

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection School Of Economics

School of Economics

---

10-2013

### Nonparametric Regression Estimation with General Parametric Error Covariance: A More Efficient Two-step Estimator

Liangjun SU

*Singapore Management University, ljsu@smu.edu.sg*

Aman ULLAH

*University of California, Riverside*

Yun WANG

*University of International Business and Economics, Beijing*

Follow this and additional works at: [https://ink.library.smu.edu.sg/soe\\_research](https://ink.library.smu.edu.sg/soe_research)



Part of the [Econometrics Commons](#)

---

#### Citation

SU, Liangjun; ULLAH, Aman; and WANG, Yun. Nonparametric Regression Estimation with General Parametric Error Covariance: A More Efficient Two-step Estimator. (2013). *Empirical Economics*. 45, (2), 1009-1024. Research Collection School Of Economics.

Available at: [https://ink.library.smu.edu.sg/soe\\_research/1421](https://ink.library.smu.edu.sg/soe_research/1421)

This Journal Article is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [library@smu.edu.sg](mailto:library@smu.edu.sg).

## Nonparametric regression estimation with general parametric error covariance: a more efficient two-step estimator

Liangjun Su · Aman Ullah · Yun Wang

**Abstract** Recently Martins-Filho and Yao (J Multivar Anal 100:309–333, 2009) have proposed a two-step estimator of nonparametric regression function with parametric error covariance and demonstrate that it is more efficient than the usual LLE. In the present paper we demonstrate that MY's estimator can be further improved. First, we extend MY's estimator to the multivariate case, and also establish the asymptotic theorem for the slope estimators; second, we propose a more efficient two-step estimator for nonparametric regression function with general parametric error covariance, and develop the corresponding asymptotic theorems. Monte Carlo study shows the relative efficiency loss of MY's estimator in comparison with our estimator in nonparametric regression with either AR(2) errors or heteroskedastic errors. Finally, in an empirical study we apply the proposed estimator to estimate the public capital productivity to illustrate its performance in a real data setting.

**Keywords** Covariance matrix · Local linear estimation · Productivity · Relative efficiency

**JEL Classification** C1 · C14 · C33

---

L. Su

School of Economics, Singapore Management University, 90 Stamford Road, Singapore, Singapore  
e-mail: [lj-su@smu.edu.sg](mailto:lj-su@smu.edu.sg)

A. Ullah (✉)

Department of Economics, University of California, Riverside, CA 92521-0427, USA  
e-mail: [aman.ullah@ucr.edu](mailto:aman.ullah@ucr.edu)

Y. Wang

School of International Trade and Economics, University of International Business and Economics,  
Beijing, China  
e-mail: [wyuncolor@gmail.com](mailto:wyuncolor@gmail.com)

## 1 Introduction

Recently there has been a growing interest in the estimation of nonparametric regression relationship by exploring the information in the error covariance. See [Lin and Carroll \(2000\)](#), [Ruckstuhl et al. \(2000\)](#), [Xiao et al. \(2003\)](#), [Su and Ullah \(2006\)](#), [Su and Ullah \(2007\)](#), [Linton and Mammen \(2008\)](#), and [Martins-Filho and Yao \(2009\)](#), among others. Except [Su and Ullah \(2006\)](#) where the errors enter the model nonparametrically, the errors in all other models exhibit a parametric correlation structure whose information can be explored to improve over the traditional nonparametric estimator.

The case considered by [\(Martins-Filho and Yao \(2009\)\)](#), MY hereafter) is fairly general. For nonparametric regression with general parametric error covariance, they have proposed a two-step estimator of nonparametric regression function and demonstrated that it is more efficient than the traditional local linear estimator (LLE). Intuitively MY gains the relative efficiency of their estimator over the LLE because the former applies the information in the off-diagonal elements of the error covariance whereas the latter fully ignores the information in the error covariance structure. Nevertheless, MY did not explore the information in the diagonal elements of the error covariance. Consequently, if these diagonal elements are not identical across observations (say when the error term is an AR process or heteroskedastic of known form), then their estimator can be further improved.

In this paper, we propose a modified estimator of MY. We demonstrate clearly that the full use of the error covariance structure can result in an asymptotically more efficient estimator than MY's estimator. The relative efficiency of our estimator over MY's is verified through simulations where the error terms in the nonparametric regression follow an AR(2) process or a heteroskedastic structure. In addition, we extend MY's estimator to the multivariate case, and also establish the asymptotic theorems for the slope estimators which are not studied by MY. In order to illustrate the applicability of our asymptotic results to popular nonparametric models, we study the asymptotic properties of our two-step estimators for seemingly unrelated regression and clustered/panel data models. Also, the practical use of the newly proposed method is demonstrated within a nonparametric panel regression model with random effects in a real data setting.

The paper is structured as follows. We introduce the MY's estimator in [Sect. 2](#) and demonstrate that it can easily be improved to achieve a more efficient estimator in [Sect. 3](#) where the asymptotic bias and variance for the two-step estimator are derived for both seemingly unrelated regression models and clustered/panel data models. A small set of simulations is conducted in [Sect. 4](#), and an empirical study on the public capital productivity is presented in [Sect. 5](#). Finally, the concluding remarks are made in [Sect. 6](#).

## 2 The MY's estimator

Consider the nonparametric regression model

$$Y_i = m(X_i) + U_i, \quad i = 1, \dots, n, \quad (1)$$

where  $X_i$  is a  $q \times 1$  vector of exogenous regressors that is continuously distributed, and  $U_i$  is an error term such that  $E(U_i) = 0$  and

$$E(U_i U_j) = \omega_{ij}(\theta_0) \text{ for some } \theta_0 \in \mathfrak{R}^p, i, j = 1, \dots, n. \quad (2)$$

Following MY, we assume for simplicity that  $\{U_i\}$  is independent of  $\{X_i\}$  but allow for time series structure in either process. In addition, we permit non-identical distributions across  $i$ 's.

Let  $\mathbf{Y} \equiv (Y_1, \dots, Y_n)'$ ,  $R_{ix} \equiv (1, (X_i - x)')'$ , and  $\mathbf{R}_x \equiv (R_{1x}, \dots, R_{nx})'$ . Let  $\delta(x) \equiv (m(x), \partial m(x)/\partial x')'$ . The conventional LLE of  $\delta(x)$  is given by

$$\widehat{\delta}_{LL, h_1}(x) = (\mathbf{R}_x' \mathbf{K}_{x, h_1} \mathbf{R}_x)^{-1} \mathbf{R}_x' \mathbf{K}_{x, h_1} \mathbf{Y} \quad (3)$$

where  $\mathbf{K}_{x, h_1} = \text{diag}(K_{h_1}(X_1 - x), \dots, K_{h_1}(X_n - x))$ ,  $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1^q$ ,  $K(\cdot)$  is a kernel function, and  $h_1$  is a bandwidth parameter. In particular, the conventional LLE of  $m(x)$  is given by

$$\widehat{m}_{LL, h_1}(x) = \mathbf{e}' (\mathbf{R}_x' \mathbf{K}_{x, h_1} \mathbf{R}_x)^{-1} \mathbf{R}_x' \mathbf{K}_{x, h_1} \mathbf{Y} \quad (4)$$

where  $\mathbf{e} \equiv (1, 0, \dots, 0)'$  denotes a  $(q+1) \times 1$  vector.

Since  $\widehat{m}_{LL, h_1}(x)$  does not explore the information in the error covariance structure, it cannot be asymptotically efficient in any sense. For this reason, MY proposes a two-step estimator of  $m(x)$  that applies the information in (2). In order to proceed, let  $\Omega(\theta)$  be an  $n \times n$  matrix with the  $(i, j)$ th element given by  $\omega_{ij}(\theta)$ . Assume that  $\Omega(\theta) = P(\theta)P(\theta)'$  for some square matrix  $P(\theta)$ . Let  $p_{ij}(\theta)$  and  $v_{ij}(\theta)$  denotes the  $(i, j)$ th element of  $P(\theta)$  and  $P(\theta)^{-1}$ , respectively. When  $\theta = \theta_0$ , the true parameter value, we frequently suppress the dependence of these matrices and their elements on  $\theta_0$  and, for example, write  $P$  for  $P(\theta_0)$  and  $v_{ij}$  for  $v_{ij}(\theta_0)$ . Let  $\mathbf{m} \equiv (m(X_1), \dots, m(X_n))'$ ,  $\mathbf{U} \equiv (U_1, \dots, U_n)'$ , and  $H \equiv \text{diag}(v_{11}^{-1}, \dots, v_{nn}^{-1})$ . Define  $\mathbf{Z} \equiv H P^{-1} \mathbf{Y} + (I_n - H P^{-1}) \mathbf{m}$  where  $I_n$  is an  $n \times n$  identity matrix. Then

$$\mathbf{Z} = \mathbf{m} + \epsilon \text{ with } \epsilon \equiv H P^{-1} \mathbf{U},$$

and it is easy to verify that  $\epsilon$  has mean 0 and covariance matrix as a diagonal matrix:

$$E(\epsilon \epsilon') = H^2 = \text{diag}(v_{11}^{-2}, \dots, v_{nn}^{-2}). \quad (5)$$

The two-step MY's estimators of  $\delta(x)$  and  $m(x)$  are given by

$$\widehat{\delta}_{MY, h_2}(x) = (\mathbf{R}_x' \mathbf{K}_{x, h_2} \mathbf{R}_x)^{-1} \mathbf{R}_x' \mathbf{K}_{x, h_2} \widehat{\mathbf{Z}}, \quad (6)$$

$$\widehat{m}_{MY, h_2}(x) = \mathbf{e}' (\mathbf{R}_x' \mathbf{K}_{x, h_2} \mathbf{R}_x)^{-1} \mathbf{R}_x' \mathbf{K}_{x, h_2} \widehat{\mathbf{Z}}, \quad (7)$$

where  $\widehat{\mathbf{Z}} \equiv H P^{-1} \mathbf{Y} + (I_n - H P^{-1}) \widehat{\mathbf{m}}_{\text{LL}, h_1}, \widehat{\mathbf{m}}_{\text{LL}, h_1} \equiv (\widehat{m}_{\text{LL}, h_1}(X_1), \dots, \widehat{m}_{\text{LL}, h_1}(X_n))'$ , and the bandwidth  $h_2$  is usually different from  $h_1$ . Clearly, here it is assumed that  $\theta_0$ , and therefore  $H$  and  $P$ , are known. When  $\theta_0$  is unknown but can be estimated by  $\widehat{\theta}$  at  $\sqrt{n}$ -rate, we can replace  $H$  and  $P$  by  $H(\widehat{\theta})$  and  $P(\widehat{\theta})$  and it is trivial to show that such a replacement will not affect the first-order asymptotic properties of  $\widehat{m}_{\text{MY}, h_2}(x)$ . Hence it is not restrictive to assume that  $\theta_0$  is known.

Let  $f_i(\cdot)$  denote the probability density function (PDF) of  $X_i$ . Let  $\bar{f}(x) \equiv \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n f_i(x)$  and  $\bar{\omega}_f(x, \theta_0) \equiv \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n v_{ii}^{-2} f_i(x)$ . In order to proceed, we make the following assumptions.

- Assumption A1** (i)  $\{\xi_i \equiv (X'_i, U_i)', i = 1, 2, \dots\}$  is a strong mixing process with mixing coefficient  $\alpha(\cdot)$  satisfying  $\sum_{j=1}^n j^a \alpha(j)^{1-2/\delta} \leq C < \infty$  for some  $a > 1 - 2/\delta$  and  $\delta > 2$ .  
(ii)  $E(U_i) = 0$  and  $\max_{1 \leq i \leq n} E|U_i|^\delta < \infty$ .  
(iii)  $f_i(\cdot)$  has compact support  $X_i$ .  $0 < \inf_{x \in \mathcal{X}} \bar{f}(x) \leq \sup_{x \in \mathcal{X}} \bar{f}(x) < \infty$  where  $X \equiv \lim_{n \rightarrow \infty} \cup_{i=1}^n X_i$ .  $f_i(\cdot)$  is Lipschitz continuous, i.e.,  $|f_i(\bar{x}) - f_i(\tilde{x})| \leq C_i \|\bar{x} - \tilde{x}\|$ , where  $\|\cdot\|$  denotes the Euclidean norm and  $\max_{1 \leq i \leq n} C_i < \infty$ . The joint PDF  $f_{i_1, \dots, i_l}(\cdot, \dots, \cdot)$  of  $X_{i_1}, \dots, X_{i_l}$  ( $2 \leq l \leq 6$ ) is bounded.

**Assumption A2**  $m(\cdot)$  is the second order uniformly continuously differentiable on  $\mathcal{X}$ .

**Assumption A3**  $K(\cdot)$  is a product kernel such that  $K(x) = \prod_{i=1}^q k(x_i)$ , where  $k(\cdot)$  is a univariate symmetric PDF with compact support  $\mathcal{K}$  such that  $|k(u) - k(v)| \leq C_k |u - v|$  for all  $u, v \in \mathcal{K}$  and some  $C_k < \infty$ .

**Assumption A4** As  $n \rightarrow \infty$ ,  $h_1/h_2 \rightarrow 0$ ,  $nh_1^q \rightarrow \infty$ , and  $nh_2^{q+4} \rightarrow c \in [0, \infty)$ .

The following theorem extends the findings in MY to the multivariate case and it also incorporates the asymptotic properties for the slope estimators.

**Theorem 1** Suppose that Assumptions A1–A4 hold. Suppose  $0 < \bar{\omega}_f(x, \theta_0) < \infty$ . Then we have

$$\sqrt{nh_2^q} D_{h_2} (\widehat{\delta}_{\text{MY}, h_2}(x) - \delta(x) - B_{\text{MY}}) \xrightarrow{d} N(0, \Omega_{\text{MY}}),$$

where

$$B_{\text{MY}} = \begin{pmatrix} \frac{\kappa_{21} h_2^2}{2} \sum_{j=1}^q \frac{\partial^2 m(x)}{\partial x_j^2} \\ \mathbf{0}_{q \times 1} \end{pmatrix}, \quad \Omega_{\text{MY}} = \begin{pmatrix} \frac{\bar{\omega}_f(x, \theta_0) (\kappa_{02})^q}{\bar{f}^2(x)} & \mathbf{0}_{1 \times q} \\ \mathbf{0}_{q \times 1} & \frac{\bar{\omega}_f(x, \theta_0) \kappa_{22} (\kappa_{02})^{q-1}}{\bar{f}^2(x) \kappa_{21}^2} I_q \end{pmatrix},$$

$D_{h_2} = \text{diag}(1, h_2, \dots, h_2)$  is a  $(q+1) \times (q+1)$  diagonal matrix, and  $\kappa_{ij} = \int z^i k(z)^j dz$  for  $i, j = 0, 1, 2$ .

The proof of the above theorem follows straightforwardly from that of Theorem 3 in MY, and is similar to that of Theorem 2 below and thus omitted. In order to obtain

the above result, a necessary condition on  $(h_1, h_2)$  is that  $h_1/h_2 \rightarrow 0$  to eliminate the first order asymptotic bias due to the first stage estimation error. Also, in order for the remainder term from the second-order Taylor expansion of  $m(X_i)$  at  $x$  vanish asymptotically, we need  $nh_2^{q+4} \rightarrow c \in [0, \infty)$ .

Let  $\widehat{\beta}_{\text{MY}, h_2}(x)$  denote the vector of the last  $q$  elements of  $\widehat{\delta}_{\text{MY}, h_2}(x)$ . Theorem 1 implies that

$$\begin{aligned} & \sqrt{nh_2^q} \left( \widehat{m}_{\text{MY}, h_2}(x) - m(x) - \frac{\kappa_{21}h_2^2}{2} \sum_{j=1}^q \frac{\partial^2 m(x)}{\partial x_j^2} \right) \\ & \xrightarrow{d} N \left( 0, \frac{\overline{\omega}_f(x, \theta_0) (\kappa_{02})^q}{\overline{f}^2(x)} \right), \text{ and} \\ & \sqrt{nh_2^q h_2} \left( \widehat{\beta}_{\text{MY}, h_2}(x) - \frac{\partial m(x)}{\partial x} \right) \xrightarrow{d} N \left( 0, \frac{\overline{\omega}_f(x, \theta_0) \kappa_{22} (\kappa_{02})^{q-1}}{\overline{f}^2(x) \kappa_{21}^2} I_q \right). \end{aligned}$$

It is easy to see that  $\widehat{m}_{\text{MY}, h_2}(x)$  shares the same asymptotic bias as the traditional LLE  $\widehat{m}_{\text{LL}, h_2}(x)$  but has smaller asymptotic variance than the latter. In order to see this, note that the asymptotic variance of  $\widehat{m}_{\text{LL}, h_2}(x)$  is given by  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \omega_{ii}(\theta_0) f_i(x) (\kappa_{02})^q / \overline{f}^2(x)$ . By the fact that for any nonsingular matrix  $A$  with inverse  $A^{-1}$ , we have  $a_{ii} a^{ii} \geq 1 \forall i$  with  $a_{ii}$  and  $a^{ii}$  being the  $i$ th diagonal elements of  $A$  and  $A^{-1}$ , respectively, we can readily show that

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \omega_{ii}(\theta_0) f_i(x) - \overline{\omega}_f(x, \theta_0) \\ & = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \left( \omega_{ii}(\theta_0) - v_{ii}^{-2} \right) f_i(x) \geq 0. \end{aligned}$$

That is,  $\widehat{m}_{\text{MY}, h_2}(x)$  is asymptotically more efficient than  $\widehat{m}_{\text{LL}, h_2}(x)$ . By the same token,  $\widehat{\beta}_{\text{MY}, h_2}(x)$  shares the same asymptotic bias as the traditional LLE of  $\partial m(x) / \partial x$  but has smaller asymptotic variance than the latter.

### 3 A more efficient two-step estimator

In this section, we first demonstrate that the MY's estimator can be improved to obtain a more efficient estimator and then consider applying our estimation method to both seemingly unrelated regression models and panel data models.

#### 3.1 A more efficient two-step estimator

As indicated in Sect. 1, MY's estimator does not use the information in the diagonal elements of the error covariance matrix  $\Omega(\theta_0)$ . So it still has a room to improve.

Apparently, the cause of the lack of efficiency of MY's estimator is due to the misuse of the diagonal matrix  $H$  in the definition of  $\mathbf{Z}$ . It turns out that we can modify the definition of  $\mathbf{Z}$  to obtain a more efficient estimator. Let  $\mathbf{Z}^* \equiv H^{-1}\mathbf{Z}$ . Then

$$\mathbf{Z}^* = H^{-1}\mathbf{m} + \epsilon^* \text{ with } \epsilon^* \equiv P^{-1}\mathbf{U}. \quad (8)$$

Clearly,  $\epsilon^*$  has mean 0 and covariance matrix as an identity matrix. We can consider the local linear estimation of  $\delta(x)$  based on the transformed equation in (8).

It is straightforward to verify that our two-step estimator of  $\delta(x)$  based on (8) is given by

$$\widehat{\delta}_{\text{SUW}, h_2}(x) \equiv (\mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \widehat{\mathbf{Z}}^* \quad (9)$$

where  $\mathbf{R}_x^* \equiv H^{-1}\mathbf{R}_x$ , and  $\widehat{\mathbf{Z}}^* \equiv P^{-1}\mathbf{Y} + (H^{-1} - P^{-1})\widehat{\mathbf{m}}_{\text{LL}, h_1}$ . Then we have the following theorem.

**Theorem 2** Suppose that Assumptions A1–A4 hold. Suppose  $\omega_f^*(x, \theta_0) \equiv \lim_{n \rightarrow \infty} n^{-1} \cdot \sum_{i=1}^n v_{ii}^2 f_i(x) \in (0, \infty)$ . Then we have

$$\sqrt{nh_2^q} D_{h_2} (\widehat{\delta}_{\text{SUW}, h_2}(x) - \delta(x) - B_{\text{SUW}}) \xrightarrow{d} N(0, \Omega_{\text{SUW}}),$$

where

$$B_{\text{SUW}} = \begin{pmatrix} \frac{\kappa_{21} h_2^2}{2} \sum_{j=1}^q \frac{\partial^2 m(x)}{\partial x_j^2} \\ \mathbf{0}_{q \times 1} \end{pmatrix} \text{ and } \Omega_{\text{SUW}} = \begin{pmatrix} \frac{(\kappa_{02})^q}{\omega_f^*(x, \theta_0)} & \mathbf{0}_{1 \times q} \\ \mathbf{0}_{q \times 1} & \frac{\kappa_{22} (\kappa_{02})^{q-1}}{\omega_f^*(x, \theta_0) \kappa_{21}^2} I_q \end{pmatrix}.$$

The proof of the above theorem is delegated to the Appendix. Theorem 2, in conjunction with Theorem 1, implies that  $\widehat{\delta}_{\text{SUW}, h_2}(x)$  shares the same asymptotic bias as MY's estimator  $\widehat{\delta}_{\text{MY}, h_2}(x)$ . In order to compare their asymptotic covariances, noting that by the Cauchy-Schwarz inequality, we have

$$\frac{1}{n^{-1} \sum_{i=1}^n v_{ii}^2 f_i(x)} \leq \frac{n^{-1} \sum_{i=1}^n v_{ii}^{-2} f_i(x)}{\{n^{-1} \sum_{i=1}^n f_i(x)\}^2},$$

which implies that  $1/\omega_f^*(x, \theta_0) \leq \overline{\omega}_f(x, \theta_0) / \bar{f}^2(x)$ . Thus, the asymptotic covariance of  $\widehat{\delta}_{\text{SUW}, h_2}(x)$  is less than that of  $\widehat{\delta}_{\text{MY}, h_2}(x)$ . That is, our two-stage estimator may have smaller asymptotic variance than MY's if a non-negligible portion of the diagonal elements is distinct from others. In other words, it pays off to explore the information in the diagonal elements of the error covariance matrix.

### 3.2 Two applications

In order to illustrate the applicability of our theorems to popular nonparametric models, we derive the asymptotic bias and variances of our two-step estimators for two popular

models, namely, seemingly unrelated regression models and panel data models. The latter has been studied in MY for the univariate case.

### 3.2.1 Seemingly unrelated regression models

We consider the seemingly unrelated regression models in which observations  $y_i^{(j)}$  are related to  $X_i^{(j)}$ , a  $q_j \times 1$  vector of exogenous regressors, as follows

$$y_i^{(j)} = m_j(X_i^{(j)}) + \varepsilon_i^{(j)}, \quad j = 1, \dots, J, \quad i = 1, \dots, n,$$

where  $X_i^{(j)}$ ,  $j = 1, \dots, J$ , can differ for different regression models,  $E(\varepsilon_i^{(j)}) = 0$ ,  $\text{Var}(\varepsilon_i^{(j)}) = \sigma_{jj}^2$ ,  $\text{Cov}(\varepsilon_i^{(s)}, \varepsilon_i^{(t)}) = \sigma_{st}$  for  $s, t = 1, \dots, J$  and  $s \neq t$ , and  $i = 1, \dots, n$ . One can stack these  $J$  regression models into the following matrix form

$$\mathbf{y} = \mathbf{m}(X) + \varepsilon$$

where  $\mathbf{y} = (y^{(1)'}, \dots, y^{(J)'})'$ ,  $\mathbf{m}(X) = (\mathbf{m}_1(X^{(1)}), \dots, \mathbf{m}_J(X^{(J)}))'$ ,  $\varepsilon = (\varepsilon^{(1)'}, \dots, \varepsilon^{(J)'})'$ ,  $y^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})'$ ,  $\mathbf{m}_j(X^{(j)}) = (m_j(X_1^{(j)}), \dots, m_j(X_n^{(j)}))'$ , and  $\varepsilon^{(j)} = (\varepsilon_1^{(j)}, \dots, \varepsilon_n^{(j)})'$ . Then we have  $E(\varepsilon) = \mathbf{0}_{nJ \times 1}$  and  $\Omega \equiv \text{Var}(\varepsilon) = \Sigma \otimes I_{n \times n}$  where  $\Sigma$  is a  $J \times J$  matrix with typical diagonal element  $\sigma_{ii}^2$  and off-diagonal element  $\sigma_{ij}$  for  $i, j = 1, \dots, J$ .

In order to simplify the notation, we will focus on the case with  $J = 2$ . Let  $\mathbf{y} \equiv (y^{(1)'}, y^{(2)'})'$ ,  $X_{i,x_j} \equiv (1, (X_i^{(j)} - x_j)')'$ ,  $\mathbf{X}_{x_j}^{(j)} \equiv (X_{1,x_j}, \dots, X_{n,x_j})'$ ,  $\mathbf{X}_x^* \equiv \begin{pmatrix} \mathbf{X}_{x_1}^{(1)} & \mathbf{0}_{n \times (q_2+1)} \\ \mathbf{0}_{n \times (q_1+1)} & \mathbf{X}_{x_2}^{(2)} \end{pmatrix}$ , and  $\Sigma \equiv \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{22}\rho \\ \sigma_{11}\sigma_{22}\rho & \sigma_{22}^2 \end{pmatrix}$ . As before, we can obtain the conventional LLE as  $\hat{\delta}_{\text{LL}} = (\mathbf{X}_x^{*'} \mathbf{K}_1 \mathbf{X}_x^*)^{-1} \mathbf{X}_x^{*'} \mathbf{K}_1 \mathbf{y}$ , where  $\mathbf{K}_1 = \text{diag}(\mathbf{K}_{11}, \mathbf{K}_{12})$ , and  $\mathbf{K}_{1j} = \text{diag}(K_{h_{1j}}(X_1^{(j)} - x_j), \dots, K_{h_{1j}}(X_n^{(j)} - x_j))$  for  $j = 1, 2$ . Similarly, assume that  $\Omega = P P'$  for some  $2n \times 2n$  matrix  $P$ . Let  $p_{ij}$  and  $v_{ij}$  denote the  $(i, j)$ th element of  $P$  and  $P^{-1}$ , respectively. Let  $H \equiv \text{diag}(v_{1,1}^{-1}, \dots, v_{2n,2n}^{-1})$ . By Cholesky decomposition we have

$$P^{-1} = \Omega^{-1/2} = \begin{pmatrix} \left( (\sigma_{11}\sqrt{1-\rho^2})^{-1} I_n - \rho \left( \sigma_{22}\sqrt{1-\rho^2} \right)^{-1} I_n \right) & \\ \mathbf{0}_{n \times n} & \sigma_{22}^{-1} I_n \end{pmatrix},$$

i.e.,  $v_{ii} = 1 / \left( \sigma_{11}\sqrt{1-\rho^2} \right)$  and  $v_{n+i,n+i} = 1/\sigma_{22}$  for  $i = 1, \dots, n$ .

Let  $\delta(x) = (m_1(x_1), \partial m_1(x_1) / \partial x_1', m_2(x_2), \partial m_2(x_2) / \partial x_2')'$  where  $x$  is a disjoint union of  $x_1$  and  $x_2$ . Let  $\mathbf{K}_2 = \text{diag}(\mathbf{K}_{21}, \mathbf{K}_{22})$  where  $\mathbf{K}_{2j} = \text{diag}(K_{h_{2j}}(X_1^{(j)} - x_j), \dots, K_{h_{2j}}(X_n^{(j)} - x_j))$  for  $j = 1, 2$ . Notice that the bandwidth  $h_{2j}$  is used in the



second step. Applying our two-step estimator to the seemingly unrelated regression models yields the following estimator of  $\delta(x)$  :

$$\widehat{\delta}_{\text{SUW}}(x) = (\mathbf{R}_x^{*\prime} \mathbf{K}_2 \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*\prime} \mathbf{K}_2 \widehat{\mathbf{Z}}^* \quad (10)$$

where  $\mathbf{R}_x^* = \text{diag}(H_1^{-1} \mathbf{X}_{x1}^{(1)}, H_2^{-1} \mathbf{X}_{x2}^{(2)})$ ,  $H_1 = \text{diag}(v_{11}^{-1}, \dots, v_{nn}^{-1})$ ,  $H_2 = \text{diag}(v_{n+1,n+1}^{-1}, \dots, v_{2n,2n}^{-1})$ , and  $\widehat{\mathbf{Z}}^*$  is analogously defined as in Sect. 3.1. Then Theorem 2 implies that

$$\widetilde{D} \left( \widehat{\delta}_{\text{SUW}}(x) - \delta(x) - B^{(\text{SUR})} \right) \xrightarrow{d} N \left( 0, \Omega^{(\text{SUR})} \right) \quad (11)$$

where  $\widetilde{D} \equiv \text{diag}(\widetilde{D}_{h21}, \widetilde{D}_{h22})$ ,  $\widetilde{D}_{h2j} = \sqrt{nh_{2j}^q} \text{diag}(1, h_{2j}, \dots, h_{2j})$  is a  $(1+q_j) \times (1+q_j)$  diagonal matrix,  $B^{(\text{SUR})} = \begin{pmatrix} B_1^{(\text{SUR})} \\ B_2^{(\text{SUR})} \end{pmatrix}$ ,  $\Omega^{(\text{SUR})} = \begin{pmatrix} \Omega_1^{(\text{SUR})} & \mathbf{0}_{(1+q_1) \times (1+q_2)} \\ \mathbf{0}_{(1+q_2) \times (1+q_1)} & \Omega_2^{(\text{SUR})} \end{pmatrix}$ ,  $B_j^{(\text{SUR})} = \begin{pmatrix} \frac{k_{21} h_{2j}^2}{2} \sum_{s=1}^{q_j} \frac{\partial^2 m_j(x_j)}{\partial x_{js}^2} \\ \mathbf{0}_{q_j \times 1} \end{pmatrix}$ ,  $\Omega_j^{(\text{SUR})} = \begin{pmatrix} \frac{(\kappa_{02})^{q_j}}{\omega_{f,j}^*(x, \theta_0)} & \mathbf{0}_{1 \times q_j} \\ \mathbf{0}_{q_j \times 1} & \frac{\kappa_{22}(\kappa_{02})^{q_j-1}}{\omega_{f,j}^*(x, \theta_0) \kappa_{21}^2} I_{q_j} \end{pmatrix}$ ,  $\omega_{f,1}^*(x, \theta_0) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n v_{ii}^2 \times f_i(x_1)$ , and  $\omega_{f,2}^*(x, \theta_0) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n v_{n+i,n+i}^2 f_i(x_2)$ , and  $x_{js}$  is the  $s$ th element of  $x_j$  for  $j = 1$  and  $2$ .

### 3.2.2 Clustered or panel data models

We consider the following one-way random effects model

$$Y_{ij} = m(X_{ij}) + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, J,$$

where  $X_{ij}$  is a  $q \times 1$  vector of exogenous variables,  $\alpha_i$  is independently and identically distributed (IID)  $(0, \sigma_\alpha^2)$ ,  $\varepsilon_{ij}$  is IID  $(0, \sigma_\varepsilon^2)$ ,  $\alpha_i$  and  $\varepsilon_{lj}$  are uncorrelated for all  $i, l = 1, 2, \dots, n$ , and  $m(\cdot)$  is an unknown smooth function. Let  $u_{ij} = \alpha_i + \varepsilon_{ij}$ ,  $u_i \equiv (u_{i1}, \dots, u_{iJ})'$ , and  $u \equiv (u_1, \dots, u_n)'$ . By assumption, we have  $\Sigma \equiv E(u_i u_i') = \sigma_\varepsilon^2 I_J + \sigma_\alpha^2 \mathbf{1}_J \mathbf{1}_J'$  and  $\Omega(\sigma_\varepsilon^2, \sigma_\alpha^2) \equiv E(u u') = I_n \otimes \Sigma$ , where  $\mathbf{1}_J$  is a  $J \times 1$  vector of ones. As in MY, assuming that  $\Omega = P P'$  for some square matrix  $P$ , then  $P^{-1} = I_n \otimes V^{-1/2}$ , where  $V^{-1/2} = (v_{ij})_{i,j=1,\dots,J}$  with  $v_{ii} \equiv v = \frac{1}{\sigma_\varepsilon} - (1 - \frac{\sigma_\varepsilon}{\sigma_1}) \frac{1}{J \sigma_\varepsilon}$  for all  $i = 1, \dots, J$ ,  $v_{ij} = v_o = -(1 - \frac{\sigma_\varepsilon}{\sigma_1}) \frac{1}{J \sigma_\varepsilon}$  for all  $i \neq j = 1, \dots, J$ , and  $\sigma_1 = \sqrt{J \sigma_\alpha^2 + \sigma_\varepsilon^2}$ .

Our two-step estimator is  $\widehat{\delta}_{\text{SUW}, h_2}(x) = (\mathbf{R}_x^{*\prime} \mathbf{K}_{x, h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*\prime} \mathbf{K}_{x, h_2} \widehat{\mathbf{Z}}^*$ , where  $\mathbf{R}_x^* \equiv H^{-1} \mathbf{R}_x$ ,  $\mathbf{R}_x \equiv (X_{x,11}, \dots, X_{x,1J}, \dots, X_{x,n1}, \dots, X_{x,nJ})$ ,  $X_{x,ij} = (1, (X_{ij} - x)')$ ,  $\mathbf{K}_{x, h_2} \equiv \text{diag}(K_{h_2}(X_{11} - x), \dots, K_{h_2}(X_{1J} - x), \dots, K_{h_2}(X_{n1} - x), \dots, K_{h_2}(X_{nJ} - x))$ , and  $\widehat{\mathbf{Z}}^*$  is analogously defined as in Sect. 3.1. Then Theorem 2 implies that

$$\sqrt{nh_2^q} D_h \left( \widehat{\delta}_{\text{SUW}, h_2}(x) - \delta(x) - B^{(\text{Panel})} \right) \xrightarrow{d} N \left( 0, \Omega^{(\text{Panel})} \right) \quad (12)$$

$$\text{where } B^{(\text{Panel})} = \begin{pmatrix} \frac{\kappa_{21} h_2^2}{2} \sum_{s=1}^q \frac{\partial^2 m(x)}{\partial x_s^2} \\ \mathbf{0}_{q \times 1} \end{pmatrix}, \Omega^{(\text{Panel})} = \begin{pmatrix} \frac{(\kappa_{02})^q}{v^2 \sum_{j=1}^J f_j(x)} & \mathbf{0}_{1 \times q} \\ \mathbf{0}_{q \times 1} & \frac{\kappa_{22} (\kappa_{02})^{q-1}}{v^2 \sum_{j=1}^J f_j(x) \kappa_{21}^2} I_q \end{pmatrix},$$

and  $f_j(\cdot)$  denotes the marginal density of  $X_{ij}$ .

#### 4 Monte Carlo simulations

Now we conduct a small set of Monte Carlo simulations to compare the finite sample performance of our estimator with that of LLE and MY. Consider the following data generating process:

$$Y_i = m(X_i) + U_i, \quad i = 1, \dots, n,$$

where the univariate random variables  $X_i$  are first generated independently from  $N(0, 1)$  and then truncated at  $\pm 3$ . We use two specifications for  $m(x)$ :  $0.5 + e^{-4x}/(1+e^{-4x})$  and  $1 - 0.9e^{-2x^2}$ , which correspond to  $m_2(x)$  and  $m_3(x)$ , respectively, in MY.

For the error terms, we consider two cases. In Case 1, we assume a time series structure for  $U_i$  and generate  $U_i$  from the following AR(2) process:  $U_i = 0.5U_{i-1} - 0.4U_{i-2} + \varepsilon_i$ , where  $\varepsilon_i$  are IID  $N(0, 1)$ . In Case 2, we assume that  $U_i$  are heteroskedastic but independent of each other, and generate  $U_i$ ,  $i = 1, \dots, \frac{n}{2}$ , as IID from  $N(0, 2)$ , and  $U_i$ ,  $i = \frac{n}{2} + 1, \dots, n$ , as IID from  $N(0, 4)$ . In the first case, only the first two diagonal elements in the square root matrix ( $P$ ) of the covariance matrix ( $\Omega$ ) of  $\mathbf{U} \equiv (U_1, \dots, U_n)'$  are distinct from others, so that the MY and SUW estimators are asymptotically equivalent and we should not observe significant difference in the finite sample performance between the two estimators. In the second case, however, the LLE and MY estimators are asymptotically equivalent and both are dominated by the SUW estimator.

For all estimators, we use the Gaussian kernel. For bandwidth sequences, we use the least-squares cross validation to choose  $h_2$ , and set  $h_1 = h_2^{5/4}$ , where  $h_1$  and  $h_2$  are used in the first and the second step estimations, respectively, for both MY and SUW estimators. The one-step LLE estimator uses  $h_2$  in the estimation.

Although we know the covariance matrix  $\Omega$  of  $\mathbf{U}$  in the simulation, we estimate it according to the AR(2) specification in Case 1 and heteroskedastic specification in Case 2. In order to be specific, we estimate the two autoregressive coefficients in the first case and the two variances in the second case. Based on the estimation of  $m(x)$  on all data points  $X_1, \dots, X_n$ , we calculate the bias, standard deviation (Std), root mean squared error (RMSE), and mean squared error (MSE) for each estimator and average them across 1000 replications. The sample sizes under our investigation are 100 and 200.

Table 1 reports the finite sample performance for the three estimators for both  $m(x)$  and  $\partial m(x)/\partial x$  in the case of AR(2) errors. First, in terms of Std and RMSE (or MSE),

**Table 1** Comparison of various estimators of  $m(x)$  and  $\partial m(x)/\partial x$  for AR(2) errors

$n$	Estimators	Specification 1				Specification 2			
		Bias	Std	RMSE	MSE	Bias	Std	RMSE	MSE
$m(x)$									
100	LLE	0.0081	0.2422	0.2681	0.0719	0.0193	0.2657	0.3031	0.0919
	MY	−0.0084	0.2141	0.2355	0.0555	0.0221	0.2386	0.2748	0.0755
	SUW	−0.0086	0.2113	0.2307	0.0532	0.0212	0.2372	0.2734	0.0748
200	LLE	−0.0009	0.1825	0.2067	0.0427	0.0231	0.1913	0.2274	0.0517
	MY	−0.0008	0.1619	0.1829	0.0335	0.0253	0.1697	0.2028	0.0411
	SUW	−0.0009	0.1600	0.1793	0.0322	0.0254	0.1689	0.2021	0.0409
$\partial m(x)/\partial x$									
100	LLE	−0.0136	0.9198	1.0064	1.0129	0.0132	1.0658	1.1774	1.3864
	MY	−0.0119	0.7799	0.8666	0.7510	0.0128	0.9214	1.0287	1.0582
	SUW	−0.0116	0.7799	0.8659	0.7497	0.0118	0.9177	1.0263	1.0532
200	LLE	−0.0151	0.7521	0.8223	0.6762	−0.0056	0.7289	0.8178	0.6689
	MY	−0.0109	0.6421	0.7129	0.5082	−0.0051	0.6099	0.6936	0.4811
	SUW	−0.0108	0.6400	0.7086	0.5022	−0.0052	0.6074	0.6911	0.4777

both MY and SUW estimators outperform the LLE estimator, and have smaller bias for estimators of  $\partial m(x)/\partial x$ , but the former tends to have slightly larger biases for estimating  $m(x)$ . Second, as expected the efficiency gain of the SUW estimator over the MY estimator is tiny and may be ignored in the AR(2) error structure. In some sense, this verifies our asymptotic theory: since only the first two diagonal elements in the  $H$  matrix in the case of AR(2) error process are different from the rest, the ratio of asymptotic variance of our estimator over that of the MY's is 1 in this case and the two estimators share the same asymptotic biases so that one expects that the two estimators behave similarly in finite samples. Noting that the more different diagonal elements in the square root matrix  $P$  of  $\Omega$ , the more efficiency gain we may have, we expect that prominent efficiency gain can be achieved only in AR( $p$ ) model with  $p \equiv p(n) \rightarrow \infty$  as  $n \rightarrow \infty$  or in ARMA( $p, q$ )-type of models.

Table 2 compares the three estimators for both  $m(x)$  and  $\partial m(x)/\partial x$  under the heteroskedastic errors.<sup>1</sup> In the presence of heteroskedasticity only, the MY estimator is identical to that of LLE so that we can compare the LLE with our estimator SUW. Obviously, SUW has improvement over LLE and thus MY in the sense of having lower Std and RMSE (or MSE). The simulation results provide a strong support that the SUW estimator is more efficient than the LLE and MY estimators by considering heterogeneity in the error structure.

<sup>1</sup> The results in Table 2 are obtained for the heteroskedastic error case with two different variances. We also did the simulations for the case with four different variances, and observed higher relative efficiency gain of SUW over MY compared to the former case.

**Table 2** Comparison of various estimators of  $m(x)$  and  $\partial m(x)/\partial x$  for heteroskedastic errors

$n$	Estimators	Specification 1				Specification 2			
		Bias	Std	RMSE	MSE	Bias	Std	RMSE	MSE
$m(x)$									
100	LLE/MY	−0.0049	0.5923	0.6708	0.4500	0.0499	0.6032	0.6437	0.4144
	SUW	−0.0119	0.5147	0.6002	0.3602	0.0467	0.5314	0.5796	0.3359
200	LLE/MY	−0.0063	0.4679	0.5489	0.3013	0.0386	0.4736	0.5845	0.3416
	SUW	−0.0078	0.4040	0.4872	0.2373	0.0402	0.4098	0.5261	0.2768
$\partial m(x)/\partial x$									
100	LLE/MY	−0.0198	1.6925	1.9024	3.6189	0.0055	1.9187	2.0116	4.0465
	SUW	−0.0173	1.4692	1.7182	2.9522	0.0044	1.7060	1.8271	3.3382
200	LLE/MY	0.0051	1.8507	2.1445	4.5991	−0.0227	1.3191	1.4887	2.2163
	SUW	0.0059	1.5985	1.9353	3.7455	−0.0135	1.1327	1.3250	1.7555

**5 Empirical application: public capital productivity**

In order to illustrate the applicability of our results in real data setting, this section conducts an empirical study, which employs a panel dataset for the U.S. 48 contiguous states over the period 1970–1986 to revisit the relationship between public capital and private sector output.<sup>2</sup> Is public-sector capital productive? What is the role for public-sector in affecting private economic performance? The debates on these questions have received extensive attention from economists. Some empirical work, for instance, [Munnell \(1990\)](#), incorporated public capital into the production function and found that the public capital played a positive and significant role in effecting the private sector output. However, some economists hold opposite conclusions which claimed that the public capital had significant and negative effects on private productivity [see, e.g., [Evans and Karras \(1994\)](#)]. In addition, another type of findings is that the contribution of the public infrastructure does not have quantitatively significant spillover effects on private sector across states, see [Holtz-Eakin \(1994\)](#) and [Baltagi and Pinnoi \(1995\)](#). All the aforementioned works are conducted within the parametric framework, and assumed a particular production function for the underlying production function, and constant elasticities of the specified models across all the states and all the years. The question arises naturally is whether or not the estimates of returns to inputs can be trusted under the above settings. As we know, nonparametric method is free from the misspecification issue; also, nonparametric regression estimation provides local estimates so that we can clearly examine a variety of the estimates of returns to inputs across all states and years.

<sup>2</sup> Details on this dataset can be found in [Munnell \(1990\)](#), [Baltagi and Pinnoi \(1995\)](#), and [Henderson and Ullah \(2008\)](#) as well.

Following Baltagi and Pinnoi (1995) and Henderson and Ullah (2008), we consider the following one-way random effects nonparametric model:

$$\log(Y_{it}) = m(\log(KG_{it}), \log(KPR_{it}), \log(L_{it}), UNEM_{it}) + \alpha_i + \varepsilon_{it},$$

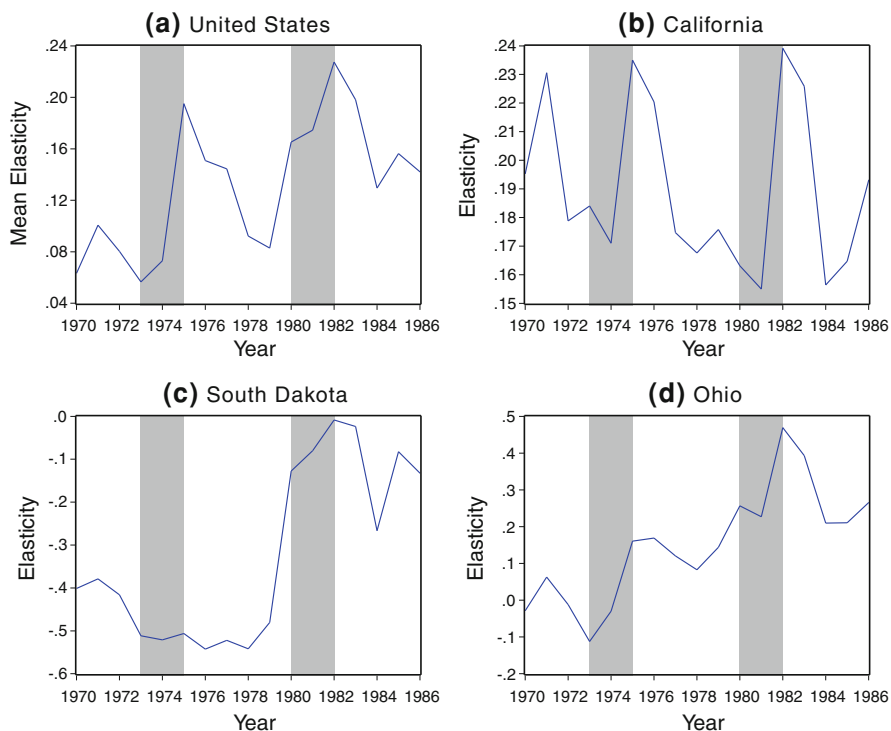
where  $i = 1, \dots, 48$ ,  $t = 1, \dots, 17$ ,  $Y_{it}$  denotes the GDP of state  $i$  in period  $t$ ,  $KG$  denotes public capital,  $KPR$  is the private capital stock estimated from the Bureau of Economics Analysis,  $L$  is employment, and  $UNEM$  stands for the unemployment rate used to control for business cycle effects as in the previous literature.

Based on the SUW LLEs, one can obtain estimates of all the first order partial derivatives of  $m$  with respect to its four arguments. Then we can calculate the estimated mean elasticities of GDP ( $Y$ ) with respect to public capital ( $KG$ ), private capital ( $KPR$ ), and employment ( $L$ ), and the estimated mean percentage increase of  $Y$  due to a unit increase of unemployment rate ( $UNEM$ ) by averaging the corresponding first order partial derivatives across all observations. The estimated mean elasticities for SUW estimators of  $KG$ ,  $KPR$ , and  $L$ , and the estimated coefficient of  $UNEM$  are 0.1314, 0.2852, 0.6326, and  $-0.0041$ , respectively. In order to obtain the standard errors for these estimates, we propose to bootstrap the data 500 times by resampling across individuals and keep the time series structure for each individual unchanged. We obtain estimates of the average elasticities and coefficients for each bootstrap data, based on which we can calculate the bootstrap standard errors for the above estimates. They are 0.0510, 0.0265, 0.0420, and 0.0036, respectively. In order to obtain these results, we use the Epanechnikov kernel and choose the second stage bandwidth  $h_2 = (h_{21}, \dots, h_{24})$  according to the Silverman rule of thumb (ROT), i.e.,  $h_{2j} = 1.06s_j n^{-1/(4+4)}$  where  $s_j$  denotes the sample Std of the  $j$ th regressor in the regression for  $j = 1, \dots, 4$ . We set  $h_{1j} = 1.06s_j n^{-1/(4+3)}$  for  $j = 1, \dots, 4$ . In addition,  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$  are estimated using the consistent estimators proposed in Ruckstuhl et al. (2000, p. 61).

Similarly, we can obtain the estimated median elasticities of  $Y$  with respect to  $KG$ ,  $KPR$ , and  $L$ , and the estimated median slope coefficient of  $UNEM$  as 0.1550, 0.2742, 0.6501, and  $-0.0027$ , respectively, and the corresponding bootstrap standard errors as 0.0433, 0.0257, 0.0339, and 0.0040, respectively. Noting that both the estimated mean and median elasticities or slope coefficients are asymptotically normally distributed, we can test whether their population true values are different from 0 based on these estimates and their corresponding standard errors using the standard normal critical values.<sup>3</sup> Obviously, the estimated mean or median elasticities of public capital, private capital, and labor are statistically significant at the 1 % nominal level, but this is not the case for the unemployment rate.

In order to summarize, the empirical findings by applying this two-step nonparametric estimation are as follows. First, the mean elasticity of public capital on states private economic growth is positive and statistically significant at the 1 % nominal

<sup>3</sup> Even though we estimate a four-dimensional nonparametric object  $m(\cdot)$  and  $48 \times 17 = 816$  observations seem not to be large enough for this purpose, our interest mainly lies in the estimation of the *average derivatives*. It is well known that the estimate of these average derivatives possesses the parametric convergence rate so that the “curse of dimensionality” may not be a problem at least in theory.



**Fig. 1** Elasticity of public capital over 1970–1986

level. In other words, the public capital has spillover effects on average across states. Even though its spillover effects are smaller than private sector capital stock but still non-negligible. Second, we find that the majority of states have positive relationship between the public capital and the private economic performance. However, a few states, for instances, Wyoming, South Dakota, North Dakota, New Mexico, Montana, have negative returns to the public capital, which are consistent with some recent studies under the nonparametric framework. One possible explanation for this is that the group of these states with negative returns to public capital may overinvest in infrastructure, meanwhile their gross state products are relatively small [see, e.g., [Henderson and Ullah \(2008\)](#)]. Third, as Fig. 1a shows the mean of returns to the public capital across all the 48 states changes over the period of 1970–1986, which implies the change of elasticity at the national level. The pattern of these changes reveals that the returns to public capital increased sharply during recessions (shaded area in Fig. 1), started to decrease when the economy stepped into recovering, and fluctuated in small magnitude during normal time. The reason behind this may be that when the economy is in recessions the private sector becomes weak, and the public sector capital turns to play a more effective role than normal periods. As a result, the private sector may gain more benefits from the government investments on the public capital during recessions than the other times. In order to show the changes of elasticity of public capital for individual states, we plot elasticities of California, South Dakota, and Ohio over 1970–1986. As shown in Fig. 1b, California has all positive returns over the time

period, and has similar pattern as Fig. 1a. Figure 1c shows that South Dakota has all negative returns to public capital during that time, but does not show any obvious pattern. The elasticity of public capital of Ohio is plotted in Fig. 1d, from which we can see that Ohio state has positive elasticity in most of the years under study, but negative elasticity in some other years. Also, similar to the changing pattern in Fig. 1a, b, the returns to public capital in Ohio sharply increased during the contraction periods.

## 6 Concluding remarks

In this paper, we propose a two-step estimator (SUW) for nonparametric regression with a general parametric error covariance that is more efficient than that of MY's. The results are applied to two popular nonparametric regression models, namely, seemingly unrelated regression models and one-way random effects model. Notice that by the transformation which we employ to obtain our two-step estimator the transformed errors has spherical parametric covariance structure. Therefore, intuitively SUW estimator should outperform those nonparametric regression estimators that fail to fully utilize the information in the error covariance. Simulations confirm the finite sample out-performance of our estimator over both LLE and MY's under both serial correlation case and heteroskedastic case. Notice that under heteroskedasticity MY's estimator degenerates to LLE as the former fails to incorporate the diagonal information in the error covariance, which is also confirmed in Monte Carlo simulations.

In order to complement the analysis of the Monte Carlo section and illustrate the applicability of our method in real data settings, an empirical study on public capital productivity puzzle is conducted. The empirical findings are consistent with the previous nonparametric studies. In general, the return to public capital is significantly positive. An interesting finding in our empirical study is that the returns to public capital may change with business cycles. The private sector may tend to gain more benefits from the government investments on the public capital during recessions than during the other time periods according to the empirical study presented here. The last remark we would like to indicate here is that the existence of random individual effects is assumed throughout the empirical study. In practice one may need test for this assumption. However, this is not of the main concern in the present paper.

## 7 Mathematical appendix: Proof of Theorem 3.1

Following MY, we can readily show that  $\widehat{\delta}_{\text{SUW},h_2}(x)$  is asymptotically equivalent to the following infeasible estimator

$$\widehat{\delta}_{\text{SUW},h_2}(x) \equiv (\mathbf{R}_x^{*'} \mathbf{K}_{x,h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x,h_2} \mathbf{Z}^* \quad (13)$$

where  $\mathbf{Z}^* \equiv P^{-1} \mathbf{Y} + (H^{-1} - P^{-1}) \mathbf{m} = H^{-1} \mathbf{m} + \epsilon^*$ . By the second order Taylor expansion around  $x$  for elements in  $\mathbf{m}$ , we have

$$\widehat{\delta}_{\text{SUW},h_2}(x) = \delta(x) + (\mathbf{R}_x^{*'} \mathbf{K}_{x,h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x,h_2} \left\{ H^{-1} \mathbf{B}_x + \epsilon^* \right\} + o_p(h_2^2)$$

where  $\mathbf{B}_x$  is a  $n \times 1$  column vector whose  $i$ th element is given by  $b_{x,i} = \frac{1}{2} (X_i - x)' \cdot m^{(2)}(x) (X_i - x)$ , and  $m^{(2)}(x)$  is the  $q \times q$  Hessian matrix of  $m(x)$ . It follows that

$$\begin{aligned} \sqrt{nh_2^q} D_{h_2} (\widehat{\delta}_{\text{SUW}, h_2}(x) - \delta(x)) &= \sqrt{nh_2^q} D_{h_2} (\mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} H^{-1} \mathbf{B}_x \\ &\quad + \sqrt{nh_2^q} D_{h_2} (\mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \mathbf{R}_x^*)^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \epsilon^* + o_p(1) \\ &\equiv B_{\text{SUW}} + V_{\text{SUW}} + o_p(1), \text{ say,} \end{aligned} \quad (14)$$

where the definitions of the bias term  $B_{\text{SUW}}$  and the variance term  $V_{\text{SUW}}$  are self-evident. Note that  $E(\epsilon^* \epsilon^{*'}) = I_{n \times n}$ .

In order to calculate the asymptotic bias, let  $S_n \equiv n^{-1} D_{h_2}^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \mathbf{R}_x^* D_{h_2}^{-1}$ . It is easy to show that

$$\begin{aligned} S_n &= n^{-1} \sum_{i=1}^n \begin{pmatrix} v_{ii}^2 & v_{ii}^2 \frac{(X_i - x)'}{h_2} \\ v_{ii}^2 \frac{X_i - x}{h_2} & v_{ii}^2 \frac{(X_i - x)(X_i - x)'}{h_2^2} \end{pmatrix} K_{h_2}(X_i - x) \\ &\xrightarrow{p} \begin{pmatrix} \omega_f^*(x, \theta_0) & 0 \\ 0 & \omega_f^*(x, \theta_0) \kappa_{21} I_q \end{pmatrix}. \end{aligned} \quad (15)$$

Similarly,

$$\begin{aligned} \frac{1}{n} D_{h_2}^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} H^{-1} \mathbf{B}_x &= \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n v_{ii}^2 \mathbf{K}_{x, h_2} b_{x,i} \\ \sum_{i=1}^n v_{ii}^2 \frac{X_i - x}{h_2} \mathbf{K}_{x, h_2} b_{x,i} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\omega_f^*(x, \theta_0) \kappa_{21} h_2^2}{2} \sum_{j=1}^q \frac{\partial^2 m(x)}{\partial x_j^2} \\ \mathbf{0}_{q \times 1} \end{pmatrix} + o_p(h_2^2). \end{aligned}$$

$$\text{It follows that } B_{\text{SUW}} = \sqrt{nh_2^q} S_n^{-1} \frac{1}{n} D_{h_2}^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} H^{-1} \mathbf{B}_x = \begin{pmatrix} \sqrt{nh_2^q} \frac{\kappa_{21} h_2^2}{2} \sum_{j=1}^q \frac{\partial^2 m(x)}{\partial x_j^2} \\ \mathbf{0}_{q \times 1} \end{pmatrix}$$

$+ o_p(h_2^2)$ .

Next, by (14)–(15) we have

$$\begin{aligned} V_{\text{SUW}} &= \sqrt{nh_2^q} S_n^{-1} \frac{1}{n} D_{h_2}^{-1} \mathbf{R}_x^{*'} \mathbf{K}_{x, h_2} \epsilon^* \\ &= \begin{pmatrix} \omega_f^*(x, \theta_0) & 0 \\ 0 & \omega_f^*(x, \theta_0) \kappa_{21} I_q \end{pmatrix}^{-1} (1 + o_p(1)) \\ &\quad \times \sqrt{n^{-1} h_2^q} D_{h_2}^{-1} \sum_{i=1}^n v_{ii} K_{h_2}(X_i - x) \begin{pmatrix} \epsilon_i^* \\ (X_i - x) \epsilon_i^* \end{pmatrix}, \end{aligned}$$

where  $\epsilon_i^*$  is the  $i$ th element of  $\epsilon^*$ . Applying the Liapounov central limit theorem yields  $V_{\text{SUW}} \xrightarrow{d} N(0, \Omega_{\text{SUW}})$ . This completes the proof of the theorem.



**Acknowledgments** We sincerely thank Badi H. Baltagi and two anonymous referees for their many insightful comments and suggestions that lead to a substantial improvement of the presentation. We all thank Carlos Martins-Filho and Daniel Henderson for discussions on the subject matter of this paper. The second author gratefully acknowledges the financial support from the Academic Senate, University of California, Riverside.

## References

- Baltagi BH, Pinnoi N (1995) Public capital stock and state productivity growth: further evidence from an error components model. *Empir Econ* 20:351–359
- Evans P, Karras G (1994) Are government activities productive? evidence from a panel of U.S. states. *Rev Econ Stat* 76(1):1–11
- Henderson D, Ullah A (2008) Nonparametric estimation in a one-way error component model: a Monte Carlo analysis. Working Paper, University of California, Riverside
- Holtz-Eakin D (1994) Public-sector capital and the productivity puzzle. *Rev Econ Stat* 76:12–21
- Lin X, Carroll RJ (2000) Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J Am Stat Assoc* 95:520–534
- Linton OB, Mammen E (2008) Nonparametric transformation to white noise. *J Econ* 142:241–264
- Martins-Filho C, Yao F (2009) Nonparametric regression estimation with general parametric error covariance. *J Multivar Anal* 100:309–333
- Munnell AH (1990) How does public infrastructure affect regional economic performance? *N Engl Econ Rev*, (September/October) 11–33
- Ruckstuhl AF, Welsh AH, Carroll RJ (2000) Nonparametric function estimation of the relationship between two repeatedly measured variables. *Stat Sin* 10:51–71
- Su L, Ullah A (2006) More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. *Econ Theory* 22:98–126
- Su L, Ullah A (2007) More efficient estimation of nonparametric panel data models with random effects. *Econ Lett* 96:375–380
- Xiao Z, Linton OB, Carroll RJ, Mammen E (2003) More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *J Am Stat Assoc* 98:980–992