

# Estimation of Nonparametric Regression Models with Measurement Error Using Validation Data

Fang Liu, Zanhua Yin\*

College of Mathematics and Computer Science, Gannan Normal University, Ganzhou, China

Email: liufang\_gnnu@163.com, \*yinzh226@163.com

**How to cite this paper:** Liu, F. and Yin, Z.H. (2017) Estimation of Nonparametric Regression Models with Measurement Error Using Validation Data. *Applied Mathematics*, 8, 1454-1463.

<https://doi.org/10.4236/am.2017.810106>

**Received:** September 17, 2017

**Accepted:** October 28, 2017

**Published:** October 31, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

We consider the problem of estimating a function  $g$  in nonparametric regression model when only some of covariates are measured with errors with the assistance of validation data. Without specifying any error model structure between the surrogate and true covariables, we propose an estimator which integrates orthogonal series estimation and truncated series approximation method. Under general regularity conditions, we get the convergence rate of this estimator. Simulations demonstrate the finite-sample properties of the new estimator.

## Keywords

Ill-Posed Inverse Problems, Measurement Errors, Nonparametric Regression, Orthogonal Series

## 1. Introduction

Consider the following nonparametric regression model of a scalar response  $Y$  on multi-covariates  $(X, Z)$

$$Y = g(X, Z) + \varepsilon, \quad (1)$$

where  $g(\cdot)$  is an unknown function and  $\varepsilon$  is a noise variable with  $E(\varepsilon|X, Z) = 0$  and  $E(\varepsilon^2) < \infty$ . It is not uncommon that  $Z$  is measured exactly but  $X$  is measured with error and instead only its surrogate variable  $W$  can be observed. Throughout we assume

$$E[\varepsilon|W, Z] = 0 \quad \text{with probability 1,} \quad (2)$$

which is always satisfied if, for example,  $W$  is a function of  $X$  and some

independent noise (see [1]).

The relationship between the true variable and the surrogate variable can be rather complicated. Misspecification of this relationship may lead to a serious misinterpretation of the data. Common solution is to use the help of validation data to infer the missing information. To be specific, one observes independent replicates  $(W_i, Z_i, Y_i)$ ,  $1 \leq i \leq N$ , of  $(W, Z, Y)$  rather than  $(X, Z, Y)$ , 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  $(X_j, W_j, Z_j)$ ,  $N+1 \leq j \leq N+n$ , of so-call validation data independent of the primary (surrogate) sample. We aim at estimating the unknown function  $g(\cdot)$  by using the surrogate data  $\{(Y_i, W_i, Z_i)\}_{i=1}^N$  and the validation data  $\{(X_j, W_j, Z_j)\}_{j=N+1}^{N+n}$ .

Recently, statistical inference based on surrogate data and a validation sample has attracted considerable attention (see [2]-[13]), and the above referenced authors developed suitable methods for different models. However, all these works mostly are concerned with the parametric or semi-parametric relationships between covariates and responses, and these approaches are difficult to generalize to nonparametric regression model. [14] and [15] proposed two nonparametric estimators for nonparametric regression model with measurement error using validation data, but their methods are not applicable to our problem since [14] assumes the response rather than the covariable is measured with error, and the method proposed by [15] applies for one-dimensional explanatory variable only.

This article is organized as follows. In Section 2 we propose a regularization-based method. Under general regularity conditions, we give the convergence rate of our estimator in Section 3. Section 4 provides some numerical results from simulation studies, whereas proofs of the theorems are presented in Appendix.

## 2. Description of the Estimator

Recall model (1) and the assumptions below it. We assume that  $X$ ,  $W$  and  $Z$  are all real-valued random variables. The extension to random vectors complicates the notation but does not affect the main ideas and results. Without loss of generality, let the supports of  $X$ ,  $W$  and  $Z$  all be contained in  $[0, 1]$  (otherwise, one can carry out monotone transformations of  $X$ ,  $W$  and  $Z$ ).

Let  $f_{XWZ}$  and  $f_{WZ}$  denote respectively the joint density of  $(X, W, Z)$  and marginal density of  $(W, Z)$ . Then, according to (2), we have

$$\begin{aligned} E(Y | W = w, Z = z) &= E[g(X, Z) | W = w, Z = z] \\ &= \int g(x, z) \frac{f_{XWZ}(x, w, z)}{f_{WZ}(w, z)} dx. \end{aligned} \quad (3)$$

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

$$L_2([0, 1]) = \left\{ \varphi: [0, 1] \rightarrow \mathcal{R}, \text{ s.t. } \|\varphi\| = \left( \int |\varphi(x)|^2 dx \right)^{1/2} < \infty \right\}.$$

Define the operator  $T_z : L_2([0,1]) \rightarrow L_2([0,1])$  as

$$(T_z \varphi_z)(w, z) = \int \varphi_z(x) f_{xwz}(x, w, z) dx,$$

where  $\varphi_z(\cdot) = \varphi(\cdot, z)$  is any function on  $L_2([0,1])$ . So that Equation (3) is equivalent to the operator equation

$$m(w, z) = (T_z g)(w, z). \tag{4}$$

According to Equation (4), the function  $g$  is the solution of a Fredholm integral equation of the first kind, and this inverse problem is known to be ill-posed and needs a regularization method. A variety of regulation schemes are available in the literature (see e.g. [16]) but we focus in this paper on the Tikhonov regularized solution:

$$g^\alpha = \arg \min_g \left[ \|T_z g - m\|^2 + \alpha \|g\|^2 \right], \tag{5}$$

where the penalization term  $\alpha > 0$  is the regularization parameter.

We define the adjoint operator  $T_z^*$  of  $T_z$

$$(T_z^* \psi)(x) = \int \psi_z(w) f_{xwz}(x, w, z) dw,$$

where  $\psi_z(w) \in L_2([0,1])$ . Then the regularized solution (5) is equivalently:

$$g^\alpha = (\alpha I + T_z^* T_z)^{-1} T_z^* m. \tag{6}$$

To obtain the estimator of  $g(x, z)$ , we consider the orthogonal series method and kernel method. Under some regularity conditions in Section 3, for each  $z \in [0,1]$ ,  $f_{xwz}(\cdot, \cdot, z)$  and  $m(\cdot, z)$  may be approximated by a truncated orthogonal series,

$$f_{xwz}^K(x, w, z) = \sum_{k=1}^K \sum_{l=1}^K d_{zkl} \phi_k(x) \phi_l(w) \text{ and } m^K(w, z) = \sum_{k=1}^K m_{zk} \phi_k(w),$$

where

$$\begin{aligned} d_{zkl} &= \iint \phi_k(x) \phi_l(w) f_{xwz}(x, w, z) dx dw \\ &= E[\phi_k(X) \phi_l(W) | Z = z] f_z(z), \end{aligned}$$

and

$$\begin{aligned} m_{zk} &= \int \phi_k(w) E(Y | W = w, Z = z) f_{wz}(w, z) dw \\ &= E[Y \phi_k(W) | Z = z] f_z(z), \end{aligned}$$

Here,  $\{\phi_k\}$  is an orthonormal basis of  $L_2([0,1])$  which may be trigonometric, polynomial, spline, wavelet, and so on. A discussion of different bases and their properties can be found in the literature (see e.g. [17], [18]). Only to be specific, here and in what follows we are considering the normalized Legendre polynomials on  $[0,1]$ , which can be obtained through the Rodrigues' formula

$$\phi_k(x) = \frac{1}{k! \sqrt{2k+1}} \frac{d^k}{dx^k} \left[ (x^2 - x)^k \right]. \tag{7}$$

The integer  $K$  is a truncation point which is the main smoothing parameter in

the approximating series.

Let  $K_h(u) = K(u/h)$  where  $K(\cdot)$  is a kernel function and  $h > 0$  is a bandwidth. We consider the following estimators,

$$\hat{d}_{zkl} = \frac{1}{nh_n} \sum_{j=N+1}^{N+n} \phi_k(X_j) \phi_l(W_j) K_{h_n}(z - Z_j),$$

and

$$\hat{m}_{zk} = \frac{1}{Nh_N} \sum_{i=1}^N Y_i \phi_k(W_i) K_{h_N}(z - Z_i).$$

Then, for each  $z \in [0, 1]$ , we have

$$\hat{f}_{XWZ}(x, w, z) = \sum_{k=1}^K \sum_{l=1}^K \hat{d}_{zkl} \phi_k(x) \phi_l(w) \text{ and } \hat{m}(w, z) = \sum_{k=1}^K \hat{m}_{zk} \phi_k(w).$$

The operators  $T_z$  and  $T_z^*$  can then be estimated by

$$(\hat{T}_z \varphi_z)(w, z) = \int \varphi_z(x) \hat{f}_{XWZ}(x, w, z) dx \text{ and } (\hat{T}_z^* \psi_z)(x, z) = \int \psi_z(w) \hat{f}_{XWZ}(x, w, z) dw.$$

Hence, for each  $z \in [0, 1]$ , the estimator of  $g(x, z)$  is

$$\tilde{g}^\alpha = (\alpha I + \hat{T}_z^* \hat{T}_z)^{-1} \hat{T}_z^* \hat{m}_z. \quad (8)$$

**Remark 2.1.** Let  $\tilde{D}_n$  be the  $K \times K$  matrix whose  $(k, l)$  element is  $\hat{d}_{zkl}$ , and  $\tilde{b}_N = (\hat{m}_{z1}, \dots, \hat{m}_{zK})^T$ . Then estimator (8) has the following form

$$\tilde{g}^\alpha(x, z) = \sum_{k=1}^K \tilde{g}_{zk} \phi_k(x), \quad (9)$$

where  $\{\tilde{g}_{zk}, k = 1, \dots, K\}$  is given by  $(\tilde{g}_{z1}, \dots, \tilde{g}_{zK})^T = (\alpha I_K + \tilde{D}_n \tilde{D}_n^T)^{-1} \tilde{D}_n \tilde{b}_N$ .

### 3. Theoretical Properties

In this section, we introduce the assumptions that will be used below to study the statistical properties of the estimator. We shall consider the following assumptions:

(A1) (i) The support of  $(X, W, Z)$  is contained in  $[0, 1]^3$ ; (ii) Conditioning on  $Z = z$ , the joint density  $f_{XWZ}$  of  $(X, W, Z)$  is square integrable w.r.t the Lebesgue measure on  $[0, 1]^2$ .

(A2) (i) The  $r$  order partial or mixed partial derivative of  $f_{XWZ}$  with respect to  $(x, w)$ , and the  $r$  order partial derivative of  $f_{XWZ}$  with respect to  $z$ , are both continuous in  $(x, w) \in [0, 1]^2$  and for each  $z \in [0, 1]$ ; (ii) The  $s$  order partial derivative of  $m(w, z)$  with respect to  $z$ , is continuous in  $w \in [0, 1]$  and for each  $z \in [0, 1]$ .

(A3)  $E(Y^2 | W = w, Z = z)$  is uniformly bounded in  $(w, z) \in [0, 1]^2$ .

(A4) The kernel function  $K(\cdot)$  is a symmetrical, twice continuously differentiable function on  $[-1, 1]$ , and  $\int_{-1}^1 u^j K(u) du = 0$  for  $j = 1, \dots, r-1$  and  $\int_{-1}^1 u^r K(u) du = c$ , with  $c \neq 0$  being some finite constant.

Assumption (A1) is sufficient condition for  $T_z$  to be a Hilbert-Schmidt operator and therefore to be compact (see [19], Theorem 2.34). As a result of

compactness, there exists a singular values decomposition. For each  $z \in [0, 1]$ , let  $\lambda_{zk}, k \geq 0$  be the sequence of singular values of  $T_z$ , then there exist the two orthonormal sequences  $\varphi_{zk}, k \geq 0$ , and  $\psi_{zl}, l \geq 0$  such that:

$$T_z \varphi_{zk} = \lambda_{zk} \psi_{zk}, T_z^* \psi_{zk} = \lambda_{zk} \varphi_{zk}; T_z^* T_z \varphi_{zk} = \lambda_{zk}^2 \varphi_{zk}, T_z T_z^* \psi_{zk} = \lambda_{zk}^2 \psi_{zk}.$$

Note that the regularization bias is

$$g - g^\alpha = \left[ I - (\alpha I + T_z^* T_z)^{-1} T_z^* T_z \right] g = \sum_{k=1}^{\infty} \frac{\alpha}{\alpha + \lambda_{zk}^2} \langle g, \varphi_{zk} \rangle \varphi_{zk}.$$

In order to control the speed of convergence to zero of the regularization bias  $g - g^\alpha$ , we introduce the following regularity space  $\Psi_\beta$  for  $\beta > 0$ :

$$\Psi_\beta = \left\{ \varphi_z \in L_2([0, 1]) \text{ such that } \sum_{k \geq 0} \frac{\langle \varphi_z, \varphi_{zk} \rangle^2}{\lambda_{zk}^{2\beta}} < +\infty \right\}.$$

We then obtain the following result by applying Proposition 3.11 in Carrasco *et al.* (2007).

**Proposition 3.1.** *Suppose Assumption (A1) hold, for each  $z \in [0, 1]$ , if  $g(\cdot, z) \in \Psi_\beta$ , then we have  $\|g(\cdot, z) - g^\alpha(\cdot, z)\|^2 = O(\alpha^{\beta \wedge 2})$ , where  $\beta \wedge 2 = \min\{\beta, 2\}$ .*

Therefore, when the regularization parameter  $\alpha$  is pushed towards zero, the smoother the function  $g$  of interest (*i.e.*  $g \in \Psi_\beta$  for larger  $\beta$ ) is, the faster the rate of convergence to zero of the regularization bias will be.

Let  $\gamma = \min\{r, s\}$  and  $\tau = 2r\gamma / [(\gamma + 1)(2r + 1)]$ , we then obtain the following convergence rate of  $\tilde{g}^\alpha(x, z)$ .

**Theorem 3.1.** *Suppose Assumptions (A1)-(A4) are satisfied. Then, for each  $z \in [0, 1]$ , if  $g(\cdot, z) \in \Psi_\beta$ , we have*

$$\begin{aligned} \|\tilde{g}^\alpha(\cdot, z) - g(\cdot, z)\|^2 = O_p \left\{ \left[ K \left( h_N^{2\gamma} + \frac{1}{Nh_N} \right) + \frac{1}{K^{2\gamma}} + K^2 \left( h_n^{2r} + \frac{1}{nh_n} \right) \right] \alpha^{-1} \right. \\ \left. + \left[ \frac{1}{K^{2r}} + K^2 \left( h_n^{2r} + \frac{1}{nh_n} \right) \right] \alpha^{(\beta-1) \wedge 0} + \alpha^{\beta \wedge 2} \right\}. \end{aligned}$$

Specially, let  $h_n = O(n^{-1/(2r+1)})$ ,  $h_N = O(N^{-1/(2\gamma+1)})$ ,  $K = O(n^{r/[(\gamma+1)(2r+1)])}$ ,  $\alpha = O(N^{-\tau/(\beta \wedge 2+1)})$  and  $\lim N/n \in (0, \infty)$ . If  $r \leq s$  or  $s < r \leq 2s(s+1)$ , then, for each  $z \in [0, 1]$ , we have

$$\|\tilde{g}^\alpha(\cdot, z) - g(\cdot, z)\|^2 = O_p \left( N^{-\frac{\tau \beta \wedge 2}{\beta \wedge 2 + 1}} \right).$$

The proofs of all the results are reported in the Appendix.

### 4. Simulation Studies

In this section, we briefly illustrate the finite-sample performance of the estimator discussed above. We compare our estimator to the standard Nadaraya-Watson estimator (denoted as  $\tilde{g}_N(x, z)$ ) base on the primary dataset  $\{(Y_i, W_i, Z_i)\}_{i=1}^N$ . In fact,  $\tilde{g}_N(x, z)$  is a gold standard in the simulation study,

even if it is practically unachievable due to measurement errors. Moreover, the performance of estimator  $g^{est}$  is assessed by using the square root of average square errors (RASE)

$$RASE = \left\{ \frac{1}{M} \sum_{s=1}^M [g^{est}(u_s) - g(u_s)]^2 \right\}^{1/2},$$

where  $u_s, s = 1, \dots, M$ , are grid points at which  $g^{est}(u_s)$  is evaluated.

We considered model (1) with the regression function being

$$g(x, z) = \frac{1}{2\pi} \exp(-0.5x^2 - 0.5z^2),$$

and  $\varepsilon$  being distributed as  $N(0, 0.1)$ . The covariate is generated according to  $(X, Z)^T \sim N(0, \Sigma)$  with  $var(X) = var(Z) = 1$  and the correlation coefficient between  $X$  and  $Z$  being 0.6, and  $W = X + \eta v, v \sim N(0, 1)$ . Results for  $\eta = 0.2, \eta = 0.4$  and  $\eta = 0.6$  are reported. Simulations were run with different validation and primary data sizes  $(n, N)$  ranging from  $(10, 30)$  to  $(50, 250)$  according to the ratio  $\rho = N/n = 3$  and  $\rho = N/n = 5$ , respectively. We generate 500 datasets for each sample size of  $(n, N)$ .

To calculate  $\tilde{g}^\alpha(x, z)$ , we used the normalized Legendre polynomials as basis and the standard normal kernel (denote  $K_0(\cdot)$ ). For  $\tilde{g}_N(x, z)$ , we used an product kernel  $K(x_1, x_2) = K_0(x_1)K_0(x_2)$ , and the bandwidth was selected by generalized cross-validation approach (GCV). For our estimator  $\tilde{g}^\alpha(x, z)$ , we used the cross-validation approach to choosing the four parameters  $h_N, h_n, K$  and  $\alpha$ . For this purpose,  $h_N, h_n$  and  $(K, \alpha)$  are selected separately as follows.

Define

$$\hat{f}_Z(z; h_n) = \frac{1}{nh_n} \sum_{j=N+1}^{N+n} K_{h_n}(z - Z_j).$$

and

$$\tilde{f}_Z(z; h_N) = \frac{1}{Nh_N} \sum_{i=1}^N K_{h_N}(z - Z_i).$$

Here, we adopt the cross-validation (CV) approach to estimate  $h_n$  by

$$\hat{h}_n = \arg \min_{h_n} \frac{1}{n} \sum_{j=N+1}^{N+n} \left\{ Z_j - \hat{f}_Z^{(-j)}(Z_j; h_n) \right\}^2,$$

where the subscript  $-j$  denotes the estimator being constructed without using the  $j$ th observation. Similarly, we get  $\hat{h}_N$ . After obtaining  $\hat{h}_N$  and  $\hat{h}_n$ , we then select  $(K, \alpha)$  by

$$(\hat{\alpha}, \hat{K}) = \arg \min_{(\alpha, K)} \frac{1}{N} \sum_{i=1}^N \left\{ Y_i - \sum_{k=1}^K \tilde{g}_{zk}^{(-i)} \phi_k(W_i) \right\}^2,$$

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

We compute the RASE at  $200 \times 200$  grid points of  $(x, z)$ . **Table 1** presents

**Table 1.** The RASE ( $\times 10^{-1}$ ) comparison for the estimators  $\tilde{g}^\alpha(x, z)$  and  $\tilde{g}_N(x, z)$ .

	$(n, N)$	$\eta = 0.2$		$\eta = 0.4$		$\eta = 0.6$	
		$\tilde{g}^\alpha(x, z)$	$\tilde{g}_N(x, z)$	$\tilde{g}^\alpha(x, z)$	$\tilde{g}_N(x, z)$	$\tilde{g}^\alpha(x, z)$	$\tilde{g}_N(x, z)$
$\rho = 3$	(10, 30)	0.4617	0.5348	0.4638	0.5423	0.4825	0.5481
	(30, 90)	0.3804	0.4997	0.4048	0.4687	0.4306	0.5111
	(50, 150)	0.3716	0.4486	0.3734	0.4644	0.3784	0.4709
$\rho = 5$	(10, 50)	0.4508	0.5068	0.4399	0.5176	0.4738	0.5206
	(30, 150)	0.3785	0.4548	0.3800	0.4624	0.3817	0.4687
	(50, 250)	0.3397	0.3636	0.3432	0.3830	0.3483	0.3835

the RASE for estimating curves  $g(x, z)$  when  $\eta = 0.2$ ,  $\eta = 0.4$  and  $\eta = 0.6$  for various sample sizes. It is obvious that our proposed estimator  $\tilde{g}^\alpha$  has much smaller RASE than  $\tilde{g}_N$ . As is expected, our proposed estimating method produces more accurate estimators than the Nadaraya-Watson estimators. Moreover, there is a drastic improvement in accuracy by using our estimator over the Nadaraya-Watson estimator; this improvement increases with  $\rho$ .

## Acknowledgments

This work was supported by GJJ160927 and Natural Science Foundation of Jiangxi Province of China under grant number 20142BAB211018.

## References

- [1] Stute, W., Xue, L. and Zhu, L. (2007) Empirical Likelihood Inference in Nonlinear Errors-in-Covariables Models with Validation Data. *Journal of the American Statistical Association*, **102**, 332-346. <https://doi.org/10.1198/016214506000000816>
- [2] Carroll, R.J. and Stefanski, L.A. (1990) Approximate Quasi-Likelihood Estimation in Models with Surrogate Predictors. *Journal of the American Statistical Association*, **85**, 652-663. <https://doi.org/10.1080/01621459.1990.10474925>
- [3] Carroll, R.J. and Wand, M.P. (1991) Semiparametric Estimation in Logistic Measurement Error Models. *Journal of the Royal Statistical Society: Series B*, **53**, 573-585.
- [4] Carroll, R.J., Gail, M.H. and Lubin, J.H. (1993) Case-Control Studied with Errors in Covariables. *Journal of the American Statistical Association*, **88**, 185-199.
- [5] Cook, J.R. and Stefanski, L.A. (1994) Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, **89**, 1314-1328. <https://doi.org/10.1080/01621459.1994.10476871>
- [6] Sepanski, J. and Lee, L.F. (1995) Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data. *Journal of the American Statistical Association*, **90**, 130-140. <https://doi.org/10.1080/01621459.1995.10476495>
- [7] Stefanski, L.A. and Buzas, J.S. (1995) Instrumental Variable Estimation in Binary Regression Measurement Error Models. *Journal of the American Statistical Association*, **90**, 541-550. <https://doi.org/10.1080/01621459.1995.10476546>
- [8] Wang, Q. and Rao, J.N.K. (2002) Empirical Likelihood-Based Inference in Linear Errors-in-Covariables Models with Validation Data. *Biometrika*, **89**, 345-358.

- <https://doi.org/10.1093/biomet/89.2.345>
- [9] Wang, Q. and Yu, K. (2007) Likelihood-Based Kernel Estimation in Semiparametric Errors-In-Covariates Models with Validation Data. *Journal of Multivariate Analysis*, **98**, 455-480.
- [10] Lü, Y.-Z., Zhang, R.-Q. and Huang, Z.-S. (2013) Estimation of Semi-Varying Coefficient Model with Surrogate Data and Validation Sampling. *Acta Mathematicae Applicatae Sinica English*, **29**, 645-660. <https://doi.org/10.1007/s10255-013-0241-3>
- [11] Xiao, Y. and Tian, Z. (2014) Dimension Reduction Estimation in Nonlinear Semiparametric Error-in-Response Models with Validation Data. *Mathematica Applicata*, **27**, 730-737.
- [12] Xu, W. and Zhu, L. (2015) Nonparametric Check for Partial Linear Errors-in-Covariates Models with Validation Data. *Annals of the Institute of Statistical Mathematics*, **67**, 793-815. <https://doi.org/10.1007/s10463-014-0476-7>
- [13] Zhang, Y. (2015) Estimation of Partially Linear Regression for Errors-in-Variables Models with Validation Data. *Springer International Publishing*, **322**, 733-742. [https://doi.org/10.1007/978-3-319-08991-1\\_76](https://doi.org/10.1007/978-3-319-08991-1_76)
- [14] Wang, Q. (2006) Nonparametric Regression Function Estimation with Surrogate Data and Validation Sampling. *Journal of Multivariate Analysis*, **97**, 1142-1161. <https://doi.org/10.1016/j.jmva.2005.05.008>
- [15] Du, L., Zou, C. and Wang, Z. (2011) Nonparametric Regression Function Estimation for Error-in-Variable Models with Validation Data. *Statistica Sinica*, **21**, 1093-1113. <https://doi.org/10.5705/ss.2009.047>
- [16] Kress, R. (1999) Linear Integral Equations. Springer, New York. <https://doi.org/10.1007/978-1-4612-0559-3>
- [17] Devroye, L. and Györfi, L. (1985) Nonparametric Density Estimation: The  $L_1$  View. John Wiley Sons, New York.
- [18] Efromovich, S. (1999) Nonparametric Curve Estimation: Methods, Theory and Applications. Springer, New York.
- [19] Carrasco, M., Florens, J.P. and Renault, E. (2007) Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization. Elsevier, North Holland, 5633-5751.
- [20] Wu, X. (2010) Exponential Series Estimator of Multivariate Densities. *Journal of Econometrics*, **156**, 354-366. <https://doi.org/10.1016/j.jeconom.2009.11.005>



## Appendix

### Proof of Theorem 3.1:

**Lemma 6.1.** *Suppose Assumptions (A1), (A2) (i) and (A4) hold. For each  $z \in [0,1]$ , we have*

$$\|\hat{f}_{XWZ} - f_{XWZ}\|^2 = O_p \left\{ K^2 \left[ (nh_n)^{-1} + h_n^{2r} \right] + K^{-2r} \right\}.$$

**Proof of Lemma 6.1.** For each  $z \in [0,1]$ , by Assumptions (A2) (i) and (A4), we have

$$\begin{aligned} E(\hat{d}_{zkl}) &= \frac{1}{h_n} \iiint \phi_k(x) \phi_l(w) K_{h_n}(z-u) f_{XWZ}(x, w, u) dx dw du \\ &= \iiint \phi_k(x) \phi_l(w) K(u) f_{XWZ}(x, w, z + uh_n) dx dw du \\ &= \iiint \phi_k(x) \phi_l(w) f_{XWZ}(x, w, z) dx dw \\ &\quad + \left\{ h_n^r \int u^r K(u) du \iiint \frac{\partial^r f_{XWZ}(x, w, z)}{\partial z^r} \phi_k(x) \phi_l(w) dx dw \right\} (1 + o(1)) \\ &= d_{zkl} + h_n^r d_{zkl}^{(r)} \int u^r K(u) du (1 + o(1)), \end{aligned}$$

where

$$d_{zkl}^{(r)} = \iiint \frac{\partial^r f_{XWZ}(x, w, z)}{\partial z^r} \phi_k(x) \phi_l(w) dx dw.$$

Note that  $\phi_k$  are orthonormal and complete basis functions on  $L_2([0,1])$ . Under Assumptions (A2) (i), for each  $z \in [0,1]$ , we have  $\partial^r f_{XWZ}(x, w, z) / \partial z^r \in L_2([0,1]^2)$ . Then, using Cauchy-Schwarz inequality,  $d_{zkl}^{(r)}$  is bounded in absolute value for each  $z \in [0,1]$ . Hence, we obtain that

$$E(\hat{d}_{zkl}) = d_{zkl} + O(h_n^r).$$

Moreover, for each  $z \in [0,1]$ , we have

$$\begin{aligned} \text{Var}(\hat{d}_{zkl}) &\leq \frac{1}{nh_n^2} E \left\{ \left[ \phi_k(X) \phi_l(W) \right]^2 K_{h_n}^2(z-Z) \right\} \\ &= \frac{1}{nh_n^2} \iiint \left[ \phi_k(x) \phi_l(w) \right]^2 K_{h_n}^2(z-u) f_{XWZ}(x, w, u) dx dw du \\ &= \frac{\|K(u)\|^2}{nh_n} \iiint \left[ \phi_k(x) \phi_l(w) \right]^2 f_{XWZ}(x, w, z) dx dw (1 + o(1)) \\ &= O[1/(nh_n)], \end{aligned}$$

where we have used the fact that  $f_{XWZ}$  is uniformly bounded on  $[0,1]^2$ .

We conclude that

$$\hat{d}_{zkl} = d_{zkl} + O(h_n^r) + O_p(1/\sqrt{nh_n}). \tag{10}$$

By the triangle inequality and Jensen inequality, we have

$$\|\hat{f}_{XWZ} - f_{XWZ}\|^2 \leq 2 \left[ \|\hat{f}_{XWZ} - f_{XWZ}^K\|^2 + \|f_{XWZ}^K - f_{XWZ}\|^2 \right].$$

Under Assumption (A2) (i), we can show that  $\|f_{XWZ}^K - f_{XWZ}\|^2 = O(K^{-2r})$  (see

Lemma A1 of [20]).

By construction of the estimator, we have

$$\left\| \hat{f}_{XWZ} - f_{XWZ}^K \right\|^2 = \sum_{k=1}^K \sum_{l=1}^K \left[ \hat{d}_{zkl} - d_{zkl} \right]^2 = O_p \left\{ K^2 \left[ (nh_n)^{-1} + h_n^{2r} \right] \right\},$$

where the last equality is due to (10). The desired result follows immediately.  $\square$

**Proof of Theorem 3.1.** Define  $\hat{A}_\alpha^z = (\alpha I + \hat{T}_z^* \hat{T}_z)^{-1}$  and  $A_\alpha^z = (\alpha I + T_z^* T_z)^{-1}$ . Notice that  $g_z^\alpha = A_\alpha^z T_z^* T_z g_z$  where  $g_z = g(\cdot, z)$ . Then we have

$$\begin{aligned} \left\| \tilde{g}^\alpha - g_z^\alpha \right\|^2 &\leq 4 \left[ \left\| \hat{A}_\alpha^z \hat{T}_z^* \right\|^2 \left\| \hat{m}_z - \hat{T}_z g_z \right\|^2 + \left\| \hat{A}_\alpha^z \hat{T}_z^* \right\|^2 \left\| \hat{T}_z - T_z \right\|^2 \left\| g_z - g_z^\alpha \right\|^2 \right. \\ &\quad \left. + \left\| \hat{A}_\alpha^z \right\|^2 \left\| \hat{T}_z^* - T_z^* \right\|^2 \left\| T_z (g_z - g_z^\alpha) \right\|^2 + \left\| g_z^\alpha - g_z \right\|^2 \right]. \end{aligned}$$

It follows from Lemma 6.1 that  $\left\| \hat{T}_z - T_z \right\|^2$  or  $\left\| \hat{T}_z^* - T_z^* \right\|^2$  are  $O_p \left\{ K^2 \left[ (nh_n)^{-1} + h_n^{2r} \right] + K^{-2r} \right\}$ . Under Assumptions (A1), we have  $\left\| g_z^\alpha - g_z \right\|^2 = O(\alpha^{\beta \wedge 2})$  and  $\left\| T_z (g_z - g_z^\alpha) \right\|^2 = O(\alpha^{(\beta+1) \wedge 2})$ . Moreover,  $\left\| \hat{A}_\alpha^z \hat{T}_z^* \right\|^2 = O_p(1/\alpha)$  and  $\left\| \hat{A}_\alpha^z \right\|^2 = O_p(1/\alpha^2)$ . The main task remained is to establish the order of the term  $\left\| \hat{m}_z - \hat{T}_z g_z \right\|^2$ . By the triangle inequality and Jensen inequality, we have

$$\left\| \hat{m}_z - \hat{T}_z g_z \right\|^2 \leq 2 \left[ \left\| \hat{m}_z - m_z \right\|^2 + \left\| (\hat{T}_z - T_z) g_z \right\|^2 \right].$$

Similar to the proof of Lemma 6.1, under Assumptions (A2)(ii), (A3) and (A4), it is easy to show that

$$\left\| \hat{m}_z - m_z \right\|^2 = O_p \left\{ K^{-2s} + K \left[ (Nh_N)^{-1} + h_N^{2\gamma} \right] \right\}.$$

Then, according to the Lemma 6.1, we have

$$\left\| \hat{m}_z - \hat{T}_z g_z \right\|^2 = O_p \left\{ K^2 \left[ h_n^{2r} + (nh_n)^{-1} \right] + K^{-2\gamma} + K \left[ (Nh_N)^{-1} + h_N^{2\gamma} \right] \right\}.$$

Let  $h_N = O\left(N^{-\frac{1}{2\gamma+1}}\right)$ ,  $h_n = O\left(n^{-\frac{1}{2r+1}}\right)$ ,  $K = O\left(n^{\frac{r}{(2r+1)(\gamma+1)}}\right)$  and  $\alpha = O\left(N^{-\frac{\tau}{\beta \wedge 2 + 1}}\right)$ , if  $r \leq s$  or  $s < r \leq 2s(s+1)$ , combining all these results, we complete the proof.  $\square$