Cost and Production Functions.* Princeton University Press, New Jersey.

SILVERMAN, B. W. (1978): "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density Function and its Derivatives," *Annals of Statistics*, 6(1), 177–184.

——— (1986): *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1 edn.

SIMON, C. P., AND L. E. BLUME (1994): *Mathematics for Economists.* W.W. Norton and Company Ltd, USA, 1 edn.

STEWART, M. B. (2005): "A Comparison of Semiparametric Estimators for the Ordered Response Model," *Computational Statistics and Data Analysis*, 49, 555–573.

STONE, C. J. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 8(8), 1348–1360.

——— (1985): "Additive Regression and other Nonparametric Models," *Annals of Statistics*, 13(2), 689–705.

——— (1986): "The Dimensionality Reduction Principle for Generalized Additive Models," *Annals of Statistics*, 14(2), 590–606.

SU, L., AND A. ULLAH (2004): "Local Polynomial Estimation of Nonparametric Simultaneous Equation Models," Unpublished Manuscript.

——— (2006): "More Efficient Estimation in Nonparametric Regression with Nonparametric Autocorrelated Errors," *Econometric Theory*, 22(1), 98–126.

TJØSTHEIM, D., AND B. H. AUESTAD (1994): "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of The American Statistical Association*, 89, 1398–1409.

TRIPATHI, G. (1998): "Nonparametric Estimation and Testing of Homogeneous Functional Forms," Unpublished Manuscript.

——— (2000): "Local Semiparametric Efficiency Bounds Under Shape Restrictions," *Econometric Theory*, 16(5), 729–739.

TRIPATHI, G., AND W. KIM (2003): "Nonparametric Estimation of Homogeneous Functions," *Econometric Theory*, 19, 640–663.

VARIAN, H. R. (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597.

WAND, M. P., AND C. JONES (1995): *Kernel Smoothing*, vol. 60 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, London, 1 edn.

WANG, Q., O. B. LINTON, AND W. HÄRDLE (2004): "Semiparametric Regression Analysis With Missing Response at Random," *Journal of the American Statistical Association*, 99(466), 334–345.

XIAO, Z., O. B. LINTON, R. J. CARROLL, AND E. MAMMEN (2003): "More Efficient Local Polynomial Estimation in Nonparametric Regression With Autocorrelated Errors," *Journal of the American Statistical Association*, 98(464), 980–992.

ZELLNER, A., AND H. RYU (1998): "Alternative Functional Forms for Production, Cost and Returns to scale Functions," *Journal of Applied Econometrics*, 13, 101–127.

ZHENG, J. X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75(2), 263–289.