Estimation of Nonparametric Multiple Regression Measurement Error Models with Validation Data

Zanhua Yin, Fang Liu

College of Mathematics and Computer Science, Gannan Normal University, Ganzhou, China
Email: yinzh226@163.com, yinzh226@nenu.edu.cn

Received 3 November 2015; accepted 27 December 2015; published 30 December 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

In this article, we develop estimation approaches for nonparametric multiple regression measurement error models when both independent validation data on covariables and primary data on the response variable and surrogate covariables are available. An estimator which integrates Fourier series estimation and truncated series approximation methods is derived without any error model structure assumption between the true covariables and surrogate variables. Most importantly, our proposed methodology can be readily extended to the case that only some of covariables are measured with errors with the assistance of validation data. Under mild conditions, we derive the convergence rates of the proposed estimators. The finite-sample properties of the estimators are investigated through simulation studies.

Keywords

Ill-Posed Inverse Problem, Linear Operator, Measurement Errors, Nonparametric Regression, Validation Data

1. Introduction

We can consider the following nonparametric regression model of a scaler response \( Y \) on an explanatory variable \( X \)

\[
Y = g(X) + \varepsilon,
\]

where \( g(\cdot) \) is assumed to be a smooth, continuous but unknown nonparametric regression function and \( \varepsilon \) is a noise variable with \( E(\varepsilon | X) = 0 \) and \( E(\varepsilon^2) < \infty \). It is not uncommon that the explanatory variable \( X \) is measured with error and instead only its surrogate variable \( W \) can be observed. In this case, one observes inde-
pendent replicates \((W_i, Y_i), 1 \leq i \leq N\), of \((W, Y)\) rather than \((X, Y)\), where the relationship between \(W_i\) and \(X_i\) may or may not be specified. If not, the missing information for the statistical inference will be taken from a sample \((W_j, X_j), N+1 \leq j \leq N+n\), of so-called validation data independent of the primary (surrogate) sample. The objective of this manuscript is to estimate the unknown function \(g(\cdot)\) via the surrogate data \(\{(Y_i, W_i)\}_{i=1}^{N}\) and the validation data \(\{(X_j, W_j)\}_{j=N+1}^{N+n}\).

A wide number of problems of similar type have attracted considerable attention in research literature over the past two decades (see, [1]-[6]). For instance, a quasi-likelihood method is intensively studied by [7]. A regression calibration approach is developed by [8] and [9] propose a method based on simulation-extrapolation (SIMEX) estimation. Other related methods include Bayesian approaches (see, [12]), semi-parametric method (see, [13] [14]), empirical likelihood method (see, [15]) and the instrumental variable method (see, [16]). Unfortunately, all these work mostly assume some parametric relationships between covariates and responses. Recently, nonparametric estimators of \(g\) have been developed by [17] and [18]. [17] develops a kernel-based approach for nonparametric regression function estimation with surrogate data and validation sampling. However, the method is not applicable for model (1) since it assumes that the response but not the covariable is measured with error. [18] proposes a nonparametric estimator which integrates local linear regression and Fourier transformation method when both explanatory and surrogate variable are scalars. Nonetheless, their method cannot be extended to multidimensional problems in which the explanatory variable vectors can consist of variables being measured with or/and without errors. For additional references and relevant topics for nonparametric regression models with measurement errors, ones may consult [19] and the references therein.

In practice, nonparametric estimation of \(g\) may not be an easy task since, as explained in Section 2, the relation that identifies \(g\) is a Fredholm equation of the first kind, i.e.

\[
Tg = m, \tag{2}
\]

which may lead to an ill-posed inverse problem. Ill-posed inverse problem related to nonparametric regression model has received considerable attention recently. [20] consider kernel-based estimators while [22] and [23] develop series or sieve estimators. However, their methods require an instrumental variable, and assume that the explanatory variable \(X\) is directly observable without errors. In this article, we propose a nonparametric estimation approach which consists of two major steps. First, we propose estimators of generalized Fourier coefficients of \(T\) and \(m\) based on surrogate and validation data. Second, we replace the infinite-dimensional operator \(T\) by the finite-dimensional approximation to avoid higher-order coefficient estimation, and hence it develops an estimator of \(g\). Furthermore, we extend this method to the case that only some of covariates are measured with errors. Under mild conditions, the consistencies of the resulting estimators are established and the convergence rates are also derived.

This article is arranged as follows. In Section 2, we first describe our estimation approach for the case that the covariates are all measured with errors. Extension to the case that only some of covariates are measured with errors will be discussed as well. We derive the convergence rates of our estimators under some regularity conditions in Section 3. Section 4 presents some numerical results from simulation studies. A brief discussion will be given in Section 5. Proofs of the theorems are presented in Appendix.

2. Methodology

We first describe our estimation approach for the case that the covariates are all measured with errors. In addition to the independent and identically distributed (i.i.d.) primary observations \(\{(Y_i, W_i)\}_{i=1}^{N}\) from model (1), assume that i.i.d. validation data \(\{(X_j, W_j)\}_{j=N+1}^{N+n}\) are also available. We shall suppose that \(X\) and \(W\) are both \(d\)-dimensional random vectors. Without loss of generality, let the supports of \(X\) and \(W\) both be contained in \(\chi = [0, 1]^d\) (otherwise, one can carry out monotone transformations of \(X\) and \(W\)).

In the following we let \(f_{XW}, f_{X}, f_{W}\) denote respectively the joint density of \((X, W)\), marginal densities of \(X\) and \(W\). Then we have

\[
E(Y | W = w) = E[g(X) | W = w] = \int_{\chi} g(x) \frac{f_{XW}(x, w)}{f_W(w)} \, dx. \tag{3}
\]
According to Equation (3), \( g \) is actually the solution to an integral equation called Fredholm equation of the first kind. Let 

\[
L_2(\chi) = \left\{ \varphi : \chi \rightarrow \mathcal{R}, \text{s.t. } \| \varphi \| = \left( \int_{\chi} |\varphi(x)|^2 \, dx \right)^{1/2} < \infty \right\}.
\]

Define the operator 

\[
T : L_2(\chi) \rightarrow L_2(\chi)
\]

by

\[
(T\varphi)(w) = \int_{\chi} \varphi(x) f_{XW}(x, w) \, dx.
\]

Hence, Equation (3) is equivalent to the operator equation

\[
m(w) = (Tg)(w).
\]

For the unknown smooth function \( g : \chi \rightarrow \mathcal{R} \), we assume that \( g \in \mathcal{H}_s \) where

\[
\mathcal{H}_s = \left\{ g \in \mathcal{W}^s(\chi) : \| g \|_{\mathcal{W}^s} < c \right\},
\]

where \( c \) is a positive and finite constant. \( \mathcal{W}^s(\chi) \) denotes the Sobolev space of smoothness \( s \geq 1 \), that is

\[
\mathcal{W}^s(\chi) = \left\{ \varphi \in L_2(\chi) : \forall \alpha = (\alpha_1, \ldots, \alpha_d), \| \varphi \|_{\mathcal{W}^\alpha} = \left( \sum_{\alpha} \left\| \frac{\partial^{\alpha}}{\partial x^{\alpha}} \varphi \right\|_{L_2(\chi)}^2 \right)^{1/2} < \infty \right\}.
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_d) \), \( \| \varphi \|_{\mathcal{W}^\alpha} = \alpha = \alpha_1 + \cdots + \alpha_d \), and the derivatives \( \partial^{\alpha} \varphi = \frac{\partial^{\alpha}}{\partial x^{\alpha}} \varphi \). Given an integer \( s \), the norm \( \| g \|_{\mathcal{W}^s} \) is

\[
\| g \|_{\mathcal{W}^s} = \left( \sum_{\alpha} \left\| \frac{\partial^{\alpha}}{\partial x^{\alpha}} g \right\|_{L_2(\chi)}^2 \right)^{1/2},
\]

Here \( \| \cdot \| \) denotes the norm on \( L_2(\chi) \).

An estimator of \( g \) can then be obtained by replacing \( T \) and \( m \) by their series estimators based on surrogate data and validation data, and solving the resultant empirical version of (4). As before, let \( \{ \phi_k, k = 1, 2, \ldots \} \) denote a complete, orthonormal sequence for \( L_2(\chi) \). Hence, we can write

\[
m(w) = \sum_{k=1}^N m_k \phi_k(w), \quad \text{and} \quad f_{XW}(x, w) = \sum_{k=1}^N d_{kl} \phi_k(x) \phi_l(w),
\]

where \( m_k \) and \( d_{kl} \) represent the generalized Fourier coefficients of \( m \) and \( f_{XW} \), respectively. Intuitively, we can obtain the estimators of \( m_k \), \( d_{kl} \) and \( f_{XW}(x, w) \) by

\[
\hat{m}_k = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N f_{XW}(x_j, w_i) \phi_k(x_j) \phi_l(w_i),
\]

\[
\hat{d}_{kl} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N f_{XW}(x_j, w_i) \phi_k(x_j) \phi_l(w_i),
\]

\[
\hat{f}_{XW}(x, w) = \sum_{k=1}^N \sum_{l=1}^N \hat{d}_{kl} \phi_k(x) \phi_l(w),
\]

respectively, where the integer \( q \) is a truncation point which is the main smoothing parameter in the approximating Fourier series. The operator \( T \) can then be consistently estimated by

\[
\hat{T}_c \varphi(w) = \int_{\chi} \varphi(x) \hat{f}_{XW}(x, w) \, dx.
\]

Define the subset of \( \mathcal{H}_s \):

\[
\mathcal{H}_s = \left\{ \varphi = \sum_{k=1}^N \phi_k \phi_k : \| \varphi \|_{\mathcal{W}^s} < c \right\}.
\]

The estimator of \( g(x) \) can be computed by

\[
\hat{g} = \arg \min_{\varphi \in \mathcal{H}_s} \| \hat{T}_c \varphi - \hat{m} \|^2.
\]

(5)
Remark 1. Let $\hat{W}_n$ be the $N \times q$ matrix whose $(i,j)$ element is $\phi_j(W_i)$ and $\hat{Y}_n = (Y_1, \cdots, Y_n)^T$ be the observed vector of $Y$ based on the surrogate data $(Y_i, W_i)^T_{i=1}^n$. Let $\hat{W}_n$ and $\hat{W}_n$, respectively, denote the $n \times q$ matrices whose $(j,k)$ elements are $\phi_k(W_j)$ and $\phi_k(X_j)$ based on the validation data. If $A_n = \frac{1}{n} \hat{W}_n^T \hat{Y}_n$, then the solution to (5) assumes the following form

$$\hat{g}(x) = \sum_{k=1}^{q} \hat{g}_k \phi_k(x)$$

(6)

where $\{\hat{g}_k, k = 1, \cdots, q\}$ is given by $(\hat{g}_1, \cdots, \hat{g}_q)^T = (A_n^T A_n)^{-1} A_n^T b_n$.

Next, we extend the estimator in (5) to nonparametric regression measurement error models with multi-covariates, that is

$$Y = g(X, Z) + \epsilon,$$

(7)

where $X$ is measured with error and $W$ being its observed surrogate variable, and $Z$ is measured without error. Let $(Y_i, W_i, Z_i)^T_{i=1}^n$ be a random sample from model (7), and $(X_j, W_j, Z_j)^T_{j=N+1}^{N+n}$ be i.i.d. validation observations. We assume that $X$ and $W$ are supported on $\{0,1\}^d$, and $Z$ is supported on $\{0,1\}^p$.

Let $f_{X,W,Z}$, $f_{X,Z}$ and $f_{W,Z}$ denote respectively the joint density of $(X,W,Z)$, marginal densities of $X$ and $W$, all conditioning on $Z = z$. Similar to (3), for any $z \in [0,1]^p$, we have

$$m(w,z) = E(Y | W = w, Z = z) f_{W,Z}(w),$$

(8)

where $m(w,z) = \int_{\mathbb{R}} \phi_\psi(x) f_{X,W,Z}(x,w) dx$, and the operator $T_\psi$ is defined by

$$T_\psi g(z) = \int_{\mathbb{R}} \phi_\psi(x) f_{X,W,Z}(x,w) dx,$$

where $\phi_\psi = \phi(\cdot, z)$ is any function on $L_2(\chi)$.

To obtain the estimator of $g(x,z)$, we set $K_h(u) = K(u/h)$ where $K$ is a kernel function and $h > 0$ is a bandwidth. Let $K_{p,k}(z) = \prod_{i=1}^{p} K_{a_i}(z)$. We consider the following estimators

$$\hat{m}_k = \frac{(Nh_n^{-p})^{-1} \sum_{j=1}^{N} \phi_j(W_j) K_{p,h_n}(z-Z_j)}{(Nh_n)^{-1} \sum_{j=1}^{N} K_{p,h_n}(z-Z_j)},$$

and

$$\hat{d}_l = \frac{(nh_n^{-p})^{-1} \sum_{j=N+1}^{N+n} \phi_j(X_j) \phi(W_j) K_{p,h_n}(z-Z_j)}{(nh_n)^{-1} \sum_{j=N+1}^{N+n} K_{p,h_n}(z-Z_j)}.$$

Then we have

$$\hat{m}(w,z) = \sum_{k=1}^{q} \hat{m}_k \phi_k(w),$$

and

$$\hat{d}(x,w) = \sum_{l=1}^{q} \hat{d}_l \phi_l(x) \phi(W_l).$$

Define the operator $\hat{T}_\psi$ by

$$(\hat{T}_\psi \phi_\psi)(w,z) = \int_{\mathbb{R}} \phi_\psi(x) \hat{f}_{X,W,Z}(x,w) dx,$$

for any $\phi_\psi \in L_2(\chi)$.

Then, for any $z \in [0,1]^p$, the estimator of $g(x,z)$ is
\[
\hat{g} = \arg\min_{\phi \in \mathcal{H}_w} \left\| \frac{1}{n} \sum \phi_k - \hat{m} \right\|^2.
\] (9)

**Remark 2.** Denote \( f_n(z) = \left(Nh^p\right)^{-1} \sum_{i=1}^{N} K_{p,h^q} \left(z - Z_i\right) \) and \( f_n(z) = \left(nh^p\right)^{-1} \sum_{j=N+1}^{\infty} K_{p,h^q} \left(z - Z_j\right) \). Let
\[
\tilde{H}_n = f_n^{-1}(z) \times \text{diag} \left(K_{p,h^q}(z - Z_1), \ldots, K_{p,h^q}(z - Z_N)\right)
\]
and
\[
\tilde{\mathcal{H}}_n = f_n^{-1}(z) \times \text{diag} \left(K_{p,h^q}(z - Z_{N+1}), \ldots, K_{p,h^q}(z - Z_{N+n})\right).
\]
If \( \tilde{A}_n = \frac{1}{nh^p} \tilde{H}_n \tilde{H}_n^\top \) and \( \tilde{b}_n = \frac{1}{Nh^p} \tilde{W}_n \tilde{H}_n \tilde{y}_n \), then the solution to (9) has the following form
\[
\hat{g}(x,z) = \sum_{k=1}^{q} \tilde{g}_k \phi_k(x),
\] (10)
where \( \{\tilde{g}_k, k = 1, \ldots, q\} \) is given by \( \left(\tilde{g}_{z1}, \ldots, \tilde{g}_{zq}\right) = (\tilde{A}_n^\top \tilde{A}_n)^{-1} \tilde{A}_n^\top \tilde{b}_n \).

**Remark 3.** If \( Z \) is discretely distributed with finite support, then \( g(x,z) \) can be estimated by (9) with \( K_{\lambda}(u) \) being replaced by \( I(u = 0) \), where \( I(\cdot) \) is the indicator function.

### 3. Theoretical Properties

In this section, we study the asymptotic properties of the estimators proposed in Section 2. We define \( \tau_n \) (\( \tau_m \)) as a sieve measure of ill-posedness (see, [23]):

\[
\tau_n = \sup_{\phi \in \mathcal{H}_w, \phi \neq 0} \frac{\|\phi\|}{\|\mathbb{I}_{\phi} - \eta\|}, \quad \tau_m = \sup_{\phi \in \mathcal{H}_w, \phi \neq 0} \frac{\|\phi\|}{\|\mathbb{I}_{\phi} - \eta\|}.
\]

First, we investigate the large-sample properties of the estimator \( \hat{g} \). For this purpose, we present the following regular conditions which are mild and can be found in [24] and [23].

**A1.** (i) The support of \((X,W)\) is contained in \( \mathcal{X}^2 \); (ii) The joint probability measure of \((Y,W)\) is absolutely continuous with respect to the product probability measure of \(Y\) and \(W\); and (iii) The support of \(W\) is a cartesian product of compact connected intervals on which \(W\) has a probability density function that is bounded away from zero.

**A2.** For each \( w \in \mathcal{X}\), the function \( E(Y^2 | W = w) \) is bounded by \( c \).

**A3.** (i) \( g \in \mathcal{H}_w \) with \( g \in \mathcal{W}^r_\infty(\mathcal{X}) \) and \( r > 1 \); and (ii) \( m(W) \) belongs to \( \mathcal{W}_\infty^r(\mathcal{X}) \) with \( r > 3/2 \).

**A4.** The set of functions \( \{\phi_k, k = 1,2,\cdots\} \) is an orthonormal, complete basis for \( L_2(\mathcal{X}) \), and bounded uniformly over \(\mathcal{X}\).

**A5.** (i) \( \lim n/N = \lambda \) for some constant \( 0 < \lambda \leq 1 \); and (ii) \( q = q(N,n) \to 0, \quad q/N \to 0, \quad q/n \to 0 \) as \( n \to \infty, \quad N \to \infty \).

**Theorem 1.** Under conditions A1-A5, as \( N \to \infty \) and \( n \to \infty \), we have
\[
\left\| \hat{g} - g \right\|_X = O_p \left( q^{-3/2} + \tau_n \times \left[ q^{-3/2} + N^{-1/2} q^{1/2} + n^{-1/2} q^{1/2} \right] \right),
\] (11)
where \( \left\| \cdot \right\|_X \) denotes \( \left\| \phi \right\|_X = \left\{ \int \phi^2(x) f_X(x) dx \right\}^{1/2} \) for any \( \phi \in L_2(\mathcal{X}) \).

In (11), the term \( q^{-3/2} \) arises from the bias of \( \hat{g} \) caused by truncating the series approximation of \( g \). The truncation bias decreases as \( s \) increases and \( g \) becomes smoother. Therefore, the smoother of \( g \) the faster the rate of convergence of \( \hat{g} \). The terms \( \tau_n N^{-1/2} q^{1/2} \) and \( \tau_n n^{-1/2} q^{1/2} \) are respectively induced by random surrogate sampling errors and random validation sampling errors in the estimates of the generalized Fourier coefficients \( \hat{g}_k \). When \( X \) is measured without error, the convergence rate of the sieve estimator of \( g \) is \( O_p \left( q^{-3/2} + N^{-1/2} q^{1/2} \right) \).

Comparing this rate to that in (11), we note that the bias part \( q^{-3/2} \) is of the same order, however, the standard deviation part blows up from \( N^{-1/2} q^{1/2} \) to \( \tau_n \times \left[ q^{-3/2} + N^{-1/2} q^{1/2} + n^{-1/2} q^{1/2} \right] \).

A more precise behaviour of the estimator can be obtained but depends on \( \tau_n \) as [23] discussed, which can be classified into mildly ill-posed case and severely ill-posed case. In the next corollary, we obtain these rates for the two particular cases.
Corollary 1. Suppose the assumptions of Theorem 1 are satisfied.

(i) Let \( \tau_n = O\left( q^{(r-s)/d} \right) \) (mildly ill-posed case) with \( r-s-1/2 > 0 \), and \( q \propto N^{d/(2r+d)} \), we have
\[
\| \hat{g} - g \|_X = O_P\left( N^{-\kappa/(2r+d)} \right);
\]
(ii) Let \( \tau_n = O\left( q^{(r-s)/d} L(q) \right) \) (severely ill-posed case) with \( r-s-1/2 > 0 \), and \( q \propto N^{d/(2r+d)} \), we have
\[
\| \hat{g} - g \|_X = O_P\left( N^{-\kappa/(2r+d)} L\left( N^{d/(2r+d)} \right) \right),
\]
where the function \( L(q) \) goes to \( \infty \) slowly such that \( L(q)/q^\varepsilon \to 0 \) for all \( \varepsilon > 0 \).

Remark 4. According to Corollary 1(i), the convergence rate becomes \( O\left( N^{-2\varepsilon/3} \right) \) when \( r = 3 \), \( s = 2 \) and \( d = 1 \). This is slower than that of the sieve estimator of a conditional mean function which can achieve the rate of convergence \( N^{-2/3} \).

Next, we study the large-sample properties of the estimator \( \hat{g}(x,z) \). For this purpose, we make the following assumptions.

B1. (i) The support of \( (X,W) \) is contained in \( \chi^2 \), and \( Z \) is supported on \( [0,1]^p \); (ii) Conditioning on \( Z = z \), the joint probability measure of \( (Y,W) \) is absolutely continuous with respect to the product probability measure of \( Y \) and \( W \) and; (iii) Conditioning on \( Z = z \), the support of \( W \) is a cartesian product of compact connected intervals on which \( W \) has a probability density function that is bounded away from zero.

B2. For each \( (w,z) \in [0,1]^{k+p} \), \( E\{Y^2 \mid W = w, Z = z\} \) is bounded by \( c \).

B3. (i) For each \( z \in [0,1]^p \), (8) has a solution \( \hat{g}(\cdot,z) \in \mathcal{H}_s \) with \( g(\cdot,z) \in \mathcal{H}_s \) and \( s > 1 \) that does not depend on \( z \) and; (ii) For each \( z \in [0,1]^p \), \( m(W,Z = z) \) belongs to \( \mathcal{H}_s \) with \( r > 1/2 \).

B4. (i) The set of functions \{\( \phi_k, k = 1,2,\cdots \}\) is an orthonormal, complete basis for \( L_b(\chi) \), and bounded uniformly over \( k \) and; (ii) The kernel function \( K \) is a symmetrical, twice continuously differentiable function on \([-1,1]\), and \( \int_{-1}^{1} u^k K(u)du = c \), with \( c \neq 0 \) being some finite constant.

B5. (i) \( N, n, h_N, h_n \) satisfy the conditions that \( Nh_N \to \infty \) and \( nh_n \to \infty \); (ii) \( h_n = c_n n^{-1/(2r+d)} \) and \( h_N = C_n N^{-1/(2r+p)} \), where \( c_n \) and \( C_n \) are constants and \( 0 < c_n, C_n < \infty \); and (iii) \( q = q(n)d/(2r+d) \) with \( \kappa = 2r/(2r+p) \) for some constant \( c_n < \infty \).

B6. (i) \( \lim nN = \lambda \) for some constant \( 0 < \lambda \leq 1 \); and (ii) \( q = q(N,n) \to \infty \), \( q/n \to 0 \), \( q/n \to 0 \) as \( n \to \infty \), \( N \to \infty \).

Theorem 2. Suppose assumptions B1 - B6 are satisfied. For each \( z \in [0,1]^p \), let \( \tau_n = O\left( q^{(r-s)/d} \right) \) with \( r-s-1/2 > 0 \), we have
\[
\| \hat{g}(x,z) - g(x,z) \|_{W^2} = O_P\left[ N^{-\kappa/(2r+d)} \right].
\]

The proofs of all the theorems are reported in Appendix.

4. Numerical Properties

In this subsection, we conducted a simulation study of the finite-sample performance of the proposed estimators. First, we choose the cosine sequence with \( \phi_1(x) = 1 \) and \( \phi_k(x) = \sqrt{2} \cos((k-1)\pi x), k = 2,\cdots \) as the complete orthonormal basis for \( L_b([0,1]) \), then get our estimators (denoted as \( \hat{g}_N(x) \) and \( \hat{g}(x,z) \)) following (6) and (10). For comparison, we consider [18] method (denoted as \( \hat{g}_N \)), and used the standard Nadaraya-Watson estimator with a Epanechnikov kernel to calculate \( \hat{g}_N \) based on the primary dataset. It should be pointed out that \( \hat{g}_N \) can serve as a gold standard in the simulation study, even though it is practically unachievable due to measurement errors. The performance of estimator \( \hat{g}^{est} \) is assessed by using the average integrated squared errors (MISE)
\[
\text{MISE} = \frac{1}{M} \sum_{s=1}^{M} \left( \hat{g}^{est}(u_s) - g(u_s) \right)^2,
\]
where \( u_s, s = 1,\cdots,M \), are grid points at which \( g^{est}(u_s) \) is evaluated.

Example 1: We considered model (1) with the regression function being
\[ g(x) = \left[ 1 - (2x - 1)^2 \right] I_{\{x \in [0,1]\}}, \]

and \( e \) being distributed as \( N(0,0.25) \). To perform this simulation, we generate \( X \) from a standard normal distribution, that is, \( X \sim N(0,1) \), and assume that \( W = \eta X + \left( 1 - \eta^2 \right)^{1/2} v \), \( v \sim N(0,1) \), and \( \eta \) is the standard deviation of the measurement error. Then, trim \( X \) and \( W \) in \( [-2.5, 2.5] \) and scale to \( [0,1] \) respectively. Only results for \( \eta = 0.7 \) and \( \eta = 0.9 \) are reported here. Simulations were run with different validation and primary data sizes \( (n,N) \) ranging from \( (10,30) \) to \( (60,300) \) according to the ratio \( \gamma = N/n = 3 \) and \( \gamma = N/n = 5 \), respectively. For each case, 1000 simulated data sets were generated for each sample size of \( (n,N) \).

It is interesting to compare our estimator \( \hat{g} \) with the estimators \( \hat{g}_D \) and \( \hat{g}_N \). Here, since our estimator \( \hat{g} \) involves the regulation parameter \( q \), we therefore present the following cross-validation (CV) selection criterion

\[ q_{CV} = \arg \min_q \left[ \frac{1}{n} \sum_{i=1}^{n} \left( Y_i - \hat{g}^{(x)}(W_i,q) \right)^2 \right], \]

where the subscript \(-i\) meant that the estimator was constructed without using the \( i \)th observation \( (Y_i,W_i) \). For \( \hat{g}_D \), [18] proposed an automatic way of choosing the smoothing parameters \( h_n, b_n \) and \( q \). For \( \hat{g}_N \), the CV approach is used for choosing bandwidth \( h_n \).

**Figure 1** shows the regression function curve \( g(x) \), and the curves of the median MISEs based on 1000 replicated estimates of \( \hat{g} , \hat{g}_D \) and \( \hat{g}_N \) with \( \eta = 0.7 \) under different sample size. From **Figure 1**, both \( \hat{g} \) and \( \hat{g}_D \) successfully capture the patterns of the true regression curves and have smaller bias than \( \hat{g}_N \). As expected, \( \hat{g}_N \) fails to produce accurate function curve estimates. In addition, it is obvious that the quality of our proposed estimator improve with the increase of sample sizes.

**Table 1** compares, for various sample sizes, the results obtained for estimating curve \( g(x) \) when \( \eta = 0.7 \) or \( \eta = 0.9 \). The estimated MISEs which were evaluated on a grid of 201 equidistant values of \( x \) in \( [0,1] \) are presented. Our results show that the estimators \( \hat{g} \) and \( \hat{g}_D \) outperform \( \hat{g}_N \). It is noteworthy that our proposed

<table>
<thead>
<tr>
<th>(n,N)</th>
<th>( \eta = 0.7 )</th>
<th>( \eta = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = 3 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10,30)</td>
<td>3.7858</td>
<td>5.1262</td>
</tr>
<tr>
<td>(30,90)</td>
<td>1.8630</td>
<td>2.2648</td>
</tr>
<tr>
<td>(60,180)</td>
<td>1.3056</td>
<td>1.8036</td>
</tr>
<tr>
<td>( \gamma = 5 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10,50)</td>
<td>2.2956</td>
<td>3.1562</td>
</tr>
<tr>
<td>(30,150)</td>
<td>1.9711</td>
<td>2.1350</td>
</tr>
<tr>
<td>(60,300)</td>
<td>1.0628</td>
<td>1.5073</td>
</tr>
</tbody>
</table>
estimator generally performs better than the estimator proposed by [18] for the resultant MISEs of \( \hat{g} \) are usually smaller. Also, the performance of \( \hat{g} \) improves (i.e. the corresponding MISEs decrease) considerably as the sample sizes increases. For any nonparametric method in measurement error regression problem, the quality of the estimator also depends on the discrepancy of the observed sample. That is, the performance of the estimator depends on the variances of measurement error. Here, we compare the results for different values of \( \eta \).

As expected, Table 1 shows that the effect of the variances on the estimator performance is obvious.

Example 2: We considered model (7) with the regression function being

\[
g(x, z) = \left[1 - (2x - 1)^2\right]^2 \exp(4x - 1.4z - 2),
\]

and \( \epsilon \) being distributed as \( N(0,0.01) \). The covariate \((X,Z)^T\) was generated from a bivariate normal distribution \( N(0,\Sigma) \) with \( \text{var}(X) = \text{var}(Z) = 1 \) and the correlation coefficient between \( X \) and \( Z \) being 0.6, and \( W = \rho X + (1 - \rho^2)^{1/2} \varepsilon, \varepsilon \sim N(0,1) \). Then, \( X, W \) and \( Z \) in \([-2.5,2.5] \) and scale to \([0,1] \) respectively.

Results for \( \rho = 0.7 \) and \( \rho = 0.9 \) are reported. Simulations were run with different validation and primary data sizes \((n,N)\) ranging from \((10,30)\) to \((60,300)\) according to the ratio \( \gamma = N/n = 3 \) and \( \gamma = N/n = 5 \), respectively. For each case, 1000 simulated data sets were generated for each sample size of \((n,N)\).

Here, we only compared our estimator \( \hat{g}(x,z) \) with the naive estimator \( \hat{g}_N \) which is the multivariate kernel regression estimator based on the primary dataset \( \{(Y_i, W_i, Z_i)\}_{i=1}^N \), since [18] method cannot be applied to multivariate cases. Here, we used the Epanechnikov kernel function \( K(x) = 0.75(1-x^2) \), \( |x| \leq 1 \) for \( \hat{g}(x,z) \) and used an product kernel \( K(x_1, x_2) = K_0(x_1)K_0(x_2) \) with \( K_0(x) = -15/8 x^2 + 9/8 |x| \leq 1 \) for \( \hat{g}_N \). For the naive estimator \( \hat{g}_N \), bandwidth selection rules were considered by [25]. For our estimator \( \hat{g}(x,z) \), we used the cross-validation approach to choosing the three parameters \( h_N, h_\eta \) and \( q \). For this purpose, \( h_\eta \) and \( (h_N,q) \) are selected separately as follows.

\[
\hat{f}_j(z;h_\eta) = \frac{1}{nh_\eta} \sum_{j=N+1}^{N+n} K_h(z-Z_j).
\]

Here, we adopt the cross-validation (CV) approach to estimate \( h_\eta \) by

\[
\hat{h}_\eta = \arg\min_{h_\eta} \frac{1}{N}\sum_{j=N+1}^{N+n} \left\{ Z_j - \hat{f}_j(z;h_\eta) \right\}^2,
\]

where the subscript \(-j\) denotes the estimator being constructed without using the \( j \)th observation. After obtaining \( \hat{h}_\eta \), we then select \( (h_N,q) \) by

\[
\left( \hat{h}_N,q \right) = \arg\min_{h_N,q} \frac{1}{N}\sum_{i=1}^{N} \left\{ Y_i - \hat{g}^{*}(W_i,Z_i;\hat{h}_N,q) \right\}^2,
\]

where the subscript \(-i\) denotes the estimator being constructed without using the \( i \)th observation \( \{(Y_i, W_i, Z_i)\} \).

We compute MISE at \( 101 \times 101 \) grid points of \((x,z)\) ranging in \([0,1] \times [0,1] \). Table 2 reports the MISE for estimating curves \( g(x,z) \) when \( \rho = 0.7 \) or \( \rho = 0.9 \) for various sample sizes. Table 2 shows that our proposed estimator substantially outperformed the naive kernel estimator \( \hat{g}_N \). It is obvious that our proposed estimator \( \hat{g} \) has much smaller MISE than \( \hat{g}_N \).

5. Discussion

In this paper, we propose a new method for estimating non-parametric regression measurement error models using surrogate data and validation sampling. The covariates are measured with errors while we do not assume any error model structure between the true covariates and the surrogate variable. Most importantly, our proposed method can be readily extended to the multi-covariates model, say, \( y = f(x,z) + \epsilon \) where \( x \) is measured with error but \( z \) is measured exactly. Numerical results show that the new estimators are promising in terms of cor-
recting the bias arising from the errors-in-variables. It generally preforms better than the approach proposed by [18].

**Acknowledgements**

This work was supported by NSFC11301245, NSFC11501126 and Natural Science Foundation of Jiangxi Province of China under grant number 20142BAB211018.

**References**


Appendix

Proof of Theorem 1

Let $T^*: L_2(\chi) \to L_2(\chi)$ denotes the adjoint operator of $T$. Under assumption A1(ii), the self-adjoint operators of $TT^*$ and $T^*T$ have the same eigenvalue sequence $\{\mu_k^2\}$ with $\mu_k^2 = \mu_k^2 \geq \cdots$. Moreover, we assume that the corresponding eigenfunctions of the operators $TT^*$ and $T^*T$ are also orthonormal basis $\{\phi_k \}$, $k = 1, 2, \cdots$, and for all $k \geq 1$

$$T\phi_k = \mu_k \phi_k, \quad T^*\phi_k = \mu_k \phi_k; \quad T^*T\phi_k = \mu_k^2 \phi_k, \quad TT^*\phi_k = \mu_k^2 \phi_k.$$

Define

$$g_n(x) = \sum_{k=1}^q g_k \phi_k(x), \quad \text{and} \quad m_n(w) = \sum_{k=1}^q m_k \phi_k(w).$$

Let $T_n$ be the operator whose kernel is

$$t_n(x, w) = \sum_{k=1}^q \sum_{l=1}^q t_{kl} \phi_k(x) \phi_l(w),$$

then $T_n g_n = m_n$. By the definition of $\mathcal{H}_n$, we have $g_n \in \mathcal{H}_n$.

Lemma 1. Under conditions A1 and A3(i) and the sieve space $\mathcal{H}_n$, we have

1) $\|T^* \{g - g_n\}\| \leq \text{const.} \times \|g - g_n\|;$

2) $\tau_n \leq 1/\mu_k$.

Lemma 2. Under conditions A1, A3(ii) and A4, we have

$$\sup_{\theta \in \mathcal{H}_n} \left\| \hat{T}_n - T \right\| \varphi = O_p\left( q^{-\epsilon/4} + q^{1/2} n^{-1/2} \right).$$

By some modifications of the proof of Theorem 2 in [23] and applying the Theorem 7 in [24], the proofs of Lemma 1 and Lemma 2 are straightforward and are omitted.

Proof of Theorem 1. By the triangle inequality, we have

$$\|\hat{g} - g\| \leq \|\hat{g} - g_n\| + \|g_n - g\|.$$  \tag{12}

By the definition of $\mathcal{H}_n$ and condition A3(i), we have

$$\|g_n - g\| = O\left(q^{-\epsilon/4}\right).$$

see e.g. [26] for Fourier series.

Next, by the definition of $\tau_n$ and the triangle inequality, we have

$$\|\hat{g} - g_n\| \leq \tau_n \|T (\hat{g} - g_n)\|.$$

We now analyze the term $\|T (\hat{g} - g_n)\|$. By the triangle inequality, we have

$$\|T (\hat{g} - g_n)\| = \left\| (T - \hat{T}_n) \hat{g} + \hat{T}_n \hat{g} - \hat{m} + \hat{m} - m + T (g - g_n) \right\|$$

$$\leq \left\| (T - \hat{T}_n) \hat{g} \right\| + \left\| \hat{T}_n \phi_k - m \right\| + \| m - m \| + \| T (g - g_n) \|.$$

By conditions A2, A4 and central limit theorem, we can show that $\|\hat{m} - \hat{m}\| = O_p\left( q/N \right)^{1/2}$. From condition A3(ii), we have $\|\hat{m} - m\| = O\left(q^{-\epsilon/4}\right)$. Hence, $\|\hat{m} - \hat{m}\| = O\left(q^{-\epsilon/4} + q^{1/2} N^{-1/2}\right)$. In addition, by the definition of $\hat{g}$ and the triangle inequality, we have

$$\|\hat{T}_n \phi_k - \hat{m}\| \leq \|\hat{T}_n \phi_k - \hat{m}\| \leq \|\hat{T}_n - T\| \phi_k + \|T (g_n - g)\| + \|m - \hat{m}\|.$$

These and Lemma 2 imply

$$\|\hat{g} - g\| \leq \tau_n \times \left\{O_p\left(q^{-\epsilon/4} + q^{1/2} N^{-1/2} + q^{1/2} n^{-1/2}\right) + O\left(\|T (g_n - g)\|\right) \right\}.$$
This and Lemma 1 imply
\[ \| \hat{g} - g \| \leq \| g - g_n \| + \tau_n \times O_p \left( q^{-r/2d} + q^{1/2} N^{-r/2} + q^{1/2} n^{-1/2} \right). \]  
(13)

The theorem follows immediately from (12)-(13).

**Proof of Theorem 2**

**Lemma 3.** For each \( z \in [0,1]^p \), define
\[ g_{z,n} (x) = \sum_{k=1}^{q} g_{z,k} \phi_k(x), \quad \text{and} \quad m_{z,n} (w) = \sum_{k=1}^{q} m_{z,k} \phi_k(w). \]

Let \( T_{z,n} \) be the operator whose kernel is
\[ t_{z,n} (x,w) = \sum_{k=1}^{q} \sum_{l=1}^{q} d_{z,k,l} \phi_k(x) \phi_l(w), \]
then \( T_{z,n} g_{z,n} = m_{z,n} \). By the definition of \( \mathcal{H}_{z,n} \), we have \( g_{z,n} \in \mathcal{H}_{z,n} \).

**Proof of Theorem 2.** For each \( z \in [0,1]^p \), by the triangle inequality, we have
\[ \| \hat{g}(x,z) - g(x,z) \| \leq \| \hat{g} - g_{z,n} \| + \| g_{z,n} - g \| \leq \tau_n \times \| T_{z,n} (\hat{g} - g_{z,n}) \| + \| g_{z,n} - g \|, \]

By assumption B3(i), it is easy to show that \( \| g_{z,n} - g \| = O(q^{-r/d}) \).

Similar to the proof of Theorem 1, we have
\[ \| T_{z,n} (\hat{g} - g_{z,n}) \| \leq \| (T_{z,n} - \hat{T}_{z,n}) \hat{g} \| + \| \hat{T}_{z,n} \hat{g} - \hat{m} \| + \| \hat{m} - m \| + \| m - g_{z,n} \|. \]

According to assumptions B2, B3(ii), B4, and B5(i), we can show that \( \| \hat{m} - E\hat{m} \| = O_p \left( q^{1/2} N^{-r/(2r+2p)} \right), \) \( \| E\hat{m} - m \| = O \left( q^{1/2} N^{-r/(2r+2p)} + q^{-r/d} \right) \). In addition, by some modifications of the proof of Lemma 2, under assumptions B1, B3(ii), B4, B5(i) and B6, we have
\[ \sup_{\phi \in \mathcal{H}_{z,n}} \| \hat{T}_{z,n} \phi - T \phi \| = O_p \left( q^{-r/d} + q^{1/2} n^{-r/(2r+2p)} \right). \]

For the term \( \| T_{z,n} (g_{z,n} - g) \| \), under assumptions B1, B3(i) and the sieve space \( \mathcal{H}_{z,n} \), we have
\[ \tau_n \times \| T_{z,n} (g_{z,n} - g) \| \leq \text{const.} \times \| g - g_{z,n} \|. \]

Combining all these results, we complete the proof. \( \square \)