Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(7): 16414–16431.
DOI:10.3934/math.2025735
Received: 17 April 2025
Revised: 12 June 2025
Accepted: 24 June 2025
Published: 22 July 2025

Research article

Asymptotic normality of parametric part in semiparametric regression in the presence of measurement error

Seçil Yalaz*

Department of Statistics, Dicle University, Science Faculty, Diyarbakır 21280, Türkiye

* **Correspondence:** Email: syalaz@dicle.edu.tr; Tel: +904122488550; Fax: +904122488300.

Abstract: In this study, we investigate a partially linear regression model where the covariate entering the nonparametric component is measured with error. A key challenge in such models is that the measurement error distribution is unknown, and this setting is further complicated by the limited availability of data; specifically, only one repeated observation of the mismeasured regressor is accessible. To address this, we propose an estimation procedure that leverages the Fourier transform, a powerful analytical tool that transforms complex convolution equations into tractable algebraic forms. By modifying an existing approach rooted in Fourier analysis, we construct a novel estimator that accommodates the unknown error distribution and efficiently utilizes the repeated measurement. We establish the asymptotic normality of the proposed estimator, demonstrating its theoretical validity and robustness. To assess its practical performance, we conduct a series of Monte Carlo simulations under various scenarios. These simulations provide strong evidence of the estimator's effectiveness in finite samples, particularly in terms of bias reduction and variance control. The methodology offers a flexible and computationally feasible framework for dealing with measurement error in semiparametric models without requiring knowledge of the error distribution. This contributes to the growing literature on measurement error models by extending nonparametric estimation techniques to more realistic and constrained data settings.

Keywords: errors in variables; partially linear model; semiparametric regression; unknown error density; asymptotic normality
Mathematics Subject Classification: 62F12, 62G08

1. Introduction

There has been an increasing interest in measurement error models for parametric and nonparametric specifications. Foundational contributions to the theory of linear measurement error models can be found in [1, 5, 8, 11, 20], while important advances for nonlinear specifications are presented in [6,9]. In particular, [9] investigated the estimation of a nonparametric regression function when the covariate is contaminated by measurement error with a known distribution. They proposed a kernel-based deconvolution estimator and established its optimal convergence properties. Building on this line of work, [14] addressed the estimation problem in semiparametric regression models under similar assumptions. Using a modified local likelihood approach, the author developed an estimator for the nonparametric component that achieves favorable theoretical guarantees.

Consider the partially linear errors in variables model for n observations:

$$\begin{aligned} y_j &= X_j^T\beta + g\left(X_j^*\right) + \Delta y_j; \quad j = 1,\dots,n \\ z_j &= X_j^* + \Delta z_j, \end{aligned} \tag{1.1}$$

where y is an $(n\times 1)$ column vector of the observations, X is a random matrix, g is an unknown function for a random variable X^*, Δy is a random error vector, and Δz is a vector of measurement errors.

When X^* is observable, by fixing the parametric component β, an estimate $\hat{g}\,(X^*,\beta)$ of the nonparametric component $g\,(.)$ can be obtained using a smoothing method. Then, $\hat{g}\,(X^*,\beta)$ was used to obtain an estimator of β, which is a solution of

$$\min \sum_{j=1}^{n} \left\{ y_j - x_j^T\beta - \hat{g}(X^*,\beta) \right\}^2 . \tag{1.2}$$

By applying the conditional expectation operator to the first equation of the Model (1.1), it could be obtained, assuming $E(\Delta y|X^*) = 0$, that

$$\begin{aligned} E(y|X^*) &= E(X^T\beta|X^*) + E(g(X^*)|X^*) + E(\Delta y|X^*) \\ &= E(X|X^*)^T\beta + g(X^*). \end{aligned}$$

Defining the conditional expectations as

$$\omega_y\,(X^*) = E(y|X^*),$$

$$\omega_X\,(X^*) = E(X|X^*),$$

the last expression could be written as

$$\omega_y\,(X^*) = \omega_X\,(X^*)^T\,\beta + g(X^*).$$

If one subtracts this from the model, it is obtained that

$$\widetilde{y} = \widetilde{X}^T\beta + \Delta y,$$

where $\widetilde{y} = y - \omega_y\,(X^*)$ and $\widetilde{X} = X - \omega_X\,(X^*)$.

Using the previous equation, the estimator of β could be obtained as,

$$\hat{\beta}_n = \left[\sum_{j=1}^{n} \widetilde{X}_j \widetilde{X}_j^T \right]^{-1} \times \sum_{j=1}^{n} \widetilde{X}_j \widetilde{y}_j.$$

When X^* is unobservable and measured with additive error Δz, we observe z instead of observing X^*. In this situation the disturbance of the measurement error Δz should be modified. If the measurement errors are coming from a known distribution, this could be managed using the kernel deconvolution method ([9, 14, 21]). [18] studied the nonparametric regression when the independent variable has an error with an unknown distribution and obtained the nonparametric estimation of conditional expectations using two error-contaminated measurements of the variable. Repeated measurements can commonly be found in data, although distributional assumptions are often not relevant in some implementations ([2–4, 10, 12, 16]).

In semiparametric errors in variables models, there does not become visible a study about the troublesome effects of the measurement errors coming from an unknown distribution. This is the gap our paper intends to fill under appropriate assumptions. Our study can be planned as follows. The semiparametric estimation method with errors coming from an unknown distribution in the nonparametric variable is proposed in Section 2. In Section 3, the asymptotic normality condition is put forward. A Monte Carlo simulation study is done to show the theoretical results in Section 4. Some summary and conclusions are ascribed in Section 5.

2. Construction of estimators

It is well established in the literature that the density of an unobserved random variable can be identified when the joint distribution of two error-contaminated measurements of that variable is available, using Kotlarski's identity ([17]). Convergence rates can be known in a consistent estimator using the empirical version of this identity ([13]). Drawing on this identity, [18] developed a kernel regression estimator to overcome the difficulties associated with nonparametric estimation when the distribution of X^* is unknown.

To introduce the procedure developed by [18], we consider the setting in which x^* is a scalar, and $\chi = x^* + \Delta\chi$ is a repeated measurement, where $\Delta\chi$ denotes a measurement error with an unknown distribution. Two repeated measurements of x^* are observed for each $j = 1, \ldots, n$ and are defined as follows:

$$\begin{aligned} \chi_j &= x^* + \Delta\chi_j, \\ z_j &= x^* + \Delta z_j. \end{aligned}$$

Then Model (1.1) turns to

$$\begin{aligned} y_j &= X_j^T\beta + g(x^*) + \Delta y_j, \\ \chi_j &= x^* + \Delta\chi_j, \\ z_j &= x^* + \Delta z_j. \end{aligned} \tag{2.1}$$

We first determine to estimate the nonparametric function $g(x^*) = E(y - X^T\beta | x^*)$ of the Model (2.1). Theoretically, if β were known, by absorbing $X^T\beta$ into y, we obtain:

$$\underbrace{y - X^T\beta}_{y^*} = g(x^*) + \Delta y.$$

This provides enough information to identify any moment of the form $E[u(y, x^*)]$ for any function

$u(y, x^*)$. Setting $u(y, x^*) = y^l K_h(x^* - \overline{x}^*)$, for $l = 0, 1$ at a given point, $\overline{x}^* \in \mathbb{R}$ qualification of moments comes from an infinite family is achieved.

Due to the success in the nonparametric errors in variables model, when the error has an unknown distribution, we derive an estimator of the nonparametric part $\hat{g}$ using the related procedure of [18]. Then, after deriving $\hat{g}(.)$, the estimator of β is introduced using the modified local-likelihood method proposed by [14].

Using the procedure defined in Section 1, let us construct consistent nonparametric estimates of $g(x^*)$ with two repeated observations. In order to be able to characterize an infinite family of moments indexed by $\overline{x}^*$ and thus to address the complexity of extending the present results to a nonparametric setting, the following assumption is needed. This can be handled elegantly by observing that the convolution operations used to compute the Nadaraya–Watson estimator are converted to simple products by the Fourier transform operation, such that the entire family of moments can be estimated by a single operation ([19]). Focusing on a kernel regression estimator, we need to make Assumption 1 to estimate $g(\overline{x}^*)$ at a given point, $\overline{x}^* \in \mathbb{R}$.

Assumption 1.

i.
$$E[\Delta y \mid X, x^*, \Delta z] = 0, \quad E[\Delta \chi \mid x^*, \Delta z] = 0.$$

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

iii. $E\left[g(x^*)^l h_n^{-1} K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right] < \infty$ for every $\overline{x}^* \in \mathbb{R}$, some sequence of bandwidths $h_n \to 0$, where $K(.)$ is a kernel function and $l = 0, 1$.

Under Assumption 1 and provided that $|E\left[e^{i\xi z}\right]| > 0$ for any finite ξ, the estimator of $g(\overline{x}^*)$ for a given $\overline{x}^* \in \mathbb{R}$ and some bandwidths $h_n \to 0$ is defined as

$$\widehat{g}(\overline{x}^*, h_n) = \frac{\widehat{M}_1(\overline{x}^*, h_n)}{\widehat{M}_0(\overline{x}^*, h_n)}, \tag{2.2}$$

where

$$\widehat{M}_l(\overline{x}^*, h_n) = \frac{1}{2\pi} \int \kappa(h_n\xi) \widehat{\phi}_l(\xi) \exp(-i\xi\overline{x}^*) \, d\xi, \quad l = 0, 1 \tag{2.3}$$

and

$$\widehat{\phi}_0(\xi) = \exp\left(\int_0^\xi \frac{i\widehat{m}_\chi(\zeta)}{\widehat{m}_1(\zeta)} d\zeta\right),$$
$$\widehat{\phi}_1(\xi) = \widehat{\phi}_0(\xi) \frac{\widehat{m}_{y^*}(\xi)}{\widehat{m}_1(\xi)},$$

where $i = \sqrt{-1}$ and $\kappa(\xi)$ is the Fourier transform of the kernel K and $\widehat{m}_a(\xi) = n^{-1} \sum_{i=1}^n a_j \exp\left(i\xi z_j\right)$ for $a = 1, \chi, y^*$.

In contrast to the well-known kernel deconvolution estimator, this estimator requires no knowledge of the distribution of the measurement error. Moreover, the convergence rate of the estimator is comparable with those of earlier estimators.

To construct a semiparametric approach using this idea for $k = 1, \ldots, p$ and $j = 1, \ldots, n$, we define the weighted observations as follows:

$$\widehat{\omega}_{x_k}(\overline{x}^*, h_n) = \frac{\widehat{\varpi}_{x_k}(\overline{x}^*, h_n)}{\widehat{M}_0(\overline{x}^*, h_n)}, \quad \widehat{\omega}_y(\overline{x}^*, h_n) = \frac{\widehat{\varpi}_y(\overline{x}^*, h_n)}{\widehat{M}_0(\overline{x}^*, h_n)},$$

where

$$\widehat{\varpi}_{x_k}\left(\overline{x}^*, h_n\right) = \frac{1}{2\pi}\int \kappa\left(h_n\xi\right)\widehat{\varrho}_{x_l}\left(\xi\right)\exp(-i\xi\overline{x}^*)\,d\xi$$

and

$$\widehat{\varpi}_{y}\left(\overline{x}^*, h_n\right) = \frac{1}{2\pi}\int \kappa\left(h_n\xi\right)\widehat{\varrho}_{y}\left(\xi\right)\exp(-i\xi\overline{x}^*)\,d\xi,$$

where $\widehat{\varrho}_{x_k}(\xi)$ and $\widehat{\varrho}_{y}(\xi)$ are expressed as

$$\widehat{\varrho}_{x_k}\left(\xi\right) = \widehat{\phi}_0\left(\xi\right)\frac{\widehat{m}_{x_k}\left(\xi\right)}{\widehat{m}_1\left(\xi\right)},$$

$$\widehat{\varrho}_{y}\left(\xi\right) = \widehat{\phi}_0\left(\xi\right)\frac{\widehat{m}_{y}\left(\xi\right)}{\widehat{m}_1\left(\xi\right)},$$

and $\widehat{\phi}_0$, $\widehat{M}_0$ and $\widehat{m}_a(\xi)$ for $a = 1, x_k, y$ which are $\widehat{m}_1$, $\widehat{m}_{x_k}$ and $\widehat{m}_y$ are described in Eq (2.3).

Let $\widehat{\beta}_n$ be the solution of the Eq (1.2) after replacing $\hat{g}(X^*, \beta)$ with $\hat{g}(z)$. Then the generalized least squares estimator $\widehat{\beta}_n$ of β can be indicated in terms of moments that involve the observable variables X, y, χ and z as

$$\widehat{\beta}_n = \left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left(\widetilde{X}^T\widetilde{y}\right), \tag{2.4}$$

where $\widetilde{y} = y - \hat{\omega}_y(z, h_n)$ and $\widetilde{X} = X - \hat{\omega}_X(z, h_n)$ for $\hat{\omega}_X(z, h_n) = \left(\hat{\omega}_{x_1}(z, h_n), ..., \hat{\omega}_{x_p}(z, h_n)\right)$.

3. Asymptotic results

In this section, we state our main results for the limit distributions of the $\widehat{\beta}_n$.

Descriptions of the ordinary smooth and super smooth error distributions are given by [9] for the case that the error distribution is known as follows:

- o Ordinary smooth of order γ: If the characteristic function $\phi_{\Delta y}$ satisfies

$$d_0|x^*|^\gamma \le \left|\phi_{\Delta y}(x^*)\right| \le d_1|x^*|^\gamma, as \quad x^* \to \infty,$$

  where d_0, d_1 are positive constants and $\gamma < 0$.
- o Super smooth of order α: If the characteristic function $\phi_{\Delta y}$ satisfies

$$d_0|x^*|^{\alpha_0}\exp\left(\delta|x^*|^\alpha\right) \le \left|\phi_{\Delta y}(x^*)\right| \le d_1|x^*|^{\alpha_1}\exp\left(\delta|x^*|^\alpha\right), as \quad x^* \to \infty,$$

  where d_0, d_1, and α are positive constants, and δ, α_0, and α_1 are constants.

Such ordinary smooth distributions are gamma distribution with $\gamma = p$ and double exponential distribution with $\gamma = 2$. Super smooth distributions are the standard normal distribution with $\delta = 2$ and the Cauchy distribution with $\delta = 1$.

New definitions covering the ordinarily smooth and super smooth cases when the error distribution is not known are required. The definition given by [18] is simultaneously described in Assumption 2 for the ordinarily smooth and super smooth cases.

Assumption 2.
The functions $\phi_0(\zeta) = E\left[e^{i\zeta x^*}\right]$, $\phi_0'(\zeta) \equiv \frac{d\phi_0(\zeta)}{d\zeta}$, $\phi_1(\zeta) = E\left[y^* e^{i\zeta x^*}\right]$ and $m_1(\zeta) = E\left[e^{i\zeta z}\right]$ satisfy

$$\left|\frac{\phi_0'(\zeta)}{\phi_0(\zeta)}\right| = O\left((1+|\zeta|)^{\gamma_r}\right) \tag{3.1}$$

for some $\gamma_r \geq 0$ and

$$\max\{|\phi_0(\zeta)|, |\phi_1(\zeta)|\} = O\left((1+|\zeta|)^{\gamma_\phi}\exp\left(\delta_\phi |\zeta|^{\alpha_\phi}\right)\right),$$
$$|m_1(\zeta)| = O\left((1+|\zeta|)^{\gamma_m}\exp\left(\delta_m |\zeta|^{\alpha_m}\right)\right)$$

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

In Assumption 2, the decay rate of $\phi_0(\zeta)$, the characteristic function of x^* , is solely influenced by how smooth the density function $f(x^*)$ is. In contrast, the decay rate of $\phi_1(\zeta)$ depends on the smoothness of the product $f(x^*)E[y \mid x^*]$. To verify the relevant expression, one would begin by establishing individual bounds for $|\phi_0(\zeta)|$ and $|\phi_1(\zeta)|$, then identify the term that decays most slowly. Combining $\phi_0(\zeta)$ and $\phi_1(\zeta)$ under a single assumption is justified, since both functions influence the estimator in a similar way. This consolidation also helps streamline the notation by reducing the number of separate asymptotic rates that need to be tracked during convergence analysis. As is typical for estimators based on deconvolution, there is a need to impose a lower bound, rather than an upper one, on a key quantity (in this case, $m_1(\zeta)$), because it appears in the denominator of the estimator's formula. Importantly, a lower bound on $m_1(\zeta)$ can be deduced from separate lower bounds on the moduli of the characteristic functions of x^* and the measurement error Δz, given that $m_1(\zeta)$ equals the product of their respective characteristic functions. Grouping these terms together again serves to lighten the notational load. While the condition on the ratio $\phi_0'(\zeta)/\phi_0(\zeta)$ might seem unfamiliar, it naturally follows from placing an upper bound on the derivative $\phi_0'(\zeta)$ and a lower bound on $\phi_0(\zeta)$. Moreover, excluding exponential terms like $\exp(\delta_r |\zeta|^{\alpha_r})$ from this condition entails little loss of generality, since both ordinarily smooth and supersmooth characteristic functions generally satisfy this property when $\gamma_r = 1$.

Assumption 3 determines the traditional finite-order kernels described in case (i) and a class of infinite-order kernels described in case (ii).

Assumption 3.

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

$$\int (x^*)^s K(x^*)\,dx^* \begin{cases} = 0, & s = 1, \ldots, \gamma_\kappa - 1 \\ \neq 0, & s = \gamma_\kappa \end{cases}$$

$$\int |x^*|^s |K(x^*)|\,dx^* < \infty, \; s = 1, \ldots, \gamma_\kappa.$$

ii. To confirm that the estimator is well-behaved, the Fourier transform of the kernel $\kappa(\xi)$ is assumed bounded and compactly supported, $\kappa(\xi) = 1$ for $|\xi| < \bar{\xi}$ for some $\bar{\xi} > 0$.

Assumption 4 introduces certain notations and conditions that are essential for deriving the asymptotic normality of the estimator.

Assumption 4.

i. Assume that the density of x^* is nonzero at $x^* = \overline{x}^*$. For a given $\overline{x}^*$, $j = 1, \ldots, n$ and $k = 1, \ldots, p$ notate

$$w_{jk}\left(\overline{x}^*\right) = E\left[x_{jk}|\overline{x}^*\right], \qquad \Lambda_{jk} = x_{jk} - w_{jk}\left(\overline{x}^*\right)$$

then $\sup_{0 \leq \overline{x}^* \leq 1} E\left(\|x_1\|^3 \left|\overline{x}^*\right.\right) < \infty$ and $E\left(\Lambda_1 \Lambda_1^T\right) = B$, where B is a positive definite matrix and $\Lambda_j = \left(\Lambda_{j1}, \ldots, \Lambda_{jp}\right)^T$.

ii. $g(.)$ and $w_{jk}(.)$ are Lipschitz continuous of order 1.

Theorem. Suppose $E\left(|\Delta y|^3 + \|\Delta z\|^3\right) < \infty$, assumptions 1–4 hold for any given $\overline{x}^*$ and any bandwidth sequence h_n satisfying either of the following items for some $\eta > 0$.

i. When the density of z is smooth, we consider the bandwidth $h_n^{-1} = O\left(n^{-\eta} n^{1/(3+2\gamma_r-2\gamma_m)}\right)$.

ii. When the density of z is supersmooth, we consider the bandwidth $h_n^{-1} = O\left(\left(\left(-\frac{(1+\eta)}{2\delta_m}\right) \ln n\right)^{-1/\alpha_m}\right)$,

where γ_r, γ_m, δ_m, and α_m are given in Assumption 2.

Then, $\widehat{\beta}_n$ is an asymptotically normal estimator:

$$\sqrt{n}\left(\widehat{\beta}_n - \beta\right) \xrightarrow{d} N\left(0, \Sigma^2 B^{-1}\right),$$

where B is given in Assumption 4(i) and $\Sigma_{j_1 j_2} = Cov\left(\widetilde{X}_{j_1}, \widetilde{X}_{j_2}\right)$.

4. Simulation studies

Section 2 derives a semiparametric regression method when the nonparametric part has consecutive measurement errors with unknown distribution. In addition, to observe whether the distribution around the parameter to which they converge while the sample size of the estimators obtained using this method converges to infinity fits the normal distribution, the asymptotic normality properties of these estimators have been comprehensively investigated in Section 3.

In this section, finite sample properties of the estimators were investigated via a Monte Carlo simulation study. To compare the convergence rates with those derived from the Nadaraya–Watson estimator and an estimator for the no measurement error case, we focused on four examples.

For the sample sizes of $n = 100, 500, 1000$, and 5000 and a model with $p = 2$, simulation results are reported. Ordinarily smooth nonparametric function $g(x^*)$, while densities of both the true regressor x^* and the measurement errors $\Delta\chi, \Delta z$ are supersmooth is considered in the first example. In the second example, we also consider the supersmooth nonparametric function $g(x^*)$, while the densities of the true regressor x^* and the measurement errors $\Delta\chi, \Delta z$ are supersmooth. The third example assumes that the measurement error and nonparametric function $g(x^*)$ are ordinarily smooth, while the true regressor x^* is supersmooth. The last example assumes that all quantities are ordinarily smooth. Table 1 lists all the examples with bandwidth choices and optimal ones considered herein (for more detailed information about the examples, see [18]).

Table 1. Examples.

1	2	3	4
$x^* \to N(0,1)$	$x^* \to N(0,1)$	$x^* \to N(0,1)$	$x^* \to Uniform[-2,2]$
$\Delta\chi, \Delta z \to N(0,0.25)$	$\Delta\chi, \Delta z \to N(0,0.25)$	$\Delta\chi, \Delta z \to L(0,0.25)$	$\Delta\chi, \Delta z \to L(0,0.25)$
$\Delta y \to N(0,0.25)$	$\Delta y \to N(0,0.25)$	$\Delta y \to N(0,0.25)$	$\Delta y \to N(0,0.25)$
$g(x^*) = S(x^*)$	$g(x^*) = erf(x^*)$	$g(x^*) = S(x^*)$	$g(x^*) = S(x^*)$
$h_n^{-1} = C(\ln n)^{1/\beta}$	$h_n^{-1} = C(\ln n)^{1/\beta}$	$h_n^{-1} = C(\ln n)^{1/\beta}$	$h_n^{-1} = Cn^{1/\left(-2\gamma_0-\gamma_\phi+\gamma_1\right)}$
$h_n^{-1} = 1.2(\ln n)^{0.25}$	$h_n^{-1} = 1.2(\ln n)^{0.25}$	$h_n^{-1} = 1.2(\ln n)^{0.25}$	$h_n^{-1} = 1.2n^{0.125}$

As examples of supersmooth and ordinarily smooth functions the normal distribution and the Laplace (or double exponential) distribution are respectively considered. The error function and the piecewise linear continuous function are respectively considered as examples of supersmooth and ordinarily smooth nonparametric regression functions. We also assume that Eq (3.1) holds for $\gamma_r = 1$. The true value of β equals $(1,2)^T$, $\sigma^2_{\Delta y} = 0.25$, $0 \le x^* \le 1$ and $X \to N_2(0, I_2)$. For the no measurement error case, the quartic kernel $(15/16)\left(1-\Delta z_j^2\right)^2 I\left(|\Delta z_j| \le 1\right)$ is used to compare the results.

Next, choose the kernel function as $K(x^*) = \frac{\sin(x^*)}{\pi x^*}$, which is an infinite-order kernel and whose Fourier transform is given by $\tilde{\kappa}(\xi) = W(\xi)\,\kappa(\xi)$, where for a given $\overline{\xi} = 0.5$

$$W(\xi) = \begin{cases} \left(1+\exp\left(0.5\left((1+\xi)^{-1} - (-\xi-0.5)^{-1}\right)\right)\right)^{-1} & , \ -1 \le \xi < -0.5 \\ 1 & , \ -0.5 \le \xi \le 0.5 \\ \left(1+\exp\left(0.5\left((1-\xi)^{-1} - (\xi-0.5)^{-1}\right)\right)\right)^{-1} & , \ 0.5 < \xi \le 1 \\ 0 & , \ otherwise. \end{cases}$$

The average values of $N = 100$ replicates of variance (Var), mean squared error (MSE), and coefficient of determination (R^2) of the three estimators are considered in four different sample sizes in Table 2.

It can be seen in Table 2 that the comparison between the proposed estimator with the Nadaraya–Watson estimator and the estimator of the no measurement error case is favorable when the proposed estimator reduces Var and MSE. The coefficient of determination of our estimator is closest to 1. A range of bandwidths from 1.0 to 2.5 is explored in increments of 0.05 to identify the bandwidth that minimizes the mean squared error. Due to the chosen bandwidths shown in Table 1, these simulation results could not be more efficient. The results are thought to be more efficient using a feasible bandwidth selection rule, which gets optimal bandwidths.

We also examine a comparison between the asymptotic distributions and the finite sample of our estimator. The abscissa is $Z = \left(Var\left(g\left(\overline{x}^*, h_n\right)\right)\right)^{-1/2}\left(g\left(\overline{x}^*, h_n\right) - E\left(g\left(\overline{x}^*, h_n\right)\right)\right)$ and the ordinate is probability in Figure 1. The normal CDF (indicated by a solid line) agrees very well with the empirical cumulative distribution function (CDF) of the estimator (indicated by a dashed line).

Table 2. Monte Carlo simulation results for the examples.

	n = 100			n = 500			n = 1000			n = 5000		
	Example 1											
	Var	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2
Fourier	0.0628	0.1487	0.9682	0.0623	0.0801	0.9847	0.0623	0.0643	0.9873	0.0625	0.0642	0.9872
Nadaraya Watson	0.5929	0.6004	0.8934	0.5120	0.5123	0.9115	0.4957	0.4958	0.9136	0.5138	0.5139	0.9080
No M. Error	0.3400	0.4346	0.8964	0.1585	0.1796	0.9653	0.1218	0.1238	0.9756	0.1106	0.1123	0.9777
	Example 2											
	Var	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2
Fourier	0.0639	0.2092	0.9545	0.0622	0.0714	0.9860	0.0622	0.0664	0.9868	0.0625	0.0626	0.9880
Nadaraya Watson	0.4573	0.4581	0.9130	0.4326	0.4351	0.9158	0.4434	0.4455	0.9173	0.4635	0.4636	0.9175
No M. Error	0.1168	0.2574	0.9438	0.1155	0.1238	0.9751	0.0975	0.1022	0.9799	0.0960	0.0961	0.9816
	Example 3											
	Var	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2
Fourier	0.0629	0.0933	0.9816	0.0622	0.0623	0.9884	0.0625	0.0660	0.9872	0.0625	0.0627	0.9878
Nadaraya Watson	0.4495	0.4580	0.9198	0.5257	0.5258	0.9093	0.5047	0.5053	0.9104	0.5051	0.5052	0.9107
No M. Error	0.4635	0.5046	0.8886	0.1736	0.1739	0.9668	0.1499	0.1534	0.9703	0.1330	0.1332	0.9742
	Example 4											
	Var	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2	**Var**	**MSE**	R^2
Fourier	0.0682	0.3290	0.9364	0.0621	0.0701	0.9866	0.0627	0.0632	0.9877	0.0624	0.0631	0.9877
Nadaraya Watson	0.5734	0.5776	0.9098	0.6330	0.6340	0.8910	0.6423	0.6427	0.8868	0.6597	0.6598	0.8866
No M. Error	0.2876	0.5231	0.8942	0.1225	0.1285	0.9756	0.1246	0.1247	0.9755	0.1182	0.1189	0.9770

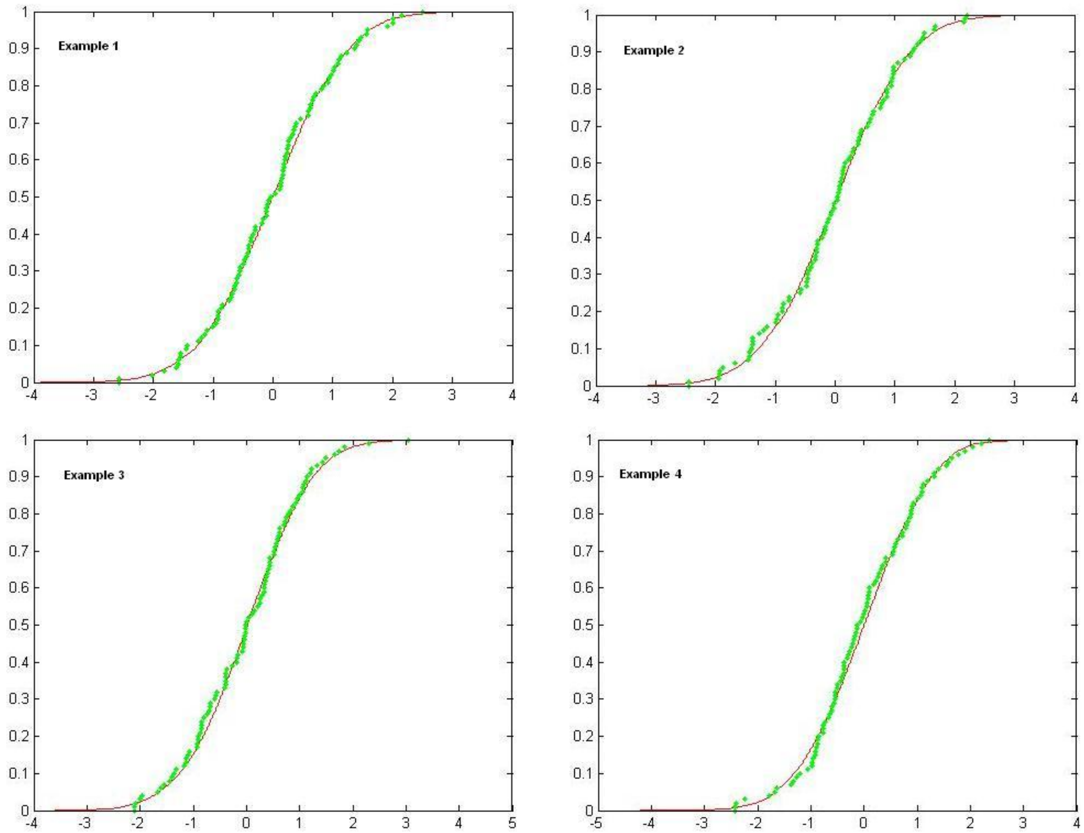

Figure 1. Comparison of the asymptotic distribution and the finite sample of the estimator.

5. Conclusions

This study determines a way to predict regression functions and densities in a partially linear regression model when the measurement error has an unknown distribution. The identification of the density of an unobserved random variable is achieved using the joint density of two error-contaminated measurements of that variable. This finding is particularly valuable because, in practice, strong distributional assumptions often do not hold, limiting the applicability of kernel deconvolution methods. However, repeated measurements are commonly available in real-world datasets where the same variable may be measured at multiple time points or reported by different sources, such as family members or both an employer and employee. Notably, the error in one of the measurements does not have to be mean-zero, which allows for the use of more general or even systematically biased repeated measurements. An estimator of β is derived using this methodology. The asymptotic normality of the proposed estimator is derived. Simulation results show that the resulting rates are comparable to those of Nadaraya–Watson estimators that employ the kernel deconvolution approach, which provides consistent estimation under a much stronger assumption when the density of the measurement error is known, and the case where there is no measurement error.

Acknowledgments

This work was funded by Scientific Research Projects Unit of Dicle University (Project No. FEN.25.010). We also thank to Prof. Dr. H. I. Tutalar and Prof. Dr. Mujgan Tez for their usefull suggestions and contributions.

Use of Generative-AI tools declaration

The author declares that Artificial Intelligence (AI) tools in the creation of this article has not been used.

Conflict of interest

The author confirms the absence of any conflicts of interest related to this study.

References

1. T. W. Anderson, Estimating linear statistical relationships, *Ann. Statist.*, **12** (1984), 1–45. http://doi.org/10.1214/aos/1176346390
2. O. Ashenfelter, A. Krueger, Estimates of the economic returns to schooling from a new sample of twins, NBER Working Paper, 1992, No. 4143.
3. M. E. Borus, G. Nestel, Response bias in reports of father's education and socioeconomic status, *J. Amer. Stat. Assoc.*, **68** (1973), 816–820. http://doi.org/10.1080/01621459.1973.10481429
4. S. Bowles, Schooling and inequality from generation to generation, *J. Polit. Econ.*, **80** (1972), 219–251. http://doi.org/10.1086/259996
5. R. J. Carroll, C. H. Spiegeman, K. G. Lan, K. T. Bailey, R. D. Abbott, On errors-in-variables for binary regression models, *Biometrika*, **71** (1984), 19–25. http://doi.org/10.1093/biomet/71.1.19
6. R. J. Carroll, D. Ruppert, L. A. Stefanski, *Measurement error in nonlinear models*, New York: Chapman and Hall, 1995.
7. H. Chen, Convergence rates for parametric components in a partly linear model, *Ann. Statist.*, **16** (1988), 136–146. http://doi.org/10.1214/aos/1176350695
8. C.-L. Cheng, J. W. Van Ness, *Statistical regression with measurement error: Kendall's library of statistics 6*, New York: Wiley, 2010.
9. J. Fan, Y. K. Truong, Nonparametric regression with errors in variables, *Ann. Statist.*, **21** (1993), 1900–1925. http://doi.org/10.1214/aos/1176349402
10. R. B. Freeman, Longitudinal analysis of the effects of trade unions, *J. Labor Econ.*, **2** (1984), 1–26. https://doi.org/10.1086/298021
11. W. A. Fuller, *Measurement error models*, New York: Wiley, 1987. http://doi.org/10.1002/9780470316665
12. J. A. Hausman, W. K. Newey, J. L. Powell, Nonlinear errors in variables: estimation of some Engel curves, *J. Econometrics*, **65** (1995), 205–233. http://doi.org/10.1016/0304-4076(94)01602-V

13. T. Li, Q. Vuong, *Nonparametric estimation of the measurement error model using multiple indicators*, *J. Multivariate Anal.*, **65** (1998), 139–165. http://doi.org/10.1006/jmva.1998.1741

14. H. Liang, Asymptotic normality of parametric part in partially linear model with measurement error in the non-parametric part, *J. Stat. Plan. Infer.*, **86** (2000), 51–62. http://doi.org/10.1016/S0378-3758(99)00093-2

15. H. Liang, W. Härdle, R. J. Carroll, Large sample theory in a semiparametric partially linear errors-in-variables model, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, 1997, Discussion paper No. 27.

16. E. R. Morey, D. M. Waldman, Measurement error in recreation demand models: the joint estimation of participation, site choice, and site characteristics, *J. Environ. Econ. Manag.*, **35** (1998), 262–276. http://doi.org/10.1006/jeem.1998.1029

17. B. L. S. P. Rao, *Identifiability in stochastic models*, Academic Press, 1992. http://doi.org/10.1016/C2009-0-29097-0

18. S. M. Schennach, Nonparametric regression in the presence of measurement error, *Economet. Theor.*, **20** (2004), 1046–1093. http://doi.org/10.1017/S0266466604206028

19. S. M. Schennach, Estimation of nonlinear models with measurement error, *Econometrica*, **72** (2004), 33–75. https://doi.org/10.1111/j.1468-0262.2004.00477.x

20. L. A. Stefanski, The effects of measurement error on parameter estimation, *Biometrika*, **72** (1985), 583–592. http://doi.org/10.2307/2336730

21. L. A. Stefanski, R. J. Carroll, *Score tests in generalized linear measurement error models*, *J. Roy. Stat. Soc. B*, **52** (1990), 345–359. https://doi.org/10.1111/j.2517-6161.1990.tb01791.x

Appendix

Throughout the study, all the limits are considered as $n \to \infty$, unless otherwise stated. Before providing the proof of the theorem, the following equations must be approved.

Assumption 1(iii) ensures that all expectations are well defined. Using Assumption 1(i) for a given $\overline{x}^* \in \mathbb{R}$, [18] demonstrated that

$$\widehat{\phi}_0(\xi) = \exp\left(\int_0^{\xi} \frac{i\widehat{m}_{\mathcal{X}}(\zeta)}{\widehat{m}_1(\zeta)} d\zeta\right)$$

and

$$\widehat{M}_0(\overline{x}^*, h_n) = \frac{1}{2\pi}\int \kappa(h_n\xi)\,\widehat{\phi}_0(\xi)\exp(-i\xi\overline{x}^*)\,d\xi.$$

Using these results for every $k \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, n\}$, the proof can be formulated as follows:

$$\begin{aligned}
\widehat{\varpi}_{x_k}(\overline{x}^*, h_n) &= \frac{1}{2\pi}\int \kappa(h_n\xi)\,\widehat{\varrho}_{x_k}(\xi)\exp(-i\xi\overline{x}^*)\,d\xi \\
&= \frac{1}{2\pi}\int \kappa(h_n\xi)\,\widehat{\phi}_0(\xi)\,\frac{\widehat{m}_{x_k}(\xi)}{\widehat{m}_1(\xi)}\exp(-i\xi\overline{x}^*)\,d\xi \\
&= \frac{1}{2\pi}\int \kappa(h_n\xi)\,\widehat{\phi}_0(\xi)\,\frac{E[x_k\exp(i\xi z)]}{E[\exp(i\xi z)]}\exp(-i\xi\overline{x}^*)\,d\xi
\end{aligned}$$

$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\widehat{\phi}_0(\xi)\frac{E[x_k\exp(i\xi x^*)]E[\exp(i\xi\Delta z)]}{\widehat{\phi}_0(\xi)E[\exp(i\xi\Delta z)]}$$
$$\times\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)E[x_k\exp(i\xi x^*)]\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\left[\int x_k f(x^*)\exp(i\xi x^*)dx^*\right]\exp(-i\xi\overline{x}^*)d\xi$$
$$= \int h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)x_k f(x^*)dx^*$$

and

$$\widehat{\varpi}_y(\overline{x}^*,h_n) = \frac{1}{2\pi}\int \kappa(h_n\xi)\widehat{\varrho}_y(\xi)\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\widehat{\phi}_0(\xi)\frac{\widehat{m}_y(\xi)}{\widehat{m}_1(\xi)}\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\widehat{\phi}_0(\xi)\frac{E[y\exp(i\xi z)]}{E[\exp(i\xi z)]}\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\widehat{\phi}_0(\xi)\frac{E[y\exp(i\xi x^*)]E[\exp(i\xi\Delta z)]}{\widehat{\phi}_0(\xi)E[\exp(i\xi\Delta z)]}$$
$$\times\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)E[E[y|x^*]\exp(i\xi x^*)]\exp(-i\xi\overline{x}^*)d\xi$$
$$= \frac{1}{2\pi}\int \kappa(h_n\xi)\left[\int E[y|x^*]f(x^*)\exp(i\xi x^*)dx^*\right]\exp(-i\xi\overline{x}^*)d\xi$$
$$= \int h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)E[y|x^*]f(x^*)dx^*.$$

Hence,

$$\widehat{\omega}_{x_k}(\overline{x}^*,h_n) = \frac{\widehat{\varpi}_{x_k}(\overline{x}^*,h_n)}{\widehat{M}_0(\overline{x}^*,h_n)}$$

and

$$\widehat{\omega}_y(\overline{x}^*,h_n) = \frac{\widehat{\varpi}_y(\overline{x}^*,h_n)}{\widehat{M}_0(\overline{x}^*,h_n)}.$$

The following lemmas serve as major steps of formulation the proof of the theorem:

Lemma 1. Let $V_1,\dots,V_n$ be independent random variables with 0 mean and $\sup_j E|V_j|^r \le C < \infty$ for $(r \ge 2)$. Let $\left\{a_{jt}, j,t = 1,\dots,n\right\}$ be a sequence of positive numbers such that $\sup_{j,t\le n}|a_{jt}| \le n^{-p_1}$ for some $0 < p_1 < 1$ and $\sum_{t=1}^{n} a_{jt} = O(n^{p_2})$ for $p_2 \ge \max(0, 2/r - p_1)$. Then,

$$\max_{1\le t\le n}\left|\sum_{j=1}^{n} a_{jt}V_j\right| = O(n^{-s}\log n), \quad s = (p_1-p_2)/2 \quad a.s.$$

Proof. The conclusion follows from [15].

Lemma 2. Suppose that Assumptions 3(i) and 4(i) hold. Then,

$$\max\left|G_k(\overline{x}^*) - \frac{E\left[G_k(\overline{x}^*)\,h_n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = o(1)\,; k = 0, \ldots, p,$$

where $G_0(.) = g(.)$ and $G_k(.) = w_{jk}(.)$ for $k = 1, \ldots, p$.

Proof. As stated before, if β were known, taking $X^T\beta$ into y as $y^* = g(x^*) + \Delta y$, the estimate of the nonparametric function is parallel to the function proposed by [18]. It can be found from the [18] that

$$\max\left|g(\overline{x}^*) - \frac{E\left[\{g(\overline{x}^*) + \Delta y\}\,h_n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = O\left(\left(-\frac{(1+\eta)}{2\delta_m}\right)\log n\right)^{\frac{-1}{\alpha_m}}.$$

If we take $V_j = \Delta y_j$ and $a_{jt} = \frac{n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)}{\widehat{M}_0(\overline{x}^*, h_n)}$ in Lemma 1, we can derive that $\sup_{1\le j\le n}\left|\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = o(1)$, and all these imply that

$$\max\left|g(\overline{x}^*) - \frac{E\left[g(\overline{x}^*)\,h_n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = o(1)\,.$$

The other part of the proof is also similar. Using Assumption 4(ii), which assumes that $w_{jk}(.)$ are Lipschitz continuous, for $c_n = n^{-1/3}\log n$, we can obtain

$$\max_{1\le j\le n}\left|w_{jk}(\overline{x}^*) - \frac{E\left[w_j(\overline{x}^*)\,h_n^{-1}K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = O(c_n)\,.$$

Finally, Lemma 2 holds for $w_{jk}(.)$ by letting $c_n \to 0$.

Lemma 3. Suppose that Assumptions 3(i) and 4(i) hold. Then,

$$\lim_{n\to\infty} n^{-1}\widetilde{X}^T\widetilde{X} = B,$$

where B is given in Assumption 4(i).

Proof. Let $\widetilde{w}_{jx_k}(\overline{x}^*) = w_{jk}(\overline{x}^*) - \widehat{\omega}_{x_k}(\overline{x}^*, h_n)$. Following from

$$\begin{aligned}
x_{jk} &= w_{jk}(\overline{x}^*) + \Lambda_{jk},\\
\widetilde{x}_{js} &= x_{js} - \widehat{\omega}_{x_s}(\overline{x}^*, h_n) = w_{js}(\overline{x}^*) + \Lambda_{js} - \widehat{\omega}_{x_s}(\overline{x}^*, h_n) = \widetilde{w}_{jx_s}(\overline{x}^*) + \Lambda_{js},\\
\widetilde{x}_{jm} &= x_{jm} - \widehat{\omega}_{x_m}(\overline{x}^*, h_n) = w_{jm}(\overline{x}^*) + \Lambda_{jm} - \widehat{\omega}_{x_m}(\overline{x}^*, h_n) = \widetilde{w}_{jx_m}(\overline{x}^*) + \Lambda_{jm},
\end{aligned}$$

the (s, m) element of $\widetilde{X}^T\widetilde{X}$, $(s, m = 1, \ldots, p)$ defined as

$$\sum_{j=1}^{n}\widetilde{x}_{js}\widetilde{x}_{jm} = \sum_{j=1}^{n}\Lambda_{js}\Lambda_{jm} + \sum_{j=1}^{n}\widetilde{w}_{jx_s}(\overline{x}^*)\Lambda_{jm} + \sum_{j=1}^{n}\widetilde{w}_{jx_m}(\overline{x}^*)\Lambda_{js}$$

$$+\sum_{j=1}^{n}\widetilde{w}_{jx_s}\left(\overline{x}^*\right)\widetilde{w}_{jx_m}\left(\overline{x}^*\right)\overset{def}{=}\sum_{j=1}^{n}\Lambda_{js}\Lambda_{jm}+\sum_{q=1}^{3}R_{nsm}^{(q)}.$$

$\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n}\Lambda_j\Lambda_j^T = B$ is a well-known property of the strong law of large numbers. Furthermore, using

$$\begin{aligned}
&E\left[\{w_s(\overline{x}^*)+\Lambda_s\}h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]\\
&=\int h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\left\{w_{js}(\overline{x}^*)+\Lambda_{js}\right\}f(x^*)\,dx^*\\
&=\frac{1}{2\pi}\int\kappa(h_n\xi)\left[\int\left\{w_{js}(\overline{x}^*)+\Lambda_{js}\right\}f(x^*)\exp(i\xi x^*)\,dx^*\right]\exp(-i\xi\overline{x}^*)\,d\xi\\
&=\frac{1}{2\pi}\int\kappa(h_n\xi)\left[\int w_{js}(\overline{x}^*)f(x^*)\exp(i\xi x^*)\,dx^*+\int\Lambda_{js}f(x^*)\exp(i\xi x^*)\ dx^*\right]\exp(-i\xi\overline{x}^*)\,d\xi\\
&=\frac{1}{2\pi}\int\kappa(h_n\xi)\left[\int w_{js}(\overline{x}^*)f(x^*)\exp(i\xi x^*)\,dx^*\right]\exp(-i\xi\overline{x}^*)\,d\xi\\
&+\frac{1}{2\pi}\int\kappa(h_n\xi)\left[\int\Lambda_{js}f(x^*)\exp(i\xi x^*)\,dx^*\right]\exp(-i\xi\overline{x}^*)\,d\xi\\
&=\frac{1}{2\pi}\int\kappa(h_n\xi)E\left[w_s\left(\overline{X}^*\right)\exp(i\xi x^*)\right]\exp(-i\xi\overline{x}^*)\,d\xi+\frac{1}{2\pi}\int\kappa(h_n\xi)E\left[\Lambda_s\exp(i\xi x^*)\right]\exp(-i\xi\overline{x}^*)\,d\xi\\
&=E\left[w_s(\overline{x}^*)h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]+E\left[\Lambda_s h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]
\end{aligned}$$

we conclude that

$$\begin{aligned}
\widetilde{w}_{jx_s}(\overline{x}^*)&=w_{js}(\overline{x}^*)-\widehat{\omega}_{x_s}(\overline{x}^*,h_n)\\
&=w_{js}(\overline{x}^*)-\frac{E\left[\{w_s(\overline{x}^*)+\Lambda_s\}h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}\\
&=w_{js}(\overline{x}^*)-\frac{E\left[w_s(\overline{x}^*)h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}-\frac{E\left[\Lambda_s h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}\\
&=\widetilde{w}_{js}(\overline{x}^*)-\frac{E\left[\Lambda_s h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}.
\end{aligned}$$

It can be derived $\frac{E\left[\Lambda_s h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}=o(1)$ by taking $a_{jt}=\frac{n^{-1}K\left(h_n^{-1}\left(x_j^*-\overline{x}^*\right)\right)}{\widehat{M}_0(\overline{x}^*,h_n)}$, $V_j=\Lambda_{js}$ in Lemma 1, and therefore, from Lemma 2, it is certain that $\widetilde{w}_{jx_s}(\overline{x}^*)=o(1)$.

This shows that $R_{nsm}^{(3)}=o(n)$, and using the Cauchy-Schwarz inequality, $R_{nsm}^{(1)}=o(n)$ and $R_{nsm}^{(2)}=o(n)$.

Proof of Theorem. We can demonstrate $\widehat{\beta}_n$ as follows:

$$\begin{aligned}
\widehat{\beta}_n&=\left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left(\widetilde{X}^T\widetilde{y}\right)=\left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left[\widetilde{X}^T\left(\widetilde{X}\beta+\widetilde{g}(\overline{x}^*)+\widetilde{\Delta y}\right)\right]\\
&=\left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left[\widetilde{X}^T\widetilde{X}\beta\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g}+\sum_{j=1}^{n}\widetilde{x}_j\Delta y_j-\sum_{j=1}^{n}\widetilde{x}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*,h_n)}\right\}\right]
\end{aligned}$$

$$= \left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left(\widetilde{X}^T\widetilde{X}\right)\beta + \left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left[\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g} + \sum_{i=j}^{n}\widetilde{x}_j\Delta y_j - \sum_{j=1}^{n}\widetilde{X}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\}\right]$$

$$= \beta + \left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left[\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g} + \sum_{j=1}^{n}\widetilde{x}_j\Delta y_j - \sum_{j=1}^{n}\widetilde{x}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\}\right].$$

Then,

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\beta}_n - \beta\right)\\
&= \sqrt{n}\left(\widetilde{X}^T\widetilde{X}\right)^{-1}\left[\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g} - \sum_{j=1}^{n}\widetilde{x}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\} + \sum_{j=1}^{n}\widetilde{x}_j\Delta y_j\right]\\
&\overset{def}{=} A(n)\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g} - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\} + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\Delta y_j\right]\\
&\overset{def}{=} A(n)\left[B(n) - C(n) + D(n)\right],
\end{aligned}$$

where $\widetilde{g} = g\left(\overline{x}^*\right) - \frac{E\left[g\left(\overline{x}^*\right)h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}$, $A(n) = n\left(\widetilde{X}^T\widetilde{X}\right)^{-1}$, $B(n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\widetilde{g}$, $C(n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\left\{\frac{E\left[\Delta y h_n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\}$ and $D(n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{x}_j\Delta y_j$.

From Lemma 3, it is evident that $A(n)$ converges to B^{-1}.

In addition, we can prove that $D(n)$ converges to normal distribution with zero mean and covariate matrix $\Sigma^2 B^{-1}$ using the central limit theorem and Lemma 3, as demonstrated in [7].

Let us show that $B(n)$ and $C(n)$ converge in probability to zero. Affirm that

$$\begin{aligned}
\sum_{j=1}^{n}\widetilde{x}_{jk}\widetilde{g} &= \sum_{j=1}^{n}\left[\widetilde{\Lambda}_{jk} + \widetilde{w}_{jk}\right]\widetilde{g}\\
&= \sum_{j=1}^{n}\left[\Lambda_{jk} + \widetilde{w}_{jk} - \frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\Lambda_k\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right]\widetilde{g}\\
&= \sum_{j=1}^{n}\Lambda_{jk}\widetilde{g} + \sum_{j=1}^{n}\widetilde{w}_{jk}\widetilde{g} - \sum_{j=1}^{n}\frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\Lambda_k\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\widetilde{g}.
\end{aligned}$$

If we consider $r = 2, V_j = \Lambda_{jk}, a_{jt} = \widetilde{g}, 1/4 < p_1 < 1/3, p_2 = 1 - p_1$ in Lemma 1 then $\left|\sum_{j=1}^{n}\Lambda_{jk}\widetilde{g}\right| = O\left(n^{-(2p_1-1)/2}\log n\right)$. Using Abel's inequality, we find $\left|\sum_{j=1}^{n}\widetilde{w}_{jk}\widetilde{g}\right| \le n\max|\widetilde{g}|\max_{1\le j\le n}\left|\widetilde{w}_{jk}\right|$. Considering $G_0 = g$ and $G_k = w_{jk}$ in Lemma 2, we obtain

$$\max|\widetilde{g}| = \max\left|g\left(\overline{x}^*\right) - \frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)g\left(\overline{x}^*\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right| = o(1) \tag{5.1}$$

and

$$\max_{1\le j\le n}\left|\widetilde{w}_{jk}\right| = \max_{1\le j\le n}\left|w_{jk}\left(\overline{x}^*\right) - \frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)w_k\left(\overline{x}^*\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right| = o(1), \tag{5.2}$$

respectively. This means that

$$\left|\sum_{j=1}^{n} \widetilde{w}_{jk}\widetilde{g}\right| \leq n \max |\widetilde{g}| \max_{1\leq j\leq n} \left|\widetilde{w}_{jk}\right| = o(1).$$

With Abel's inequality, we obtain

$$\sum_{j=1}^{n} \frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\widetilde{g} \leq n \max |\widetilde{g}| \max_{1\leq j\leq n} \left|\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right|. \tag{5.3}$$

In Lemma 1, considering $r = 3, V_j = \Lambda_{jk}$, $a_{jt} = \frac{n^{-1}K\left(h_n^{-1}(x^*-\overline{x}^*)\right)}{\widehat{M}_0(\overline{x}^*,h_n)}$, $p_1 = 2/3$, $p_2 = 0$, we can obtain

$$\max_{1\leq j\leq n} \left|\frac{E\left[\Lambda_k K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| = O\left(n^{-1/3}\log n\right), k = 1, \ldots, p \quad a.s. \tag{5.4}$$

Using Eqs (5.1) and (5.4), it can be easily observed that

$$\sum_{j=1}^{n} \frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\widetilde{g} \leq o(1).$$

With these equations, it is decided that $B(n) = o(1)$.

Using the definition of $C(n)$, we obtain

$$\begin{aligned}
\sqrt{n}C(n) &= \sum_{j=1}^{n} \left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\widetilde{x}_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\Delta y_j \\
&= \sum_{j=1}^{n} \left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\Delta y_j + \sum_{j=1}^{n} \left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\widetilde{w}_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\Delta y_j \\
&- \sum_{j=1}^{n} \left[E\left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\frac{K\left(h_n^{-1}(x^* - \overline{x}^*)\right)}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\right]\Delta y_j.
\end{aligned}$$

If we consider $r = 2, V_j = \Delta y_j, a_{jt} = \frac{E\left[K\left(h_n^{-1}(x^*-\overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*,h_n)}$, $1/4 < p_1 < 1/3$ and $p_2 = 1 - p_1$ in Lemma 1, we can obtain

$$\left|\sum_{j=1}^{n} \left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Lambda_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\Delta y_j\right| = O\left(n^{-(2p_1-1)/2}\log n\right).$$

Using Abel's inequality, we obtain

$$\left|\sum_{j=1}^{n} \left\{\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\widetilde{w}_k\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right\}\Delta y_j\right| \leq n \max_{1\leq j\leq n} \left|\frac{E\left[K\left(h_n^{-1}(x^* - \overline{x}^*)\right)\Delta y\right]}{\widehat{M}_0(\overline{x}^*, h_n)}\right| \max_{j\leq n} \left|\widetilde{w}_{jk}\right|.$$

In Lemma 1, considering $r = 3$, $V_j = \Delta y_j$, $a_{jt} = \frac{n^{-1}K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}$,$p_1 = 2/3$, $p_2 = 0$, we can obtain

$$\max_{1\le j\le n}\left|\frac{E\left[\Delta y K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right| = O\left(n^{-1/3}\log n\right), a.s. \tag{5.5}$$

Equations (5.2) and (5.5), imply that

$$\left|\sum_{j=1}^{n}\left\{\frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\widetilde{w}_k\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\}\Delta y_j\right| \le o\left(n^{1/2}\right).$$

From Eqs (5.4) and (5.5), it is evident that the following term is also $o\left(n^{1/2}\right)$,

$$\left|\sum_{j=1}^{n}\left[E\left\{\frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\Lambda_k\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\frac{K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right\}\right]\Delta y_j\right|$$
$$\le n\max_{j\le n}\left|\frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\Delta y\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right|\max_{j\le n}\left|\frac{E\left[K\left(h_n^{-1}\left(x^* - \overline{x}^*\right)\right)\Lambda_k\right]}{\widehat{M}_0\left(\overline{x}^*, h_n\right)}\right|.$$

We can now conclude that $C(n) = o(1)$.