

NONPARAMETRIC REGRESSION IN THE PRESENCE OF MEASUREMENT ERROR

SUSANNE M. SCHENNACH
University of Chicago

We introduce a nonparametric regression estimator that is consistent in the presence of measurement error in the explanatory variable when one repeated observation of the mismeasured regressor is available. The approach taken relies on a useful property of the Fourier transform, namely, its ability to convert complicated integral equations into simple algebraic equations. The proposed estimator is shown to be asymptotically normal, and its rate of convergence in probability is derived as a function of the smoothness of the densities and conditional expectations involved. The resulting rates are often comparable to kernel deconvolution estimators, which provide consistent estimation under the much stronger assumption that the density of the measurement error is known. The finite-sample properties of the estimator are investigated through Monte Carlo experiments.

1. INTRODUCTION

1.1. Motivation

The bias resulting from the presence of measurement error in the explanatory variables is a common problem in regression analysis. Although numerous solutions to this problem have been derived for parametric regression models, the corresponding problem in nonparametric specifications has remained relatively unexplored.

Some aspects of the nonparametric errors-in-variables problem have been previously investigated. The problem of estimating the density of an unobserved variable when this variable is measured with error and when the density of the error is known has received considerable attention in the literature. In this setting, the so-called kernel deconvolution estimator (for a review of the extensive literature, see, e.g., Carroll and Hall, 1988; Liu and Taylor, 1989; Carroll, Ruppert, and Stefanski, 1995) has been shown to reach the optimal rate of convergence (Fan, 1991b). The problem of the nonparametric estimation of a regression function when the independent variable is measured with an error drawn

This work was made possible in part through financial support from the National Science Foundation via grant SES-0214068. The author is grateful to the referees and the co-editor for their helpful comments. Address correspondence to Susanne M. Schennach, Department of Economics, University of Chicago, 1126 E. 59 St., Chicago, IL 60637; smschenn@alum.mit.edu.

from a known distribution has also been studied. In this case, a kernel regression estimator based on kernel deconvolution is known to achieve optimal convergence rates (Fan and Truong, 1993). A more challenging problem is the estimation of densities and regression functions when the independent variable is measured with an error drawn from an *unknown* distribution. Thanks to an identity due to Kotlarski (see Rao, 1992, p. 21), the identification of the density of an unobserved random variable is possible when the joint density of two error-contaminated measurements of that variable is known. Li and Vuong (1998) show that the empirical version of this identity leads to a consistent estimator with known convergence rates.

In contrast to the nonparametric density estimation problem, the nonparametric estimation of conditional expectations under similar conditions has so far remained unsolved. This is the gap our paper intends to fill by extending the traditional Nadaraya–Watson kernel regression estimator to allow for the independent variable to be contaminated with an error of an unknown distribution. We show that the availability of two error-contaminated measurements of the independent variable is all that is needed to achieve identification. The usefulness of this result stems from the observation that although distributional assumptions are often not appropriate in applications, thus precluding the use of kernel deconvolution estimators, repeated measurements can frequently be found in data sets (Ashenfelter and Krueger, 1994; Hausman, Newey, and Powell, 1995; Morey and Waldman, 1998; Bowles, 1972; Borus and Nestel, 1973; Freeman, 1984).¹ For instance, a given quantity may be repeatedly measured over time, or the same quantity may be reported by different people, such as different family members or an employer and an employee. The error on one of the measurements does not need to have zero mean, thus expanding the set of valid repeated measurements to more general indicators, or to repeated measurements that exhibit a systematic drift.

Our analysis not only derives the convergence rate of the proposed estimator but also provides its asymptotic distribution. The asymptotic properties of the estimator are analyzed through various analytical examples, and its finite-sample properties are investigated through Monte Carlo simulations that illustrate the bias-correcting power of our estimator. All proofs can be found in the Appendix.

1.2. Background

To understand the difficulties faced in nonparametric estimation in the presence of measurement error, it is instructive to recall the well-known solution to the simpler problem of finding the density of an unobserved variable x^* given an imperfect measurement z (for a review, see Carroll et al., 1995):

$$z = x^* + \Delta z. \quad (1)$$

The measurement error Δz is usually assumed to be independent from x^* and to be drawn from a known density. It is well known that the density of z is

given by the convolution of the density of x^* with the density of Δz . Thanks to the convolution theorem, this relationship can be concisely expressed using characteristic functions:

$$m(\nu) = \phi(\nu)\psi(\nu), \quad (2)$$

where $\phi(\nu)$, $m(\nu)$, and $\psi(\nu)$, respectively, denote the characteristic functions of x^* , z , and Δz . We can therefore identify the characteristic function of interest, $\phi(\nu)$, through

$$\phi(\nu) = \frac{m(\nu)}{\psi(\nu)}, \quad (3)$$

where $m(\nu)$ can be estimated by the Fourier transform of a nonparametric estimator of the density of z , such as a kernel estimator. The problem with this procedure arises from the fact that, under mild assumptions (such as assuming that the density of Δz is continuous), $\psi(\nu)$ vanishes as $\nu \rightarrow \infty$, so that this operation is not well defined for all ν . Hence, merely replacing $m(\nu)$ by a consistent estimate $\hat{m}(\nu)$ may not yield a consistent estimate of $\phi(\nu)$, because small errors on $\hat{m}(\nu)$ are magnified by the arbitrarily large factor $1/\psi(\nu)$. This is the well-known ill-defined inverse problem that occurs when one tries to invert a convolution operation. The so-called kernel deconvolution estimator (Carroll et al., 1995; Fan, 1991b) addresses this problem by estimating $m(\nu)$ using a kernel whose Fourier transform, $\kappa(\nu)$, is compactly supported. This ensures that the estimated characteristic function $\hat{m}(\nu)$ is also compactly supported, which in turn guarantees that the numerator of equation (3) will vanish well before the denominator causes the ratio to diverge.

It is clear that truncating the characteristic function of z in this fashion introduces a bias. To obtain a consistent estimator, the support of $\kappa(\nu)$ is allowed to expand as sample size grows in such a way that the total integrated noise over all frequencies in the support of $\kappa(\nu)$ decreases. The faster $\psi(\nu) \rightarrow 0$ as $\nu \rightarrow \infty$, the more slowly the support of $\kappa(\nu)$ can expand with sample size and the slower the convergence rate. This is the fundamental difficulty associated with nonparametric estimation in the presence of measurement error. As the smoothness of the density of the measurement error increases, the characteristic function $\psi(\nu)$ goes to zero increasingly rapidly as $\nu \rightarrow \infty$ and the convergence rate worsens. The smoothness of the density of x^* also plays a role in determining the convergence rate. The bias introduced by the truncation of $m(\nu)$ at a finite frequency is a function of the magnitude of the rate of decay of $m(\nu)$ as $\nu \rightarrow \infty$. The smoother the density of x^* , the faster its Fourier transform $m(\nu)$ decays as $\nu \rightarrow \infty$, and the faster the bias decreases as the kernel bandwidth shrinks.

The literature focusing on kernel deconvolution estimators typically describes the smoothness of a density in terms of the asymptotic rate of decay of its Fourier transform as frequency ν goes to infinity. The basis for such a description is that the number of derivatives of a density that are continuous is directly

related to the asymptotic behavior of its Fourier transform as $\nu \rightarrow \infty$. This leads to the traditional distinction between “ordinarily smooth” functions (which admit a finite number of continuous derivatives and whose Fourier transform decays as $|\nu|^\gamma$, $\gamma < 0$) and “supersmooth” functions (which admit an infinite number of continuous derivatives and whose Fourier transform decays as $\exp(\alpha|\nu|^\beta)$, $\alpha < 0, \beta > 0$). Examples of ordinarily smooth functions are gamma, uniform, and double exponential, and normal and Cauchy are supersmooth functions.

The kernel deconvolution estimator exhibits a wide variety of convergence rates depending on the smoothness of the densities involved. Whenever the densities of x^* and of Δz are ordinarily smooth, the kernel deconvolution will exhibit a rate of convergence of the form n^{-c} for some $c > 0$ where n is sample size. The situation degrades significantly when the density of Δz is supersmooth while the density of x^* remains ordinarily smooth. The convergence rate is then of the form $(\ln n)^{-c}$ for some $c > 0$, which is slower than any negative power of n .

The problem solved in this paper is more challenging than the one described above. First, we focus on a kernel regression estimator rather than a kernel density estimator. Second, we assume the density of the measurement error to be unknown.

2. ESTIMATION PROCEDURE

Our task is to find a function $g(x^*): \mathbb{R} \mapsto \mathbb{R}$ such that

$$y = g(x^*) + \Delta y \quad \text{where } E[\Delta y|x^*] = 0. \tag{4}$$

We consider x^* a scalar to simplify the exposition, although a multivariate extension is clearly possible.² When both y and x^* are observed, a natural candidate for this task is the well-known Nadaraya–Watson kernel estimator of $g(\tilde{x}^*)$ at a given point \tilde{x}^*

$$\hat{g}(\tilde{x}^*, h) = \frac{n^{-1} \sum_{l=1}^n y_l K_h(x_l^* - \tilde{x}^*)}{n^{-1} \sum_{l=1}^n K_h(x_l^* - \tilde{x}^*)}, \tag{5}$$

where x_l^* and y_l for $l = 1 \dots n$ denote the data points and the kernel $K_h(\cdot)$ is of the form

$$K_h(x^*) = \frac{1}{h} K\left(\frac{x^*}{h}\right) \tag{6}$$

and h is the bandwidth parameter. The problem we are facing is that x^* is not observed. As shown in Schennach (2004) the availability of two repeated measurements of x^*

$$x = x^* + \Delta x, \tag{7}$$

$$z = x^* + \Delta z, \tag{8}$$

provides enough information to identify any moment of the form $E[u(y, x^*)]$ for any function $u(y, x^*)$. Because the probability limit (at constant bandwidth h) of the Nadaraya–Watson kernel estimator is the ratio

$$g(\tilde{x}^*, h) = \frac{E[yK_h(x^* - \tilde{x}^*)]}{E[K_h(x^* - \tilde{x}^*)]}, \tag{9}$$

a similar technique can be applied here, setting $u(y, x^*) = y^k K_h(x^* - \tilde{x}^*)$, for $k = 0, 1$. The extension of the existing results to a nonparametric setting nevertheless requires additional steps to handle the fact that we need to characterize an infinite family of moments, indexed by \tilde{x}^* . Fortunately, this complication can be elegantly handled by observing that the convolution operations involved in computing the Nadaraya–Watson estimator are converted into simple products through the Fourier transform operation, enabling the whole family of moments to be estimated in a single operation. The formal result that permits identification is summarized in the following set of assumptions and associated theorem. Throughout the paper, we will take the convention that integrals without explicit bounds are taken over the whole real line.

Assumption 1.

$$\begin{aligned} E[\Delta y | x^*, \Delta z] &= 0, \\ E[\Delta x | x^*, \Delta z] &= 0; \end{aligned} \tag{10}$$

Δz and x^* are mutually independent.

Assumption 2. $E[|x^*|]$, $E[|\Delta x|]$, and $E[|y|]$ are finite.

Assumption 3. $E[y^k h^{-1} K(h^{-1}(x^* - \tilde{x}^*))] < \infty$ for all \tilde{x}^* , any $h > 0$, and $k = 0, 1$.

THEOREM 1. *Under Assumptions 1–3, and provided $|E[e^{i\xi z}]| > 0$ for any finite ξ , the function*

$$g(\tilde{x}^*, h) = \frac{E[yh^{-1}K(h^{-1}(x^* - \tilde{x}^*))]}{E[h^{-1}K(h^{-1}(x^* - \tilde{x}^*))]}, \tag{11}$$

for $\tilde{x}^* \in \mathbb{R}$ and $h \geq 0$, can be expressed solely in terms of moments that involve the observable variables y , x , and z :

$$g(\tilde{x}^*, h) = \frac{M_1(\tilde{x}^*, h)}{M_0(\tilde{x}^*, h)}, \tag{12}$$

where, for $k = 0, 1$,

$$M_k(\tilde{x}^*, h) = \frac{1}{2\pi} \int \kappa(h\xi) \phi_k(\xi) \exp(-i\xi\tilde{x}^*) d\xi \tag{13}$$

and where $\phi_k(\xi) \equiv E[y^k \exp(i\xi x^*)]$ is given by³

$$\phi_0(\xi) = \exp\left(\int_0^\xi \frac{im_x(\zeta)}{m_1(\zeta)} d\zeta\right), \tag{14}$$

$$\phi_1(\xi) = \phi_0(\xi) \frac{m_y(\xi)}{m_1(\xi)}, \tag{15}$$

where $i = \sqrt{-1}$ and $\kappa(\xi)$ is the Fourier transform of the kernel $K(x^*)$ and

$$m_a(\xi) = E[a \exp(i\xi z)] \quad \text{for } a = 1, x, y. \tag{16}$$

Note that knowledge of the moments $m_a(\xi)$, for $a = 1, x, y$, which involve observable variables only, is sufficient to identify $g(\tilde{x}^*, h)$. Because the moments $m_a(\xi)$ can be estimated from the corresponding sample averages, we propose the following estimator.

DEFINITION 1. Let (x_i, y_i, z_i) , for $i = 1, \dots, n$ denote a sample of size n . For a given $\tilde{x}^* \in \mathbb{R}$ and some sequence of bandwidths $h_n \rightarrow 0$, let

$$\hat{g}(\tilde{x}^*, h_n) = \frac{\hat{M}_1(\tilde{x}^*, h_n)}{\hat{M}_0(\tilde{x}^*, h_n)}, \tag{17}$$

where, for $k = 0, 1$,

$$\hat{M}_k(\tilde{x}^*, h_n) = \frac{1}{2\pi} \int \kappa(h_n \xi) \hat{\phi}_k(\xi) \exp(-i\xi\tilde{x}^*) d\xi, \tag{18}$$

$$\hat{\phi}_0(\nu) = \exp\left(\int_0^\nu \frac{i\hat{m}_x(\zeta)}{\hat{m}_1(\zeta)} d\zeta\right), \tag{19}$$

$$\hat{\phi}_1(\nu) = \frac{\hat{m}_y(\zeta)}{\hat{m}_1(\zeta)} \hat{\phi}_0(\nu), \tag{20}$$

and where, for $a = 1, x, y$,

$$\hat{m}_a(\zeta) = n^{-1} \sum_{j=1}^n a_j \exp(i\zeta z_j). \tag{21}$$

An interesting property of this estimator is that it reduces to the Nadaraya–Watson estimator in the absence of measurement error (i.e., when $z = x = x^*$). Indeed, in that case, $\hat{m}_x(\zeta) = \hat{m}_z(\zeta) = n^{-1} \sum_{j=1}^n z_j \exp(i\zeta z_j) = d\hat{m}_1(\zeta)/d\zeta$, and equation (19) can be integrated analytically to yield $\hat{\phi}_0(\nu) = m_1(\nu)$, thus imply-

ing that equation (20) becomes $\hat{\phi}_1(\nu) = \hat{m}_y(\zeta)$. With these equalities in mind, equation (18) then defines the Fourier representation of the numerator and the denominator of the Nadaraya–Watson estimator.

To ensure that the proposed estimator is well behaved, we need to make the following assumption.

Assumption 4. The Fourier transform of the kernel, $\kappa(\xi)$, is (i) bounded and (ii) compactly supported (without loss of generality, we consider the support to be $[-1, 1]$).

The boundedness of $\kappa(\xi)$ is a very weak requirement because any kernel $K(z)$ violating it would necessarily fail to be absolutely integrable. The assumption of compact support of $\kappa(\xi)$ is commonly made in the derivation of the asymptotic properties of kernel deconvolution estimators (Fan and Truong, 1993). The need for this assumption arises from the fact that the estimator involves a division by an asymptotically vanishing characteristic function. Under very mild smoothness requirements, characteristic functions decay to zero as frequency increases toward infinity. A compactly supported kernel (in Fourier representation) explicitly makes the frequency range considered in a given sample finite, ensuring that the divergence is kept under control.

The restriction of compact support (in Fourier representation) poses few problems in practice, because one can take any given kernel $K(x^*)$ and construct a modified kernel $\tilde{K}(x^*)$ that exhibits most of the properties of the original kernel, while possessing a compact support in Fourier representation. This is achieved by computing the Fourier transform $\kappa(\xi)$ of the original kernel $K(x^*)$ and multiplying it by a “windowing” function $W(\xi)$ that vanishes beyond a given frequency:

$$\tilde{\kappa}(\xi) = W(\xi)\kappa(\xi). \tag{22}$$

Judicious choice of a windowing function will ensure that the modified kernel $\tilde{\kappa}(\xi)$ keeps most of the properties of the original kernel. For instance, a windowing function such as

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

for any $\bar{\xi} \in]0, 1[$ **(23)**

will leave the order of the kernel unaffected, because the windowing function is constant in the neighborhood of the origin. The fact that this windowing function is infinitely many times differentiable will guarantee that the modified kernel $\tilde{K}(x^*)$ decays faster than any power of x^* as $|x^*| \rightarrow \infty$ (provided that the original kernel $K(x^*)$ had this property).

3. ASYMPTOTIC PROPERTIES

This section is organized as follows. To facilitate the analysis of the asymptotic properties of the proposed estimator $\hat{g}(\tilde{x}^*, h_n)$, we first provide a linear representation of this estimator, denoted $\bar{g}(\tilde{x}^*, h_n)$, that will be shown to be asymptotically equivalent to $\hat{g}(\tilde{x}^*, h_n)$. This linearization serves two purposes. First, it will enable the derivation of the convergence rate of the estimator using techniques that are analogous to the standard bias and variance decomposition used in the context of conventional kernel estimators. Second, a linear representation is essential to establish the asymptotic normality of the estimator.

3.1. Linearization

In this section, we will provide very general results that summarize the properties of a linearized estimator $\bar{g}(\tilde{x}^*, h_n)$ that will be used to establish the asymptotic properties of $\hat{g}(\tilde{x}^*, h_n)$. The form of the estimator prompts for two levels of linearization. First, as is commonly done in the analysis of nonparametric conditional expectation kernel estimators, the ratio of $\hat{M}_1(\tilde{x}^*, h_n)$ and $\hat{M}_0(\tilde{x}^*, h_n)$ in equation (17) is expanded in a Taylor series up to first order. Second, unlike the usual Nadaraya–Watson estimator and kernel deconvolution estimators, $\hat{M}_0(\tilde{x}^*, h)$ and $\hat{M}_1(\tilde{x}^*, h)$ themselves take the form of nonlinear functionals of the data generating process. It is thus convenient to carry out the linearization a step further by calculating the Fréchet derivative of $\hat{M}_0(\tilde{x}^*, h)$ and $\hat{M}_1(\tilde{x}^*, h)$ with respect to the estimated moment $\hat{m}_a(\zeta)$ for $a = 1, x, y$ in the vicinity of $\hat{m}_a(\zeta) = m_a(\zeta)$. The following definition gives a linearized version $\bar{g}(\tilde{x}^*, h)$ of the estimator $\hat{g}(\tilde{x}^*, h)$.⁴

DEFINITION 2. For $\tilde{x}^* \in \mathbb{R}$ and $h > 0$, let

$$\begin{aligned} \bar{g}(\tilde{x}^*, h) = & g(\tilde{x}^*, h) + \frac{1}{M_0(\tilde{x}^*, 0)} (\bar{M}_1(\tilde{x}^*, h) - M_1(\tilde{x}^*, h)) \\ & - \frac{M_1(\tilde{x}^*, 0)}{(M_0(\tilde{x}^*, 0))^2} (\bar{M}_0(\tilde{x}^*, h) - M_0(\tilde{x}^*, h)), \end{aligned} \tag{24}$$

where, for $k = 0, 1$, $M_k(\tilde{x}^*, h)$ is given by equation (13),

$$\bar{M}_k(\tilde{x}^*, h) = M_k(\tilde{x}^*, h) + \sum_{l=1, x, y} \int U_l^k(\zeta, \tilde{x}^*, h) (\hat{m}_l(\zeta) - m_l(\zeta)) d\zeta, \tag{25}$$

$$U_1^0(\zeta, \tilde{x}^*, h) = - \frac{iA_0(\zeta, \tilde{x}^*, h)m_x(\zeta)}{(m_1(\zeta))^2}, \tag{26}$$

$$U_x^0(\zeta, \tilde{x}^*, h) = \frac{iA_0(\zeta, \tilde{x}^*, h)}{m_1(\zeta)},$$

$$U_y^0(\zeta, \tilde{x}^*, h) = 0,$$

$$U_1^1(\zeta, \tilde{x}^*, h) = \frac{iA_1(\zeta, \tilde{x}^*, h)m_x(\zeta)}{(m_1(\zeta))^2} + \frac{C_1(\zeta, \tilde{x}^*, h)}{m_1(\zeta)},$$

$$U_x^1(\zeta, \tilde{x}^*, h) = \frac{iA_1(\zeta, \tilde{x}^*, h)}{m_1(\zeta)},$$

$$U_y^1(\zeta, \tilde{x}^*, h) = \frac{C_0(\zeta, \tilde{x}^*, h)}{m_1(\zeta)},$$

$$A_k(\zeta, \tilde{x}^*, h) = \begin{cases} (2\pi)^{-1} \int_{\zeta}^{\infty} \exp(-i\nu\tilde{x}^*)\kappa(-h\nu)\phi_k(\nu) d\nu & \text{if } \zeta \geq 0 \\ (2\pi)^{-1} \int_{\zeta}^{-\infty} \exp(-i\nu\tilde{x}^*)\kappa(-h\nu)\phi_k(\nu) d\nu & \text{if } \zeta < 0, \end{cases} \tag{27}$$

$$C_k(\zeta, \tilde{x}^*, h) = (2\pi)^{-1} \exp(-i\zeta\tilde{x}^*)\kappa(-h\zeta)\phi_k(\zeta).$$

The advantage of the linear representation provided by Definition 2 is that it is possible to decompose the error $\bar{g}(\tilde{x}^*, h) - g(\tilde{x}^*, 0)$ into well-defined “bias” and “variance” terms, as given by Lemma 1, which follows.

Assumption 5. $[y_i, x_i, z_i, x_i^*, \Delta y_i, \Delta x_i, \Delta z_i]$ for $i = 1 \dots n$ is an independent and identically distributed (i.i.d.) sequence.

Assumption 6. $E[y^{2-j}|z|^j] < \infty, E[x^{2-j}|z|^j] < \infty$, for $j = 0, 1$.

Assumption 7. The density of x^* is nonzero at $x^* = \tilde{x}^*$.

LEMMA 1. Under Assumptions 1–7, for $\tilde{x}^* \in \mathbb{R}$ and $h > 0$,

$$\bar{g}(\tilde{x}^*, h) - g(\tilde{x}^*, 0) = b(\tilde{x}^*, h) + v(\tilde{x}^*, h), \tag{28}$$

where the bias term $b(\tilde{x}^*, h)$ and the variance term $v(\tilde{x}^*, h)$ are given by

$$b(\tilde{x}^*, h) = g(\tilde{x}^*, h) - g(\tilde{x}^*, 0), \tag{29}$$

$$v(\tilde{x}^*, h) = \bar{g}(\tilde{x}^*, h) - g(\tilde{x}^*, h) \tag{30}$$

and where $v(\tilde{x}^*, h)$ satisfies $E[v(\tilde{x}^*, h)] = 0$ and

$$E[(v(\tilde{x}^*, h))^2] = n^{-1}\Omega(\tilde{x}^*, h),$$

where

$$\Omega(\tilde{x}^*, h) = \frac{\Sigma_{11}(\tilde{x}^*, h)}{(M_0(\tilde{x}^*, h))^2} + \frac{M_1^2(\tilde{x}^*, h)\Sigma_{00}(\tilde{x}^*, h)}{(M_0(\tilde{x}^*, h))^4} - 2 \frac{M_1(\tilde{x}^*, h)\Sigma_{10}(\tilde{x}^*, h)}{(M_0(\tilde{x}^*, h))^3} \tag{31}$$

$$\Sigma_{k_1 k_2}(\tilde{x}^*, h) = n^{-1} \sum_{l_1=1, x, y} \sum_{l_2=1, x, y} \iint U_{l_1}^{k_1}(\zeta, \tilde{x}^*, h) V_{l_1 l_2}(\zeta, \xi) (U_{l_2}^{k_2}(\xi, \tilde{x}^*, h))^\dagger d\zeta d\xi \tag{32}$$

for $k_1, k_2 = 0, 1$, where $U_l^k(\zeta, \tilde{x}^*, h)$ is given in Definition 2, where $M_k(\tilde{x}^*, h)$ for $k = 0, 1$ is given in Theorem 1, and where \dagger denotes complex conjugation, and

$$V_{l_1 l_2}(\zeta, \xi) = m_{(l_1 l_2)}(\zeta - \xi) - m_{l_1}(\zeta)m_{l_2}(-\xi) \quad \text{for } l_1, l_2 = 1, x, y. \tag{33}$$

Under our assumptions, the expectation and the variance of $\bar{g}(\tilde{x}^*, h) - g(\tilde{x}^*, 0)$ are well defined, even though the corresponding moments of $\hat{g}(\tilde{x}^*, h) - g(\tilde{x}^*, 0)$ may not exist. As long as the remainder $\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n)$ can be shown to be asymptotically negligible in probability, the mean and the variance of $\bar{g}(\tilde{x}^*, h)$ can be interpreted as the mean and the variance of the limiting distribution of $\hat{g}(\tilde{x}^*, h)$, whether or not the first two moments of $\hat{g}(\tilde{x}^*, h_n)$ are bounded. This situation is not unique, as these observations apply to any estimator involving ratios of random quantities. To ascertain that the linear approximation $\bar{g}(\tilde{x}^*, h_n)$ is appropriate, the following lemma provides the order of the remainder of the linearization of $\hat{g}(\tilde{x}^*, h_n)$ and also the order of the statistical fluctuations in $\bar{g}(\tilde{x}^*, h_n)$. This result is included for completeness, but it is not essential for the reader to master it to understand the main results of the subsequent sections.

LEMMA 2. Let Assumptions 1–7 hold and let, for $\phi_0(\zeta)$, $\phi_1(\zeta)$, and $m_1(\zeta)$ as in Theorem 1,

$$\begin{aligned} \bar{U}(h_n) = \int |m_1(\zeta)|^{-1} \sum_{k=0}^1 \left(\left(1 + \frac{|\phi'_0(\zeta)|}{|\phi_0(\zeta)|} \right) \right. \\ \times \int_{|\xi|}^{\infty} |\kappa(-h_n \xi)| |\phi_k(\xi)| d\xi \\ \left. + |\kappa(-h_n \zeta)| |\phi_k(\zeta)| \right) d\zeta, \end{aligned} \tag{34}$$

$$\lambda(h_n) = (1 + h_n^{-1}) \left(\sup_{\xi \in [-h_n^{-1}, h_n^{-1}]} \frac{|\phi'_0(\xi)|}{|\phi_0(\xi)|} \right) \left(\sup_{\xi \in [-h_n^{-1}, h_n^{-1}]} |m_1(\xi)|^{-1} \right) \tag{35}$$

for $\phi'_0(\xi) \equiv d\phi_0(\xi)/d\xi$.⁵ If the sequence h_n is such that (i) $M_k(\tilde{x}^*, h_n) \rightarrow M_k(\tilde{x}^*, 0)$ for $k = 0, 1$, (ii) $\bar{U}(h_n)n^{-1/2} \rightarrow 0$, and (iii) $\lambda(h_n)n^{-1/2+\epsilon} \rightarrow 0$ for some $\epsilon > 0$, then

$$\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = O_p(\bar{U}(h_n)n^{-1/2}), \tag{36}$$

$$\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n) = O_p(\bar{U}(h_n)n^{-1/2}\lambda(h_n)n^{-1/2+\epsilon}). \tag{37}$$

If, in addition, (iii) $(\Omega(\tilde{x}^*, h_n))^{-1/2}\bar{U}(h_n)\lambda(h_n)n^{-1/2+\epsilon} \rightarrow 0$, then

$$\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))(1 + o_p(1)). \tag{38}$$

The quantity $\bar{U}(h_n)$ is defined so that it bounds any of the quantities defined in equations (26) that enter the expression of the asymptotic variance of the estimator, whereas $\lambda(h_n)$ bounds the remainder terms from the linearization performed in Definition 2. As expected, the preceding stochastic expansion is written in terms of successive powers of $n^{-1/2}$, with the exception that the second term is proportional to $n^{-1+\epsilon}$ instead of n^{-1} , because bounding the second remainder term involves uniformly bounding various random functions, which slows the rate down by a factor n^ϵ .

In the proof of our convergence rate and asymptotic normality results, we will subsequently verify that the hypotheses of Lemma 2 are implied by more primitive regularity conditions. The first conclusion of the Lemma (equations (36) and (37)) will be sufficient to obtain the convergence rate of the estimator. Indeed, if it can be shown that $\lambda(h_n)n^{-1/2+\epsilon} \rightarrow 0$, the convergence rate is then simply given by $O_p(\bar{U}(h_n)n^{-1/2})$. Because $O_p(\bar{U}(h_n)n^{-1/2})$ is an upper bound on the convergence rate, which may or may not be binding, the second, slightly stronger conclusion of Lemma 2 (equation (38)) will be needed to obtain the limiting distribution of the estimator. The basic intuition behind the additional condition (iii) is that, for the $O_p(\bar{U}(h_n)n^{-1/2}\lambda(h_n)n^{-1/2+\epsilon})$ nonlinear remainder to have no effect on the limiting distribution, it must be asymptotically negligible relative to the exact standard deviation of $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)$, which is given by $(\Omega(\tilde{x}^*, h_n))^{1/2}n^{-1/2}$, by Lemma 1.

3.2. Regularity Conditions

We now provide primitive regularity conditions that will enable us to derive explicit convergence rates. These regularity conditions take the form of smoothness restrictions imposed via constraints on the tail behavior of various Fourier transforms. To specify the regularity conditions, we employ the following convenient notation.

DEFINITION 3. *An expression of the form $f(\zeta) \leq g(\zeta)$ for $f, g : \mathbb{R} \mapsto \mathbb{R}$ indicates that there exists a constant $C > 0$, independent of ζ , such that $f(\zeta) \leq Cg(\zeta)$ for all $\zeta \in \mathbb{R}$ (and similarly for \geq). Analogously, $a_n \leq b_n$ for two sequences a_n, b_n indicates that there exists a constant C independent of n such that $a_n \leq Cb_n$ for all $n \in \mathbb{N}$.*

The literature focusing on “kernel deconvolution estimators” (see, e.g., Carroll et al., 1995) and related estimators (Fan and Truong, 1993) traditionally distinguishes between “ordinarily smooth” functions (whose Fourier transform decays as $|\zeta|^\gamma$, $\gamma < 0$ as $|\zeta| \rightarrow \infty$) and “supersmooth” functions (whose Fourier transform decays as $\exp(\alpha|\zeta|^\beta)$, $\alpha < 0, \beta > 0$ as $|\zeta| \rightarrow \infty$). For the benefit of conciseness, our regularity conditions are given in terms of expressions of the form $(1 + |\zeta|)^\gamma \exp(\alpha|\zeta|^\beta)$, thereby simultaneously covering the ordinarily smooth and supersmooth cases.

Assumption 8. The functions $\phi_0(\zeta) = E[e^{i\zeta x^*}]$, $\phi'_0(\zeta) \equiv d\phi_0(\zeta)/d\zeta$, $\phi_1(\zeta) = E[ye^{i\zeta x^*}]$, and $m_1(\zeta) = E[e^{i\zeta z}]$ satisfy

$$\left| \frac{\phi'_0(\zeta)}{\phi_0(\zeta)} \right| \leq (1 + |\zeta|)^{\gamma_r} \tag{39}$$

for some $\gamma_r \geq 0$ and

$$\max\{|\phi_0(\zeta)|, |\phi_1(\zeta)|\} \leq (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}), \tag{40}$$

$$|m_1(\zeta)| \geq (1 + |\zeta|)^{\gamma_m} \exp(\alpha_m |\zeta|^{\beta_m}) \tag{41}$$

for some $\gamma_\phi, \gamma_m \in \mathbb{R}$, $\alpha_\phi \leq 0$, $\alpha_m \leq 0$, $\beta_\phi \geq 0$, $\beta_m \geq 0$ such that $\gamma_\phi \beta_\phi \geq 0$ and $\gamma_m \beta_m \geq 0$.

A few remarks are in order. While the rate of decay of $\phi_0(\zeta)$, the characteristic function of x^* , is entirely determined by the smoothness of the density $f(x^*)$ of x^* , the rate of decay of $\phi_1(\zeta)$ is governed by the smoothness of $f(x^*)E[y|x^*]$. Verifying equation (40) would first involve finding bounds on $|\phi_0(\zeta)|$ and $|\phi_1(\zeta)|$ individually before taking the most slowly decaying term. Regrouping $\phi_0(\zeta)$ and $\phi_1(\zeta)$ in a single assumption is possible without loss of generality, because both quantities enter the expression of the estimator in a similar fashion. This grouping is also notationally convenient, as it will reduce the number of independent orders of magnitude that have to be considered when determining the convergence rates of the estimator.

As is always the case in deconvolution-type estimators, one quantity (here $m_1(\zeta)$) needs to be bounded below (in equation (41)), instead of above, because it appears in a denominator in the expression of the estimator. Note that equation (41) is implied by separate lower bounds on the modulus of the characteristic functions of x^* and Δz because $m_1(\zeta) = E[e^{i\zeta z}] = E[e^{i\zeta x^*}]E[e^{i\zeta \Delta z}]$. The grouping of $E[e^{i\zeta x^*}]$ and $E[e^{i\zeta \Delta z}]$ is also aimed at reducing the notational burden. Although the constraint on the ratio $\phi'_0(\zeta)/\phi_0(\zeta)$ imposed by equation (39) may appear unusual, it is clear that it is implied by a more familiar upper bound on $|\phi'(\zeta)|$ and a lower bound on $|\phi(\zeta)|$. The absence of a term of the form $\exp(\alpha_r |\zeta|^{\beta_r})$ in equation (39) results in very little loss of generality, because all common ordinarily smooth and supersmooth functions are such that equation (39) holds for $\gamma_r = 1$.

Before we can derive the convergence rate of the estimator, we also need to characterize the type of kernel $K(x^*)$ used. While most studies of measurement error in nonparametric settings focus either on finite-order kernels (Fan, 1991b; Fan and Truong, 1993) or on infinite-order kernels (Politis and Romano, 1999; Li and Vuong, 1998), we will consider both finite- and infinite-order kernels. The traditional finite-order kernels we consider are defined in Assumption 9.

Assumption 9. $\int K(x^*) dx^* = 1$ and, for some integer $\gamma_\kappa > 0$,

$$\int (x^*)^j K(x^*) dx^* = 0 \quad \text{for } j = 1, \dots, \gamma_\kappa - 1, \tag{42}$$

$$\int (x^*)^{\gamma_\kappa} K(x^*) dx^* \neq 0, \tag{43}$$

$$\int |x^*|^j |K(x^*)| dx^* < \infty \quad \text{for } j = 1, \dots, \gamma_\kappa. \tag{44}$$

We also consider the following class of “infinite-order” kernels.

Assumption 10. The Fourier transform of the kernel, $\kappa(\xi)$, is such that $\kappa(\xi) = 1$ for $|\xi| < \bar{\xi}$ for some $\bar{\xi} > 0$.

Assumption 10 allows for a kernel of the form

$$K(x^*) = \frac{\sin(x^*)}{\pi x^*}, \tag{45}$$

which is particularly suited to the Fourier representation because its Fourier transform is 1 in the $[-1, 1]$ interval and zero elsewhere. This type of kernel has previously been used in other Fourier-based estimators (Li and Vuong, 1998) and amounts to truncating the Fourier transform above a given frequency. When both $E[y|x^*]$ and the density of x^* are infinitely many times differentiable, an infinite-order kernel will guarantee that the bias goes to zero faster than any power of the bandwidth. The bias could then, for instance, be an exponentially decaying function of the inverse bandwidth h^{-1} .

3.3. Rate of Convergence in Probability

The procedure to determine the asymptotic rates of pointwise convergence in probability can be outlined as follows.

1. Bound the bias of the linearized estimator in terms of the bounds given in Assumption 8.
2. Bound the variance of the linearized estimator in terms of the bounds given in Assumption 8.

3. Find the sequence of bandwidths h_n that makes the order of the bias squared and of the variance equal and verify that the higher order terms are asymptotically negligible, so that the asymptotic properties of the estimator $\hat{g}(\tilde{x}^*, h_n)$ can be obtained from the properties of its linearization $\bar{g}(\tilde{x}^*, h_n)$.

Step 1. Calculation of the bias $g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)$. We distinguish two cases, depending on whether the kernel used satisfies Assumption 9 or Assumption 10. In the following two lemmas, recall that the parameters γ_ϕ , α_ϕ , and β_ϕ , defined in Assumption 8, describe the smoothness of the density $f(x^*)$ of x^* and of the conditional expectation $E[y|x^*]$ by specifying that their Fourier transforms both decay at least as fast as $\zeta^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi})$ as frequency $\zeta \rightarrow \infty$.

LEMMA 3. *Under Assumptions 1–8, if the kernel is of order γ_κ , as defined by Assumption 9, then the bias satisfies*

$$|g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)| = O((h_n^{-1})^{\gamma_b} \exp(\alpha_b (h_n^{-1})^{\beta_b})) = O((h_n^{-1})^{\gamma_b}),$$

where $\alpha_b = 0$, $\beta_b = 0$, and

$$\gamma_b = \begin{cases} -\gamma_\kappa & \text{if } \alpha_\phi \neq 0 \\ \text{smallest } \gamma_b \in \mathbb{Z} \text{ such that } \gamma_b \geq -\gamma_\kappa \text{ and } \gamma_b > \gamma_\phi + 1 & \text{if } \alpha_\phi = 0. \end{cases} \quad (46)$$

LEMMA 4. *Under Assumptions 1–8, if the kernel satisfies Assumption 10 for some constant $\bar{\xi}$, then the bias satisfies*

$$|g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)| = O((h_n^{-1})^{\gamma_b} \exp(\alpha_b (h_n^{-1})^{\beta_b})), \quad (47)$$

where

$$\alpha_b = \alpha_\phi \bar{\xi}^{\beta_\phi}, \quad (48)$$

$$\beta_b = \beta_\phi, \quad (49)$$

$$\gamma_b = \gamma_\phi + 1. \quad (50)$$

In short, when a finite-order kernel is used, the rate of decrease of the bias is controlled either by the order of the kernel γ_κ or by the smoothness of $f(x^*)$ and $E[y|x^*]$, whichever is more limiting. In particular, when both $f(x^*)$ and $E[y|x^*]$ are supersmooth, so that $\alpha_\phi \neq 0$, it is the order of the kernel that determines the rate of decrease of the bias. When an infinite-order kernel is used, only the smoothness of $f(x^*)$ and $E[y|x^*]$ matters. Note that the bias term is identical to that of a traditional kernel estimator that would be used if x^* were perfectly observed, because, via equations (12) and (13), the bias can be expressed entirely in terms of $\phi_k(\zeta)$ for $k = 0, 1$ and the kernel, which are nonrandom measurement error-free quantities.

Step 2. Calculation of the order of the variance term $\bar{g}(\tilde{x}^*, h) - g(\tilde{x}^*, h)$.

LEMMA 5. Under Assumptions 1–8, the variance term satisfies

$$\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = O_p(n^{-1/2}(h_n^{-1})^{\gamma_v} \exp(\alpha_v(h_n^{-1})^{\beta_v})), \tag{51}$$

where

$$\gamma_v = 2 + \gamma_\phi - \gamma_m + \gamma_r, \tag{52}$$

$$(\alpha_v, \beta_v) = \begin{cases} (-\alpha_m, \beta_m) & \text{if } \beta_\phi < \beta_m \\ (\alpha_\phi - \alpha_m, \beta_m) & \text{if } \beta_\phi = \beta_m. \end{cases} \tag{53}$$

Note that the order of the variance term is determined not only by the smoothness of $f(x^*)$ and $E[y|x^*]$ (through $\gamma_\phi, \alpha_\phi, \beta_\phi,$ and γ_r) but also by the smoothness of the density of the measurement error Δz (through the terms $\gamma_m, \alpha_m,$ and β_m). It is important to point out that the variance term increases much faster as $h \rightarrow 0$ (at constant n) than that of a standard kernel estimator with perfectly observed variables (whose variance term is $O_p((h_n n)^{-1/2})$). Combined with the fact that the bias term is unchanged, as indicated in step 1, this implies that the achievable convergence rates will generally be slower than for a conventional kernel estimator.

Step 3. Determination of the rate of decrease of the bandwidth that offers the best trade-off between bias squared and variance. To obtain explicit rates of convergence, we need to distinguish various cases, based on the values of $\beta_b,$ which characterizes the rate of convergence of the bias term as the bandwidth shrinks, and $\beta_v,$ which characterizes the rate of divergence of the variance term as the bandwidth shrinks (at constant sample size). Both β_b and β_v represent an “exponent of supersmoothness,” that is, the constant β in an expression of the form $(h_n^{-1})^\gamma \exp(\alpha(h_n^{-1})^\beta)$.

THEOREM 2. Under Assumptions 1–8 and either Assumption 9 or 10, the optimal bandwidth choices and the corresponding convergence rates in probability of the estimator can be expressed in terms of the constants $\alpha_b, \beta_b, \gamma_b, \alpha_v, \beta_v, \gamma_v$ defined by Lemmas 3–5. Let $\epsilon > 0$ be arbitrarily small, let C_1, C_2 be some positive constants, and let \tilde{x}^* be given.

Case 1. If $\beta_v > \beta_b > 0$

$$h_n^{-1} = (C_1 \ln n + C_2 (\ln n)^{\beta_b/\beta_v})^{1/\beta_v}, \tag{54}$$

$$(\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)) = O_p\left(\exp\left(\left(\alpha_b + \epsilon\right)\left(\frac{\ln n}{2\alpha_v}\right)^{\beta_b/\beta_v}\right)\right). \tag{55}$$

Case 2. If $\beta_v > 0$ and $\beta_b = 0$ (with $\alpha_b = 0$ and $\gamma_b < 0$)

$$h_n^{-1} = (C_1 \ln n + C_2 \ln \ln n)^{1/\beta_v}, \tag{56}$$

$$(\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)) = O_p((\ln n)^{\gamma_b/\beta_v}). \tag{57}$$

Case 3. If $\beta_b = \beta_v \neq 0$

$$h_n^{-1} = C_1 (\ln n)^{1/\beta_b}, \tag{58}$$

$$(\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)) = O_p(n^{\alpha_b/(2\alpha_v - 2\alpha_b) + \epsilon}). \tag{59}$$

Case 4. If $\beta_b = \beta_v = 0$ (with $\alpha_b = \alpha_v = 0$ and $\gamma_b < 0$)

$$h_n^{-1} = C_1 n^{1/(2\gamma_v - 2\gamma_b)}, \tag{60}$$

$$(\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)) = O_p(n^{\gamma_b/(2\gamma_v - 2\gamma_b)}). \tag{61}$$

A few remarks are in order. First, it can be verified (see the proof of Theorem 2 in the Appendix) that the bandwidth sequences given above are such that conditions (i) and (ii) of Lemma 2 hold, thus implying that the nonlinear remainders are indeed negligible and that our simple bias-variance decomposition is justified. Second, the arbitrarily small ϵ was introduced to drastically simplify the calculations and the statement of the results at the expense of a very small loss in precision. Third, it is impossible to have $\beta_b > \beta_v$ because $\beta_b = \beta_\phi$, $\beta_v = \beta_m$, and

$$\begin{aligned} (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) &\geq |\phi_0(\zeta)| = |E[e^{i\zeta x^*}]| \geq |E[e^{i\zeta x^*}]| |E[e^{i\zeta \Delta z}]| \\ &\geq |E[e^{i\zeta z}]| \geq |m_1(\zeta)| \geq (1 + |\zeta|)^{\gamma_m} \exp(\alpha_m |\zeta|^{\beta_m}). \end{aligned} \tag{62}$$

The convergence rate of the proposed estimator varies substantially as a function of the smoothness of the densities and the conditional expectations involved. An important trend to observe among these rates is that large values of β_b (indicating a rapidly decreasing bias as $h \rightarrow 0$) and small values of β_v (indicating a slowly increasing variance as $h \rightarrow 0$) are desirable. The convergence rates obtained are typically slower than that of the Nadaraya–Watson kernel estimator used when the variables are perfectly observed. This limitation is not an artifact of our estimation procedure: it has also been observed in the simpler Fan and Truong estimator, which is known to be optimal under stronger assumptions than ours (see Fan and Truong, 1993). The different cases will be discussed—and compared to Fan and Truong’s findings—in more detail in Section 4.

Although we have focused on pointwise convergence rates, our results also provide information regarding global convergence rates. The upper bounds on the pointwise bias and variance (and of the nonlinear remainder terms) are in

fact independent of \tilde{x}^* . If the density of x^* is bounded away from zero over some finite interval $[a, b]$, it is straightforward to show that

$$\left(\int_a^b |\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)|^p W(\tilde{x}^*) d\tilde{x}^* \right)^{1/p} \tag{63}$$

converges to zero in probability at the same rate as the pointwise rates derived earlier for any bounded weighting function $W(\tilde{x}^*)$ and any $p \in [1, 2]$. However, rates of uniform convergence in probability do not follow directly from the results presented above.

3.4. Asymptotic Normality

To establish the asymptotic normality of the proposed estimator, we need to introduce a few additional assumptions. First, we need assumptions that are commonly made whenever a central limit theorem for triangular arrays is invoked (see, e.g., Härdle and Linton, 1994, Theorem 2; Andrews, 1991, Assumption A).

Assumption 11. There exists $C > 0$ such that $E[|x|^{2+\delta}|z] \leq C, E[|y|^{2+\delta}|z] \leq C, \text{Var}[x|z] \geq C,$ and $\text{Var}[y|z] \geq C$ for all z .⁶

The remaining assumptions are used to ensure that the condition $(\Omega(\tilde{x}^*, h_n))^{-1/2} \bar{U}(h_n) \lambda(h_n) n^{-1/2+\epsilon} \rightarrow 0$ in Lemma 2 holds, so that the higher order remainder terms are asymptotically negligible relative to the standard deviation of the linearized estimator $\bar{g}(\tilde{x}^*, h_n)$. The main obstacle to overcome is the necessity to find a lower bound for the variance $\Omega(\tilde{x}^*, h_n)$ of the estimator. The difficulty of obtaining such a result is noted by Fan (1991a) in his study of the limiting distribution of the kernel deconvolution estimator. Fan’s solution to this problem is simply to assume that the tails of the various Fourier transforms entering the estimator are not only bounded by some function of the form $\zeta^\gamma \exp(\alpha|\zeta|^\beta)$ but are asymptotically equal (as $|\zeta| \rightarrow \infty$) to such functional form, thereby limiting the set of allowed functions. Our solution to this problem is similar in spirit to Fan’s but considerably expands the range of possible behavior toward infinity by employing the concept of functions that are “well behaved at infinity,” as described by Lighthill (1962). The following definition formalizes this notion.⁷

DEFINITION 4. Let \mathbb{W} be the set of all functions $\psi: \mathbb{R} \mapsto \mathbb{R}$ such that (i) $\psi(\zeta)$ is absolutely integrable in every finite interval and (ii) $\int_{|\zeta| \geq T} |\psi(\zeta) - \Psi(\zeta)| d\zeta < \infty$ for some $T \in \mathbb{R}^+$ and some function $\Psi(\zeta)$ that can be written as a finite linear combination of finite products of functions of the form $|\zeta|^{c^+}, \text{sgn}(\zeta)|\zeta|^{c^+}, \ln|\zeta|, \sin(c\zeta), \cos(c\zeta), \exp(c\zeta^\gamma)$ with $c^+ \in \mathbb{R}^+, c \in \mathbb{R},$ and $\gamma \in \mathbb{N}^+.$

Assumption 12. For a given \tilde{x}^* , the functions $\psi_{kl}(\zeta) = (d/d\zeta)U_l^k(\zeta, \tilde{x}^*, 0)$, for $k = 0, 1$ and $l = 1, x, y$ and for $U_l^k(\zeta, \tilde{x}^*, h)$ given in equation (26), belong to \mathbb{W} .

For simplicity, we do not state Assumption 12 in terms of elementary quantities such as $m_1(\zeta)$ and $\phi_k(\zeta)$, but it is clear that Assumption 12 is only a few algebraic manipulations away from being a primitive condition. We need to constrain the derivative of $U_l^k(\zeta, \tilde{x}^*, 0)$ to rule out counterexamples where the density of z arbitrarily far away from the point of evaluation \tilde{x}^* could have a nonvanishing influence on $\hat{g}(\tilde{x}^*, h_n)$ asymptotically, making it difficult to characterize the behavior of the variance as $n \rightarrow \infty$.

The following condition requires the distribution of z to be supported on \mathbb{R} , which is usually the case in deconvolution problems because distributions that have a nonvanishing characteristic function (as imposed by equation (41) in Assumption 8) rarely have compact support.

Assumption 13. $f(z) > 0$ for all $z \in \mathbb{R}$.

Finally, we need to impose a few constraints that would be very difficult to state in a more primitive fashion. However, these assumptions are not very restrictive because the counterexamples violating them are somewhat contrived.

Assumption 14. $\text{Var}[\bar{g}(\tilde{x}^*, h_n)] \geq \max_{a=1,x,y} \max_{k=0,1} \text{Var}[\int U_a^k(\zeta, \tilde{x}^*, h_n) \times ae^{i\zeta z} d\zeta]$.

This assumption merely states that the variance of the estimator is of an order no less than any term in its asymptotic representation. This constraint can only be violated if two or more of the terms $\int U_a^k(\zeta, \tilde{x}^*, h_n)ae^{i\zeta z} d\zeta$ happen to cancel out asymptotically, which is unlikely because each term depends on different random quantities.

Assumption 15. For $C_k(\xi, \tilde{x}^*, h_n)$ as in equation (27), $\int_{\zeta}^{\infty} |C_k(\xi, \tilde{x}^*, h_n)| d\xi \leq |\int_{\zeta}^{\infty} C_k(\xi, \tilde{x}^*, h_n) d\xi|$ for $k = 0, 1$, for all $\zeta \in \mathbb{R}^+$ and all $n \in \mathbb{N}$.

This assumption requires that $\int_{\zeta}^{\infty} |C_k(\xi, \tilde{x}^*, h_n)| d\xi$ and $|\int_{\zeta}^{\infty} C_k(\xi, \tilde{x}^*, h_n) d\xi|$ be of the same order. It precludes $C_k(\xi, \tilde{x}^*, h_n)$ from having an oscillatory behavior (as ξ varies) such that a precise cancellation would occur between the values of $C_k(\xi, \tilde{x}^*, h_n)$ at different ξ during the integration. The cancellation would have to occur for all ζ and n sufficiently large and be such that the order of $\int_{\zeta}^{\infty} C_k(\xi, \tilde{x}^*, h_n) d\xi$ would be affected.

Assumptions 12–15 imply condition $(\Omega(\tilde{x}^*, h_n))^{-1/2} \bar{U}(h_n) \lambda(h_n) n^{-1/2+\epsilon} \rightarrow 0$ in Lemma 2, thus establishing the required asymptotic negligibility of the nonlinear remainder terms. If it is possible to calculate $(\Omega(\tilde{x}^*, h_n))^{-1/2} \times \bar{U}(h_n) \lambda(h_n) n^{-1/2+\epsilon}$ directly and verify that it goes to 0 asymptotically, then Assumptions 12–15 can be avoided altogether.⁸ We are now ready to state our asymptotic normality result.

THEOREM 3. *Under Assumptions 1–8 and 11–15, for any given \tilde{x}^* and any sequence h_n satisfying*

$$\begin{aligned}
 h_n^{-1} &\leq n^{-\eta} n^{1/(3+2\gamma_r-2\gamma_m)} \quad \text{if the density of } z \text{ is smooth } (\beta_m = 0) \\
 h_n^{-1} &\leq \left(\left(-\frac{(1+\eta)}{2\alpha_m} \right) \ln n \right)^{1/\beta_m} \quad \text{if the density of } z \text{ is supersmooth } (\beta_m \neq 0)
 \end{aligned}
 \tag{64}$$

for some $\eta > 0$, we have

$$n^{1/2}(\Omega(\tilde{x}^*, h_n))^{-1/2}(\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0) - b(\tilde{x}^*, h_n)) \xrightarrow{d} N(0, 1),
 \tag{65}$$

where $\Omega(\tilde{x}^*, h_n)$ and $b(\tilde{x}^*, h_n)$ are given in Lemma 1.

4. EXAMPLES

Section 3.3 derives the convergence rates of the proposed estimator under very general conditions. We now focus on specific examples that will allow us to compare these convergence rates with those derived for the estimator proposed by Fan and Truong (1993), which is the most closely related to ours. Fan and Truong’s estimator extends the standard kernel deconvolution estimators used for density estimation in the presence of a measurement error drawn from a known distribution to the case of nonparametric regressions. The estimator presented here accomplishes a more difficult task than Fan and Truong’s because it considers the density of the measurement error unknown, relying instead on two error-contaminated measurements of the unobserved regressor. Hence, it would come as no surprise if the kernel deconvolution rates were better. The comparison is nevertheless instructive, because it quantifies the precision loss incurred by relaxing the distributional assumptions regarding the measurement error.

We consider four examples. We first study the “difficult” deconvolution problem that consists of estimating an ordinarily smooth conditional expectation ($E[y|x^*]$) when the density of both the true regressor x^* and the measurement error Δz are supersmooth. This problem is difficult because a supersmooth measurement error strongly damps out the high-frequency components of $E[y|x^*]$ and of the density of x^* . Inverting this operation involves the amplification of these damped-out components, an operation that necessarily causes a substantial magnification of the statistical noise. In standard kernel deconvolution estimators, this situation gives rise to extremely slow convergence rates, and it is instructive to verify that the situation does not degrade further when the distribution of the measurement error is unknown. The second example shows that this slow convergence problem is avoided when the conditional expectation $E[y|x^*]$ is supersmooth as well. The third example assumes the density of the measurement error is ordinarily smooth, a situation that avoids the slow con-

vergence problem for the kernel deconvolution estimator but, as we will see, not for our estimator. The final example completes the analysis by showing that when all quantities are ordinarily smooth, the slow convergence problem is avoided.

Table 1 summarizes the assumptions made in each of the four cases considered. A few remarks are in order. In each case, we assume that the order of the kernel is sufficiently large so that the smoothness of $E[y|x^*]$ and of the density of x^* (and not the order of the kernel) is the factor limiting the rate at which the bias goes to zero. We also assume that equation (39) holds with $\gamma_r = 1$. Table 1 also summarizes the convergence rates obtained by applying Theorem 2 in each of the four examples considered. We will now discuss the significance of these results.

In Example 1, the rates are entirely comparable to those obtained by Fan and Truong (1993) for kernel deconvolution estimators. They found rates of the form $(\ln n)^{k/\beta}$ where k is the number of continuous derivatives that $g(x^*)$ possesses. Because a function whose Fourier transform behaves asymptotically as $\zeta^{-(k+1+\epsilon)}$ necessarily has k continuous derivatives, it is clear that the rates are comparable. The rates differ by ϵ , because Fan and Truong formulate their regularity conditions in terms of derivatives whereas we formulate them in terms of the asymptotic behavior of Fourier transforms. Formulating our regularity condi-

TABLE 1. Convergence rates obtained under given regularity assumptions

Example	Assumptions	Bandwidth choices/rates
1	$\exp(\alpha_0 \zeta ^\beta) \leq \phi_0(\zeta) \leq \exp(\alpha_0 \zeta ^\beta)$ $\exp(\alpha_\psi \zeta ^\beta) \leq \psi(\zeta) \leq \exp(\alpha_\psi \zeta ^\beta)$ $ \phi_1(\zeta) \leq (1 + \zeta)^{\gamma_1}$	$h_n^{-1} = C(\ln n)^{1/\beta}$ $O_p((\ln n)^{(\gamma_1+1)/\beta})$
2	$\exp(\alpha_0 \zeta ^\beta) \leq \phi_0(\zeta) \leq \exp(\alpha_0 \zeta ^\beta)$ $\exp(\alpha_\psi \zeta ^\beta) \leq \psi(\zeta) \leq \exp(\alpha_\psi \zeta ^\beta)$ $ \phi_1(\zeta) \leq \exp(\alpha_1 \zeta ^\beta)$	$h_n^{-1} = C(\ln n)^{1/\beta}$ $O_p(n^{-\alpha_1/(2\alpha_0+2\alpha_\psi+\epsilon)})$
3	$\exp(\alpha_0 \zeta ^\beta) \leq \phi_0(\zeta) \leq \exp(\alpha_0 \zeta ^\beta)$ $(1 + \zeta)^{\gamma_\psi} \leq \psi(\zeta) \leq (1 + \zeta)^{\gamma_\psi}$ $ \phi_1(\zeta) \leq (1 + \zeta)^{\gamma_1}$	$h_n^{-1} = C(\ln n)^{1/\beta}$ $O_p((\ln n)^{(\gamma_1+1)/\beta})$
4	$(1 + \zeta)^{\gamma_0} \leq \phi_0(\zeta) \leq (1 + \zeta)^{\gamma_0}$ $(1 + \zeta)^{\gamma_\psi} \leq \psi(\zeta) \leq (1 + \zeta)^{\gamma_\psi}$ $ \phi_1(\zeta) \leq (1 + \zeta)^{\gamma_1}$	$h_n^{-1} = Cn^{1/(-2\gamma_0-\gamma_\psi+\gamma_1)}$ $O_p(n^{(\gamma_1+1)/(-2\gamma_0-\gamma_\psi+\gamma_1)})$

Note: $\psi(\zeta) = E[\exp(i\zeta\Delta z)]$.

tions in terms of derivatives would yield results identical to Fan and Truong's. It is remarkable that under the assumptions leading to the worst-case convergence rates for kernel deconvolution estimators, the assumption of a known measurement error distribution can be relaxed without bringing the convergence rate down further.

Example 2 shows that the slow convergence rate problem can be alleviated if the unknown regression function $g(x^*)$ is supersmooth and if an "infinite-order" kernel is used. This situation ensures that the bias term goes to zero faster than any power of h , which is sufficient to convert a convergence rate of the form $(\ln n)^\gamma$ to a rate of the form n^γ for $\gamma < 0$. More generally, relatively fast convergence rates can be achieved with infinite-order kernels whenever case 3 of Section 3.3 applies. Caution is, however, advised when using high-order kernels. They are known not to perform as well in finite samples as their asymptotic properties would suggest (see Härdle and Linton, 1994). The origin of the problem is that a high-order kernel must necessarily take negative values over a portion of its support, which makes it likely for the denominator of the Nadaraya–Watson kernel estimator to approach zero, even at a point where the true density is bounded away from zero.

In Example 3, making the density of the measurement error Δz ordinarily smooth instead of supersmooth does not improve the convergence rates relative to Example 1. This is in sharp contrast to the behavior of kernel deconvolution estimators, whose convergence rates are of the form n^k under the same assumptions. The reason for this distinction is that the only characteristic function appearing in the denominator of a kernel deconvolution estimator is that of the measurement error Δz , whereas in our estimator, it is the characteristic function of z that appears in the denominator. The density of z is supersmooth if either the density of the true regressor x^* or of the measurement error Δz is supersmooth. Hence, a supersmooth density for x^* will also cause our estimator to converge slowly.

In Example 4, it is seen that when the density of x^* is made ordinarily smooth as well, the slow convergence problem is avoided, as expected. The resulting rates are not necessarily identical to those of Fan and Truong's kernel deconvolution estimator, but the rates at least take the form of a negative power of n , indicating that the distributional assumptions regarding the measurement error can be relaxed without an undue increase in the statistical noise.

5. MONTE CARLO SIMULATIONS

We now investigate the finite-sample properties of the proposed estimator through various Monte Carlo simulations. The designs are chosen so as to illustrate the examples of Section 4, summarized in Table 1, which cover the most common combinations of smooth and supersmooth distributions and conditional expectations. As an example of a supersmooth distribution, the normal distribution with variance σ^2 naturally comes to mind. Its characteristic func-

tion has a tail of the form $\exp(-(\sigma^2/2)|\zeta|^2)$. As an example of an ordinarily smooth distribution, we consider the Laplace (or double exponential) distribution with mean μ and variance σ^2 denoted by $L(\mu, \sigma^2)$ and defined as

$$\frac{1}{\sigma\sqrt{2}} \exp(-\sigma|t - \mu|\sqrt{2}) \tag{66}$$

for any $t \in \mathbb{R}$. The tail of the characteristic function of a Laplace density is of the form $|\zeta|^{-2}$.

Our example of a supersmooth regression function is the error function

$$E[y|x^*] = \frac{2}{\sqrt{\pi}} \int_0^{x^*} e^{-t^2} dt \equiv \text{erf}(x^*) \tag{67}$$

having a Fourier transform decaying at the rate $|\zeta|^{-1} \exp(-\frac{1}{4}|\zeta|^2)$ as $|\zeta| \rightarrow \infty$. Finally, our example of an ordinarily smooth regression function is a piecewise linear continuous function with a discontinuous first derivative

$$E[y|x^*] = S(x^*) \equiv \begin{cases} -1 & \text{if } x^* < -1 \\ x^* & \text{if } x^* \in [-1, 1] \\ 1 & \text{if } x^* > 1, \end{cases} \tag{68}$$

whose Fourier transform decays as $|\zeta|^{-2}$. To simplify comparisons, both functions are normalized to have the same range and a similar general shape, so that any difference in the results can be attributed to their difference in smoothness. All simulations proceed by drawing 500 samples of 1,000, 2,000, or 8,000 observations from the distributions given in Table 2. Table 2 also provides the theoretical convergence rate in each case, obtained by substituting the appropriate smoothness parameters in the expressions of Table 1. The distribution of Δy is never altered, because it has little impact on the asymptotic properties of

TABLE 2. Monte Carlo simulation designs

Example	x^*	$\Delta x, \Delta z$	Δy	$E[y x^*]$	Convergence rate
1	$N(0,1)$	$N(0,1/4)$	$N(0,1/4)$	$S(x^*)$	$O_p((\ln n)^{-1/2})$
2	$N(0,1)$	$N(0,1/4)$	$N(0,1/4)$	$\text{erf}(x^*)$	$O_p(n^{-2/5+\epsilon})$
3	$N(0,1)$	$L(0,1/4)$	$N(0,1/4)$	$S(x^*)$	$O_p((\ln n)^{-1/2})$
4	$L(0,1)$	$L(0,1/4)$	$N(0,1/4)$	$S(x^*)$	$O_p(n^{-1/4})$

the estimator except for a trivial scaling of some of the components of the asymptotic variance. For each sample, the variables x, y, z are constructed through

$$y = E[y|x^*] + \Delta y, \quad (69)$$

$$x = x^* + \Delta x, \quad (70)$$

$$z = x^* + \Delta z. \quad (71)$$

The variables (y, x, z) are used as an input for our estimator, and the variables (y, x) are fed into the Nadaraya–Watson estimator. We also construct an (infeasible) Nadaraya–Watson estimator from the variables (y, x^*) for comparative purposes. For all three estimators, an infinite-order kernel whose Fourier transform is given by equation (23) with $\bar{\xi} = \frac{1}{2}$ is used. In this fashion, the kernel is never the factor limiting the convergence rate. For each sample, we keep track of the value of the estimated function at a given point (here, $x^* = 1$) and use it to calculate the bias squared, the variance, and the sum of the two, the mean square error. A set of bandwidths ranging from 1.0 to 2.5 is scanned in increments of 0.05 in search of the bandwidth minimizing the mean square error.⁹ Of course, this method of locating the optimal bandwidth relies on our knowledge of the true regression function. Although this is appropriate for the purpose of investigating the properties of an estimator, a feasible bandwidth selection rule would be a useful tool to develop.

Table 3 compares the bias squared, the variance, and the mean square error of the three estimators considered as a function of bandwidth for a sample size of 1,000. For conciseness, only a subset of the bandwidths considered is shown. The rightmost column gives all quantities evaluated at the optimal bandwidth (which may lie between two of the bandwidths listed in the previous columns). A few important features can be consistently observed throughout the four examples considered.

In comparison with the Nadaraya–Watson estimator, our estimator is clearly very effective at reducing the bias. More specifically, it is clear that the bias of the Nadaraya–Watson estimator does not converge to zero with decreasing bandwidth but instead settles to a nonzero value. In contrast, the bias of our estimator decreases by orders of magnitude over the range of bandwidths sampled, as the bandwidth decreases. Our estimator's residual bias is attributable to the fact that we are performing a nonparametric estimation, so that a fully unbiased estimation is impossible. In fact, it can readily be seen that, at a given bandwidth, the bias of our estimator is very close to the bias of the infeasible Nadaraya–Watson estimator using the uncontaminated regressor x^* , thus indicating that our estimator does not appear to introduce additional bias at the sample size considered. Of course, because the variance of our estimator is larger than the infeasible Nadaraya–Watson estimator, a larger bandwidth must be used, and the resulting bias, evaluated at the optimal bandwidth, is slightly larger than in the error-free case.

TABLE 3. Monte Carlo simulation results for the examples

Example 1								
Fourier								(optimal)
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	2.0500
Bias squared	0.0004	0.0039	0.0053	0.0125	0.0284	0.0592	0.1073	0.0331
Variance	0.4696	0.1978	0.2514	0.0995	0.0211	0.0030	0.0017	0.0133
Mean square error	0.4700	0.2018	0.2567	0.1120	0.0495	0.0623	0.1091	0.0464
Nadaraya–Watson								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.4000
Bias squared	0.0853	0.0815	0.0829	0.0903	0.1089	0.1399	0.1820	0.0820
Variance	0.0237	0.0043	0.0024	0.0017	0.0013	0.0012	0.0010	0.0030
Mean square error	0.1090	0.0858	0.0853	0.0920	0.1102	0.1411	0.1831	0.0850
No measurement error								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.1000
Bias squared	0.0027	0.0045	0.0054	0.0109	0.0274	0.0598	0.1078	0.0037
Variance	0.0043	0.0021	0.0018	0.0015	0.0011	0.0010	0.0008	0.0027
Mean square error	0.0070	0.0066	0.0073	0.0124	0.0285	0.0608	0.1086	0.0064
Example 2								
Fourier								(optimal)
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	2.1500
Bias squared	0.0056	0.0012	0.0007	0.0001	0.0042	0.0183	0.0461	0.0111
Variance	0.5137	0.1919	0.2399	0.0958	0.0202	0.0028	0.0016	0.0049
Mean square error	0.5194	0.1931	0.2406	0.0960	0.0244	0.0211	0.0476	0.0159
Nadaraya–Watson								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.4500
Bias squared	0.0346	0.0309	0.0313	0.0357	0.0470	0.0668	0.0948	0.0310
Variance	0.0306	0.0042	0.0022	0.0015	0.0013	0.0011	0.0010	0.0025
Mean square error	0.0651	0.0351	0.0335	0.0372	0.0482	0.0679	0.0958	0.0335
No measurement error								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.7500
Bias squared	0.0002	0.0006	0.0007	0.0000	0.0038	0.0186	0.0463	0.0000
Variance	0.0037	0.0024	0.0018	0.0014	0.0010	0.0009	0.0008	0.0014
Mean square error	0.0039	0.0030	0.0025	0.0014	0.0049	0.0195	0.0471	0.0014
Example 3								
Fourier								(optimal)
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.7000
Bias squared	0.0045	0.0016	0.0044	0.0098	0.0258	0.0576	0.1050	0.0082
Variance	0.1569	0.2382	0.0164	0.0080	0.0033	0.0019	0.0015	0.0093
Mean square error	0.1614	0.2397	0.0208	0.0179	0.0291	0.0595	0.1065	0.0175
Nadaraya–Watson								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.4500
Bias squared	0.0569	0.0625	0.0626	0.0716	0.0925	0.1260	0.1705	0.0619
Variance	0.0289	0.0078	0.0025	0.0017	0.0013	0.0011	0.0009	0.0028
Mean square error	0.0858	0.0703	0.0651	0.0734	0.0938	0.1270	0.1714	0.0647
No measurement error								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.3500
Bias squared	0.0027	0.0044	0.0052	0.0107	0.0272	0.0595	0.1074	0.0045
Variance	0.0059	0.0032	0.0025	0.0015	0.0010	0.0009	0.0008	0.0029
Mean square error	0.0086	0.0076	0.0077	0.0122	0.0282	0.0604	0.1082	0.0075
Example 4								
Fourier								(optimal)
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.4500
Bias squared	0.0140	0.0033	0.0046	0.0397	0.0990	0.1691	0.2410	0.0017
Variance	0.1450	0.0198	0.0076	0.0037	0.0022	0.0015	0.0011	0.0091
Mean square error	0.1590	0.0231	0.0123	0.0434	0.1012	0.1706	0.2421	0.0108
Nadaraya–Watson								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.1000
Bias squared	0.0925	0.0966	0.1172	0.1531	0.2005	0.2544	0.3105	0.0921
Variance	0.0038	0.0025	0.0018	0.0014	0.0011	0.0010	0.0008	0.0030
Mean square error	0.0963	0.0990	0.1191	0.1545	0.2016	0.2554	0.3113	0.0951
No measurement error								
Bandwidth	1.0000	1.2500	1.5000	1.7500	2.0000	2.2500	2.5000	1.4000
Bias squared	0.0098	0.0032	0.0047	0.0400	0.0993	0.1693	0.2410	0.0002
Variance	0.0047	0.0035	0.0022	0.0014	0.0010	0.0008	0.0007	0.0026
Mean square error	0.0145	0.0067	0.0069	0.0414	0.1003	0.1702	0.2417	0.0029

The bias reduction made possible by the proposed estimator comes at the expense of an increased variance relative to the Nadaraya–Watson estimator based on mismeasured regressors. However, the decrease in the bias more than offsets the increase in the variance, so that the mean square error we obtain is still better than for the Nadaraya–Watson estimator.

It is instructive to observe the estimator’s behavior as a function of the smoothness of the various densities and conditional expectations considered. The asymptotic theory presented earlier predicts the convergence rate, which can be directly compared with the change in the mean square error at the optimal bandwidth, as a function of sample size for each of the examples considered (see Table 4). The fifth column of Table 4, labeled “ MSE_{8000}/MSE_{2000} ,” reports the ratio of mean square error at a sample size of 8,000 relative to the mean square error at sample size 2,000. We focus on these sample sizes because the differences between the various examples are more readily seen at large sample sizes. In Examples 1 and 3, where the convergence rate should be slow (i.e., a negative power of the log of sample size), convergence is indeed much slower than for Examples 2 and 4, where the convergence rate should be fast (i.e., a negative power of sample size). Moreover, the decrease in mean square error predicted by asymptotic theory (obtained by squaring the rates given in Table 2 and shown

TABLE 4. Monte Carlo simulation results as a function of sample size

Sample size	1,000	2,000	8,000	MSE_{8000}/MSE_{2000}	Theory
Example 1					
Bandwidth	2.0500	1.5500	1.5500		
Bias squared	0.0331	0.0072	0.0061		
Variance	0.0133	0.0061	0.0024		
Mean square error	0.0464	0.0133	0.0084	0.631	0.721
Example 2					
Bandwidth	2.1500	1.8000	1.7500		
Bias squared	0.0111	0.0003	0.0000		
Variance	0.0049	0.0028	0.0009		
Mean square error	0.0159	0.0031	0.0009	0.290	0.330
Example 3					
Bandwidth	1.7000	1.5500	1.6500		
Bias squared	0.0082	0.0058	0.0078		
Variance	0.0093	0.0062	0.0021		
Mean square error	0.0175	0.0120	0.0099	0.825	0.721
Example 4					
Bandwidth	1.4500	1.4000	1.4000		
Bias squared	0.0017	0.0002	0.0002		
Variance	0.0091	0.0055	0.0013		
Mean square error	0.0108	0.0057	0.0016	0.280	0.500

in the last column of Table 4) is an excellent predictor of the actual decrease in three out of the four examples. Note that the systematic changes in bandwidth as a function of sample size are difficult to distinguish from the inherent simulation noise, because bandwidth variations are much smaller than the changes in mean square error, as predicted by Theorem 2.

Monte Carlo simulations can also be used to verify the applicability of the asymptotic distribution in a finite sample. The designs described in Table 2 are again used, with the mean square minimizing bandwidths given in Table 3 and a sample size of 1,000. For each sample, we keep track of the value of the estimated function at a given point ($x^* = 1.0$) and the estimated variance at that point obtained with equations (31) and (32) by replacing all expected values by sample averages. The point estimates are then standardized, that is, demeaned by the average of the point estimates and normalized by the average of the estimated pointwise variance. Figure 1 shows the empirical cumulative distribution function (c.d.f.) of the standardized point estimates p_i for $i = 1, \dots, 500$, obtained by sorting the p_i in increasing order and by joining the points $(p_i, (i - 1)/499)$ by lines. The resulting empirical c.d.f. (jagged lines in Figure 1) agrees very well with the normal c.d.f. predicted by asymptotic theory (shown as a smooth line in Figure 1).

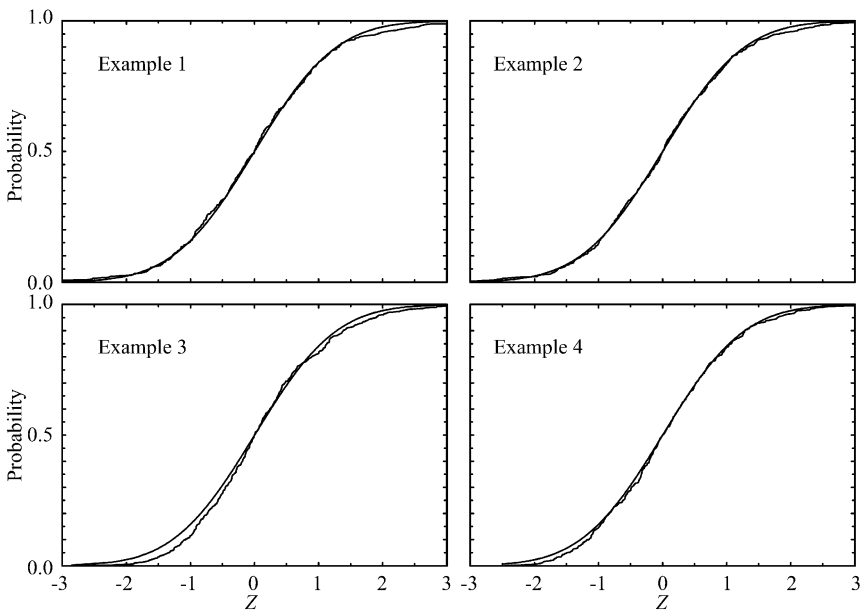


FIGURE 1. Comparison between the finite-sample and the asymptotic distributions of the estimator. The abscissa is $Z = (\text{Var}[\hat{g}(0.5, 2)])^{-1/2}(\hat{g}(0.5, 2) - E[\hat{g}(0.5, 2)])$.

6. CONCLUSION

This paper presents a new kernel-based nonparametric estimator that extends the conventional Nadaraya–Watson kernel estimator to cover the case of an error-ridden regressor. We show that identification is achievable when one repeated measurement of the error-contaminated regressor is available. One remarkable property of our estimator is that it requires no knowledge of the distribution of the measurement error, contrary to the popular kernel deconvolution estimator. The convergence rate and the asymptotic distribution of the proposed estimator are derived. A series of examples illustrates the main factors determining the convergence rate and enables us to compare the convergence rates we obtain with those of earlier estimators. Various Monte Carlo simulations are used to investigate the finite-sample properties of the estimator.

NOTES

1. Freeman's data set (the January 1977 Employer Employee Matched Sample, Current Population Survey) contains wages reported by employers and employees, which are perfect examples of repeated measurements.
2. As in any nonparametric regression, the well-known "curse of dimensionality" of course limits the number of dimensions that can be handled in practice.
3. Equation (14) is similar to an identity derived by Kotlarski (see Rao, 1992, p. 21), but our proof of this result requires weaker independence assumptions. In particular, we do not require independence between Δx and x^* and between Δx and Δz .
4. The calculation of the Fréchet derivative can be found in the proof of Lemma 2 in the Appendix.
5. Note that the ratio $|\phi'_0(\zeta)|/|\phi_0(\zeta)|$ entering the definitions of $\lambda(h_m)$ and $\bar{U}(h_m)$ can equivalently be written as $|m_x(\zeta)|/|m_1(\zeta)|$ because $|m_x(\zeta)|/|m_1(\zeta)| = |E[xe^{i\zeta z}]/E[e^{i\zeta z}]| = |E[x^*e^{i\zeta z}]/E[e^{i\zeta z}]| = (|E[x^*e^{i\zeta x^*}]/E[e^{i\zeta x^*}]|)(|E[e^{i\zeta \Delta z}]/E[e^{i\zeta \Delta z}]|) = |E[x^*e^{i\zeta x^*}]/E[e^{i\zeta x^*}]| = |\phi'_0(\zeta)|/|\phi_0(\zeta)|$.
6. The familiar condition $E[|K(x^*)|^{2+\delta}] < \infty$, which is helpful to show the asymptotic normality of standard kernel estimators, is of no use in establishing the asymptotic normality of our more complex estimator. In any case, Assumption 4 implies that $E[|K(x^*)|^{2+\delta}] < \infty$.
7. We expand Lighthill's definition by allowing for exponential tails, which is essential to handle supersmooth functions.
8. And the term $n^{1/(3+2\gamma_r-2\gamma_m)}$ in equation (64) can be replaced by $n^{1/(2+2\gamma_r-2\gamma_m)}$.
9. For less than 0.5% of the samples drawn, numerical issues associated with near division by zero in equations (19) and (20) were observed for a few of the smallest bandwidths sampled. To simplify the reporting of the results as a function of bandwidth, these draws were discarded and new draws were made so that the total number of samples kept remains 500. Of course, when studying any given sample, practitioners would simply never choose such a small bandwidth. The problem only occurs because we are performing Monte Carlo simulations and wish to report averages over replications as a function of bandwidth.

REFERENCES

- Andrews, D.W.K. (1991) Asymptotic normality of series estimators for nonparametric and semi-parametric regression models. *Econometrica* 59, 307–345.
- Ashenfelter, O. & A.B. Krueger (1994) Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review* 84, 1157–1173.

- Borus, M.E. & G. Nestel (1973) Response bias in reports of father's education and socioeconomic status. *Journal of the American Statistical Association* 68, 816–820.
- Bowles, S. (1972) Schooling and inequality from generation to generation. *Journal of Political Economy* 80, S219–S251.
- Carroll, R. & P. Hall (1988) Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83, 1184–1186.
- Carroll, R.J., D. Ruppert, & L.A. Stefanski (1995) *Measurement Error in Nonlinear Models*. Chapman and Hall.
- Fan, J. (1991a) Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics* 53, series A, pt. 1, 97–110.
- Fan, J. (1991b) On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics* 19, 1257–1272.
- Fan, J. & Y.K. Truong (1993) Nonparametric regression with errors in variables. *Annals of Statistics* 21, 1900–1925.
- Freeman, R.B. (1984) Longitudinal analysis of the effects of trade unions. *Journal of Labor Economics* 2, 1–26.
- Gel'fand, I.M. & G.E. Shilov (1964) *Generalized Functions*. Academic Press.
- Härdle, W. & O. Linton (1994) Applied nonparametric methods. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics* vol. IV, pp. 2295–2339. Elsevier Science.
- Hausman, J., W. Newey, & J. Powell (1995) Nonlinear errors in variables: Estimation of some Engel curves. *Journal of Econometrics* 65, 205–233.
- Li, T. & Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* 65, 139–165.
- Lighthill, M.J. (1962) *Introduction to Fourier Analysis and Generalized Function*. Cambridge University Press.
- Liu, M. & R. Taylor (1989) A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics* 17, 427–438.
- Morey, E.R. & D.M. Waldman (1998) Measurement error in recreation demand models: The joint estimation of participation, site choice, and site characteristics. *Journal of Environmental Economics and Management* 35, 262–276.
- Politis, D.N. & J.P. Romano (1999) Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis* 68, 1–25.
- Rao, P. (1992) *Identifiability in Stochastic Models*. Academic Press.
- Schennach, S.M. (2004) Estimation of nonlinear models with measurement error. *Econometrica* 72, 33–75.

APPENDIX: PROOFS

Proof of Theorem 1. The result can be shown by direct substitution. Assumption 2 ensures that all expectations are well defined. First, observe that equation (14) indeed provides the value of $\phi_0(\xi)$, by using Assumption 1:

$$\begin{aligned} & \exp\left(\int_0^\xi \frac{im_x(\zeta)}{m_1(\zeta)} d\zeta\right) \\ &= \exp\left(\int_0^\xi \frac{iE[x \exp(i\zeta z)]}{E[\exp(i\zeta z)]} d\zeta\right) \end{aligned} \quad (\text{A.1})$$

$$= \exp \left(\int_0^\xi \frac{E[ix^* \exp(i\zeta x^*) \exp(i\zeta \Delta z)] + iE[\Delta x \exp(i\zeta x^*) \exp(i\zeta \Delta z)]}{E[\exp(i\zeta x^*) \exp(i\zeta \Delta z)]} d\zeta \right) \tag{A.2}$$

$$= \exp \left(\int_0^\xi \frac{E[ix^* \exp(i\zeta x^*)]E[\exp(i\zeta \Delta z)] + iE[E[\Delta x | x^*, \Delta z] \exp(i\zeta x^*) \exp(i\zeta \Delta z)]}{E[\exp(i\zeta x^*)]E[\exp(i\zeta \Delta z)]} d\zeta \right) \tag{A.3}$$

$$= \exp \left(\int_0^\xi \frac{E[ix^* \exp(i\zeta x^*)]}{E[\exp(i\zeta x^*)]} d\zeta \right) \tag{A.4}$$

$$= \exp \left(\int_0^\xi \frac{d}{d\zeta} \ln E[\exp(i\zeta x^*)] d\zeta \right) \tag{A.5}$$

$$= E[\exp(i\xi x^*)]/E[1] \tag{A.6}$$

$$= \phi_0(\xi). \tag{A.7}$$

Letting $f(x^*)$ be the density of x^* , one can then show that $M_1(\tilde{x}^*, h)$ and $M_0(\tilde{x}^*, h)$, respectively, provide the numerator and the denominator of the Nadaraya–Watson estimator. In what follows, we use the independence between x^* and Δz and the fact that

$$E[y \exp(i\xi z)] = E[E[y | x^*, \Delta z] \exp(i\xi x^*) \exp(i\xi \Delta z)] \tag{A.8}$$

$$= E[E[g(x^*) | x^*, \Delta z] \exp(i\xi x^*) \exp(i\xi \Delta z)] \tag{A.9}$$

$$= E[g(x^*) \exp(i\xi x^*) \exp(i\xi \Delta z)] \tag{A.10}$$

$$= E[g(x^*) \exp(i\xi x^*)] E[\exp(i\xi \Delta z)] \tag{A.11}$$

$$= E[y \exp(i\xi x^*)] E[\exp(i\xi \Delta z)], \tag{A.12}$$

$$M_1(\tilde{x}^*, h) = \frac{1}{2\pi} \int \kappa(h\xi) \phi_0(\xi) \frac{m_y(\xi)}{m_1(\xi)} \exp(-i\xi \tilde{x}^*) d\xi \tag{A.13}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) E[\exp(i\xi x^*)] \frac{E[y \exp(i\xi z)]}{E[\exp(i\xi z)]} \exp(-i\xi \tilde{x}^*) d\xi \tag{A.14}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) E[\exp(i\xi x^*)] \frac{E[y \exp(i\xi x^*)] E[\exp(i\xi \Delta z)]}{E[\exp(i\xi x^*)] E[\exp(i\xi \Delta z)]} \times \exp(-i\xi \tilde{x}^*) d\xi \tag{A.15}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) E[y \exp(i\xi x^*)] \exp(-i\xi \tilde{x}^*) d\xi \tag{A.16}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) E[E[y|x^*] \exp(i\xi x^*)] \exp(-i\xi \tilde{x}^*) d\xi \tag{A.17}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) \left(\int E[y|x^*] f(x^*) \exp(i\xi x^*) dx^* \right) \exp(-i\xi \tilde{x}^*) d\xi \tag{A.18}$$

$$= \int h^{-1} K(h^{-1}(x^* - \tilde{x}^*)) (E[y|x^*] f(x^*)) dx^* \tag{A.19}$$

by Parseval's identity

$$= E[h^{-1} K(h^{-1}(x^* - \tilde{x}^*)) E[y|x^*]] \tag{A.20}$$

$$= E[y h^{-1} K(h^{-1}(x^* - \tilde{x}^*))], \tag{A.21}$$

$$M_0(\tilde{x}^*, h) = \frac{1}{2\pi} \int \kappa(h\xi) \phi(\xi) \exp(-i\xi \tilde{x}^*) d\xi \tag{A.22}$$

$$= \frac{1}{2\pi} \int \kappa(h\xi) \left(\int f(x^*) \exp(i\xi x^*) dx^* \right) \exp(-i\xi \tilde{x}^*) d\xi \tag{A.23}$$

$$= \int h^{-1} K(h^{-1}(x^* - \tilde{x}^*)) f(x^*) dx^* \tag{A.24}$$

by Parseval's identity

$$= E[h^{-1} K(h^{-1}(x^* - \tilde{x}^*))]. \tag{A.25}$$

■

Proof of Lemma 1. The fact that $E[v(\tilde{x}^*, h)] = 0$ follows from equation (25) and the fact that $E[\hat{m}_a(\xi)] = n^{-1} \sum_{j=1}^n E[a_j e^{i\xi z_j}] = E[a_j e^{i\xi z_j}] = m_a(\xi)$ for $a = 1, x, y$. Finally, to calculate $E[v^2(\tilde{x}^*, h)]$ we note that, by equation (25),

$$E[(\bar{M}_{k_1}(\tilde{x}^*, h) - M_{k_1}(\tilde{x}^*, h))(\bar{M}_{k_2}(\tilde{x}^*, h) - M_{k_2}(\tilde{x}^*, h))] \tag{A.26}$$

$$= E[(\bar{M}_{k_1}(\tilde{x}^*, h) - M_{k_1}(\tilde{x}^*, h))(\bar{M}_{k_2}(\tilde{x}^*, h) - M_{k_2}(\tilde{x}^*, h))^\dagger] \tag{A.27}$$

$$= \sum_{l_1=1,x,y} \sum_{l_2=1,x,y} \iint U_{l_1}^{k_1}(\xi, \tilde{x}^*, h) E[(\hat{m}_{l_1}(\xi) - m_{l_1}(\xi))(\hat{m}_{l_2}(\xi) - m_{l_2}(\xi))^\dagger] \times (U_{l_2}^{k_2}(\xi, \tilde{x}^*, h))^\dagger d\xi d\xi \tag{A.28}$$

$$= \sum_{l_1=1,x,y} \sum_{l_2=1,x,y} \iint U_{l_1}^{k_1}(\xi, \tilde{x}^*, h) V_{l_1 l_2}(\xi, \xi) (U_{l_2}^{k_2}(\xi, \tilde{x}^*, h))^\dagger d\xi d\xi \tag{A.29}$$

$$= \Sigma_{k_1 k_2}(\tilde{x}^*, h), \tag{A.30}$$

where $V_{l_1 l_2}(\xi, \xi) = E[(\hat{m}_{l_1}(\xi) - m_{l_1}(\xi))(\hat{m}_{l_2}(\xi) - m_{l_2}(\xi))^\dagger] = E[\hat{m}_{l_1}(\xi) \hat{m}_{l_2}^\dagger(\xi)] - m_{l_1}(\xi) m_{l_2}^\dagger(\xi) = E[l_{1j} l_{2j} e^{i\xi z_j} e^{-i\xi z_j}] - m_{l_1}(\xi) m_{l_2}^\dagger(\xi) = m_{(l_1 l_2)}(\xi - \xi) - m_{l_1}(\xi) m_{l_2}^\dagger(-\xi)$. Equation (31) then follows directly from squaring equation (24), taking its expectation, and using the expression for $\Sigma_{k_1 k_2}(\tilde{x}^*, h)$ just derived. ■

LEMMA 6. If a_j and z_j are sequences of i.i.d. real-valued random variables such that $E[a_j^2] < \infty$ and $E[|a_j||z_j|] < \infty$, then, for any $u, U \geq 0$ and $\varepsilon > 0$,

$$\sup_{\zeta \in [-Un^u, Un^u]} |\hat{m}_a(\zeta) - m_a(\zeta)| = o_p(n^{-1/2+\varepsilon}), \tag{A.31}$$

where $\hat{m}_a(\zeta) = n^{-1} \sum_{j=1}^n a_j \exp(i\zeta z_j)$ and $m_a(\zeta) = E[a \exp(i\zeta z)]$.

Proof. See Lemma 6 in Schennach (2004).

Proof of Lemma 2. To compute the Fréchet derivative of $\hat{M}_k(\bar{x}^*, h_n)$ for $k = 0, 1$ with respect to the estimated moment $\hat{m}_a(\zeta)$ for $a = 1, x, y$ in the vicinity of $\hat{m}_a(\zeta) = m_a(\zeta)$, we first note a few simple results. A ratio of two random functions $\hat{m}_a(\zeta)$ for $a = 1, x, y$ and $\hat{m}_1(\zeta)$ can be exactly written as

$$\frac{m_a(\zeta) + \delta\hat{m}_a(\zeta)}{m_1(\zeta) + \delta\hat{m}_1(\zeta)} = q_a(\zeta) + \delta\hat{q}_a(\zeta), \tag{A.32}$$

where $\delta m_a(\zeta) = \hat{m}_a(\zeta) - m_a(\zeta)$, $q_a(\zeta) = m_a(\zeta)/m_1(\zeta)$ and where $\delta\hat{q}_a(\zeta)$ can be written in two alternative ways: Either

$$\delta\hat{q}_a(\zeta) = \left(\frac{\delta\hat{m}_a(\zeta)}{m_1(\zeta)} - \frac{m_a(\zeta)\delta\hat{m}_1(\zeta)}{(m_1(\zeta))^2} \right) \left(1 + \frac{\delta\hat{m}_1(\zeta)}{m_1(\zeta)} \right)^{-1} \tag{A.33}$$

or

$$\delta\hat{q}_a(\zeta) = \delta_1 \hat{q}_a(\zeta) + \delta_2 \hat{q}_a(\zeta), \tag{A.34}$$

$$\delta_1 \hat{q}_a(\zeta) = \frac{\delta\hat{m}_a(\zeta)}{m_1(\zeta)} - \frac{m_a(\zeta)\delta\hat{m}_1(\zeta)}{(m_1(\zeta))^2}, \tag{A.35}$$

$$\begin{aligned} \delta_2 \hat{q}_a(\zeta) &= \frac{m_a(\zeta)}{m_1(\zeta)} \left(\frac{\delta\hat{m}_1(\zeta)}{m_1(\zeta)} \right)^2 \left(1 + \frac{\delta\hat{m}_1(\zeta)}{m_1(\zeta)} \right)^{-1} \\ &\quad - \frac{\delta\hat{m}_a(\zeta)}{m_1(\zeta)} \frac{\delta\hat{m}_1(\zeta)}{m_1(\zeta)} \left(1 + \frac{\delta\hat{m}_1(\zeta)}{m_1(\zeta)} \right)^{-1}. \end{aligned} \tag{A.36}$$

Similarly, for $Q_x(\xi) = \int_0^\xi [im_x(\zeta)/m_1(\zeta)] d\zeta$, $\delta\hat{Q}_x(\xi) = \int_0^\xi [i\hat{m}_x(\zeta)/\hat{m}_1(\zeta)] d\zeta - \int_0^\xi [im_x(\zeta)/m_1(\zeta)] d\zeta$, and some random function $\delta\bar{Q}_x(\xi)$ such that $|\delta\bar{Q}_x(\xi)| \leq |\delta\hat{Q}_x(\xi)|$,

$$\exp(Q_x(\xi) + \delta\hat{Q}_x(\xi)) = \exp(Q_x(\xi)) \left(1 + \delta\bar{Q}_x(\xi) + \frac{1}{2} \exp(\delta\bar{Q}_x(\xi)) (\delta\hat{Q}_x(\xi))^2 \right). \tag{A.37}$$

Substituting expansions (A.32) and (A.37) into

$$\begin{aligned} \hat{M}_k(\bar{x}^*, h_n) - M_k(\bar{x}^*, h_n) &= \frac{1}{2\pi} \int \kappa(h_n \xi) \left(\frac{\hat{m}_y^k(\xi)}{\hat{m}_1(\xi)} \exp \left(\int_0^\xi \frac{i\hat{m}_x(\zeta)}{\hat{m}_1(\zeta)} d\zeta \right) \right. \\ &\quad \left. - \frac{m_y^k(\xi)}{m_1(\xi)} \exp \left(\int_0^\xi \frac{im_x(\zeta)}{m_1(\zeta)} d\zeta \right) \right) d\xi \end{aligned} \tag{A.38}$$

for $k = 0, 1$ and keeping the terms linear in $\delta\hat{m}_1(\zeta)$, $\delta\hat{m}_x(\zeta)$, or $\delta\hat{m}_y(\zeta)$ gives the linearization of $\hat{M}_k(\tilde{x}^*, h_n)$, denoted $\bar{M}_k(\tilde{x}^*, h_n)$:

$$\begin{aligned} & \bar{M}_0(\tilde{x}^*, h_n) - M_0(\tilde{x}^*, h_n) \\ &= \frac{1}{2\pi} \int \kappa(h_n \xi) \frac{m_{y^k}(\xi)}{m_1(\xi)} \exp\left(\int_0^\xi \frac{im_x(\zeta)}{m_1(\zeta)} d\zeta\right) \\ & \quad \times \int_0^\xi \left(\frac{i\delta\hat{m}_x(\zeta)}{m_1(\zeta)} - \frac{im_x(\zeta)\delta\hat{m}_1(\zeta)}{(m_1(\zeta))^2}\right) d\zeta d\xi \\ & \quad + \frac{1}{2\pi} \int \kappa(h_n \xi) \left(\frac{\delta\hat{m}_{y^k}(\xi)}{m_1(\xi)} - \frac{m_{y^k}(\xi)\delta\hat{m}_1(\xi)}{(m_1(\xi))^2}\right) \exp\left(\int_0^\xi \frac{im_x(\zeta)}{m_1(\zeta)} d\zeta\right) d\xi \quad (\text{A.39}) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2\pi} \int \kappa(h_n \xi) \phi_k(\xi) \int_0^\xi \left(\frac{i\delta\hat{m}_x(\zeta)}{m_1(\zeta)} - \frac{im_x(\zeta)\delta\hat{m}_1(\zeta)}{(m_1(\zeta))^2}\right) d\zeta d\xi \\ & \quad + \frac{1}{2\pi} \int \kappa(h_n \xi) \left(\frac{\delta\hat{m}_{y^k}(\xi)}{m_1(\xi)} \phi_0(\xi) - \frac{\delta\hat{m}_1(\xi)}{m_1(\xi)} \phi_k(\xi)\right) d\xi. \quad (\text{A.40}) \end{aligned}$$

By making use of the identity $\int_{-\infty}^\infty \int_0^\xi f(\xi, \zeta) d\zeta d\xi = \int_0^\infty \int_\xi^\infty f(\xi, \zeta) d\xi d\zeta + \int_{-\infty}^0 \int_\xi^- f(\xi, \zeta) d\xi d\zeta$ for any absolutely integrable function f , we obtain

$$\begin{aligned} \bar{M}_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, h_n) &= \int A_k(\xi, \tilde{x}^*, h_n) \left(\frac{i\delta\hat{m}_x(\xi)}{m_1(\xi)} - \frac{im_x(\xi)\delta\hat{m}_1(\xi)}{(m_1(\xi))^2}\right) d\xi \\ & \quad + \frac{1}{2\pi} \int C_0(\xi, \tilde{x}^*, h_n) \frac{\delta\hat{m}_{y^k}(\xi)}{m_1(\xi)} d\xi \\ & \quad - \frac{1}{2\pi} \int C_k(\xi, \tilde{x}^*, h_n) \frac{\delta\hat{m}_1(\xi)}{m_1(\xi)} d\xi \quad (\text{A.41}) \end{aligned}$$

$$= \sum_{l=1, x, y} \int U_l^k(\xi, \tilde{x}^*, h) \delta\hat{m}_l(\xi) d\xi. \quad (\text{A.42})$$

The order of $n^{-1/2}(\bar{M}_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, h_n))$ (in probability) can be found through its variance $\Sigma_{kk}(\tilde{x}^*, h)$ given by Lemma 1:

$$\Sigma_{kk}(\tilde{x}^*, h) = \sum_{l_1=1, x, y} \sum_{l_2=1, x, y} \iint U_{l_1}^k(\zeta, \tilde{x}^*, h) V_{l_1 l_2}(\zeta, \xi) (U_{l_2}^k(\xi, \tilde{x}^*, h))^\dagger d\zeta d\xi \quad (\text{A.43})$$

$$\leq \sum_{l_1=1, x, y} \sum_{l_2=1, x, y} \iint |U_{l_1}^k(\zeta, \tilde{x}^*, h)| |V_{l_1 l_2}(\zeta, \xi)| |U_{l_2}^k(\xi, \tilde{x}^*, h)| d\zeta d\xi, \quad (\text{A.44})$$

where, by Assumptions 5 and 6,

$$|V_{l_1 l_2}(\zeta, \nu)| = |m_{(l_1 l_2)}(\zeta - \xi) - m_{l_1}(\zeta) m_{l_2}(-\xi)| \quad (\text{A.45})$$

$$\leq E[|l_1 l_2| |e^{i(\zeta - \xi)z}|] + E[|l_1| |e^{i\zeta z}|] E[|l_2| |e^{-i\xi z}|] \leq 1. \quad (\text{A.46})$$

It follows that

$$\Sigma_{kk}(\tilde{x}^*, h) \leq \sum_{l_1=1,x,y} \sum_{l_2=1,x,y} \iint |U_{l_1}^k(\zeta, \tilde{x}^*, h)| |U_{l_2}^k(\xi, \tilde{x}^*, h)| d\zeta d\xi \tag{A.47}$$

$$= \left(\sum_{l=1,x,y} \int |U_l^k(\zeta, \tilde{x}^*, h)| d\zeta \right)^2 \tag{A.48}$$

and therefore that

$$\begin{aligned} \bar{M}_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, h_n) &= O_p \left(n^{-1/2} \sum_{l=1,x,y} \int |U_l^k(\zeta, \tilde{x}^*, h)| d\zeta \right) \\ &= O_p(n^{-1/2} \bar{U}(h_n)), \end{aligned} \tag{A.49}$$

where $\bar{U}(h_n)$, given in the statement of the lemma, has been explicitly constructed to bound any of the $\int |U_l^k(\zeta, \tilde{x}^*, h)| d\zeta$ terms (up to a multiplicative constant). By equation (30), equation (A.49) implies equation (36) in the statement of the lemma, provided that Assumption 7 holds.

To establish equation (37), we substitute expansions (A.32) and (A.37) into

$$\begin{aligned} \hat{M}_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, h_n) \\ = \int \kappa(h_n \xi) \left(\frac{\hat{m}_{y,k}(\xi)}{\hat{m}_1(\xi)} \exp \left(\int_0^\xi \frac{i \hat{m}_x(\zeta)}{\hat{m}_1(\zeta)} d\zeta \right) - \phi_k(\xi) \right) d\xi \end{aligned} \tag{A.50}$$

for $k = 0, 1$ and remove the terms linear in $\delta \hat{m}_a(\zeta)$ for $a = 1, x, y$. We then find that $\hat{M}_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, h_n)$ can be written as $|\hat{M}_k(\tilde{x}^*, h_n) - \bar{M}_k(\tilde{x}^*, h_n)| \leq \sum_{j=1}^7 R_j$ where

$$R_1 = \int_0^\infty |\kappa(h_n \xi)| |\delta_1 \hat{q}_{y,k}(\xi)| |\phi_1(\xi)| \left(\int_0^\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta \right) d\xi, \tag{A.51}$$

$$R_2 = \int_0^\infty |\kappa(h_n \xi)| |\delta_2 \hat{q}_{y,k}(\xi)| |\phi_1(\xi)| d\xi, \tag{A.52}$$

$$R_3 = \int_0^\infty |\kappa(h_n \xi)| |\delta_2 \hat{q}_{y,k}(\xi)| |\phi_1(\xi)| \int_0^\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta d\xi, \tag{A.53}$$

$$R_4 = \int_0^\infty |\kappa(h_n \xi)| |q_{y,k}(\xi)| |\phi_1(\xi)| \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi, \tag{A.54}$$

$$R_5 = \int_0^\infty |\kappa(h_n \xi)| |\delta \hat{q}_{y,k}(\xi)| |\phi_1(\xi)| \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi, \tag{A.55}$$

$$R_6 = \int_0^\infty |\kappa(h_n \xi)| |q_{y,k}(\xi)| |\phi_1(\xi)| \frac{1}{2} \exp(|\delta \bar{Q}_x(\xi)|) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi, \tag{A.56}$$

$$R_7 = \int_0^\infty |\kappa(h_n \xi)| |\delta \hat{q}_{y,k}(\xi)| |\phi_1(\xi)| \frac{1}{2} \exp(|\delta \bar{Q}_x(\xi)|) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi. \tag{A.57}$$

These terms can then be bounded in terms of $\lambda(h_n)$, $\bar{U}(h_n)$ (given in the statement of the lemma), and $\hat{\Phi}_n = \max_{b=1,x,y} \sup_{\xi \in [-h_n^{-1}, h_n^{-1}]} |\hat{m}_b(\xi) - m_b(\xi)|$, where the supremum can be taken over $[-h_n^{-1}, h_n^{-1}]$ because $\kappa(h_n \xi)$ vanishes outside that interval. By Lemma 6, $\hat{\Phi}_n = O_p(n^{-1/2+\epsilon})$ for any $\epsilon > 0$. Also, we note that $\sup_{\xi \in [-h_n^{-1}, h_n^{-1}]} \hat{\Phi}_n / |m_1(\xi)| \leq \hat{\Phi}_n \lambda(h_n^{-1}) = n^{-1/2+\epsilon} \lambda(h_n^{-1}) = o_p(1)$. Now, for $k = 0, 1$, we have

$$R_1 \leq \int_0^\infty |\kappa(h_n \xi)| \left(\frac{1}{|m_1(\xi)|} + \frac{|m_{y^k}(\xi)|}{|m_1(\xi)|^2} \right) \hat{\Phi}_n |\phi_0(\xi)| \left(\int_0^\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta \right) d\xi \tag{A.58}$$

$$\leq \lambda(h_n) \hat{\Phi}_n \int_0^\infty |\kappa(h_n \xi)| \left(1 + \frac{|m_{y^k}(\xi)|}{|m_1(\xi)|} \right) |\phi_0(\xi)| \left(\int_0^\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta \right) d\xi \tag{A.59}$$

$$= \lambda(h_n) \hat{\Phi}_n \int_0^\infty \int_\xi^\infty |\kappa(h_n \xi)| \left(1 + \frac{|m_{y^k}(\xi)|}{|m_1(\xi)|} \right) |\phi_0(\xi)| d\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta \tag{A.60}$$

$$= \lambda(h_n) \hat{\Phi}_n \int_0^\infty \int_\xi^\infty |\kappa(h_n \xi)| (|\phi_0(\xi)| + |\phi_k(\xi)|) d\xi |\delta_1 \hat{q}_x(\zeta)| d\zeta \tag{A.61}$$

$$\leq \lambda(h_n) \hat{\Phi}_n^2 \int_0^\infty \int_\xi^\infty |\kappa(h_n \xi)| (|\phi_0(\xi)| + |\phi_k(\xi)|) d\xi \left(1 + \frac{|m_x(\zeta)|}{|m_1(\zeta)|} \right) \frac{1}{|m_1(\zeta)|} d\zeta \tag{A.62}$$

$$\leq \lambda(h_n) \hat{\Phi}_n^2 \bar{U}(h_n). \tag{A.63}$$

The remaining terms can be similarly bounded:

$$R_2 \leq \int_0^\infty |\kappa(h_n \xi)| \left| \frac{|m_{y^k}(\xi)|}{|m_1(\xi)|^2} \frac{1}{|m_1(\xi)|} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} + \frac{1}{|m_1(\xi)|^2} \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} |\phi_0(\xi)| \right| d\xi \tag{A.64}$$

$$\leq \lambda(h_n) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \int_0^\infty |\kappa(h_n \xi)| \frac{1}{|m_1(\xi)|} \left| \frac{|m_{y^k}(\xi)|}{|m_1(\xi)|} + 1 \right| |\phi_0(\xi)| d\xi \tag{A.65}$$

$$\leq \lambda(h_n) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \left(\int_0^\infty \frac{|\kappa(h_n \xi)| |\phi_k(\xi)|}{|m_1(\xi)|} d\xi + \int_0^\infty \frac{|\kappa(h_n \xi)| |\phi_0(\xi)|}{|m_1(\xi)|} d\xi \right) \tag{A.66}$$

$$= \lambda(h_n) \hat{\Phi}_n^2 \bar{U}(h_n) (1 + o_p(1)), \tag{A.67}$$

$$R_3 \leq \lambda(h_n) \hat{\Phi}_n \int_0^\infty |\kappa(h_n \xi)| |\delta_2 \hat{q}_y(\xi)| |\phi_0(\xi)| d\xi \tag{A.68}$$

$$= \lambda(h_n) \hat{\Phi}_n R_2 = o_p(1) R_2, \tag{A.69}$$

$$R_4 = \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi \tag{A.70}$$

$$\leq \lambda(h_n) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \int_0^\infty \frac{\int_\zeta^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| d\xi}{|m_1(\zeta)|} d\zeta \tag{A.71}$$

$$= \lambda(h_n) \hat{\Phi}_n^2 (1 + o_p(1)) \bar{U}(h_n), \tag{A.72}$$

$$R_5 \leq \int_0^\infty |\kappa(h_n \xi)| \left(\frac{1}{|m_1(\zeta)|} + \frac{|m_{y^k}(\zeta)|}{|m_1(\zeta)|^2} \right) \hat{\Phi}_n |1 + o_p(1)|^{-1} |\phi_0(\xi)| \times \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi \tag{A.73}$$

$$= \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} \int_0^{h_n^{-1}} |\kappa(h_n \xi)| \left(1 + \frac{|m_{y^k}(\zeta)|}{|m_1(\zeta)|} \right) |\phi_0(\xi)| \times \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi \tag{A.74}$$

$$= \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} \left(\int_0^\infty |\kappa(h_n \xi)| |\phi_0(\xi)| \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi + \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \int_0^\xi |\delta_2 \hat{q}_x(\zeta)| d\zeta d\xi \right) \tag{A.75}$$

$$= \lambda(h_n) \hat{\Phi}_n (1 + o_p(1)) R_4 = o_p(1) R_4, \tag{A.76}$$

$$R_6 \leq \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \frac{1}{2} \exp \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi \tag{A.77}$$

$$\leq \frac{1}{2} \exp(o_p(1)) \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) d\xi \tag{A.78}$$

$$\leq \frac{1}{2} \exp(o_p(1)) \lambda(h_n) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \times \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \left(\int_0^\xi \left(\frac{1}{|m_1(\zeta)|} + \frac{|m_x(\zeta)|}{|m_1(\zeta)|^2} \right) d\zeta \right) d\xi \tag{A.79}$$

$$= \frac{1}{2} \exp(o_p(1)) \lambda(h_n) \hat{\Phi}_n^2 |1 + o_p(1)|^{-1} \times \int_0^\infty \int_\zeta^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| d\xi \left(\frac{1}{|m_1(\zeta)|} + \frac{|m_x(\zeta)|}{|m_1(\zeta)|^2} \right) d\zeta \tag{A.80}$$

$$= O_p(1) \lambda(h_n) \hat{\Phi}_n^2 \bar{U}(h_n), \tag{A.81}$$

$$R_7 \leq \int_0^\infty |\kappa(h_n \xi)| \left(1 + \frac{|m_{y^k}(\zeta)|}{|m_1(\zeta)|} \right) \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} |\phi_0(\xi)| \times \frac{1}{2} \exp \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi \tag{A.82}$$

$$\leq \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} \times \int_0^\infty |\kappa(h_n \xi)| \left(1 + \frac{|m_{y^k}(\zeta)|}{|m_1(\zeta)|} \right) |\phi_0(\xi)| \times \exp \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi \tag{A.83}$$

$$\leq \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} \int_0^\infty |\kappa(h_n \xi)| |\phi_0(\xi)| \times \exp \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi + A \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} \int_0^\infty |\kappa(h_n \xi)| |\phi_k(\xi)| \times \exp \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right) \left(\int_0^\xi |\delta \hat{q}_x(\zeta)| d\zeta \right)^2 d\xi \tag{A.84}$$

$$\leq \lambda(h_n) \hat{\Phi}_n |1 + o_p(1)|^{-1} R_6 \tag{A.85}$$

$$= o_p(1) R_6. \tag{A.86}$$

It then follows that $|\hat{M}_k(\tilde{x}^*, h_n) - \bar{M}_k(\tilde{x}^*, h_n)| \leq \lambda(h_n) \hat{\Phi}_n^2 \bar{U}(h_n) = O_p(\lambda(h_n) \times \bar{U}(h_n) n^{-1+\delta})$ for some $\delta > 0$. By a standard Taylor expansion of the ratio $\hat{M}_1(\tilde{x}^*, h_n) / \hat{M}_0(\tilde{x}^*, h_n)$ around $\hat{M}_k(\tilde{x}^*, h_n) = M_k(\tilde{x}^*, h_n)$ for $k = 0, 1$, we have

$$\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n) = \frac{\hat{M}_1(\tilde{x}^*, h_n)}{\hat{M}_0(\tilde{x}^*, h_n)} - \frac{\bar{M}_1(\tilde{x}^*, h_n)}{\bar{M}_0(\tilde{x}^*, h_n)} \tag{A.87}$$

$$= \frac{\hat{M}_1(\tilde{x}^*, h_n) - \bar{M}_1(\tilde{x}^*, h_n)}{\hat{M}_0} - \frac{\dot{M}_1(\hat{M}_0(\tilde{x}^*, h_n) - \bar{M}_0(\tilde{x}^*, h_n))}{\dot{M}_0^2} \tag{A.88}$$

for some \dot{M}_k lying between $\hat{M}_k(\tilde{x}^*, h_n)$ and $\bar{M}_k(\tilde{x}^*, h_n)$. Because (i) we have just shown that $\hat{M}_k(\tilde{x}^*, h_n) \xrightarrow{p} \bar{M}_k(\tilde{x}^*, h_n)$ and $\bar{M}_k(\tilde{x}^*, h_n) \xrightarrow{p} M_k(\tilde{x}^*, h_n)$, (ii) $M_k(\tilde{x}^*, h_n) \rightarrow M_k(\tilde{x}^*, 0)$ by assumption, and (iii) $M_1(\tilde{x}^*, 0)$ is bounded and $M_0(\tilde{x}^*, 0)$ is bounded away from zero by assumption, it follows that \hat{M}_0^{-1} and \dot{M}_1 / \dot{M}_0^2 converge in probability to finite quantities and therefore $\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n)$ is of the same order as $\hat{M}_k(\tilde{x}^*, h_n) - \bar{M}_k(\tilde{x}^*, h_n)$, thus implying equation (37).

To establish the second conclusion of the theorem, we note that, because $n^{1/2}(\Omega(\tilde{x}^*, h_n))^{-1/2}(\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)) = O_p(1)$, we can write

$$(\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n)) = O_p(1)O_p(\bar{U}(h_n)n^{-1/2}\lambda(h_n)n^{-1/2+\epsilon}) \tag{A.89}$$

$$= n^{1/2}(\Omega(\tilde{x}^*, h_n))^{-1/2}(\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))O_p(\bar{U}(h_n)n^{-1/2}\lambda(h_n)n^{-1/2+\epsilon}) \tag{A.90}$$

$$= (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))O_p((\Omega(\tilde{x}^*, h_n))^{-1/2}\bar{U}(h_n)\lambda(h_n)n^{-1/2+\epsilon}). \tag{A.91}$$

Then,

$$\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)) + (\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n)) \tag{A.92}$$

$$= (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)) + (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)) \times O_p((\Omega(\tilde{x}^*, h_n))^{-1/2}\bar{U}(h_n)\lambda(h_n)n^{-1/2+\epsilon}) \tag{A.93}$$

$$= (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))(1 + O_p((\Omega(\tilde{x}^*, h_n))^{-1/2}\bar{U}(h_n)\lambda(h_n)n^{-1/2+\epsilon})) \tag{A.94}$$

$$= (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))(1 + o_p(1)) \tag{A.95}$$

because $(\Omega(\tilde{x}^*, h_n))^{-1/2}\bar{U}(h_n)\lambda(h_n)n^{-1/2+\epsilon} \rightarrow 0$ by assumption. ■

Proof of Lemma 3. First, by equation (13), we have, for $k = 0, 1$,

$$M_k(\tilde{x}^*, h) - M_k(\tilde{x}^*, 0) = \frac{1}{2\pi} \int (\kappa(h\xi) - 1)\phi_k(\xi)\exp(-i\xi\tilde{x}^*) d\xi. \tag{A.96}$$

Expanding the Fourier transform of the kernel in a Taylor series up to order γ , we obtain

$$\begin{aligned} &|M_k(\tilde{x}^*, h_n) - M_k(\tilde{x}^*, 0)| \\ &= \left| 2 \int_0^\infty (1 - \kappa(h_n\xi))\phi_k(\xi)\exp(-i\xi\tilde{x}^*) d\xi \right| \tag{A.97} \\ &= \left| 2 \int_0^\infty \left(\sum_{i=1}^{\gamma-1} \frac{1}{i!} \kappa^{(i)}(0)(h_n\xi)^i + \frac{1}{\gamma!} \kappa^{(\gamma)}(\tilde{\xi})(h_n\xi)^\gamma \right) \phi_k(\xi)\exp(-i\xi\tilde{x}^*) d\xi \right| \\ &\qquad\qquad\qquad \text{for some } \tilde{\xi} \in [0, \xi]. \tag{A.98} \end{aligned}$$

Now let the order γ be chosen as follows. If α_ϕ (defined in Assumption 8) is nonzero, then let $\gamma = \gamma_\kappa$, the order of the kernel. If $\alpha_\phi = 0$, then let γ be the largest integer such that $\gamma \leq \gamma_\kappa$ and $\gamma < -\gamma_\phi - 1$. With this choice of γ , equation (A.98) simplifies to

$$|M_k(\tilde{x}^*, h) - M_k(\tilde{x}^*, 0)| = h_n^\gamma \left| 2 \int_0^\infty \frac{1}{\gamma!} \kappa^{(\gamma)}(\tilde{\xi})\xi^\gamma \phi_k(\xi)\exp(-i\xi\tilde{x}^*) d\xi \right| \tag{A.99}$$

because all terms where $i < \gamma$ vanish, by the definition of the order of a kernel. Furthermore, we have

$$|M_k(\tilde{x}^*, h) - M_k(\tilde{x}^*, 0)| \leq h_n^\gamma \frac{2}{\gamma!} \int_0^\infty |\kappa^{(\gamma)}(\tilde{\xi})| |\xi|^\gamma |\phi_k(\xi)| d\xi \tag{A.100}$$

$$\leq h_n^\gamma \frac{2}{\gamma!} \left(\max_{\tilde{\xi} \in \mathbb{R}} |\kappa^{(\gamma)}(\tilde{\xi})| \right) \int_0^\infty |\xi|^\gamma |\phi_k(\xi)| d\xi, \tag{A.101}$$

where $\max_{\tilde{\xi} \in \mathbb{R}} |\kappa^{(\gamma)}(\tilde{\xi})|$ is finite by equation (44) of Assumption 9. The term $\int_0^\infty |\xi|^\gamma |\phi_k(\xi)| d\xi$ is finite also, because our choice of γ guarantees that the integrand decays to zero faster than ξ^{-1} . Then, by a standard Taylor expansion of the ratio $M_1(\tilde{x}^*, h_n)/M_0(\tilde{x}^*, h_n)$ around $M_1(\tilde{x}^*, h_n) = M_1(\tilde{x}^*, 0)$ and $M_0(\tilde{x}^*, h_n) = M_0(\tilde{x}^*, 0)$, the convergence rate of $g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)$ is $O(h_n^\gamma)$ also. ■

LEMMA 7. For $\zeta \geq 0$, if $\gamma > 0$, $\alpha < 0$, $\beta > 0$ or if $\gamma \in \mathbb{R}$, $\alpha = \beta = 0$, then

$$\int_\zeta^\infty (1 + \xi)^\gamma \exp(\alpha\xi^\beta) d\xi \leq (1 + \xi)^{1+\gamma} \exp(\alpha\xi^\beta). \tag{A.102}$$

Proof. The case where $\alpha = \beta = 0$ is trivial. If $\alpha < 0$ and $\beta > 0$, Lemma 4.2 in Li and Vuong (1998) shows that, for $\gamma > 0$ and $\zeta > 0$, $\int_\zeta^\infty \xi^\gamma \exp(\alpha\xi^\beta) d\xi = O(\xi^{1+\gamma-\beta} \exp(\alpha\xi^\beta))$, thus implying the result because $\xi^{1+\gamma-\beta} \exp(\alpha\xi^\beta) \leq \xi^{1+\gamma} \exp(\alpha\xi^\beta)$. ■

Proof of Lemma 4. From equation (A.96), we have

$$|M_k(\tilde{x}^*, h) - M_k(\tilde{x}^*, 0)| = \left| 2 \int_0^\infty (1 - \kappa(h\xi)) \phi_k(\xi) \exp(-i\xi\tilde{x}^*) d\xi \right| \tag{A.103}$$

$$= \left| 2 \int_{\bar{\xi}h^{-1}}^\infty (1 - \kappa(h\xi)) \phi_k(\xi) \exp(-i\xi\tilde{x}^*) d\xi \right| \tag{A.104}$$

$$\leq 2 \int_{\bar{\xi}h^{-1}}^\infty (1 + |\kappa(h\xi)|) |\phi_k(\xi)| d\xi \tag{A.105}$$

$$\leq \int_{\bar{\xi}h^{-1}}^\infty |\phi_k(\xi)| d\xi \quad \text{by the boundedness of } \kappa(h\xi) \tag{A.106}$$

$$\leq \int_{\bar{\xi}h^{-1}}^\infty (1 + \xi)^{\gamma_\phi} \exp(\alpha_\phi \xi^{\beta_\phi}) d\xi \tag{A.107}$$

$$\leq (1 + \bar{\xi}h^{-1})^{\gamma_\phi+1} \exp(\alpha_\phi (\bar{\xi}h^{-1})^{\beta_\phi}). \tag{A.108}$$

Then, by a standard Taylor expansion of the ratio $M_1(\tilde{x}^*, h_n)/M_0(\tilde{x}^*, h_n)$ around $M_1(\tilde{x}^*, h_n) = M_1(\tilde{x}^*, 0)$ and $M_0(\tilde{x}^*, h_n) = M_0(\tilde{x}^*, 0)$, the convergence rate of $g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)$ is $O((1 + h^{-1})^{\gamma_\phi+1} \exp(\alpha_\phi (\bar{\xi}h^{-1})^{\beta_\phi}))$ also. ■

LEMMA 8. For $\zeta \geq 0$, if $\beta \geq 0$ and if $(1 + \xi)^\gamma \exp(\alpha\xi^\beta)$ is increasing in ξ ,

$$\int_0^\zeta (1 + \xi)^\gamma \exp(\alpha\xi^\beta) d\xi \leq (1 + \xi)^{1+\gamma} \exp(\alpha\xi^\beta). \tag{A.109}$$

Proof. $\int_0^\zeta (1 + \xi)^\gamma \exp(\alpha \xi^\beta) d\xi \leq \zeta (1 + \zeta)^\gamma \exp(\alpha \zeta^\beta) \leq (1 + \zeta)^{1+\gamma} \exp(\alpha \zeta^\beta)$ ■

Proof of Lemma 5. By Lemma 2, the order of the variance term is $O_p(\bar{U}(h_n)n^{-1/2})$, where

$$\begin{aligned} \bar{U}(h_n) &= \int |m_1(\zeta)|^{-1} \sum_{k=0}^1 \left(\left(1 + \frac{|\phi'_0(\zeta)|}{|\phi_0(\zeta)|} \right) \right. \\ &\quad \left. \times \int_{|\xi|}^\infty |\kappa(-h_n \xi)| |\phi_k(\xi)| d\xi + |\kappa(-h_n \zeta)| |\phi_k(\zeta)| \right) d\zeta, \end{aligned} \tag{A.110}$$

where, for $k = 0, 1$,

$$(2\pi)^{-1} \int_\zeta^{-\infty} |\kappa(-h\xi)| |\phi_k(\xi)| d\xi \leq (2\pi)^{-1} 1(|\zeta| \leq h^{-1}) \int_\zeta^{-\infty} |\phi_k(\xi)| d\xi \tag{A.111}$$

$$\leq 1(|\zeta| \leq h^{-1}) \left\{ \int_\zeta^{-\infty} (1 + |\xi|)^{\gamma_\phi} \exp(\alpha_\phi |\xi|^{\beta_\phi}) d\xi \right\} \tag{A.112}$$

$$= 1(|\zeta| \leq h^{-1}) (1 + |\zeta|)^{1+\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) \tag{A.113}$$

and

$$(2\pi)^{-1} |\kappa(-h\zeta)| |\phi_k(\zeta)| \leq 1(|\zeta| \leq h^{-1}) (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}). \tag{A.114}$$

It follows that

$$\begin{aligned} \bar{U}(h_n) &\leq \int (1 + |\zeta|)^{-\gamma_m} \exp(-\alpha_m |\zeta|^{\beta_m}) \\ &\quad \times \sum_{k=0}^1 \left((1 + (1 + |\zeta|)^{\gamma_r}) 1(|\zeta| \leq h_n^{-1}) (1 + |\zeta|)^{1+\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) \right. \\ &\quad \left. + 1(|\zeta| \leq h_n^{-1}) (1 + |\zeta|)^{\gamma_\phi} \exp(\alpha_\phi |\zeta|^{\beta_\phi}) \right) d\zeta \end{aligned} \tag{A.115}$$

$$\leq \int 1(|\zeta| \leq h_n^{-1}) (1 + |\zeta|)^{1+\gamma_\phi+\gamma_r-\gamma_m} \exp(\alpha_\phi |\zeta|^{\beta_\phi} - \alpha_m |\zeta|^{\beta_m}) d\zeta \tag{A.116}$$

$$\leq (1 + h_n^{-1})^{2+\gamma_\phi+\gamma_r-\gamma_m} \exp(\alpha_\phi h_n^{-\beta_\phi} - \alpha_m h_n^{-\beta_m}). \tag{A.117}$$

Hence, $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = O_p(n^{-1/2}(1 + h_n^{-1})^{2+\gamma_\phi+\gamma_r-\gamma_m} \exp(\alpha_\phi h_n^{-\beta_\phi} - \alpha_m h_n^{-\beta_m}))$. ■

Proof of Theorem 2. We make use of the order of the bias $g(\tilde{x}^*, h_n) - g(\tilde{x}^*, 0)$ provided by Lemmas 3 and 4 and of the order of the variance term $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)$ provided by Lemma 5. To check that the higher order term $\hat{g}(\tilde{x}^*, h_n) - \bar{g}(\tilde{x}^*, h_n)$ does not affect the rates obtained by considering the first-order terms only, we observe that,

by Lemma 2, the upper bound on $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)$ provided by Lemma 5 holds for $\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)$ also, if we can show that $\lambda(h_n)n^{-(1/2)+\epsilon} = o(1)$ for some ϵ .

We consider each subcase of the theorem separately. Let R_n be such that $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = O_p(R_n)$. Throughout the proof, let $\epsilon, \epsilon_1, \epsilon_2, \dots$ denote arbitrarily small positive numbers.

Case 1. $\beta_v > \beta_b$. If the bandwidth h_n is chosen to be

$$(h_n^{-1})^{\beta_v} = \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right) + \frac{(\alpha_b + 2\epsilon_b)}{(\alpha_v + \epsilon_v)} \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \tag{A.118}$$

for some $\epsilon_v, \epsilon_b > 0$, the bias and the variance are of the same order and the convergence rate is

$$R_n \leq (h^{-1})^{\gamma_b} \exp(\alpha_b(h_n^{-1})^{\beta_b}) + n^{-1/2}(h^{-1})^{\gamma_v} \exp(\alpha_v(h_n^{-1})^{\beta_v}) \tag{A.119}$$

$$\leq \exp((\alpha_b + \epsilon_b)(h_n^{-1})^{\beta_b}) + n^{-1/2} \exp((\alpha_v + \epsilon_v)(h_n^{-1})^{\beta_v}) \tag{A.120}$$

$$\begin{aligned} &\leq \exp\left((\alpha_b + \epsilon_b) \left(\left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right) - \frac{1}{(\alpha_v + \epsilon_v)} \left(\frac{(\alpha_b - \epsilon_b)^{\beta_v/\beta_b}}{2(\alpha_v + \epsilon_v)} \ln n \right)^{\beta_b/\beta_v} \right)^{\beta_b/\beta_v} \right) \\ &\quad + n^{-1/2} \exp\left(\left(\frac{1}{2} \ln n \right) + (\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) \end{aligned} \tag{A.121}$$

$$\begin{aligned} &\leq \exp\left((\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) \\ &\quad + n^{-1/2} \exp\left(\left(\frac{1}{2} \ln n \right) + (\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) \end{aligned} \tag{A.122}$$

$$\begin{aligned} &= \exp\left((\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) + \exp\left((\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) \\ &\tag{A.123} \end{aligned}$$

$$= 2 \exp\left((\alpha_b + 2\epsilon_b) \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} \right)^{\beta_b/\beta_v} \right) \tag{A.124}$$

$$\leq \exp\left((\alpha_b + \epsilon) \left(\frac{\ln n}{2\alpha_v} \right)^{\beta_b/\beta_v} \right) \text{ for some } \epsilon > 0. \tag{A.125}$$

Now, to check the negligibility of the higher-order terms, we verify that $\lambda(h_n)n^{-1/2+\epsilon_1} = o(1)$ for some suitably chosen $\epsilon_1 > 0$. Noting that $\beta_v = \beta_m$ and $\alpha_v = -\alpha_m$ if $\beta_v > \beta_b$, we have

$$\lambda(h_n)n^{-1/2+\epsilon_1} = (1 + h_n^{-1})^{1+\gamma_r-\gamma_m} \exp(-\alpha_m(h_n^{-1})^{\beta_m})n^{-1/2+\epsilon_1} \tag{A.126}$$

$$\leq \exp((-\alpha_m + \epsilon_2)(h_n^{-1})^{\beta_m})n^{-1/2+\epsilon_1} \quad \text{for any } \epsilon_2 > 0 \tag{A.127}$$

$$\leq \exp\left(\frac{(-\alpha_m + \epsilon_2)\ln n}{2(\alpha_v + \epsilon_v)}\right)n^{-1/2+\epsilon_1} \tag{A.128}$$

$$\leq \exp\left(\frac{(-\alpha_m + \epsilon_2)\ln n}{2(-\alpha_m + \epsilon_v)}\right)n^{-1/2+\epsilon_1} \tag{A.129}$$

$$\leq n^{-\epsilon_3+1/2}n^{-1/2+\epsilon_1} \quad \text{for some } \epsilon_3 > 0 \tag{A.130}$$

$$= n^{\epsilon_1-\epsilon_3} \tag{A.131}$$

$$= o(1) \quad \text{if } \epsilon_1 < \epsilon_3. \tag{A.132}$$

Case 2. $\beta_b = 0$ (and $\gamma_b < 0$) and $\beta_v > 0$. For some $\epsilon_v > 0$, let

$$(h_n^{-1})^{\beta_v} = \frac{\ln n}{2(\alpha_v + \epsilon_v)} + \frac{\gamma_b \ln \ln n}{(\alpha_v + \epsilon_v)}. \tag{A.133}$$

Then,

$$R_n = (h_n^{-1})^{\gamma_b} + n^{-1/2}(h_n^{-1})^{\gamma_v} \exp(\alpha_v(h_n^{-1})^{\beta_v}) \tag{A.134}$$

$$\leq (h_n^{-1})^{\gamma_b} + n^{-1/2} \exp((\alpha_v + \epsilon_v)(h_n^{-1})^{\beta_v}) \tag{A.135}$$

$$\leq \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} + \frac{\gamma_b \ln \ln n}{(\alpha_v + \epsilon_v)}\right)^{\gamma_b/\beta_b} + n^{-1/2} \exp\left(\frac{1}{2} \ln n + \gamma_b \ln \ln n\right) \tag{A.136}$$

$$\leq \left(\frac{\ln n}{2(\alpha_v + \epsilon_v)} + \frac{\gamma_b \ln \ln n}{(\alpha_v + \epsilon_v)}\right)^{\gamma_b/\beta_b} + (\ln n)^{\gamma_b} \tag{A.137}$$

$$\leq (\ln n)^{\gamma_b/\beta_b} \tag{A.138}$$

and

$$\lambda(h_n)n^{-1/2+\epsilon_1} \leq \exp((-\alpha_m + \epsilon_2)(h_n^{-1})^{\beta_m})n^{-1/2+\epsilon_1} \quad \text{for any } \epsilon_2 > 0 \tag{A.139}$$

$$\leq \left((-\alpha_m + \epsilon_2) \frac{\ln n}{2(\alpha_v + \epsilon_v)}\right)n^{-1/2+\epsilon_1} \tag{A.140}$$

$$\leq \left(\frac{(-\alpha_m + \epsilon_2)}{2(-\alpha_m + \epsilon_v)} \ln n\right)n^{-1/2+\epsilon_1} \tag{A.141}$$

$$\leq n^{1/2-\epsilon_3}n^{-1/2+\epsilon_1} \quad \text{for some } \epsilon_3 > 0 \tag{A.142}$$

$$= o(1) \quad \text{if } \epsilon_1 < \epsilon_3. \tag{A.143}$$

Case 3. $\beta_b = \beta_v \neq 0$. For some $\epsilon > 0$, let

$$h_n^{-1} = \left(\frac{\ln n}{2(\alpha_v - \alpha_b + \epsilon)} \right)^{1/\beta_b}. \tag{A.144}$$

Then,

$$R_n = (h_n^{-1})^{\gamma_b} \exp(\alpha_b (h_n^{-1})^{\beta_b}) + n^{-1/2} (h_n^{-1})^{\gamma_v} \exp(\alpha_v (h_n^{-1})^{\beta_b}) \tag{A.145}$$

$$\leq \exp((\alpha_b + \epsilon_b) (h_n^{-1})^{\beta_b}) + n^{-1/2} \exp((\alpha_v + \epsilon_v) (h_n^{-1})^{\beta_b}) \tag{A.146}$$

$$\begin{aligned} &\leq \exp\left((\alpha_b + \epsilon_b) \left(\frac{\ln n}{2(\alpha_v - \alpha_b + \epsilon)} \right) \right) \\ &\quad + n^{-1/2} \exp\left((\alpha_v + \epsilon_v) \left(\frac{\ln n}{2(\alpha_v - \alpha_b + \epsilon)} \right) \right) \end{aligned} \tag{A.147}$$

$$\begin{aligned} &\leq \exp\left(\left(\frac{\alpha_b}{2(\alpha_v - \alpha_b)} + \epsilon_1 \right) \ln n \right) \\ &\quad + n^{-1/2} \exp\left(\left(\frac{\alpha_v}{2(\alpha_v - \alpha_b)} + \epsilon_2 \right) \ln n \right) \quad \text{for } \epsilon_1, \epsilon_2 > 0 \end{aligned} \tag{A.148}$$

$$\leq n^{\alpha_b/(2\alpha_v - 2\alpha_b) + \epsilon_1} + n^{-1/2} n^{\alpha_v/(2\alpha_v - 2\alpha_b) + \epsilon_2} \tag{A.149}$$

$$\leq n^{\alpha_b/(2\alpha_v - 2\alpha_b) + \epsilon_1} + n^{\alpha_b/(2\alpha_v - 2\alpha_b) + \epsilon_2} \tag{A.150}$$

$$\leq n^{\alpha_b/(2\alpha_v - 2\alpha_b) + \epsilon} \quad \text{for some } \epsilon > 0, \tag{A.151}$$

$$\lambda(h_n) n^{-1/2 + \epsilon_1} \leq \exp((-\alpha_m + \epsilon_2) (h_n^{-1})^{\beta_m}) n^{-1/2 + \epsilon_1} \quad \text{for any } \epsilon_2 > 0 \tag{A.152}$$

$$\leq \exp\left(\frac{(-\alpha_m + \epsilon_2) \ln n}{2(\alpha_v - \alpha_b + \epsilon)} \right) n^{-1/2 + \epsilon_1} \tag{A.153}$$

$$\leq \exp\left(\frac{(-\alpha_m + \epsilon_2) \ln n}{2(\alpha_\phi - \alpha_m - \alpha_\phi + \epsilon)} \right) n^{-1/2 + \epsilon_1} \tag{A.154}$$

$$= \exp\left(\frac{(-\alpha_m + \epsilon_2) \ln n}{2(-\alpha_m + \epsilon)} \right) n^{-1/2 + \epsilon_1} \tag{A.155}$$

$$= n^{1/2 - \epsilon_4} n^{-1/2 + \epsilon_1} \quad \text{for some } \epsilon_4 > 0 \text{ if } \epsilon_2 < \epsilon \tag{A.156}$$

$$= n^{\epsilon_1 - \epsilon_4} \tag{A.157}$$

$$= o(1) \quad \text{for some } \epsilon_1 < \epsilon_4. \tag{A.158}$$

Case 4. $\beta_b = \beta_v = 0$ (and $\alpha_b = \alpha_v = 0$ and $\gamma_b < 0$). Let $h_n^{-1} = n^{1/(2\gamma_v - 2\gamma_b)}$. Then,

$$R_n \leq (h_n^{-1})^{\gamma_b} + n^{-1/2}(h_n^{-1})^{\gamma_v} \tag{A.159}$$

$$\leq n^{\gamma_b/(2\gamma_v - 2\gamma_b)} + n^{-1/2 + \gamma_v/(2\gamma_v - 2\gamma_b)} \tag{A.160}$$

$$\leq n^{\gamma_b/(2\gamma_v - 2\gamma_b)}, \tag{A.161}$$

$$\lambda(h_n)n^{-1/2 + \epsilon} = (1 + h_n^{-1})^{1 + \gamma_r - \gamma_m} n^{-1/2 + \epsilon_1} \tag{A.162}$$

$$\leq (n^{1/(2\gamma_v - 2\gamma_b)})^{1 + \gamma_r - \gamma_m} n^{-1/2 + \epsilon_1} \tag{A.163}$$

$$= n^{(1 + \gamma_r - \gamma_m)/(2\gamma_v - 2\gamma_b)} n^{-1/2 + \epsilon_1}. \tag{A.164}$$

Noting that $\gamma_b \geq \gamma_\phi + 1$, $\gamma_v = 2 + \gamma_\phi - \gamma_m + \gamma_r$, and $\gamma_m < 0$, the exponent of n can be written as

$$\frac{(1 + \gamma_r - \gamma_m)}{2(\gamma_v - \gamma_b)} - \frac{1}{2} + \epsilon_1 = \frac{(1 + \gamma_r - \gamma_m)}{2(2 + \gamma_\phi - \gamma_m + \gamma_r - \gamma_b)} - \frac{1}{2} + \epsilon_1 \tag{A.165}$$

$$\leq \frac{(1 + \gamma_r - \gamma_m)}{2(2 + \gamma_\phi - \gamma_m + \gamma_r - \gamma_\phi + 1)} - \frac{1}{2} + \epsilon_1 \tag{A.166}$$

$$= \frac{(1 + \gamma_r - \gamma_m)}{2(3 + \gamma_r - \gamma_m)} - \frac{1}{2} + \epsilon_1 \tag{A.167}$$

$$< 0 \quad \text{for } \epsilon_1 \text{ sufficiently small.} \tag{A.168}$$

■

LEMMA 9. Let $K_n(z)$ be a sequence of real-valued nonrandom functions of a real variable, let a_j and z_j be i.i.d. sequences with a_j satisfying $E[a_j^{2+\delta} | z_j = z] \leq C$ for some $C, \delta > 0$ for all z and $\text{Var}[a_j | z_j = z] \geq C$ for some $C > 0$ and for all z , and let

$$\sigma_n = (E[a_j^2 K_n^2(z_j)] - (E[a_j K_n(z_j)])^2)^{1/2}. \tag{A.169}$$

If $\inf_{n \geq N} \sigma_n > 0$ for some $N \in \mathbb{N}^+$ and $\sup_{z \in \mathbb{R}} |dK_n(z)/dz| = O(n^{(3/2) - \delta})$ for some $\delta > 0$, then

$$\sigma_n^{-1} n^{-1/2} \sum_{j=1}^n (a_j K_n(z_j) - E[a_j K_n(z_j)]) \xrightarrow{d} N(0, 1). \tag{A.170}$$

Proof. Let $Z_{nj} = a_j K_n(z_j)$. The proof consists in verifying that Z_{nj} satisfies the hypothesis of the Lindeberg–Feller central limit theorem for triangular arrays. Indeed, the Z_{n1}, \dots, Z_{nn} are i.i.d. by assumption, and it remains to be shown that the Lindeberg condition holds: for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \sigma_n^{-2} E[1(|Z_{nj}| \geq \epsilon \sigma_n n^{1/2}) Z_{nj}^2] \rightarrow 0. \tag{A.171}$$

First, noting that $1(ab \geq c) \leq 1(a \geq c^\eta) + 1(b \geq c^{1-\eta})$ for any $a, b, c \in \mathbb{R}^+$ and any $\eta \in]0, 1[$, we can write

$$\sigma_n^{-2} E[1(|a_j K_n(z_j)| \geq \epsilon \sigma_n n^{1/2}) |a_j K_n(z_j)|^2] \leq T_1 + T_2 \tag{A.172}$$

where

$$T_1 = \sigma_n^{-2} E[1(|a_j| \geq \varepsilon^\eta \sigma_n^\eta n^{\eta/2}) |a_j K_n(z_j)|^2], \tag{A.173}$$

$$T_2 = \sigma_n^{-2} E[1(|K_n(z_j)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) |a_j K_n(z_j)|^2]. \tag{A.174}$$

Then, $T_1 = \sigma_n^{-2} E[E[1(|a_j| \geq \varepsilon^\eta \sigma_n^\eta n^{\eta/2}) a_j^2 |z_j] |K_n(z_j)|^2] \leq \sigma_n^{-2} E[(\varepsilon^\eta \sigma_n^\eta n^{\eta/2})^{-\delta} \times |K_n(z_j)|^2]$ because $E[a_j^{2+\delta} |z_j = z] \leq C$ (i.e., $E[1(|a_j| \geq c) a_j^2 |z_j] \leq E[1(|a_j| \geq c) \times (a_j/c)^\delta a_j^2 |z_j] \leq c^{-\delta} E[1(|a_j| \geq c) a_j^{2+\delta} |z_j] \leq c^{-\delta} E[a_j^{2+\delta} |z_j] \leq c^{-\delta}$). Noting that

$$\sigma_n^2 = E[a_j^2 K_n^2(z_j)] - (E[a_j K_n(z_j)])^2 \tag{A.175}$$

$$= E[E[a_j^2 |z_j] K_n^2(z_j)] - (E[a_j K_n(z_j)])^2 \tag{A.176}$$

$$= E[(E[a_j^2 |z_j] - (E[a_j |z_j])^2) K_n^2(z_j)] + E[(E[a_j |z_j])^2 K_n^2(z_j)] - (E[E[a_j |z_j] K_n(z_j)])^2 \tag{A.177}$$

$$= E[\text{Var}[a_j |z_j] K_n^2(z_j)] + \text{Var}[E[a_j |z_j] K_n(z_j)] \tag{A.178}$$

$$\geq E[K_n^2(z_j)] + \text{Var}[E[a_j |z_j] K_n(z_j)] \tag{A.179}$$

$$\geq E[K_n^2(z_j)], \tag{A.180}$$

we have $T_1 \leq (E[K_n^2(z_j)])^{-1} E[(\varepsilon^\eta \sigma_n^\eta n^{\eta/2})^{-\delta} K_n^2(z_j)] = (\varepsilon^\eta \sigma_n^\eta n^{\eta/2})^{-\delta} (E[K_n^2(z_j)])^{-1} \times E[K_n^2(z_j)] = (\varepsilon^\eta \sigma_n^\eta n^{\eta/2})^{-\delta} \rightarrow 0$. Also, $T_2 = \sigma_n^{-2} E[E[a_j^2 |z_j] 1(|K_n(z_j)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) K_n^2(z_j)] \leq C \sigma_n^{-2} E[1(|K_n(z_j)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) K_n^2(z_j)] \leq E[1(|K_n(z_j)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) K_n^2(z_j)]$.

Let $s_n = \sup_{z \in \mathbb{R}} |dK_n(z)/dz|$. For a given value of $E[K_n^2(z_j)]$ the maximum value of $E[1(|K_n(z_j)| \geq C) K_n^2(z_j)]$ for some $C > 0$ is obtained when the support of $K_n(z)$ is inside the support of the distribution of z and when $K_n(z)$ is triangular,

$$K_n(z) = \begin{cases} 0 & \text{if } |z| \geq l_n \\ s_n(l_n - |z|) & \text{if } |z| < l_n, \end{cases} \tag{A.181}$$

where $l_n = (3\sigma_n^2/(2s_n^2))^{1/3}$. Then,

$$I_n = \int 1(|K_n(z)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) |K_n(z)|^2 f(z) dz \tag{A.182}$$

$$\geq \int 1(|K_n(z)| \geq \varepsilon^{1-\eta} \sigma_n^{1-\eta} n^{(1-\eta)/2}) |K_n(z)|^2 dz = 2\sigma_n^{-2} \max \left\{ \int_{(\varepsilon \sigma_n n^{1/2})^{1-\eta} s_n^{-1}}^{l_n} s_n^2 z^2 dz, 0 \right\} \tag{A.183}$$

$$= \frac{2s_n^2}{3\sigma_n^2} \max \left\{ \left(l_n^3 - \frac{(\varepsilon^3 \sigma_n^3 n^{3/2})^{(1-\eta)}}{s_n^3} \right), 0 \right\} = \max \left\{ \left(1 - \frac{2(\varepsilon^3 \sigma_n n^{3/2})^{(1-\eta)}}{3s_n} \right), 0 \right\}, \tag{A.184}$$

where

$$\frac{2(\varepsilon^3 \sigma_n n^{3/2})^{(1-\eta)}}{3s_n} \geq \frac{2(\varepsilon^3 \sigma_n n^{3/2})^{(1-\eta)}}{3n^{(3/2)-\delta}} \tag{A.185}$$

$$= \frac{2\varepsilon^{3(1-\eta)} \sigma_n^{(1-\eta)} n^{(3/2)-\delta/2}}{3n^{(3/2)-\delta}} \quad \text{for } \eta = \delta \tag{A.186}$$

$$= \frac{2}{3} \varepsilon^{3(1-\eta)} \sigma_n^{(1-\eta)} n^{\delta/2} \rightarrow \infty, \tag{A.187}$$

and it follows that $I_n \rightarrow 0$, as desired. ■

LEMMA 10. *If $\psi \in \mathbb{W}$, then $\lim_{|z| \rightarrow \infty} p(z) = 0$, where $p(z)$ is the inverse Fourier transform of $\psi(\zeta)$.*

Proof. This result is Theorem 18 in Lighthill (1962) with the trivial modification that the Fourier transform is replaced by the inverse Fourier transform and with the slight extension that allows the tail behavior of the function $\psi(\zeta)$ to be exponential (see Definition 4). This extension is straightforward because Lighthill’s proof proceeds by writing $\psi(\zeta) = \Psi(\zeta) + (\psi(\zeta) - \Psi(\zeta))$ where $(\psi(\zeta) - \Psi(\zeta))$ can be handled using the Riemann–Lebesgue lemma. By the assumption that $\psi \in \mathbb{W}$, the function $\Psi(\zeta)$ can be chosen such that its inverse Fourier transform, $p_\infty(z)$, can be calculated analytically and be shown to satisfy $\lim_{|z| \rightarrow \infty} p_\infty(z) = 0$. All that is needed to allow for more flexible choices of tail behavior than initially employed by Lighthill is to find functions $\Psi(\zeta)$ whose inverse Fourier transform can be calculated analytically and have the appropriate tail behavior. Using the techniques described in Gel’fand and Shilov, (1964, Example 5, p. 169) the inverse Fourier transform of exponentials of the form $\exp(c\zeta^\gamma)$ for $c \in \mathbb{R}$ and $\gamma \in \mathbb{N}^+$ can be shown to be

$$\sum_{j=0}^{\infty} \frac{(ic)^{\gamma j} \delta^{(\gamma j)}(z)}{j!}, \tag{A.188}$$

where $\delta^{(k)}(\zeta)$ denotes the k th derivative of Dirac’s delta distribution. This distribution clearly vanishes as $|z| \rightarrow \infty$, as required. Note that, although such a distribution does not belong to the class of the so-called tempered distributions, it does belong to the wider class of distributions that forms the dual of compactly supported infinitely differentiable test functions (i.e., the so-called Type K distributions of Gel’fand and Shilov). ■

Proof of Theorem 3. According to the second conclusion of Lemma 2, to have $\hat{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n) = (\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n))(1 + o_p(1))$ we need to show that $(\Omega(\tilde{x}^*, h_n))^{-1/2} \bar{U}(h_n) \lambda(h_n) n^{-1/2+\epsilon} \rightarrow 0$ for some $\epsilon > 0$. We proceed by finding a lower bound on $\Omega(\tilde{x}^*, h_n)$ and relating it to $\bar{U}(h_n)$. First, by Assumption 14, $\text{Var}[\bar{g}(\tilde{x}^*, h_n)] \equiv \Omega(\tilde{x}^*, h_n) \geq \max_{a=1,x,y} \max_{k=0,1} \text{Var}[T_{k,a,n}]$, where $T_{k,a,n}$, for $k = 0, 1$ and $a = 1, x, y$, is given by

$$T_{k,a,n} = \int U_a^k(\zeta, \tilde{x}^*, h_n) \hat{m}_a(\zeta) d\zeta \tag{A.189}$$

$$= \int U_l^k(\zeta, \tilde{x}^*, h_n) \left(n^{-1} \sum_{j=1}^n a_j e^{i\zeta z_j} \right) d\zeta \tag{A.190}$$

$$= n^{-1} \sum_{j=1}^n a_j K_{k,a,n}(z_j), \tag{A.191}$$

where

$$K_{k,a,n}(z_j) = \int U_a^k(\zeta, \tilde{x}^*, h_n) e^{i\zeta z_j} d\zeta. \tag{A.192}$$

Then,

$$\text{Var}[a_j K_{k,a,n}(z_j)] \geq E[a_j^2 (K_{k,a,n}(z_j))^2] \tag{A.193}$$

$$= \int (K_{k,a,n}(z))^2 E[a^2 | z] f(z) dz \tag{A.194}$$

$$\geq \int_{z \in I_{a,k}} (K_{k,a,n}(z))^2 E[a^2 | z] f(z) dz \tag{A.195}$$

for any finite interval $I_{a,k}$ not reduced to a point. By Assumptions 11 and 13, $\inf_{z \in I_{a,k}} E[a^2 | z] f(z) \geq C > 0$, and we have

$$\text{Var}[a_j K_{k,a,n}(z_j)] \geq \int_{z \in I_{a,k}} (K_{k,a,n}(z))^2 dz. \tag{A.196}$$

We now show that $\int_{z \in \mathbb{R}} \mathbb{1}_{I_{a,k}}(K_{k,a,n}(z))^2 dz$ remains bounded as $n \rightarrow \infty$, thus implying that $\int (K_{k,a,n}(z))^2 dz$ diverges at the same rate as $\int_{z \in I_{a,k}} (K_{k,a,n}(z))^2 dz$. First, $\lim_{n \rightarrow \infty} K_{k,a,n}(z) \equiv K_{k,a,\infty}(z)$ is the inverse Fourier transform of $U_a^k(\zeta, \tilde{x}^*, 0)$ and, by the moment theorem, the inverse Fourier transform of $(d/d\zeta)U_l^k(\zeta, \tilde{x}^*, 0)$ is $izK_{k,a,\infty}(z)$. Because $(d/d\zeta)U_l^k(\zeta, \tilde{x}^*, 0)$ belongs to \mathbb{W} by Assumption 12, we can apply Lemma 10 to conclude that $\lim_{|z| \rightarrow \infty} |z| |K_{k,a,\infty}(z)| = 0$. Therefore, there exist constants $A, C > 0$ such that $|K_{k,a,\infty}^2(z)| \leq A|z|^{-2}$ for $|z| \geq C$ and $k = 0, 1$ and $l = 1, x, y$. It is therefore impossible for $\int_{z \in \mathbb{R} \setminus I_{a,k}} (K_{k,a,n}(z))^2 dz$ to become unbounded as $n \rightarrow \infty$ if $I_{a,k}$ is chosen to be $[-C, C]$. We can then write

$$\text{Var}[a_j K_{k,a,n}(z_j)] \geq \int (K_{k,a,n}(z))^2 dz. \tag{A.197}$$

By Parseval's identify and the fact that $U_l^k(\zeta, \tilde{x}^*, h_n)$ vanishes for $|\zeta| \geq h_n^{-1}$, we have

$$\int (K_{k,a,n}(z))^2 dz = \int |U_l^k(\zeta, \tilde{x}^*, h_n)|^2 d\zeta \tag{A.198}$$

$$= \int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)|^2 d\zeta. \tag{A.199}$$

By the Cauchy–Schwartz inequality,

$$\left(\int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)| d\zeta \right)^2 \leq \left(\int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)|^2 d\zeta \right) \left(\int_{|\zeta| \leq h_n^{-1}} 1 d\zeta \right),$$

which becomes, upon rearrangement,

$$\int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)|^2 d\zeta \geq \left(\int_{|\zeta| \leq h_n^{-1}} 1 d\zeta \right)^{-1} \left(\int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)| d\zeta \right)^2 \tag{A.200}$$

$$= 2h_n \left(\int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)| d\zeta \right)^2. \tag{A.201}$$

Collecting equations (A.197), (A.198), and (A.200), we have

$$\Omega(\tilde{x}^*, h_n) \geq h_n \left(\max_{a=1, x, y} \max_{k=0, 1} \int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)| d\zeta \right)^2. \tag{A.202}$$

We then observe that, by equation (34) and Assumption 15,

$$\bar{U}(h_n) = \int |m_1(\zeta)|^{-1} \sum_{k=0}^1 \left(\left(1 + \frac{|m_x(\zeta)|}{|m_1(\zeta)|} \right) \int_{|\zeta|}^{\infty} |C_k(\zeta, \tilde{x}^*, h_n)| d\xi + |C_k(\zeta, \tilde{x}^*, h_n)| \right) d\zeta \tag{A.203}$$

$$\leq \int |m_1(\zeta)|^{-1} \sum_{k=0}^1 \left(\left(1 + \frac{|m_x(\zeta)|}{|m_1(\zeta)|} \right) \left| \int_{|\zeta|}^{\infty} C_k(\zeta, \tilde{x}^*, h_n) d\xi \right| + |C_k(\zeta, \tilde{x}^*, h_n)| \right) d\zeta \tag{A.204}$$

$$\leq \max_{a=1, x, y} \max_{k=0, 1} \int_{|\zeta| \leq h_n^{-1}} |U_l^k(\zeta, \tilde{x}^*, h_n)| d\zeta. \tag{A.205}$$

Combining equations (A.202) and (A.203) yields

$$(\Omega(\tilde{x}^*, h_n))^{1/2} \geq h_n^{1/2} \bar{U}(h_n), \tag{A.206}$$

thus implying that $h_n^{1/2} \lambda(h_n) n^{-1/2+\epsilon} \rightarrow 0$ for some $\epsilon > 0$ is a sufficient condition for the asymptotic negligibility of the higher order terms, which we can now verify.

If $\alpha_m = 0$, then $h_n^{1/2} \lambda(h_n) n^{-1/2+\epsilon} = (1 + h_n^{-1})^{1/2} (1 + h_n^{-1}) (1 + h_n^{-1})^{\gamma_r - \gamma_m} n^{-1/2+\epsilon} = (1 + h_n^{-1})^{3/2 + \gamma_r - \gamma_m} n^{-1/2+\epsilon} \leq (n - \eta_n)^{1/(3+2\gamma_r - 2\gamma_m)}^{3/2 + \gamma_r - \gamma_m} n^{-1/2+\epsilon} \leq n^{-\eta(3/2 + \gamma_r - \gamma_m)} \times n^\epsilon = o(1)$ for $\epsilon > 0$ sufficiently small.

If $\alpha_m \neq 0$, then $h_n^{1/2} \lambda(h_n) n^{-1+\epsilon} \leq (1 + h_n^{-1})^{3/2 + \gamma_r - \gamma_m} \exp(-\alpha_m (h_n^{-1})^{\beta_m}) n^{-1/2+\epsilon} \leq \exp(-\alpha_m (1 + \epsilon_2) (h_n^{-1})^{\beta_m}) n^{-1/2+\epsilon} \leq \exp((\alpha_m (1 + \epsilon_2) (1 + \eta) / 2\alpha_m) \ln n) n^{-1/2+\epsilon} = n^{-1/2(\epsilon_2 + \eta + \eta\epsilon_2)} n^\epsilon = o(1)$ for some $\epsilon_2 > 0$ and for $\epsilon > 0$ sufficiently small.

We have now shown that the limiting distribution of $\hat{g}(\tilde{x}^*, h_n)$ is the same as that of $\bar{g}(\tilde{x}^*, h_n)$. To obtain the limiting distribution of $\bar{g}(\tilde{x}^*, h_n)$, we note that $\bar{g}(\tilde{x}^*, h_n) - g(\tilde{x}^*, h_n)$ is a finite linear combination of the kernel-type estimators $T_{k,a,n}$ defined in equation (A.191) using the kernel $K_{k,a,n}(z)$ defined by equation (A.192). The asymptotic normality of $T_{k,a,n}$ can be shown using Lemma 9, provided that we can show that $\sup_{z \in \mathbb{R}} |dK_{k,a,n}(z)/dz| = O(n^{(3/2)-\delta})$ for some $\delta > 0$. By the moment theorem, this requirement is satisfied if $\int |U_l^k(\zeta, \tilde{x}^*, h_n)| |\zeta| d\zeta = O(n^{(3/2)-\delta})$ for $k = 0, 1$ and $a = 1, x, y$. Using the same techniques as in the proof of Lemma 5, we have

$$\int |U_a^k(\zeta, \tilde{x}^*, h_n)| |\zeta| d\zeta \leq (1 + h_n^{-1})^\gamma \exp(\alpha(h_n^{-1})^\beta), \tag{A.207}$$

where $\gamma = 3 + \gamma_\phi - \gamma_m + \gamma_r$ and

$$(\alpha, \beta) = \begin{cases} (-\alpha_m, \beta_m) & \text{if } \beta_\phi < \beta_m \\ (\alpha_\phi - \alpha_m, \beta_m) & \text{if } \beta_\phi = \beta_m. \end{cases} \tag{A.208}$$

If $\alpha_m = 0$, then $\alpha = 0$ and we have

$$\int |U_a^k(\zeta, \tilde{x}^*, h_n)| |\zeta| d\zeta \leq (h_n^{-1})^{3+\gamma_\phi-\gamma_m+\gamma_r} \tag{A.209}$$

$$\leq (n^{-\eta} n^{1/(3+2\gamma_r-2\gamma_m)})^{3+\gamma_\phi-\gamma_m+\gamma_r} \tag{A.210}$$

$$= n^{-\eta(3+\gamma_\phi-\gamma_m+\gamma_r)} n^{(3+\gamma_\phi-\gamma_m+\gamma_r)/(3+2\gamma_r-2\gamma_m)} = o(n) \tag{A.211}$$

because (i) $(3 + \gamma_\phi - \gamma_m + \gamma_r) > 0$ because $\gamma_\phi \geq -\gamma_m$ and $\gamma_r \geq 0$ and (ii)

$$\begin{aligned} \frac{3 + \gamma_\phi + \gamma_r - \gamma_m}{3 + 2\gamma_r - 2\gamma_m} &= \frac{3 + 2\gamma_r - 2\gamma_m - \gamma_r + \gamma_m + \gamma_\phi}{3 + 2\gamma_r - 2\gamma_m} \\ &= 1 - \frac{\gamma_r - \gamma_m - \gamma_\phi}{3 + 2\gamma_r - 2\gamma_m} < 1. \end{aligned} \tag{A.212}$$

If $\alpha_m \neq 0$, then

$$\int |U_a^k(\zeta, \tilde{x}^*, h_n)| |\zeta| d\zeta \leq \exp((\alpha + \epsilon_4)(h_n^{-1})^\beta) \quad \text{for some } \epsilon_4 > 0 \tag{A.213}$$

$$\leq \exp\left(-\frac{(\alpha + \epsilon_4)(1 + \eta)}{2\alpha_m} \ln n\right) \tag{A.214}$$

$$\leq \exp\left(\frac{(2\alpha_m - \epsilon_4)(1 + \eta)}{2\alpha_m} \ln n\right) \tag{A.215}$$

$$\leq \exp((1 + \epsilon_5) \ln n) \leq n^{1+\epsilon_5} \quad \text{for some } \epsilon_5 < \frac{1}{2}. \tag{A.216}$$

Hence, the hypotheses of Lemma 9 are verified, and the $T_{k,a,n}$ are asymptotically normal. The expectation and the variance of $\tilde{g}(\tilde{x}^*, h_n)$ can then be calculated as in Lemma 1. ■