Journal of Multivariate Analysis 105 (2012) 164–175

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Difference based ridge and Liu type estimators in semiparametric regression models[☆]

Esra Akdeniz Duran[a,1], Wolfgang Karl Härdle[b], Maria Osipenko[c,*]

[a] Department of Statistics, Gazi University, Turkey
[b] Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Germany
[c] CASE, School of Business and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Germany

ARTICLE INFO

Article history:
Received 4 March 2011
Available online 7 September 2011

AMS subject classifications:
62G08
62J07

Keywords:
Difference based estimator
Differencing estimator
Differencing matrix
Liu estimator
Liu type estimator
Multicollinearity
Ridge regression estimator
Semiparametric model

ABSTRACT

We consider a difference based ridge regression estimator and a Liu type estimator of the regression parameters in the partial linear semiparametric regression model, $y = X\beta + f + \varepsilon$. Both estimators are analyzed and compared in the sense of mean-squared error. We consider the case of independent errors with equal variance and give conditions under which the proposed estimators are superior to the unbiased difference based estimation technique. We extend the results to account for heteroscedasticity and autocovariance in the error terms. Finally, we illustrate the performance of these estimators with an application to the determinants of electricity consumption in Germany.

1. Introduction

Semiparametric partial linear models have received considerable attention in statistics and econometrics. They have a wide range of applications, from biomedical studies to economics. In these models, some explanatory variables have a linear effect on the response while others are entering nonparametrically. Consider the semiparametric regression model:

$$y_i = x_i^\top \beta + f(t_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where y_i's are observations at t_i, $0 \le t_1 \le t_2 \le \cdots \le t_n \le 1$ and $x_i^\top = (x_{i1}, x_{i2}, \ldots, x_{ip})$ are known p-dimensional vectors with $p \le n$. In many applications, t_i's are values of an extra univariate "time" variable at which responses y_i are observed. In the case $t_i \in \mathbb{R}^k$, $t_i = (t_{1i}, \ldots, t_{ki})^\top$, the triples $(y_1, x_1, t_1), \ldots, (y_n, x_n, t_n)$ should be ordered using one of the algorithms mentioned in [30], Appendix A, or in [8, Section 2.2].

☆ This research was supported by Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk'.
* Corresponding author.
E-mail addresses: esraakdeniz@gmail.com (E. Akdeniz Duran), maria.osipenko@wiwi.hu-berlin.de (M. Osipenko).
1 Dr. Esra Akdeniz Duran was a research associate at Humboldt-Universität zu Berlin, Germany during this research.

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.08.018

In Eq. (1), $\beta = (\beta_1, \ldots, \beta_p)^\top$ is an unknown p-dimensional parameter vector, $f(\cdot)$ is an unknown smooth function and ε's are independent and identically distributed random errors with $\mathrm{E}(\varepsilon|x, t) = 0$ and $\mathrm{Var}(\varepsilon|x, t) = \sigma^2$. We shall call $f(t)$ the smooth part of the model and assume that it represents a smooth unparameterized functional relationship.

The goal is to estimate the unknown parameter vector β and the nonparametric function $f(t)$ from the data $\{y_i, x_i, t_i\}_{i=1}^n$. In vector/matrix notation, (1) is written as

$$y = X\beta + f + \varepsilon \tag{2}$$

where $y = (y_1, \ldots, y_n)^\top, X = (x_1, \ldots, x_n), f = \{f(t_1), \ldots, f(t_n)\}^\top, \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$.

Semiparametric models are by design more flexible than standard linear regression models since they combine both parametric and nonparametric components. There exist various goodness-of-fit tests to identify the nonparametric part in this kind of models; see [8] and the references therein. Estimation techniques for semiparametric partially linear models are based on different nonparametric regression procedures. The most important approaches to estimate β and f are given in [12,4,7,6,5,14,24,15,33].

In practice, researchers often encounter the problem of multicollinearity. In case of multicollinearity, we know that the $(p \times p)$ matrix $X^\top X$ has one or more small eigenvalues; the estimates of the regression coefficients can therefore have large variances: the least squares estimator performs poorly in this case. Hoerl and Kennard [17] proposed the ridge regression estimator and it has become the most common method to overcome this particular weakness of the least squares estimator. For the purpose of this paper, we will employ the biased estimator that was proposed by Liu [20] to combat the multicollinearity. The Liu estimator combines the Stein [26] estimator with the ridge regression estimator; see also [1,13].

The condition number is a measure of multicollinearity. If $X^\top X$ is ill-conditioned with a large condition number, the ridge regression estimator or Liu estimator can be used to estimate β, [21]. We consider difference based ridge and Liu type estimators in comparison to the unbiased difference based approach. We give theoretical conditions that determine superiority among the estimation techniques in the mean squared error matrix sense.

We use data on monthly electricity consumption and its determinants (income, electricity and gas prices, temperature) for Germany. The purpose is to understand electricity consumption as a linear function of income and price and a nonlinear function of temperature: semiparametric approach is therefore necessary here. The data reveal a high condition number of 20.5; we therefore expect a more precise estimation with Ridge or Liu type estimators. We show how our theoretically derived conditions can be implemented for a given data set and be used to determine the appropriate biased estimation technique.

The paper is organized as follows. In Section 2, the model and the differencing estimator is defined. We introduce difference based ridge and Liu type estimators in Section 3. In Section 4, the differencing estimator proposed by Yatchew [30] and the difference based Liu type estimator are compared in terms of the mean squared error. In Section 5, both biased regression methodologies in semiparametric regression models are compared in terms of the mean squared error. Section 6 relaxes the assumption of i.i.d. errors and replicates the results of the previous sections in the presence of heteroscedasticity and autocorrelation. Section 7 gives a real data example to show the performance of the proposed estimators.

2. The model and differencing estimator

In this section, we introduce a difference based technique for the estimation of the linear coefficient vector in a semiparametric regression. This technique has been used to remove the nonparametric component in the partially linear model by various authors (e.g. [30,32,19,3]).

Consider the semiparametric regression model (2). Let $d = (d_0, d_1, \ldots, d_m)^\top$ be an $m + 1$ vector where m is the order of differencing and $d_0, d_1, \ldots, d_m$ are differencing weights that minimize

$$\sum_{k=1}^{m} \left(\sum_{j=1}^{m-k} d_j d_{k+j} \right)^2,$$

such that

$$\sum_{j=0}^{m} d_j = 0 \quad \text{and} \quad \sum_{j=0}^{m} d_j^2 = 1 \tag{3}$$

are satisfied.

Let us define the $(n - m) \times n$ differencing matrix D to have first and last rows $(d^\top, 0_{n-m-1}^\top)$, $(0_{n-m-1}^\top, d^\top)$ respectively, with i-th row $(0_i, d^\top, 0_{n-m-i-1}^\top)$, $i = 1, \ldots, (n - m - 1)$, where 0_r indicates an r-vector of all zero elements

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m & 0 \\ 0 & 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m \end{pmatrix}.$$

Applying the differencing matrix to (2) permits direct estimation of the parametric effect. Eubank et al. [6] showed that the parameter vector in (2) can be estimated with parametric efficiency. If f is an unknown function with bounded first derivative, then Df is essentially 0, so that applying the differencing matrix we have

$$\begin{aligned} Dy &= DX\beta + Df + D\varepsilon \approx DX\beta + D\varepsilon \\ \widetilde{y} &\approx \widetilde{X}\beta + \widetilde{\varepsilon} \end{aligned} \tag{4}$$

where $\widetilde{y} = Dy$, $\widetilde{X} = DX$ and $\widetilde{\varepsilon} = D\varepsilon$. Constraints (3) ensure that the nonparametric effect is removed and $\mathrm{Var}(\widetilde{\varepsilon}) = \mathrm{Var}(\varepsilon) = \sigma^2$. With (4), a simple differencing estimator of the parameter β in the semiparametric regression model results:

$$\begin{aligned} \widehat{\beta}_{(0)} &= \{(DX)^\top (DX)\}^{-1} (DX)^\top Dy \\ &= \left(\widetilde{X}^\top \widetilde{X}\right)^{-1} \widetilde{X}^\top \widetilde{y}. \end{aligned} \tag{5}$$

Thus, differencing allows one to perform inferences on β as if there were no nonparametric component f in model (2), [9].

We will also use the modified estimator of σ^2 proposed by Eubank et al. [7]

$$\widehat{\sigma}^2 = \frac{\widetilde{y}^\top (I - P^\perp)\widetilde{y}}{\mathrm{tr}\{D^\top (I - P^\perp) D\}} \tag{6}$$

with $P^\perp = \widetilde{X}(\widetilde{X}^\top \widetilde{X})^{-1}\widetilde{X}^\top$, I $(p \times p)$ identity matrix and $\mathrm{tr}(\cdot)$ denoting the trace function for a square matrix.

3. Difference based ridge and Liu type estimator

As an alternative to $\widehat{\beta}_{(0)}$ in (5), [27] propose:

$$\widehat{\beta}_{(1)}(k) = (\widetilde{X}^\top \widetilde{X} + kI)^{-1} \widetilde{X}^\top \widetilde{y}, \quad k \geq 0;$$

here k is the ridge-biasing parameter selected by the researcher. We call $\widehat{\beta}_{(1)}(k)$ a difference based ridge regression estimator of the semiparametric regression model.

From the least squares perspective, the coefficients β are chosen to minimize

$$(\widetilde{y} - \widetilde{X}\beta)^\top (\widetilde{y} - \widetilde{X}\beta). \tag{7}$$

Adding to the least squares objective (7) a penalizing function of the squared norm $\|\eta\widehat{\beta}_{(0)} - \beta\|^2$ for the vector of regression coefficients, yields a conditional objective:

$$L = (\widetilde{y} - \widetilde{X}\beta)^\top (\widetilde{y} - \widetilde{X}\beta) + (\eta\widehat{\beta}_{(0)} - \beta)^\top (\eta\widehat{\beta}_{(0)} - \beta). \tag{8}$$

Minimizing (8) with respect to β, we obtain the estimator $\widehat{\beta}_{(2)}(\eta)$ an alternative to $\widehat{\beta}_{(0)}$ in (5):

$$\widehat{\beta}_{(2)}(\eta) = (\widetilde{X}^\top \widetilde{X} + I)^{-1} (\widetilde{X}^\top \widetilde{y} + \eta\widehat{\beta}_{(0)}), \tag{9}$$

where η, $0 \leq \eta \leq 1$, is a biasing parameter and when $\eta = 1$, $\widehat{\beta}_{(2)}(\eta) = \widehat{\beta}_{(0)}$. The formal resemblance between (9) and the Liu estimator motivated [1,18,29] to call it the difference based Liu type estimator of the semiparametric regression model.

4. Mean squared error matrix (MSEM) comparison of $\widehat{\boldsymbol{\beta}}_{(0)}$ with $\widehat{\boldsymbol{\beta}}_{(2)}(\boldsymbol{\eta})$

In this section, the objective is to examine the difference of the mean square error matrices of $\widehat{\beta}_{(0)}$ and $\widehat{\beta}_{(2)}(\eta)$. We note that for any estimator $\widetilde{\beta}$ of β, its mean squared error matrix (MSEM) is defined as $\mathrm{MSEM}(\widetilde{\beta}) = \mathrm{Cov}(\widetilde{\beta}) + \mathrm{Bias}(\widetilde{\beta})\,\mathrm{Bias}(\widetilde{\beta})^\top$, where $\mathrm{Cov}(\widetilde{\beta})$ denotes the variance–covariance matrix and $\mathrm{Bias}(\widetilde{\beta}) = \mathsf{E}(\widetilde{\beta}) - \beta$ is the bias vector. The expected value of $\widehat{\beta}_{(2)}(\eta)$ can be written as

$$\mathsf{E}\{\widehat{\beta}_{(2)}(\eta)\} = \beta - (1 - \eta)(\widetilde{X}^\top \widetilde{X} + I)^{-1}\beta.$$

The bias of the $\widehat{\beta}_{(2)}(\eta)$ is given as

$$\mathrm{Bias}\{\widehat{\beta}_{(2)}(\eta)\} = -(1 - \eta)(\widetilde{X}^\top \widetilde{X} + I)^{-1}\beta. \tag{10}$$

Denoting $F_\eta = (\widetilde{X}^\top \widetilde{X} + I)^{-1}(\widetilde{X}^\top \widetilde{X} + \eta I)$ and observing F_η and $(\widetilde{X}^\top \widetilde{X})^{-1}$ are commutative, we may write $\widehat{\beta}_{(2)}(\eta)$ as

$$\begin{aligned} \widehat{\beta}_{(2)}(\eta) &= F_\eta \widehat{\beta}_{(0)} = F_\eta (\widetilde{X}^\top \widetilde{X})^{-1} \widetilde{X}^\top \widetilde{y} \\ &= (\widetilde{X}^\top \widetilde{X})^{-1} F_\eta \widetilde{X}^\top \widetilde{y}. \end{aligned}$$

Setting $S = (D^\top \widetilde{X})^\top (D^\top \widetilde{X})$ and $U = (\widetilde{X}^\top \widetilde{X})^{-1}$ we may write $\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ as

$$\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\} = \sigma^2 F_\eta USUF_\eta^\top, \tag{11}$$

$$\mathrm{Cov}(\widehat{\beta}_{(0)}) = \sigma^2 USU. \tag{12}$$

Using (11) and (12), the difference $\Delta_1 = \mathrm{Cov}(\widehat{\beta}_{(0)}) - \mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ can be expressed as

$$\begin{aligned}\Delta_1 &= \sigma^2 \left(USU - F_\eta USUF_\eta^\top\right)\\ &= \sigma^2 F_\eta \{F_\eta^{-1} USU (F_\eta^\top)^{-1} - USU\} F_\eta^\top \\ &= \sigma^2 (1-\eta^2)(U^{-1}+I)^{-1} \left\{ \frac{1}{1+\eta}(US+SU) + USU \right\} (U^{-1}+I)^{-1}. \end{aligned} \tag{13}$$

Let $\tau = \frac{1}{1+\eta} > 0$, $M = USU$, $N = US + SU$. Since $M = L^\top L$ and $\mathrm{rank}(L) = p < n - m$, then M is a $(p \times p)$ positive definite matrix, where $L = D^\top \widetilde{X}(\widetilde{X}^\top \widetilde{X})^{-1}$ and $N = US + SU$ is a symmetric matrix. Thus, we may write (13) as

$$\begin{aligned}\Delta_1 &= \sigma^2(1-\eta^2) H (M + \tau N) H \\ &= \sigma^2(1-\eta^2) H (Q^\top)^{-1} (Q^\top M Q + \tau Q^\top N Q) Q^{-1} H \\ &= \sigma^2(1-\eta^2) H (Q^\top)^{-1} (I + \tau E) Q^{-1} H,\end{aligned}$$

where $I + \tau E = \mathrm{diag}(1+\tau e_{11}, \ldots, 1+\tau e_{pp})$ and $H = (U^{-1}+I)^{-1}$. Since M is a positive definite and N is a symmetric matrix, a nonsingular matrix Q exists such that $Q^\top MQ = I$ and $Q^\top NQ = E$; here E is a diagonal matrix and its diagonal elements are the roots of the polynomial equation $|M^{-1}N - eI| = 0$ (see [11, pp. 408] and [16, pp. 563]) and since $N = US + SU \neq 0$, there is at least one diagonal element of E that is nonzero. Let $e_{ii} < 0$ for at least one i; then positive definiteness of $I + \tau E$ is guaranteed by

$$0 < \tau < \min_{e_{ii}<0} \left| \frac{1}{e_{ii}} \right|. \tag{14}$$

Hence $1 + \tau e_{ii} > 0$ for all $i = 1, \ldots, p$ and therefore $I + \tau E$ is a positive definite matrix. Consequently, Δ_1 becomes a positive definite matrix, as well. It is now evident that the estimator $\widehat{\beta}_{(2)}(\eta)$ has a smaller variance compared with the estimator $\widehat{\beta}_{(0)}$ if and only if (14) is satisfied.

Next, we give necessary and sufficient conditions for the difference based Liu type estimator $\widehat{\beta}_{(2)}(\eta)$ to be superior to $\widehat{\beta}_{(0)}$ in the mean squared error matrix (MSEM) sense.

The proof of the next theorem requires the following lemma.

Lemma 4.1 (*Farebrother [10]*). *Let A be a positive definite $(p \times p)$ matrix, b a $(p \times 1)$ nonzero vector and δ a positive scalar. Then $\delta A - bb^\top$ is non-negative if and only if $b^\top A^{-1} b \le \delta$.*

Let us compare the performance of $\widehat{\beta}_{(2)}(\eta)$ with the differencing estimator $\widehat{\beta}_{(0)}$ with respect to the MSEM criterion. In order to do that, define $\Delta_2 = \mathrm{MSEM}(\widehat{\beta}_{(0)}) - \mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}$. Observe that

$$\mathrm{MSEM}(\widehat{\beta}_{(0)}) = \mathrm{Cov}(\widehat{\beta}_{(0)}) = \sigma^2 USU \tag{15}$$

and

$$\mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\} = \sigma^2 F_\eta USUF_\eta^\top + (1-\eta)^2 (U^{-1}+I)^{-1} \beta\beta^\top (U^{-1}+I)^{-1}. \tag{16}$$

Then from (15) and (16) one derives

$$\begin{aligned}\Delta_2 &= \sigma^2 F_\eta \{F_\eta^{-1} USU (F_\eta^\top)^{-1} - USU\} F_\eta^\top - (1-\eta)^2 (U^{-1}+I)^{-1} \beta\beta^\top (U^{-1}+I)^{-1}, \\ &= H \left\{\sigma^2 (1-\eta^2)(M+\tau N) - (1-\eta)^2 \beta\beta^\top \right\} H, \\ &= (1-\eta)^2 H \left\{ \sigma^2 \frac{1+\eta}{1-\eta} (M + \tau N) - \beta\beta^\top \right\} H.\end{aligned}$$

Applying Lemma 4.1 and assuming condition (14) to be satisfied, we see Δ_2 is positive definite if and only if

$$\beta^\top (M+\tau N)^{-1} \beta \le \sigma^2 \frac{1+\eta}{1-\eta}, \quad 0 < \eta < 1.$$

Now we may state the following theorem.

Theorem 4.1. *Consider the two estimators $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(0)}$ of β. Let $W = \frac{1+\eta}{1-\eta}(M + \tau N)$ be a positive definite matrix. Then the biased estimator $\widehat{\beta}_{(2)}(\eta)$ is MSEM superior to $\widehat{\beta}_{(0)}$ if and only if*

$$\beta^\top W^{-1} \beta \le \sigma^2.$$

5. MSEM comparison of $\widehat{\beta}_{(1)}(k)$ and $\widehat{\beta}_{(2)}(\eta)$

Let us now compare the MSEM performance of

$$\begin{aligned}\widehat{\beta}_{(1)}(k) &= (\widetilde{X}^\top\widetilde{X}+kI)^{-1}\widetilde{X}^\top\widetilde{y}\\ &= S_k\widetilde{X}^\top Dy\\ &= A_1y \end{aligned} \quad (17)$$

with

$$\begin{aligned}\widehat{\beta}_{(2)}(\eta) &= (\widetilde{X}^\top\widetilde{X}+I)^{-1}(\widetilde{X}^\top y+\eta\widehat{\beta}_{(0)})\\ &= (\widetilde{X}^\top\widetilde{X})^{-1}(\widetilde{X}^\top\widetilde{X}+I)^{-1}(\widetilde{X}^\top\widetilde{X}+\eta I)\widetilde{X}^\top\widetilde{y}\\ &= UF_\eta\widetilde{X}^\top Dy\\ &= A_2y. \end{aligned} \quad (18)$$

The MSEM of the difference based ridge regression estimator $\widehat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned}\mathrm{MSEM}\{\widehat{\beta}_{(1)}(k)\} &= \mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\}+\mathrm{Bias}\{\widehat{\beta}_{(1)}(k)\}\,\mathrm{Bias}\{\widehat{\beta}_{(1)}(k)\}^\top\\ &= S_k(\sigma^2S+k^2\beta\beta^\top)S_k^\top\\ &= \sigma^2(A_1A_1^\top)+d_1d_1^\top,\end{aligned}$$

where $S_k=(\widetilde{X}^\top\widetilde{X}+kI)^{-1}$ and $d_1=\mathrm{Bias}\{\widehat{\beta}_{(1)}(k)\}=-kS_k\beta$; see [27]. The MSEM in (16) may be written as

$$\mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}=\sigma^2(A_2A_2^\top)+d_2d_2^\top,$$

with $d_2=\mathrm{Bias}\{\widehat{\beta}_{(2)}(\eta)\}=-(1-\eta)(U^{-1}+I)^{-1}\beta$.

Define

$$\Delta_3=\mathrm{MSEM}\{\widehat{\beta}_{(1)}(k)\}-\mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}=\sigma^2(A_1A_1^\top-A_2A_2^\top)+(d_1d_1^\top-d_2d_2^\top). \quad (19)$$

For the following proofs we employ the following lemma.

Lemma 5.1 (*Trenkler and Toutenburg [28]*). *Let $\widetilde{\beta}_{(j)}=A_jy, j=1,2$ be the two linear estimators of β. Suppose the difference $\mathrm{Cov}(\widetilde{\beta}_{(1)})-\mathrm{Cov}(\widetilde{\beta}_{(2)})$ of the covariance matrices of the estimators $\widetilde{\beta}_{(1)}$ and $\widetilde{\beta}_{(2)}$ is positive definite. Then $\mathrm{MSEM}(\widetilde{\beta}_{(1)})-\mathrm{MSEM}(\widetilde{\beta}_{(2)})$ is positive definite if and only if $d_2^\top\{\mathrm{Cov}(\widetilde{\beta}_{(1)})-\mathrm{Cov}(\widetilde{\beta}_{(2)})+d_1d_1^\top\}^{-1}d_2<1$.*

Theorem 5.1. *The sampling variance of $\widehat{\beta}_{(2)}(\eta)$ is smaller than that of $\widehat{\beta}_{(1)}(k)$, if and only if $\lambda_{\min}(G_2^{-1}G_1)>1$, where $\lambda_{\min}$ is the minimum eigenvalue of $G_2^{-1}G_1$ and $G_j=\sigma^2A_jA_j^\top$, $j=1,2$.*

Proof. Consider the difference

$$\begin{aligned}\Delta^* &= \mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\}-\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}\\ &= \sigma^2(A_1A_1^\top-A_2A_2^\top),\\ &= G_1-G_2\end{aligned}$$

with $G_1=(D^\top\widetilde{X}W_kU)^\top=V^\top V$, $W_k=I+kU$ and $G_2=(\widetilde{X}F_\eta^\top U)^\top(\widetilde{X}F_\eta^\top U)$. Since $\mathrm{rank}(V)=p<n-m$, G_1 is a $(p\times p)$ positive definite matrix and G_2 is a symmetric matrix. Hence, a nonsingular matrix O exists such that $O^\top G_1O=I$ and $O^\top G_2O=\Lambda$, with Λ diagonal matrix with diagonal elements roots λ of the polynomial equation $|G_1-\lambda G_2|=0$ (see [16, p. 563] or [25, p. 160]). Thus, we may write $\Delta^*=(O^\top)^{-1}(O^\top G_1O-O^\top G_2O)O^{-1}=(O^\top)^{-1}(\Lambda-I)O^{-1}$ or $O^\top\Delta^*O=\Lambda-I$. If G_1-G_2 is positive definite, then $O^\top G_1O-O^\top G_2O=\Psi-I$ is positive definite. Hence $\lambda_i-1>0$, $i=1,2,\ldots,p$, so we get $\lambda_{\min}(G_2^{-1}G_1)>1$.

Now let $\lambda_{\min}(G_2^{-1}G_1)>1$ hold. Furthermore, with G_2 positive definite and G_1 symmetric, we have $\lambda_{\min}<\frac{\nu^\top G_1\nu}{\nu^\top G_2\nu}<\lambda_{\max}$ for all nonzero $(p\times 1)$ vectors ν, so G_1-G_2 is positive definite; see [23, p. 74]. It is obvious that $\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}-\mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\}$ is positive definite for $0\le\eta\le 1$, $k\ge 0$ if and only if $\lambda_{\min}(G_2^{-1}G_1)>1$. □

Theorem 5.2. *Consider $\widehat{\beta}_{(1)}(k)=A_1y$ and $\widehat{\beta}_{(2)}(\eta)=A_2y$ of β. Suppose that the difference $\mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\}-\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ is positive definite. Then*

$$\Delta_3=\mathrm{MSEM}\{\widehat{\beta}_{(1)}(k)\}-\mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}$$

is positive definite if and only if

$$d_2^\top\{\sigma^2(A_1A_1^\top-A_2A_2^\top)+d_1d_1^\top\}^{-1}d_2<1$$

with $A_1=S_k\widetilde{X}^\top D$, $A_2=UF_\eta\widetilde{X}^\top D$.

Proof. The difference between the MSEMs of $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned}\Delta_3 &= \mathrm{MSEM}\{\widehat{\beta}_{(1)}(k)\} - \mathrm{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}\\ &= \sigma^2(A_1A_1^\top - A_2A_2^\top) + (d_1d_1^\top - d_2d_2^\top)\\ &= \mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\} - \mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\} + (d_1d_1^\top - d_2d_2^\top).\end{aligned}$$

Applying Lemma 5.1 yields the desired result. □

It should be noted that all results reported above are based on the assumption that k and η are non-stochastic. The theoretical results indicate that the $\widehat{\beta}_{(2)}(\eta)$ is not always better than the $\widehat{\beta}_{(1)}(k)$, and vice versa. For practical purposes, we have to replace these unknown parameters by some suitable estimators.

6. The heteroscedasticity and correlated error case

Up to this point, independent errors with equal variance were assumed. The error term might also exhibit autocorrelation. To account for these effects, we extend the results in this section and consider the more general case of heteroscedasticity and autocovariance in the error terms.

Consider now observations $\{y_t, x_t, t_t\}_{t=1}^T$ and the semiparametric partial linear model $y_t = x_t^\top\beta + f(t_t) + \varepsilon_t,\ t = 1, \ldots, T$. Let $\mathsf{E}(\varepsilon\varepsilon^\top|x, t) = \Omega$ not necessarily diagonal. To keep the structure of the errors for later inference, we define an $(n \times n)$ permutation matrix P as in [32]. Consider a permutation:

$$\begin{pmatrix} 1 & t_{(1)}\\ \cdots & \cdots\\ i & t_{(i)}\\ \cdots & \cdots\\ n & t_{(n)}\end{pmatrix}$$

where $i = 1, \ldots, n$ is the index of the ordered nonparametric variable and $t_{(i)} = 1, \ldots, T$ corresponding time index of the observations. Then P is defined for $i, j = 1, \ldots, n$:

$$P_{ij} = \begin{cases}1, & j = t_{(i)}\\ 0, & \text{otherwise.}\end{cases}$$

We can now rewrite the model after reordering and differencing:

$$DPy = DPX\beta + DPf(x) + DP\varepsilon, \qquad \mathsf{E}(\varepsilon\varepsilon^\top|x, t) = \Omega. \tag{20}$$

Then, with $\widetilde{X} = DPX$ and $\widetilde{y} = DPy$ from (20), $\widehat{\beta}_{(0)}$ is given:

$$\widehat{\beta}_{(0)} = (\widetilde{X}^\top\widetilde{X})^{-1}\widetilde{X}^\top\widetilde{y} \tag{21}$$

with

$$\begin{aligned}\mathrm{Cov}(\widehat{\beta}_{(0)}) &= (\widetilde{X}^\top\widetilde{X})^{-1}\widetilde{X}^\top DP\Omega D^\top P^\top\widetilde{X}(\widetilde{X}^\top\widetilde{X})^{-1}\\ &= U\widetilde{X}^\top DP\Omega D^\top P^\top\widetilde{X}U.\end{aligned} \tag{22}$$

We will use a heteroscedasticity and autocovariance consistent estimator described in [22] for the interior matrix of (22), which is in our case:

$$DP\widehat{\Omega D^\top}P^\top = \{\widehat{DP\varepsilon}(\widehat{DP\varepsilon})^\top\} \odot \left\{\sum_{\ell=0}^{\mathcal{L}}\left(1 - \frac{\ell}{\mathcal{L}+1}\right)H^\ell\right\} \tag{23}$$

with $\widehat{DP\varepsilon} = \widetilde{y} - \widetilde{X}\widehat{\beta}_{(0)}$, $\odot$ denoting the elementwise matrix product, $\mathcal{L}$ the maximum lag of nonzero autocorrelation in the errors and H^0 the identity matrix. Let L_ℓ be a matrix with ones on the ℓth diagonal; then $H^\ell,\ \ell = 1, \ldots \mathcal{L}$ are such that:

$$H_{ij}^\ell = \begin{cases}0, & \text{if } \{DP(L_\ell + L_\ell^\top)D^\top P^\top\}_{ij} = 0,\\ 1, & \text{otherwise and } i, j = 1, \ldots, p.\end{cases}$$

Plugging (23) in (22), we obtain a consistent estimator for $\mathrm{Cov}(\widehat{\beta}_{(0)})$; see [31] for details.

Denoting $\widetilde{S} = \widetilde{X}^\top DP\Omega D^\top P^\top\widetilde{X}$, we can write down $\mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\}$ and $\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ in model (20).

$$\mathrm{Cov}\{\widehat{\beta}_{(1)}(k)\} = S_k\widetilde{S}S_k \tag{24}$$

$$\mathrm{Cov}\{\widehat{\beta}_{(2)}(\eta)\} = F_\eta U\widetilde{S}UF_\eta. \tag{25}$$

Using (22) and (25), the difference $\Delta_1 = \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(2)}(\eta)\}$ can be expressed as

$$\begin{aligned}\Delta_1 &= \left(U\widetilde{S}U - F_\eta U\widetilde{S}UF_\eta^\top\right)\\ &= F_\eta\{F_\eta^{-1}U\widetilde{S}U(F_\eta^\top)^{-1} - U\widetilde{S}U\}F_\eta^\top\\ &= (1-\eta^2)(U^{-1}+I)^{-1}\left\{\frac{1}{1+\eta}(U\widetilde{S}+\widetilde{S}U)+U\widetilde{S}U\right\}(U^{-1}+I)^{-1},\end{aligned} \tag{26}$$

with $\tau = \frac{1}{1+\eta} > 0$, $\widetilde{M} = U\widetilde{S}U$, $\widetilde{N} = U\widetilde{S} + \widetilde{S}U$. Since $\widetilde{M}$ is a $(p \times p)$ positive definite matrix and $\widetilde{N}$ is a symmetric matrix, a nonsingular matrix T exists such that $T^\top\widetilde{M}T = I$ and $T^\top\widetilde{N}T = \widetilde{E}$; here $\widetilde{E}$ is a diagonal matrix and its diagonal elements are the roots of the polynomial equation $|\widetilde{M}^{-1}\widetilde{N} - \widetilde{e}I| = 0$ (see [11, pp. 408] and [16, pp. 563]) and we may write (26) as

$$\begin{aligned}\Delta_1 &= (1-\eta^2)H(\widetilde{M}+\tau\widetilde{N})H\\ &= (1-\eta^2)H(T^\top)^{-1}(T^\top\widetilde{M}T+\tau T^\top\widetilde{N}T)T^{-1}H\\ &= (1-\eta^2)H(T^\top)^{-1}(I+\tau\widetilde{E})T^{-1}H,\end{aligned}$$

where $I + \widetilde{\tau E} = \text{diag}(1+\tau\widetilde{e}_{11}, \ldots, 1+\tau\widetilde{e}_{pp})$ and $H = (U^{-1}+I)^{-1}$. Since $\widetilde{N} = U\widetilde{S} + \widetilde{S}U \neq 0$, there is at least one diagonal element of $\widetilde{E}$ that is nonzero.

Let $\widetilde{e}_{ii} < 0$ for at least one i; then positive definiteness of $I + \tau\widetilde{E}$ is guaranteed by

$$0 < \tau < \min_{\widetilde{e}_{ii}<0}\left|\frac{1}{\widetilde{e}_{ii}}\right|. \tag{27}$$

Hence $1+\tau\widetilde{e}_{ii} > 0$ for all $i = 1, \ldots, p$ and therefore $I+\tau\widetilde{E}$ is a positive definite matrix. Consequently, Δ_1 becomes a positive definite matrix, as well. It is now evident that the estimator $\widehat{\beta}_{(2)}(\eta)$ has a smaller variance compared with the estimator $\widehat{\beta}_{(0)}$ if and only if (27) is satisfied.

With

$$\begin{aligned}\Delta_1' &= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(1)}(k)\}\\ &= k^2S_k\left\{\frac{1}{k}(U\widetilde{S}+\widetilde{S}U)+U\widetilde{S}U\right\}S_k\\ &= k^2S_k\left(\frac{1}{k}\widetilde{N}+\widetilde{M}\right)S_k\end{aligned}$$

and analogous argumentation as above obtained for $\widehat{\beta}_{(1)}(k)$:

$$0 < \frac{1}{k} < \min_{\widetilde{e}_{ii}<0}\left|\frac{1}{\widetilde{e}_{ii}}\right|. \tag{28}$$

The next theorem extends the results of Theorem 3.1 in [27] and Theorem 4.1 of Section 4 to the more general case of (20).

Theorem 6.1. *Consider the estimators* $\widehat{\beta}_{(i)}(x),\ i = \{1,2\};\ x = \{k,\eta\}$ *and* $\widehat{\beta}_{(0)}$ *of* β. *Let* $W_1 = \widetilde{M} + \tau\widetilde{N}$, $W_2 = \frac{1+\eta}{1-\eta}(\widetilde{M}+\tau\widetilde{N})$ *be positive definite (alternative: assume that (27) and (28) hold). Then the biased estimator* $\widehat{\beta}_{(i)}(x)$ *is MSEM superior to* $\widehat{\beta}_{(0)}$ *if and only if*

$$\beta^\top W_i^{-1}\beta \leq 1.$$

Proof. Consider the differences

$$\begin{aligned}\Delta_2 &= \text{MSEM}(\widehat{\beta}_{(0)}) - \text{MSEM}\{\widehat{\beta}_{(2)}(\eta)\}\\ &= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(2)}(\eta)\} - \text{Bias}\{\widehat{\beta}_{(2)}(\eta)\}\,\text{Bias}\{\widehat{\beta}_{(2)}(\eta)\}^\top\\ &= F_\eta\{F_\eta^{-1}U\widetilde{S}U(F_\eta^\top)^{-1} - U\widetilde{S}U\}F_\eta^\top - (1-\eta)^2(U^{-1}+I)^{-1}\beta\beta^\top(U^{-1}+I)^{-1}\\ &= (1-\eta)^2H\left\{\frac{1+\eta}{1-\eta}(\widetilde{M}+\tau\widetilde{N}) - \beta\beta^\top\right\}H\\ &= (1-\eta)^2H\left(W_2 - \beta\beta^\top\right)H.\end{aligned}$$

$$\begin{aligned}
\Delta_2' &= \text{MSEM}(\widehat{\beta}_{(0)}) - \text{MSEM}\{\widehat{\beta}_{(1)}(k)\} \\
&= \text{Cov}(\widehat{\beta}_{(0)}) - \text{Cov}\{\widehat{\beta}_{(1)}(k)\} - \text{Bias}\{\widehat{\beta}_{(1)}(k)\}\,\text{Bias}\{\widehat{\beta}_{(1)}(k)\}^\top \\
&= S_k\{k(\widetilde{S}U + U\widetilde{S}) + k^2 U\widetilde{S}U - k^2\beta\beta^\top\}S_k \\
&= k^2 S_k\left(\frac{1}{k}\widetilde{N} + \widetilde{M} - \beta\beta^\top\right)S_k \\
&= k^2 S_k(W_1 - \beta\beta^\top)S_k.
\end{aligned}$$

With Lemma 4.1, the assertion follows. □

Theorem 6.1 gives conditions under which the biased estimator $\widehat{\beta}_{(i)}(x),\ i = \{1, 2\};\ x = \{k, \eta\}$ is superior to $\widehat{\beta}_{(0)}$ in the presence of heteroscedasticity and autocorrelation in the data.

Note that for comparison of the biased estimators Theorem 5.1 can be extended straight forwardly to the general case by exchanging G_1 and G_2 by $\widetilde{G}_1 = \widetilde{A}_1\Omega\widetilde{A}_1^\top$ and $\widetilde{G}_2 = \widetilde{A}_2\Omega\widetilde{A}_2^\top$ correspondingly, with $\widetilde{A}_1 = S_k\widetilde{X}^\top DP,\ \widetilde{A}_2 = UF_\eta\widetilde{X}^\top DP$. Hence, the sampling variance of $\widehat{\beta}_{(2)}(\eta)$ is always smaller than that of $\widehat{\beta}_{(1)}(k)$, if and only if $\lambda_{\min}(\widetilde{G}_2^{-1}\widetilde{G}_1) > 1$, where $\lambda_{\min}$ is the minimum eigenvalue of $\widetilde{G}_2^{-1}\widetilde{G}_1$.

Now, we give a generalized version of Theorem 5.2.

Theorem 6.2. *Consider $\widehat{\beta}_{(1)} = \widetilde{A}_1 y$ and $\widehat{\beta}_{(2)} = \widetilde{A}_2 y$ of β. Suppose that the difference $\text{Cov}\{\widehat{\beta}_{(1)}\} - \text{Cov}\{\widehat{\beta}_{(2)}\}$ is positive definite. Then*

$$\Delta_3 = \text{MSEM}(\widehat{\beta}_{(1)}) - \text{MSEM}(\widehat{\beta}_{(2)})$$

is positive definite if and only if

$$d_2^\top(\widetilde{A}_1\Omega\widetilde{A}_1^\top - \widetilde{A}_2\Omega\widetilde{A}_2^\top + d_1 d_1^\top)^{-1} d_2 < 1.$$

Proof. The difference between the MSEMs of $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(1)}(k)$ is given by

$$\begin{aligned}
\Delta_3 &= \text{MSEM}(\widehat{\beta}_{(1)}) - \text{MSEM}(\widehat{\beta}_{(2)}) \\
&= \widetilde{A}_1\Omega\widetilde{A}_1^\top - \widetilde{A}_2\Omega\widetilde{A}_2^\top + d_1 d_1^\top - d_2 d_2^\top \\
&= \text{Cov}(\widehat{\beta}_{(1)}) - \text{Cov}(\widehat{\beta}_{(2)}) + d_1 d_1^\top - d_2 d_2^\top.
\end{aligned}$$

Applying Lemma 5.1 yields the desired result. □

We note that in order to use the criteria above, one has to estimate the parameters. The estimation of Ω is thereby the most challenging. However, as long as estimator (23) is available, all considered criteria can be evaluated on the real data and can be used for practical purposes.

7. Determinants of electricity demand

The empirical study example is motivated by the importance of explaining variation in electricity consumption. Since electricity is a non-storable good, electricity providers are interested in understanding and hedging demand fluctuations.

Electricity consumption is known to be influenced negatively by the price of electricity and positively by the income of the consumers. As electricity is frequently used for heating and cooling, the effect of the air temperature must also be present. Both heating by low temperatures and cooling by high temperatures result in higher electricity consumption and motivate the use of a nonparametric specification for the temperature effect. Thus we consider the semiparametric regression model defined in (1)

$$y = f(t) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{13} x_{13} + \varepsilon, \tag{29}$$

where y is the log monthly electricity consumption per person (aggregated electricity consumption was divided by population interpolated linearly from quarterly data), t is cumulated average temperature index for the corresponding month taken as average of 20 German cities computed from the data of German weather service (Deutscher Wetterdienst), x_1 is the log GDP per person interpolated linearly from quarterly data, detrended and deseasonalized and x_2 is the log rate of electricity price to the gas price, detrended. The data for 199601-201009 comes from EUROSTAT. Reference prices for electricity were computed as an average of electricity tariffs for consumer groups IND-Ie and HH-Dc, for gas—IND-I3-2 and HH-D3 with reference period 2005S1. Time series of prices were obtained by scaling with electricity price or correspondingly gas price indices. $x_3, x_4, \ldots, x_{13}$ are dummy variables for the monthly effects.

The model in (29) includes both parametric effects and a nonparametric effect. The only nonparametric effect is implied by the temperature variable. From Fig. 1, we can see that the effect of t on y is likely to be nonlinear, while the effects of other variables are roughly linear. The dummy variables enter into the linear part in the specification of the semiparametric regression as well.

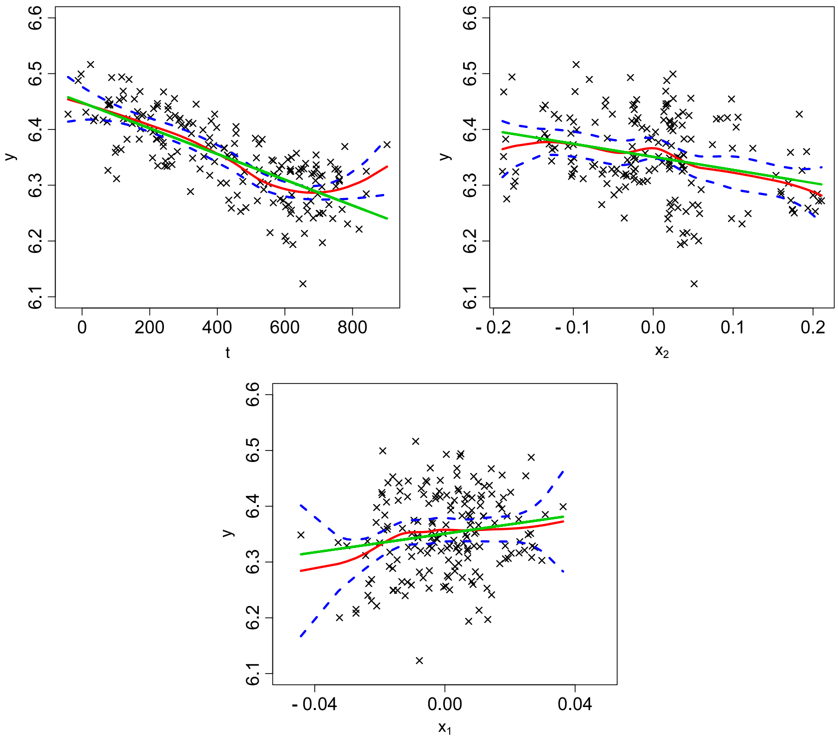

Fig. 1. Plots of individual exp. variables vs. dependent variable, linear fit (green), local polynomial fit (red), 95% confidence bands (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We note that the condition number of $X^\top X$ of these explanatory variables is 20.5, which justifies the use of $\widehat{\beta}_{(1)}(k)$ and $\widehat{\beta}_{(2)}(\eta)$; see [2].

Throughout the paper, we use fifth-order differencing ($m = 5$). Results for other orders of differencing were similar.

The admissible regions for the biasing parameters η and k for MSEM superiority were $\eta \geq 0.923$ and $k \leq 0.0085$. These bounds were determined using the estimated parameters and the inequalities from Theorem 4.1 and Theorem 3.1 in [27], respectively. Under more general assumptions on Ω and resulting heteroscedasticity and autocovariance consistent Newey–West covariance estimator, defined in (23), the admissible region for η (Theorem 6.1 and restriction (27)) was shrinked to $\eta \geq 0.927$. For $\widehat{\beta}_{(1)}(k)$, no admissible values of k were found, since admissible $k \geq 1.57$ of (28) do not satisfy the condition of Theorem 6.1 (see Table 2).

Alternatively, we used a scalar mean squared error (SMSE), defined as the trace of the corresponding MSEM, to compare the estimators. The bounds for k and η can then be calculated only numerically using a grid on [0, 1] for the biasing parameters and determining the regions where SMSEs of the proposed estimators are lower. SMSE superiority of $\widehat{\beta}_{(1)}(k)$ and $\widehat{\beta}_{(2)}(\eta)$ over $\widehat{\beta}_{(0)}$ under general Ω is given for $k \leq 0.0267$ and $\eta \geq 0.384$ compared to $k \leq 0.0123$ and $\eta \geq 0.708$ by standard assumptions; see Fig. 2 which depicts SMSE of the estimators and the corresponding η and k under standard and general assumptions. Thus the SMSE superiority intervals for η and k become even larger in the case of the general form of Ω.

Our computations here are performed with R 2.10.1 and the codes are available on www.quantlet.org.

Results of different estimation procedures can be found in Table 1. We note that regardless of the estimator type, the effect of income is positive and the effect of relative price is negative as expected from an economic perspective, as in [4]. However, the R^2 obtained by difference based methods is higher and SMSE lower for Liu type and ridge difference based estimator. The values of biasing parameters for which conditions of Theorems 5.1 and 5.2 are satisfied are given in Table 3. The superiority of $\widehat{\beta}_{(2)}(\eta)$ over $\widehat{\beta}_{(1)}(k)$ is assured for the zone of values marked by plus.

Returning to our semiparametric specification, we may now remove the estimated parametric effect from the dependent variable and analyze the nonparametric effect. We use a local linear estimator of f to model the nonparametric effect of temperature. The resulting plots are presented in Fig. 3 where we also include the linear effect. We notice that all differencing procedures result in similar estimators of f, regardless of notable differences in the coefficients of the linear part. The estimator of f is consistent with findings e.g. of [4] for US electricity data.

In both specifications, f is different from the linear effect and therefore including temperature as a linear effect is misleading.

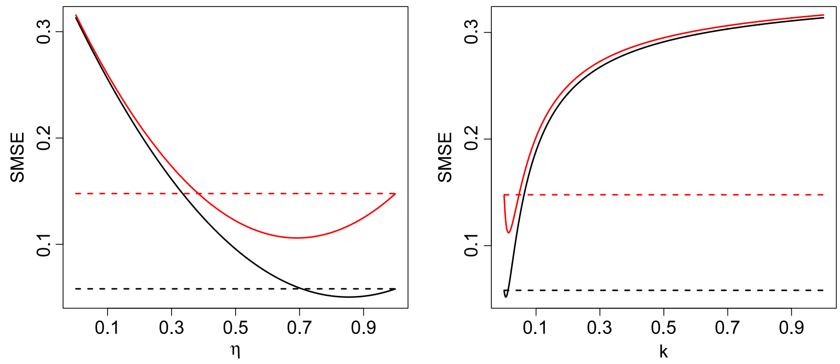

Fig. 2. SMSE of $\widehat{\beta}_{(2)}(\eta)$ in dependence of η (left) and $\widehat{\beta}_{(1)}(k)$ in dependence of k (right) against that of $\widehat{\beta}_{(0)}$ (dashed) under standard assumptions (black) and under generalized assumptions (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of OLS, difference based and Liu type difference based estimations.

	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{(0)}$	$\widehat{\beta}_{(1)}(10^{-3})$	$\widehat{\beta}_{(2)}(0.95)$
x_1	0.634	0.578*	0.550*	0.562*
x_2	−0.152***	−0.160***	−0.158***	−0.161***
x_3	0.030***	0.030*	0.030*	0.030*
x_4	−0.043***	−0.040**	−0.040**	−0.040**
x_5	0.011	0.031	0.031	0.031
x_6	−0.051**	−0.014	−0.013	−0.014
x_7	−0.054*	−0.014	−0.013	−0.014
x_8	−0.079**	−0.065	−0.064	−0.065
x_9	−0.036	−0.037	−0.036	−0.037
x_{10}	−0.052	−0.044	−0.043	−0.044
x_{11}	−0.049	−0.013	−0.012	−0.013
x_{12}	−0.000	0.040	0.040	0.040
x_{13}	−0.001	0.016	0.016	0.016
t	$-13 \cdot 10^{-5}$***	–	–	–
R^2	0.729	0.749	0.749	0.749

* Indicates significance on 10%.
** Indicates significance on 5%.
*** Indicates significance on 1%.

Table 2
Standard errors of the estimators in comparison to Newey–West standard errors for the effects of x_1 (income) and x_2 (relative price).

$\widehat{\Omega}$	$\widehat{\beta}_{(0)}$		$\widehat{\beta}_{(1)}(10^{-3})$		$\widehat{\beta}_{(2)}(0.95)$	
	$\widehat{\sigma}^2 I$	$\widehat{\Omega}_{NW}$	$\widehat{\sigma}^2 I$	$\widehat{\Omega}_{NW}$	$\widehat{\sigma}^2 I$	$\widehat{\Omega}_{NW}$
x_1	0.215	**0.347**	0.209	**0.337**	0.205	**0.215**
x_2	0.034	**0.047**	0.034	**0.047**	0.034	**0.034**
SMSE	0.058	**0.148**	0.056	**0.141**	0.054	**0.058**

8. Conclusion

We proposed a difference based Liu type estimator and a difference based ridge regression estimator for the partial linear semiparametric regression model.

The results show that in case of multicollinearity, the proposed estimator, $\widehat{\beta}_{(2)}(\eta)$ is superior to the difference based estimator $\widehat{\beta}_{(0)}$. We gave bounds on the value of η which ensure the superiority of the proposed estimator. The two biased estimators $\widehat{\beta}_{(2)}(\eta)$ and $\widehat{\beta}_{(1)}(k)$ for different values of η and k can be compared in terms of MSEM with the theoretical results above.

Finally, an application to electricity consumption has been provided to show properties of the proposed estimator based on the mean square error criterion. We could estimate the linear effects of the linear determinants as well as the nonparametric effect f of a cumulated average temperature index.

Thus, the theoretical results obtained allow us to tackle the problem of multicollinearity in real applications of semiparametric models. Moreover, we are able to get estimators of the linear effects with lower standard errors by tuning parameters k and η accordingly.

Table 3
Admissible biasing parameters η and k marked by plus if they satisfy conditions of Theorems 5.1 and 5.2, i.e. $\widehat{\beta}_{(2)}(\eta)$ is superior to $\widehat{\beta}_{(1)}(k)$.

$\eta \cdot 10^2$	$k \cdot 10^4$												
	1	2	3	4	5	6	7	8	9	10	11	12	13
9.23–9.23	−	−	−	−	−	−	−	−	−	−	−	−	−
9.24–9.24	+	−	−	−	−	−	−	−	−	−	−	−	−
9.25–9.25	+	+	−	−	−	−	−	−	−	−	−	−	−
9.26–9.26	+	+	+	−	−	−	−	−	−	−	−	−	−
9.27–9.27	+	+	+	+	−	−	−	−	−	−	−	−	−
9.28–9.28	+	+	+	+	+	−	−	−	−	−	−	−	−
9.29–9.30	+	+	+	+	+	+	−	−	−	−	−	−	−
9.31–9.31	+	+	+	+	+	+	+	−	−	−	−	−	−
9.32–9.32	+	+	+	+	+	+	+	+	−	−	−	−	−
9.34–9.35	+	+	+	+	+	+	+	+	+	−	−	−	−
9.36–9.37	+	+	+	+	+	+	+	+	+	+	−	−	−
9.38–9.39	+	+	+	+	+	+	+	+	+	+	+	−	−
9.40–9.43	+	+	+	+	+	+	+	+	+	+	+	+	−
9.44–9.56	+	+	+	+	+	+	+	+	+	+	+	+	+
9.57–9.61	+	+	+	+	+	+	+	+	+	+	+	+	−
9.62–9.65	+	+	+	+	+	+	+	+	+	+	+	−	−
9.66–9.69	+	+	+	+	+	+	+	+	+	+	−	−	−
9.70–9.72	+	+	+	+	+	+	+	+	+	−	−	−	−
9.73–9.76	+	+	+	+	+	+	+	+	−	−	−	−	−
9.77–9.79	+	+	+	+	+	+	+	−	−	−	−	−	−
9.80–9.82	+	+	+	+	+	+	−	−	−	−	−	−	−
9.83–9.85	+	+	+	+	+	−	−	−	−	−	−	−	−
9.86–9.88	+	+	+	+	−	−	−	−	−	−	−	−	−
9.89–9.91	+	+	+	−	−	−	−	−	−	−	−	−	−
9.92–9.94	+	+	−	−	−	−	−	−	−	−	−	−	−
9.95–9.97	+	−	−	−	−	−	−	−	−	−	−	−	−
9.98–9.99	−	−	−	−	−	−	−	−	−	−	−	−	

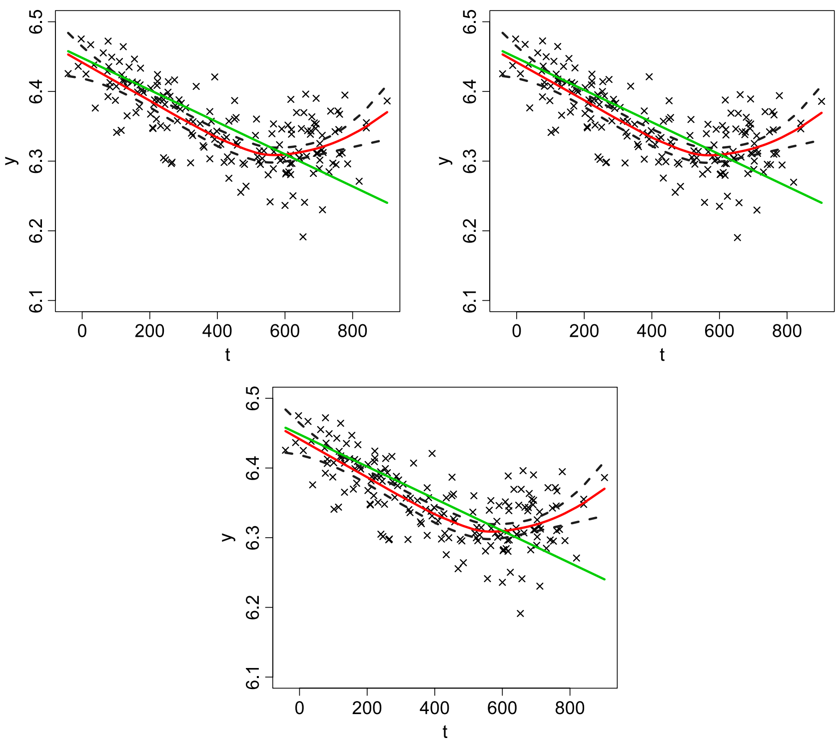

Fig. 3. Estimated f nonlinear effect of t on y via differenced based (left), Liu-type differenced based (right) and difference-based ridge (center) approaches.

References

[1] F. Akdeniz, S. Kaçiranlar, On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE, Communications in Statistics Theory and Methods 24 (7) (1995) 1789–1797.
[2] D. Belsley, E. Kuh, R. Welsch, Regression Diagnostics, Wiley, New York, 1980.

[3] L. Brown, M. Levine, Variance estimation in nonparametric regression via the difference sequence method, Annals of Statistics 35 (2007) 2219–2232.
[4] R.F. Engle, C. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, Journal of American Statistical Association 81 (1986) 310–320.
[5] R. Eubank, Nonparametric Regression and Spline Smoothing, Marcel Dekker, New York, 1999.
[6] R. Eubank, E. Kambour, J. Kim, K. Klipple, C. Reese, M. Schimek, Kernel smoothing in partial linear models, Journal of the Royal Statistical Society Series B 50 (3) (1988) 413–436.
[7] R. Eubank, E. Kambour, J. Kim, K. Klipple, C. Reese, M. Schimek, Estimation in partially linear models, Computational Statistics and Data Analysis 29 (1998) 27–34.
[8] J. Fan, L. Huang, Goodness-of-fit tests for parametric regression models, Journal of American Statistical Association 96 (2001) 640–652.
[9] J. Fan, Y. Wu, Semiparametric estimation of covariance matrices for longitudinal data, Journal of American Statistical Association 103 (2008) 1520–1533.
[10] R. Farebrother, Further results on the mean square error of ridge regression, Journal of the Royal Statistical Society Series B 38 (1976) 248–250.
[11] F. Graybill, Matrices with Applications in Statistics, Duxbury Classic, 1983.
[12] P. Green, C. Jennison, A. Seheult, Analysis of field experiments by least squares smoothing, Journal of the Royal Statistical Society Series B 47 (1985) 299–315.
[13] M. Gruber, Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators, Marcell Dekker, Inc., New York, 1985
[14] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Physika Verlag, Heidelberg, 2000.
[15] W. Härdle, M. Müller, S. Sperlich, A. Werwatz, Nonparametric and Semiparametric Models, Springer Verlag, Heidelberg, 2004.
[16] D. Harville, Matrix Algebra from a Statistician's Perspective, Springer Verlag, New York, 1997.
[17] A. Hoerl, R. Kennard, Ridge regression:biased estimation for orthogonal problems, Technometrics 12 (1970) 55–67.
[18] M. Hubert, P. Wijekoon, Improvement of the Liu estimation in linear regression model, Statistical Papers 47 (3) (2006) 471–479.
[19] K. Klipple, R. Eubank, Difference-based variance estimators for partially linear models, Festschrift in honor of Distinguished Professor Mir Masoom Ali on the occasion of his retirement, 2007, pp. 313–323.
[20] K. Liu, A new class of biased estimate in linear regression, Communications in Statistics Theory and Methods 22 (1993) 393–402.
[21] K. Liu, Using Liu type estimator to combat multicollinearity, Communications in Statistics Theory and Methods 32 (5) (2003) 1009–1020.
[22] W. Newey, K. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55 (3) (1987) 703–708.
[23] C. Rao, Linear Statistical Inference and its Applications, Wiley, New York, 1973.
[24] D. Ruppert, M. Wand, R. Carroll, R. Gill, Semiparametric Regression, Cambridge University Press, 2003.
[25] J. Schott, Matrix Analysis for Statistics, second ed., Wiley Inc., New Jersey, 2005.
[26] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: Proc. Third Berkeley Symp. Math. Statist. Prob., vol. 1, 1956, pp. 197–206.
[27] G. Tabakan, F. Akdeniz, Difference-based ridge estimator of parameters in partial linear model, Statistical Papers 51 (2010) 357–368.
[28] G. Trenkler, H. Toutenburg, Mean square matrix comparisons between two biased estimators-an overview of recent results, Statistical Papers 31 (1990) 165–179.
[29] H. Yang, J. Xu, An alternative stochastic restricted Liu estimator in linear regression, Statistical Papers 50 (2009) 639–647.
[30] A. Yatchew, An elementary estimator of the partial linear model, Economics Letters 57 (1997) 135–143.
[31] A. Yatchew, Differencing methods in nonparametric regression: simple techniques for the applied econometrician, 1999. http://www.economics.utoronto.ca/yatchew/.
[32] A. Yatchew, Semiparametric Regression for the Applied Econometrician, Cambridge University Press, 2003.
[33] J. You, G. Chen, Y. Zhou, Statistical inference of partially linear regression models with heteroscedastic errors, Journal of Multivariate Analysis 98 (8) (2007) 1539–1557.