

Orthogonal Series Estimation of Nonparametric Regression Measurement Error Models with Validation Data

Zanhua Yin

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

How to cite this paper: Yin, Z.H. (2017) Orthogonal Series Estimation of Nonparametric Regression Measurement Error Models with Validation Data. *Applied Mathematics*, **8**, 1820-1831.

https://doi.org/10.4236/am.2017.812130

Received: November 28, 2017 Accepted: December 26, 2017 Published: December 29, 2017

Copyright © 2017 by author 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/

CC O Open Access

Abstract

In this article we study the estimation method of nonparametric regression measurement error model based on a validation data. The estimation procedures are based on orthogonal series estimation and truncated series approximation methods without specifying any structure equation and the distribution assumption. The convergence rates of the proposed estimator are derived. By example and through simulation, the method is robust against the misspecification of a measurement error model.

Keywords

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

1. Introduction

Let Y be a scalar response variable and X be an explanatory variable in regression. We consider the nonparametric regression model

$$Y = g\left(X\right) + \varepsilon \tag{1}$$

where $g(\cdot)$ is an unknown nonparametric regression function, ε is a noise variable, and given X the errors $\varepsilon = Y - g(X)$ are assumed to be independent and identically distributed. We consider the model (1) with explanatory variable X measured with error and Y measured exactly. That is, instead of the true X, the surrogate variable W is observed. Throughout we assume

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

which is always satisfied if, for example, W is a function of X and some independent noise (see e.g. [1]).

Nonparametric regression model (1) in presence of errors in covariables has attracted considerable attention in the literature, and is by now well understood. See Carroll *et al.* [2] for an excellent source of references for various approaches. However, all these works mostly focus on specifying error model structure between the true variables X and the surrogate variables W (e.g. the classical error structure and the Berkson error structure). In practice, the relationship between the surrogate variables and the true variables can be rather complicated compared to the classical or Berkson error structural equations usually assumed. This situation presents serious difficulties in making valid statistical inferences. Common solution is to use the help of validation data to infer the missing information about relationship between W and X.

We consider settings where some validation data are available for relating X and W. To be specific, we assume that independent validation data (W_j, X_j) , $N+1 \le j \le N+n$ are available in addition to the independent primary data $\{(Y_i, W_i)\}_{i=1}^N$. Recently, several approaches to statistical inference based on surrogate data and a validation sample are available (see, for example, [1], [3]-[12] and among others). But these approaches do not applicable for handling nonparametric regression measurement error model with the availability of a validation data set. Actually, the models considered by the above referenced authors are some parametric or semiparametric models, and the model (1) is a nonparametric one. With the help of validation data, [13], [14] and [15] developed estimation methods for the nonparametric regression model (1) with measurement error. However, [13] assumes that the response Y but not the covariable X is measured with error; The method proposed by [14] cannot be extended to the subject assume explanatory variable X is a vector; The approach

In this paper, without specifying any structural equations, an orthogonal series method is proposed to estimate g with the help of validation data. As explained in Section 2, we estimate g by solving the following Fredholm equation of the first kind,

$$Tg = m \tag{3}$$

Here, we propose orthogonal series estimator of T using the validation data. Using a similar approach, we estimate m based on primary data set. Then an estimator of g is obtained by Tikhonov regularization method.

This paper is arranged as follows. In Section 2, we define an orthogonal series estimation method. In Section 3, we state the convergence rates of the proposed estimator. Simulation results are reported in Section 4 and a brief discussion is given in Section 5. Proofs of the theorems are presented in **Appendix**.

2. Model and Series Estimation

2.1. Model

Recall model (1) and the assumptions below it. Assume that in addition to the primary data set consisting of N independent and identically distributed obser-

vations $\{(Y_i, W_i)\}_{i=1}^N$ from model (1), validation consisting of *n* independent and identically distributed observations $\{(X_j, W_j)\}_{j=N+1}^{N+n}$ are available. Furthermore, we suppose that *X* and *W* are both 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* and *W* both be contained in [0,1] (otherwise, one can carry out monotone transformations of *X* and *W*).

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

$$E(Y | W = w) = E[g(X) | W = w] = \int g(x) \frac{f_{XW}(x, w)}{f_W(w)} dx$$
(4)

Let $m(w) = E(Y | W = w) f_W(w)$ and

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

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

$$(T\varphi)(w) = \int \varphi(x) f_{XW}(x, w) dx$$

So that Equation (4) is equivalent to the operator equation

$$m(w) = (Tg)(w) \tag{5}$$

According to Equation (5), 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:

$$g^{\alpha} = \arg\min_{g} \left[\left\| Tg - m \right\|^2 + \alpha \left\| g \right\|^2 \right]$$
(6)

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

We define the adjoint operator T^* of T

$$(T^*\psi)(x) = \int \psi(w) f_{XW}(x,w) dw$$

where $\psi(w) \in L_2([0,1])$. Then the regularized solution (6) is equivalently:

$$g^{\alpha} = \left(\alpha I + T^*T\right)^{-1} T^* m \tag{7}$$

2.2. Orthogonal Series Estimation

In order to estimate the solution (7), we need to estimate T, T^* and m. In this paper, we consider the orthogonal series method. Under some regularity conditions in Section 4.1, the density function $f_{XW}(x,w)$ and m(w) may be approximated with any wished accuracy by a truncated orthogonal series,

$$f_{XW}^{K}(x,w) = \sum_{k=1}^{K} \sum_{l=1}^{K} d_{kl} \phi_{k}(x) \phi_{l}(w) \text{ and } m^{K}(w) = \sum_{k=1}^{K} m_{k} \phi_{k}(w)$$

where

$$d_{kl} = \iint f_{XW}(x, w) \phi_k(x) \phi_l(w) dx dw \text{ and } m_k = \int m(w) \phi_k(w) dw$$

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]). 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

$$\phi_k(x) = \frac{1}{k!\sqrt{2k+1}} \frac{d^k}{dx^k} \left[\left(x^2 - x \right)^k \right]$$
(8)

The integer K is a truncation point which is the main smoothing parameter in the approximating series, and d_{kl} and m_k represent the generalized Fourier coefficients of f_{XW} and m_k respectively.

Note that $d_{kl} = E\left[\phi_k(X)\phi_l(W)\right]$ and $m_k = E\left[Y\phi_k(W)\right]$. Intuitively, we can obtain the estimators of d_{kl} , $f_{XW}(x,w)$, m_k and m(w) by

$$\hat{d}_{kl} = \frac{1}{n} \sum_{j=N+1}^{N+n} \phi_k (X_j) \phi_l (W_j), \quad \hat{f}_{XW} (x, w) = \sum_{k=1}^{K} \sum_{l=1}^{K} \hat{d}_{kl} \phi_k (x) \phi_l (w)$$
$$\hat{m}_k = \frac{1}{N} \sum_{i=1}^{N} Y_i \phi_k (W_i) \text{ and } \hat{m}(w) = \sum_{k=1}^{K} \hat{m}_k \phi_k (w)$$

respectively. The operators T and T^* can then be consistently estimated by

$$(\hat{T}\varphi)(w) = \int \varphi(x) \hat{f}_{XW}(x,w) dx$$
 and $(\hat{T}^*\psi)(x) = \int \psi(w) \hat{f}_{XW}(x,w) dw$

Conclude that, the estimator of g(x) is obtained by

$$\hat{g}^{\alpha} = \left(\alpha I + \hat{T}^* \hat{T}\right)^{-1} \hat{T}^* \hat{m}$$
(9)

3. Theoretical Properties

The main objective of this section is to derive the statistical properties of the estimator proposed in Section 2.2. For this purpose, we assume:

Assumption 1. 1) The support of (X,W) is contained in $[0,1]^2$; 2) The joint density of (X,W) is square integrable w.r.t the Lesbegue measure on $[0,1]^2$.

This is sufficient condition for T to be a Hilbert-Schmidt operator and therefore to be compact (see [18]). As a result of compactness, there exists a singular values decomposition. Let $\lambda_k, k \ge 0$ be the sequence of the nonzero singular values of T and the two orthonormal sequences $\varphi_k, k \ge 0$, and $\psi_l, l \ge 0$ such that (see [16]):

$$T\varphi_k = \lambda_k \varphi_k, T^* \psi_k = \lambda_k \varphi_k; T^* T\varphi_k = \lambda_k^2 \varphi_k, TT^* \psi_k = \lambda_k^2 \psi_k, \text{ for } k \ge 0$$

We define Φ_{β} as a β -regularity space for $\beta > 0$:

$$\Phi_{\beta} = \left\{ \varphi \in L_2([0,1]) \text{ such that } \sum_{k \ge 0} \frac{\left\langle \varphi, \varphi_k \right\rangle}{\lambda_k^{2\beta}} < +\infty \right\}$$

Here and blow, we denote by $\langle \cdot, \cdot \rangle$ the scalar product in $L^2([0,1])$.

Assumption 2. We have $g \in \Phi_{\beta}$ for some $\beta > 0$.

We then obtain the following result (see [18]).

Proposition 3.1. Suppose Assumptions 1 and 2 hold, then we have $\|g - g^{\alpha}\|^2 = O(\alpha^{\beta \wedge 2})$, where $\beta \wedge 2 = \min{\{\beta, 2\}}$.

In order to obtain the rate of convergence for $\|\hat{g}^{\alpha} - g\|^2$, we impose the following additional conditions:

Assumption 3. 1) The joint density f_{XW} is r-times continuously differentiable on $[0,1]^2$; 2) The function $m(\cdot)$ is s-times continuously differentiable on [0,1].

Assumption 4. The function $E(Y^2 | W = w)$ is bounded uniformly on [0,1].

Assumption 5. 1) $\lim n/N = \mu \in [0,\infty)$; 2) $\alpha \to 0$, $K = K_{(N,n)} \to \infty$, $K/N \to 0$, $K^2/n \to 0$ as $n \to \infty$, $N \to \infty$.

Theorem 3.1. Suppose Assumptions 1 - 5 hold. Let $\gamma = \min\{r, s\}$, then we have

$$\left\|\hat{g}^{\alpha} - g\right\|^{2} = O_{p}\left[\frac{1}{\alpha} \times \left(\frac{K}{N} + \frac{1}{K^{2\gamma}} + \frac{K^{2}}{n}\right) + \alpha^{\beta \wedge 2}\right]$$
(10)

In (10), the term $K^{-2\gamma}$ arises from the bias of \hat{g} caused by truncating the series approximation of f_{XW} and m. The truncation bias decreases as γ increases. The terms $N^{-1}K$ and $n^{-1}K^2$ are respectively induced by random surrogate sampling errors and random validation sampling errors in the estimates of the generalized Fourier coefficients m_k and d_{kl} . By Theorem 3.1, it is easy to obtain the following corollary.

Corollary 3.1. Suppose the assumptions of Theorem 3.1 are satisfied. Let $K = O\left(n^{1/(2\gamma+2)}\right)$ and $\alpha = O\left(n^{-\gamma/\left[(\gamma+1)(\beta\wedge 2+1)\right]}\right)$, then we have

$$\left\|\hat{g}^{\alpha}-g\right\|^{2}=O_{P}\left(n^{-\kappa\frac{\beta\wedge2}{\beta\wedge2+1}}\right)$$

where $\kappa = \gamma / (\gamma + 1)$.

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

4. Simulation Studies

In this section, we conducted simulation studies of the finite-sample performance of the proposed estimators. First, for comparison, we consider the standard Nadaraya-Watson estimator base on the primary dataset $\{(Y_i, W_i)\}_{i=1}^N$ (denoted as \hat{g}_N). 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. Second, The performance of estimator g^{est} is assessed by using the square root of average square errors (RASE)

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

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

1) $g(x) = \phi_{0,1.5}(4x) + \phi_{1,2}(4x) + \phi_{2.5}(4x), \varepsilon \sim N(0,0.2),$ 2) $g(x) = 5\sin(2x)\exp(-16x^2/50), \varepsilon \sim N(0,0.2),$

where $\phi_{\mu,\sigma}$ is the density of an Normal (μ,σ^2) variable. To perform this simulation, we generate W from f_W and δ from f_{δ} , and put $X = W + \delta$. The densities f_W and f_{δ} , chosen in the beta family, are

$$f_{w}(w) = \frac{(1 - w^{2}/4)}{2B(1/2, 2)} I(w \in [-2, 2])$$
$$f_{\delta}(u) = \frac{(1 - u^{2})^{\rho_{\delta}}}{B(1/2, \rho_{\delta} + 1)} I(u \in [-1, 1])$$

where we chose $\rho_{\delta} = 1$, $\rho_{\delta} = 3$ and $\rho_{\delta} = 5$ (in fact, the greater the value of ρ_{δ} , the smaller the variance of δ). Simulations were run with different validation and primary data sizes (n, N) ranging from (20, 60) to (50, 250) according to the ratio $\kappa = N/n = 3$ and $\kappa = N/n = 5$, respectively. For each case, 500 simulated data sets were generated for each sample size of (n, N).

To implement our method (9), the regularization parameter α and truncating parameter K should be chosen. Here, we estimate α and K by minimizing the following two-dimensional cross-validation score selection criterion

$$\operatorname{CV}(\alpha, K) = \sum_{i=1}^{N} \left\{ Y_i - \sum_{k=1}^{K} \hat{g}_k^{(-i)} \phi_k(W_i) \right\}^2$$

where $\hat{g}_k^{(-i)}$ are the solutions based on (9), after deleting the *i*th primary observation (Y_i, W_i) . In addition, for the naive estimator \hat{g}_N , we used the standard normal kernel, and the bandwidth was selected by leave-one-out CV approach. In all graphs, to illustrate the performance of an estimator, we show the estimated curves corresponding to the first (Q1), second (Q2) and third (Q3) quartiles of the ordered RASEs. The target curve is always represented by a solid curve.

Figure 1 shows the regression function curve, the quartile curves of 500 estimates $\hat{g}^{\alpha}(x)$ under different values of ρ_{δ} for sample size (N,n) = (90,30), in the example (a). From this figure, we clearly see that the proposed estimator $\hat{g}^{\alpha}(x)$ appeared to perform very well in this study. Taking the measurement error levels into account, as the variances of δ decrease, $\hat{g}^{\alpha}(x)$ tends to have smaller bias at the peaks of the regression curve.

Figure 2 illustrates the way in which the estimator improves as sample size increases. We compare the results obtained when estimating curve (b) under different settings of sample size (N,n) for $\rho_{\delta} = 3$. We see clearly that, as the sample size increases, the quality of the estimators improves significantly.

Table 1 compares, for various sample sizes, the results obtained for estimating curves (a) and (b) when $\rho_{\delta} = 1$, $\rho_{\delta} = 3$ and $\rho_{\delta} = 5$. The estimated RASEs which were evaluated at 2⁷ grid points of *x* are presented. Our results show that



Figure 1. Estimation of regression function (a) for samples of size (N,n) = (90,30), when $\rho_{\delta} = 1$ (left panel), $\rho_{\delta} = 3$ (middle panel) and $\rho_{\delta} = 5$ (right panel). The solid curve is the target curve.

				(a)				
Curve	K	(n,N)	$\rho_{\delta} = 1$		$\rho_{s} = 3$		$\rho_{\delta} = 5$	
			$\hat{g}^{a}(x)$	$\hat{g}_{\scriptscriptstyle N}(x)$	$\hat{g}^{\alpha}(x)$	$\hat{g}_{\scriptscriptstyle N}(x)$	$\hat{g}^{\alpha}(x)$	$\hat{g}_{\scriptscriptstyle N}(x)$
	<i>κ</i> = 3	(20, 60)	1.2136	2.0637	1.0597	1.9101	1.0537	1.7948
		(30 ,90)	0.9319	1.7127	0.8452	1.5598	0.8012	1.4746
		(50, 150)	0.7002	1.3890	0.6397	1.3086	0.5590	1.2341
	κ = 5	(20, 100)	0.8332	1.5661	0.7981	1.4378	0.7266	1.3667
		(30, 150)	0.7511	1.3930	0.6546	1.3074	0.5635	1.2514
		(50, 250)	0.5373	1.1940	0.4824	1.0696	0.3993	1.0033
				(b)				
Curve	K	(n,N)	$\rho_{\delta} = 1$		$\rho_{\delta} = 3$		$\rho_{\delta} = 5$	
			$\hat{g}^{a}(x)$	$\hat{g}_{N}(x)$	$\hat{g}^{\alpha}(x)$	$\hat{g}_{N}(x)$	$\hat{g}^{\alpha}(x)$	$\hat{g}_{N}(x)$
	к = 3	(20, 60)	9.1934	22.3251	8.5573	21.9729	7.2736	21.7931
		(30, 90)	9.0040	20.7143	7.4141	20.0651	5.2429	19.9803
		(50, 150)	8.5422	19.4815	6.5897	18.6286	3.7987	18.5734
	κ = 5	(20, 100)	9.1158	20.3010	6.8672	19.7296	4.4416	19.5391
		(30, 150)	8.9352	19.3803	6.9529	18.6887	4.0148	18.5831
		(50, 250)	7.7743	18.5282	5.3232	18.3610	2.5054	18.2475

Table 1. The RASE comparison for the estimators	$\hat{g}^{\alpha}(x)$) and \hat{g}_N	(x). Let	$\kappa = N/n$
---	-----------------------	-------------------	----------	----------------



Figure 2. Estimation of regression function (b) for $\rho_{\delta} = 3$, when (N,n) = (60,20) (left panel), (N,n) = (90,30) (middle panel) and (N,n) = (150,50) (right panel). The solid curve is the target curve.

the estimator \hat{g}^{α} outperforms \hat{g}_N . Also, the performance of \hat{g}^{α} improves (*i.e.*, the corresponding RASEs 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 ρ_{δ} . As expected, **Table 1** shows that the effect of the variances on the estimator performance is obvious.

5. Discussion

In this paper, we have proposed a new method for estimating non-parametric regression models when the explanatory variable is measured with error under the assumption that a proper validation data set is available. The validation data set allows us to estimate joint density f_{XW} of the true variable and the surrogate variable via an orthogonal series method. In practice, our proposed method can be extended to multidimensional cases in which X may be a p-variate explanatory variable. When the dimension of X and hence of W is large, the curse of dimensionality may occur because of the multivariate density estimation of f_{XW} . In this case, exponential series estimator proposed by [19] ensures the positiveness of the estimated density. After obtaining the exponential series estimator of f_{XW} , we can obtain results similar to those in the previous sections. Asymptotic theory in this setting still needs to be pursued in the further research.

Acknowledgements

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

References

- 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. <u>https://doi.org/10.1198/016214506000000816</u>
- [2] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006) Measurement Error in Nonlinear Models. 2nd Edition, Chapman and Hall CRC Press, Boca Raton. https://doi.org/10.1201/9781420010138
- [3] 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. <u>https://doi.org/10.1080/01621459.1990.10474925</u>
- [4] Carroll, R.J. and Wand, M.P. (1991) Semiparametric Estimation in Logistic Measurement Error Models. *Journal of the Royal Statistical Society B*, **53**, 573-585.
- [5] Sepanski, J.H. and Lee, L.F. (1995) Semiparametric Estimation of Nonlinear Errors-in-Variables Models with Validation Study. *Journal of Nonparametric Statistics*, 4, 365-394. <u>https://doi.org/10.1080/10485259508832627</u>
- [6] 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. <u>https://doi.org/10.1080/01621459.1994.10476871</u>
- [7] 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.
- [8] 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. <u>https://doi.org/10.1080/01621459.1995.10476546</u>
- [9] Lv, 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. <u>https://doi.org/10.1007/s10255-013-0241-3</u>
- [10] Xiao, Y. and Tian, Z. (2014) Dimension Reduction Estimation in Nonlinear Semiparametric Error-in-Response Models with Validation Data. *Mathematica Applicata*, 27, 730-737.
- [11] Yu, S.H. and Wang, D.H. (2014) Empirical Likelihood for First-Order Autoregressive Error-in-Variable of Models with Validation Data. *Communications in Statistics Theory Methods*, 43, 1800-1823. <u>https://doi.org/10.1080/03610926.2012.679763</u>
- [12] Xu, W. and Zhu, L. (2015) Nonparametric Check for Partial Linear Errors-in-Covariables Models with Validation Data. Annals of the Institute of Statistical Mathematics, 67, 793-815. <u>https://doi.org/10.1007/s10463-014-0476-7</u>
- [13] Wang, Q. (2006) Nonparametric Regression Function Estimation with Surrogate Data and Validation Sampling. Journal of Multivariate Analysis, 97, 1142-1161. <u>https://doi.org/10.1016/j.jmva.2005.05.008</u>
- [14] Du, L., Zou, C. and Wang, Z. (2015) Nonparametric Regression Function Estimation for Error-in-Variable Models with Validation Data. *Open Journal of Statistics*, 5, 808-819. <u>https://doi.org/10.4236/ojs.2015.57080</u>
- [15] Yin, Z. and Liu, F. (2011) Orthogonal Series Estimation of Nonparametric Regression Measurement Error Models with Validation Data. *Statistica Sinica*, 21, 1093-1113. <u>https://doi.org/10.5705/ss.2009.047</u>
- [16] Darolles, S., Florens, J.P., Renault, E. and Kress, R. (1999) Linear Integral Equations. Springer, New York.
- [17] Efromovich, S. (1999) Nonparametric Curve Estimation: Methods, Theory and Ap-

plications. Springer, New York.

- [18] 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.
- [19] Wu, X. (2010) Exponential Series Estimator of Multivariate Densities. Journal of Econometrics, 156, 354-366. <u>https://doi.org/10.1016/j.jeconom.2009.11.005</u>
- [20] Groetsch, C. (1984) The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Pitman, London.

Appendix

Proofs of Theorem 3.1 and Corollary 3.1:

We first present some Lemmas that are need to prove the main theorem.

Lemma 7.1. Suppose Assumptions 1 and 3(1) hold. Then:

1)
$$\left\|\hat{T} - T\right\|_{HS}^{2} = O_{P}\left(K^{-2r} + n^{-1}K^{2}\right);$$

2) $\left\|\hat{T}^{*} - T^{*}\right\|_{HS}^{2} = O_{P}\left(K^{-2r} + n^{-1}K^{2}\right).$

where $\left\|\cdot\right\|_{HS}$ denotes the Hilbert-Schmidt norm, *i.e.*:

$$\left\|\hat{T} - T\right\|_{HS}^{2} = \iint \left[\hat{f}_{XW}\left(x, w\right) - f_{XW}\left(x, w\right)\right]^{2} \mathrm{d}x \mathrm{d}w$$

Proof of Lemma 7.1. According to Lemma A1 of Wu [19], we have

$$\left\| f_{XW} - f_{XW}^{K} \right\|^{2} = O\left(K^{-2r}\right)$$

Note that the Legendre polynomials ϕ_k in (8) are orthonormal and complete on $L_2([0,1])$. Then

$$\left\|\hat{f}_{XW} - f_{XW}^{K}\right\|^{2} = \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\hat{d}_{kl} - d_{kl}\right)^{2}$$

By $E\hat{d}_{kl} = d_{kl}$, we have

$$E\left\{\left|\sum_{k=1}^{K}\sum_{l=1}^{K} \left(\hat{d}_{kl} - d_{kl}\right)^{2}\right|\right\} = \sum_{k=1}^{K}\sum_{l=1}^{K} Var\left(\hat{d}_{kl}\right)$$
$$\leq \frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{K} E\left[\phi_{k}\left(X\right)\phi_{l}\left(W\right)\right]^{2}$$
$$= O\left[\frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{K} \left\|\phi_{k}\left(x\right)\phi_{l}\left(w\right)\right\|^{2}\right]$$
$$= O\left(K^{2}/n\right)$$

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

By Chebyshev's inequality, then we have $\|\hat{f}_{XW} - f_{XW}^{\kappa}\|^2 = O_p(\kappa^2/n)$. Then the desired result follows immediately.

Lemma 7.2. Suppose Assumptions 1, 3 and 4 hold. Let $\gamma = \min\{r, s\}$, then

$$\left\|\hat{m} - \hat{T}g\right\|^2 = O_P\left(N^{-1}K + K^{-2\gamma} + n^{-1}K^2\right)$$

Proof of Lemma 7.2. Note that Tg = m. By the triangle inequality and Jensen inequality, we have

$$\left\|\hat{m} - \hat{T}g\right\|^{2} \le 2\left[\left\|\hat{m} - m\right\|^{2} + \left\|\left(T - \hat{T}\right)g\right\|^{2}\right]$$

If $g \in L_2([0,1])$, Lemma 7.1 gives $\left\| \left(T - \hat{T}\right)g \right\|^2 = O_p\left(K^{-2r} + n^{-1}K^2\right)$. According to the proof of Lemma 7.1, under Assumptions 3(2) and 4, we can show that $\left\| \hat{m} - m \right\|^2 = O_p\left(K^{-2s} + N^{-1}K\right)$. Then we obtain the result in Lemma 7.2.

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

$$\hat{g}^{\alpha} - g = \hat{A}_{\alpha} \left(\hat{T}^* \hat{m} - \hat{T}^* \hat{T}g \right) + \left(\hat{A}_{\alpha} \hat{T}^* \hat{T}g - A_{\alpha} T^* Tg \right) + \left(g^{\alpha} - g \right)$$

The second right-hand side term can itself be decomposed into two components:

$$\hat{A}_{\alpha}\hat{T}^{*}\hat{T}g - A_{\alpha}T^{*}Tg = \hat{A}_{\alpha}\left(\hat{T}^{*}\hat{T} - T^{*}T\right)g + \left(\hat{A}_{\alpha} - A_{\alpha}\right)T^{*}Tg$$

Actually, since in this case $\hat{A}_{\alpha} = (\alpha I + \hat{T}^* \hat{T})^{-1}$ and $A_{\alpha} = (\alpha I + T^* T)^{-1}$, the identity $B^{-1} - C^{-1} = B^{-1} (C - B) C^{-1}$ gives:

$$\hat{A}_{\alpha} - A_{\alpha} = -\hat{A}_{\alpha} \left(\hat{T}^* \hat{T} - T^* T \right) A_{\alpha}$$

Thus,

$$\hat{A}_{\alpha}\hat{T}^{*}\hat{T}g - A_{\alpha}T^{*}Tg = \hat{A}_{\alpha}\left(\hat{T}^{*}\hat{T} - T^{*}T\right)\left(g - g^{\alpha}\right)$$

From the properties of norm, we have

$$\left\|\hat{g}^{\alpha} - g\right\|^{2} \leq 3 \left[\left\|\hat{A}_{\alpha}\hat{T}^{*}(\hat{m} - \hat{T}g)\right\|^{2} + \left\|\hat{A}_{\alpha}(\hat{T}^{*}\hat{T} - T^{*}T)(g - g^{\alpha})\right\|^{2} + \left\|g^{\alpha} - g\right\|^{2}\right]$$

Let us consider the first term, we have

$$\left\|\hat{A}_{\alpha}\hat{T}^{*}\left(\hat{m}-\hat{T}g\right)\right\|^{2} \leq \left\|\hat{A}_{\alpha}\hat{T}^{*}\right\|^{2}\left\|\hat{m}-\hat{T}g\right\|^{2}$$

The first norm $\|\hat{A}_{\alpha}\hat{T}^*\|^2 = \|(\alpha I + \hat{T}^*\hat{T})^{-1}\hat{T}^*\|^2$ is equal to the larger eigen value of the operator. These eigen values converges to $\lambda_k/(\alpha + \lambda_k^2)$ and are then smaller than $1/\alpha$. It follows from Lemma 7.2 that

$$\left\|\hat{A}_{\alpha}\hat{T}^{*}\left(\hat{m}-\hat{T}g\right)\right\|^{2} = O_{p}\left[\alpha^{-1}\left(N^{-1}K+K^{-2\gamma}+n^{-1}K^{2}\right)\right]$$
(11)

Next, we consider the term $\left\|\hat{A}_{\alpha}(\hat{T}^{*}\hat{T}-T^{*}T)(g-g^{\alpha})\right\|^{2}$. Note that $\hat{T}^{*}\hat{T}-T^{*}T=\hat{T}^{*}(\hat{T}-T)+(\hat{T}^{*}-T^{*})T$

Then

$$\begin{aligned} & \left\| \hat{A}_{\alpha} \left(\hat{T}^{*} \hat{T} - T^{*} T \right) \left(g - g^{\alpha} \right) \right\|^{2} \\ & \leq 2 \bigg[\left\| \hat{A}_{\alpha} \hat{T}^{*} \right\|^{2} \left\| \hat{T} - T \right\|^{2} \left\| g - g^{\alpha} \right\|^{2} + \left\| \hat{A}_{\alpha} \right\|^{2} \left\| \hat{T}^{*} - T^{*} \right\|^{2} \left\| T \left(g - g^{\alpha} \right) \right\|^{2} \bigg] \end{aligned}$$

We have $\|\hat{A}_{\alpha}\hat{T}^*\|^2 = O_P(1/\alpha)$, and $\|\hat{A}_{\alpha}\|^2 = O_P(1/\alpha^2)$ (see [20]). According to Lemma 7.1, we have $\|\hat{T} - T\|^2$ or $\|\hat{T}^* - T^*\|^2$ are $O_P(K^{-2r} + n^{-1}K^2)$.

By Proposition 3.1:

$$\left\|g^{\alpha} - g\right\|^{2} = O\left(\alpha^{\beta \wedge 2}\right) \tag{12}$$

The term $T(g-g^{\alpha})$ identical to $\alpha A_{\alpha}T^*g$, is the regularity bias of T^*g equal to $O(\alpha^{(\beta+1)\wedge 2})$.

Therefore, we have

$$\left\|\hat{A}_{\alpha}\left(\hat{T}^{*}\hat{T}-T^{*}T\right)\left(g-g^{\alpha}\right)\right\|^{2}=O_{P}\left[\alpha^{(\beta-1)\wedge0}\left(K^{-2r}+n^{-1}K^{2}\right)\right]$$
(13)

Combining (11), (12) and (13) gives the desired result of Theorem 3.1.

Proof of Corollary 3.1. By Theorem 3.1, the proof of Corollary 3.1 is straightforward and is omitted.