Febrero-Bande, M., Galeano, P., & González-Manteiga, W. (2019). Estimation, imputation and prediction for the functional linear model with scalar response with responses missing at random. *Computational Statistics & Data Analysis*, 131, pp. 91-103.

DOI:10.1016/j.csda.2018.07.006

Functional principal component regression and functional partial least squares regression: an overview and a comparative study

Manuel Febrero-Bande[1], Pedro Galeano[2], and Wenceslao González-Manteiga[1]

[1]Universidade de Santiago de Compostela

[2]Universidad Carlos III de Madrid

Abstract

Functional data analysis is a field of growing importance in Statistics. In particular, the functional linear model with scalar response is surely the model that has attracted more attention in both theoretical and applied research. Two of the most important methodologies used to estimate the parameters of the functional linear model with scalar response are functional principal component regression and functional partial least squares regression. We provide an overview of estimation methods based on these methodologies and discuss their advantages and disadvantages. We emphasize that the role played by the functional principal components and by the functional partial least squares components that are used in estimation appears to be very important to estimate the functional slope of the model. A functional version of the best subset selection strategy usual in multiple linear regression is also analyzed. Finally, we present an extensive comparative simulation study to compare the performance of all the considered methodologies that may help practitioners in the use of the functional linear model with scalar response.

Keywords: Cross-validation; Eigenfunctions; Eigenvalues; Functional linear model; Functional principal components; Functional partial least squares.

1 Introduction

The collection of functional data is becoming progressively frequent in fields such as chemometrics, climatology, economy, image analysis, linguistics, meteorology, and many other areas. As a consequence, there is a recent interest in methods dealing with functional data that include, among many others, functional principal component analysis, linear models in which the predictors and/or the response are of a functional nature, functional analysis of variance and functional outlier detection. Ramsay and Silverman (2002, 2006), Ferraty and Vieu (2006) and Horváth and Kokoszka (2012) are excellent summaries of methods and case studies for handling functional data from different approaches.

Particularly, the functional linear model with scalar response in which a functional random variable is used to predict a real random variable has attracted considerable attention in the literature. Several procedures have been proposed to estimate the parameters of the model including functional principal component regression and functional partial least squares regression, which are the focus of this paper.

Functional principal component regression has been considered by many authors including Cardot, Ferraty and Sarda (1999, 2003), Hall and Hosseini-Nasab (2006), Cai and Hall (2006), Hall and Horowitz (2007), Cardot, Mas and Sarda (2007), and Ferraty, González-Manteiga, Martínez-Calvo, and Vieu (2012), among others. To obtain an estimate of the functional slope of the model, the standard functional principal component regression estimation method regresses the response on the principal component scores linked with the largest eigenvalues of the functional predictor covariance operator. However, the functional slope estimate obtained in this way may have a large variability even for large sample sizes. To achieve more stable slope estimates, Hall and Horowitz (2007) and Ferraty et al. (2012) have proposed, under different perspectives, a regularized functional principal component regression estimation method based on introducing a small perturbation in the eigenvalues of the functional predictor covariance operator. This regularized estimator can be seen as a member of the family of estimators proposed by Cardot, Mas and Sarda (2007).

On the other hand, functional partial least squares regression was proposed by Preda and Saporta (2005) and posteriorly analyzed in Reiss and Ogden (2007), Aguilera, Escabias, Preda and Saporta (2010) and Delaigle and Hall (2012), among others. The functional partial least squares regression method iteratively produces a sequence of orthogonal functions, as the functional principal components are, but with maximum predictive performance. Then, to obtain a functional slope estimate, the response is regressed on the projections of the predictor on the sequence of orthogonal functions previously produced.

The purpose of this paper is twofold. The first aim is to provide with an overview of estimation methods based on functional principal component regression and functional partial least squares regression and to discuss their advantages and disadvantages. In particular, we focus on the role played by the functional principal components and the functional partial least squares components that are used in the considered estimation methods. We point out that the selection of components used in these methodologies appears to be a major task in real data analysis if we want to obtain sensible estimates of the functional slope of the model. Particularly, the undertaken analysis leads to an estimator based in functional principal components regression that takes into account the response when choosing the scores used in estimation. This estimator can be seen as a functional version of the best subset selection strategy frequently used in multiple linear regression. The advantages and disadvantages of this estimator are also discussed. See Cardot and Sarda (2011) and Mas and Pumo (2011) for another surveys on functional linear regression, although from a different perspective.

The second purpose of this paper is to compare the performance of all the considered methodologies by means of an extensive simulation study to provide advice to practitioners on the use of the functional linear model with scalar response for real-world applications. The comparison is performed in terms of the mean square error of the functional slope estimate as well as the mean square prediction error of the fitted model. It is important to note that a thorough understanding of available methods for model estimation and their challenges can be of broad interest because there is no existing standard approach to estimating the model in practical situations. The methods and results presented in this paper would permit practitioners to handle real-world functional data properly.

The rest of this paper is structured as follows. Section 2 briefly presents the functional linear model with scalar response. Section 3 introduces functional principal component regression that leads to three different estimation procedures. Section 4 introduces functional partial least squares regression that leads to another estimation procedure. Section 5 presents a large simulation study to evaluate the performance of the four analyzed estimation procedures. Finally, Section 6 concludes.

2 The functional linear model with scalar response

The functional linear model with scalar response establishes a linear relationship between a functional random variable and a real random variable. Firstly, we present the functional framework under which the functional random variable is defined and, then, we introduce the real random variable and the model.

Let $L^2(T)$ be the separable Hilbert space of functions η defined on the closed interval $T = [a, b] \subset \mathbb{R}$ satisfying $\int_T \eta^2(s)\, ds < \infty$ and let $\langle \eta, \kappa \rangle = \int_T \eta(s)\, \kappa(s)\, ds$ be the usual inner product of functions η and κ defined on $L^2(T)$. The inner product induces the $L^2(T)$ norm given by $\|\eta\|_2 = \langle \eta, \eta \rangle^{1/2}$, for all $\eta \in L^2(T)$. Let $\chi \in L^2(T)$ be a square integrable functional random variable, i.e., $E\left[\|\chi\|^2\right] < \infty$ and such that $\chi(t)$ is the value of the function at point $t \in T$. Thus, the functional random variable χ has a mean function, denoted by μ_χ, such that $\mu_\chi(t) = E[\chi(t)]$, for all $t \in T$, and a positive definite covariance function, denoted by c_χ, such that:

$$c_\chi(s, t) = Cov[\chi(s), \chi(t)] = E[(\chi(s) - \mu_\chi(s))(\chi(t) - \mu_\chi(t))], \tag{1}$$

for all $s, t \in T$. The covariance function c_χ in (1) allows the covariance operator of χ, denoted by Γ_χ, to be defined as:

$$\Gamma_\chi(\eta)(t) = \int_T c_\chi(s, t)\, \eta(s)\, ds,$$

for all $t \in T$. Note that the covariance operator Γ_χ can be written in terms of the inner product as $\Gamma_\chi(\eta) = E[\langle \chi - \mu_\chi, \eta \rangle (\chi - \mu_\chi)]$. In particular, as Γ_χ is positive definite, there exists a sequence of positive eigenvalues of Γ_χ, denoted by $a_1 > a_2 > \cdots > 0$, and a set of orthonormal eigenfunctions of Γ_χ, denoted by $\psi_1, \psi_2, \ldots$ such that $\Gamma_\chi(\psi_k) = a_k \psi_k$, for $k = 1, 2, \ldots$ On the other hand, let y be a real random variable defined on the same probability space that χ. We assume that y has a mean, denoted by m_y, such that $m_y = E[y]$, and a positive variance, denoted by σ_y^2, such that $\sigma_y^2 = E\left[(y - m_y)^2\right]$.

The functional linear model with scalar response establishes that the relationship between the functional random variable χ and the real random variable y is given by:

$$y = m_y + \langle \chi - \mu_\chi, \beta \rangle + e = m_y + \int_T (\chi(t) - \mu_\chi(t)) \beta(t) + e, \quad (2)$$

where β, the functional slope of the model, is a square integrable function, i.e., $\|\beta\|^2 < \infty$, and e is an error real random variable with mean 0, finite variance σ_e^2, and independent of χ. Consequently, the model in (2) assumes that the conditional mean and variance of y given χ are given by $E[y|\chi] = m_y + \langle \chi - \mu_\chi, \beta \rangle$ and $Var[y|\chi] = \sigma_e^2$, respectively.

In the following, we assume that we observe $\{(\chi_i, y_i), \ i = 1, \ldots, n\}$, a sample of independent and identically distributed random variables drawn from the pair (χ, y). The goal is to estimate the parameters of the model in (2), specifically the functional slope β, from the observed sample. Next, Sections 3 and 4 summarize and analyze estimation methods based on functional principal component and functional partial least squares regression.

3 Estimation through functional principal components

This section presents methods for estimating the parameters of the functional linear model with scalar response based on functional principal components.

3.1 Functional principal component regression

The eigenfunctions of the predictor covariance operator Γ_χ, $\psi_1, \psi_2, \ldots$, form a complete orthonormal basis in $L^2(T)$ that allows the Karhunen-Loève expansion of χ to be written in terms of the elements of the basis as:

$$\chi = \mu_\chi + \sum_{k=1}^{\infty} s_k \psi_k, \quad (3)$$

where $s_k = \langle \chi - \mu_\chi, \psi_k \rangle$, for $k = 1, 2 \ldots$, are called the functional principal component scores. These are uncorrelated random variables with mean 0 and variance a_k. Similarly, the functional slope β in (2) can be written in terms of $\psi_1, \psi_2, \ldots$ as:

$$\beta = \sum_{k=1}^{\infty} b_k \psi_k, \quad (4)$$

where $b_k = \langle \beta, \psi_k \rangle$, for $k = 1, 2 \ldots$ Thus, the functional linear model with scalar response in (2) can be rewritten using the expansions in (3) and (4) as:

$$y = m_y + \sum_{k=1}^{\infty} b_k s_k + e, \tag{5}$$

in which the scalar response y is written as an infinite linear combination of $s_1, s_2, \ldots$ From (5) and taking into account that β is square integrable and thus $\sum_{k=1}^{\infty} b_k^2 < \infty$, it is not difficult to see that the coefficients b_k are given by:

$$b_k = \frac{c_{y,s_k}}{a_k}, \tag{6}$$

(see, Lemma 8.1 in Horváth and Kokoszka, 2012), where $c_{y,s_k} = Cov\left[y, s_k\right]$, for $k = 1, 2 \ldots$, that allows β to be written in terms of c_{y,s_k} and the pairs (a_k, ψ_k), for $k = 1, 2 \ldots$, as:

$$\beta = \sum_{k=1}^{\infty} \frac{c_{y,s_k}}{a_k} \psi_k. \tag{7}$$

The quality of the functional predictor χ to explain linearly the real response y can be measured in terms of the coefficient of determination, denoted by R^2, and defined as the proportion of the scalar response variance explained by the functional predictor, i.e.:

$$R^2 = \frac{Var\left[E\left[y|\chi\right]\right]}{\sigma_y^2}. \tag{8}$$

Now, from (5) and (6), it is possible to show that:

$$Var\left[E\left[y|\chi\right]\right] = \sum_{k=1}^{\infty} a_k b_k^2 = \sum_{k=1}^{\infty} \frac{c_{y,s_k}^2}{a_k},$$

which leads to,

$$R^2 = \frac{1}{\sigma_y^2} \sum_{k=1}^{\infty} \frac{c_{y,s_k}^2}{a_k} = \sum_{k=1}^{\infty} r_{y,s_k}^2, \tag{9}$$

where $r_{y,s_k} = Cor\left[y, s_k\right]$, for $k = 1, 2 \ldots$ Consequently, the quality of χ to explain linearly y is determined by the linear relationships between the real response y and the principal component scores $s_1, s_2, \ldots$

3.2 Standard functional principal component regression estimation

Given $\{(\chi_i, y_i),\ i = 1, \ldots, n\}$, a sample of independent and identically distributed random variables drawn from the pair (χ, y), the standard functional principal component regression estimation method consists on truncating the infinite sum in (7) after k_n terms, where $1 \leq k_n \leq k_{\max}$ is a certain threshold, and on estimating the unknown quantities in the first k_n terms in (7) with their sample counterparts. Thus, μ_χ is estimated with the sample mean of $\chi_1, \ldots, \chi_n$, given by:

$$\widehat{\mu}_\chi = \frac{1}{n} \sum_{i=1}^{n} \chi_i,$$

Γ_χ is estimated with the sample covariance operator of $\chi_1, \ldots, \chi_n$, given by:

$$\widehat{\Gamma}_\chi (\eta) = \frac{1}{n} \sum_{i=1}^{n} \langle \chi_i - \widehat{\mu}_\chi, \eta \rangle (\chi_i - \widehat{\mu}_\chi), \tag{10}$$

for all $\eta \in L^2(T)$, $\psi_1, \ldots, \psi_{k_n}$ and $a_1, \ldots, a_{k_n}$ are estimated with the eigenfunctions and eigenvalues of $\widehat{\Gamma}_\chi$ in (10), denoted by $\widehat{\psi}_1, \ldots, \widehat{\psi}_{k_n}$ and $\widehat{a}_1, \ldots, \widehat{a}_{k_n}$, respectively, and $s_1, \ldots, s_{k_n}$ are replaced with the sample scores given by $\widehat{s}_k = (\widehat{s}_{1k}, \ldots, \widehat{s}_{nk})'$, where $\widehat{s}_{ik} = \left\langle \chi_i - \widehat{\mu}_\chi, \widehat{\psi}_k \right\rangle$, for $i = 1, \ldots, n$ and $k = 1, 2, \ldots, k_n$. Consequently, the standard functional principal component regression estimate of β is given by:

$$\widehat{\beta}_S = \sum_{k=1}^{k_n} \widehat{b}_{k,S} \widehat{\psi}_k = \sum_{k=1}^{k_n} \frac{\widehat{c}_{y,s_k}}{\widehat{a}_k} \widehat{\psi}_k, \tag{11}$$

where $\widehat{c}_{y,s_k}$ is the sample covariance between the responses and the k-th sample scores, i.e.:

$$\widehat{c}_{y,s_k} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{m}_y) \widehat{s}_{ik},$$

and $\widehat{m}_y$ is the sample mean of $y_1, \ldots, y_n$.

Once β has been estimated through (11), fitted values of the responses and residuals can be obtained. First, if $Y = (y_1, \ldots, y_n)'$ is the vector of responses, then, the vector of fitted responses is $\widehat{Y}_S = (\widehat{y}_{S,1}, \ldots, \widehat{y}_{S,n})'$, where $\widehat{y}_{S,i} = \widehat{m}_y - \left\langle \chi_i - \widehat{\mu}_\chi, \widehat{\beta}_S \right\rangle$, that, in matrix form, is given by $\widehat{Y}_S = H_S Y$, where H_S is the $n \times n$ hat matrix given by:

$$H_S = \frac{1}{n}\left(1_n 1_n' + \sum_{k=1}^{k_n} \frac{\widehat{s}_k \widehat{s}_k'}{\widehat{a}_k}\right),$$

and 1_n is the $n \times 1$ column vector of ones. Second, the vector of residuals is given by:

$$\widehat{e}_S = Y - \widehat{Y}_S = (I_n - H_S)\, Y,$$

where I_n is the $n \times n$ identity matrix, that allows the error variance σ_e^2 to be estimated through the functional residual variance, given by:

$$\widehat{\sigma}_{e,S}^2 = \frac{\widehat{e}_S' \widehat{e}_S}{n - k_n - 1}. \tag{12}$$

Finally, the coefficient of determination in (8) can be estimated with the sample coefficient of determination:

$$\widehat{R}_S^2 = \sum_{k=1}^{k_n} \widehat{a}_k \widehat{b}_{k,S}^2 = \sum_{k=1}^{k_n} \frac{\widehat{c}_{y,s_k}^2}{\widehat{a}_k} = \sum_{k=1}^{k_n} \widehat{r}_{y,s_k}^2$$

where $\widehat{r}_{y,s_k}$ is the sample correlation between the responses and the k-th sample scores.

The finite sample behavior of the estimator $\widehat{\beta}_S$ in (11) can be analyzed through the conditional mean square error of $\widehat{\beta}_S$ as well as the mean square prediction error of the fitted model. First, the conditional mean square error of $\widehat{\beta}_S$ was derived by Hall and Hosseini-Nasab (2006) and Ferraty et al. (2012), and is given by:

$$E\left[\left\|\beta - \widehat{\beta}_S\right\|^2 |\chi_1, \ldots, \chi_n\right] = \frac{\sigma_e^2}{n} \sum_{k=1}^{k_n} \frac{1}{\widehat{a}_k} + \left\|\widehat{R}_{k_n}\right\|^2, \tag{13}$$

where $\widehat{R}_{k_n} = \sum_{k=k_n+1}^{\infty} \left\langle \beta, \widehat{\psi}_k \right\rangle \widehat{\psi}_k$. Additionally, Theorem 5 in Hall and Hosseini-Nasab (2006) gives conditions under which the ratio between $MSE\left(\widehat{\beta}_S\right)$ and $\frac{\sigma_e^2}{n} \sum_{k=1}^{k_n} \frac{1}{a_k} + \|R_{k_n}\|^2$, where $R_{k_n} = \sum_{k=k_n+1}^{\infty} \langle \beta, \psi_k \rangle \psi_k$, converges to 1 when $n \to \infty$. Note that the first term in the right hand side of (13) is the contribution of the variability of $\widehat{\beta}_S$ while the second term is the contribution of the square bias of $\widehat{\beta}_S$. Thus, there is a trade-off between bias and variability of $\widehat{\beta}_S$ in terms of the threshold k_n. On the one hand, $\widehat{\beta}_S$ is a biased estimator of β, although the larger k_n, the smaller the bias. On the other hand, the larger k_n, the larger the variability of $\widehat{\beta}_S$. This is because the eigenvalues $\widehat{a}_k$ decrease rapidly towards 0. Consequently, the behavior of $\widehat{\beta}_S$ strongly depends on the threshold k_n because if k_n is small enough, $\widehat{\beta}_S$ may be largely biased, while if k_n is large enough, $\widehat{\beta}_S$ may be too rough. Therefore, k_n controls the smoothness of $\widehat{\beta}_S$. Second, the conditional mean square prediction error of the fitted model is given by:

$$E\left[(y - \widehat{y}_{S,n+1})^2 \,|\chi_1, \ldots, \chi_{n+1}\right] = \sigma_e^2 + \frac{\sigma_e^2}{n}\left(1 + \sum_{k=1}^{k_n} \frac{\widehat{s}_{n+1,k}^2}{\widehat{a}_k}\right) + \left\langle \chi_{n+1} - \widehat{\mu}_\chi, \widehat{R}_{k_n} \right\rangle^2, \qquad (14)$$

where $\widehat{y}_{S,n+1} = \widehat{m}_y - \left\langle \chi_{n+1} - \widehat{\mu}_\chi, \widehat{\beta}_S \right\rangle$ and $\widehat{s}_{n+1,k} = \left\langle \chi_{n+1} - \widehat{\mu}_\chi, \widehat{\psi}_k \right\rangle$ is the k-th principal component score corresponding to χ_{n+1}. The MSPE in (14) is slightly different than the one that appears in Theorem 1 in Ferraty et al (2012) because (14) includes the variability due to the estimation of the means of both the response and the predictor variables, that was not considered in Ferraty et al (2012). The MSPE in (14) also depends on k_n because the larger k_n, the larger the second term in the right hand side of (14) although the smaller the third term.

The asymptotic behavior of the estimator $\widehat{\beta}_S$ in (11) has been analyzed in several papers. Cardot, Ferraty and Sarda (1999) showed that under certain assumptions on the eigenvalues $a_1, a_2, \ldots$ $\widehat{\beta}_S$ converges in probability and almost surely to β. Here, the convergence assumptions holds if k_n is a sequence of integers which converges to ∞ slowly enough as a function of n. Cardot, Ferraty and Sarda (2003) showed the convergence in probability of $\widehat{\beta}_S$ after smoothing it with a B-spline approximation. Hall and Hosseini-Nasab (2006), besides obtaining the conditional mean square error of $\widehat{\beta}_S$, pointed out the importance of the spacings of the eigenvalues $a_1, a_2, \ldots$ in the behavior of (14). Particularly, Hall and Hosseini-Nasab (2006) showed that $\widehat{\beta}_S$ may have problems if, along $a_1, a_2, \ldots$, there are from time to time very closely spaced eigenvalues. Moreover, these authors give a simple sufficient condition on k_n for $\widehat{\beta}_S$, based on spacings of $a_1, a_2, \ldots$, to be consistent for β. Cai and Hall (2006) focus on the prediction of the scalar response rather than in estimating β and showed that if the threshold k_n in $\widehat{\beta}_S$ is optimal to estimate β, then $\widehat{\beta}_S$ will usually be oversmoothed for predicting the response y. Therefore, these authors suggest to undersmooth $\widehat{\beta}_S$ to obtain optimal predictions. Finally, Hall and Horowitz (2007) give optimal convergence rates of $\widehat{\beta}_S$, assuming that k_n tends to ∞ at a certain rate and certain conditions on the spacings of $a_1, a_2, \ldots$ and the coefficients $b_1, b_2, \ldots$

3.3 Regularized functional principal component regression estimation

The regularized functional principal component regression estimation method tries primarily to avoid the effect due to the presence of small eigenvalues in (11) and is given by:

$$\widehat{\beta}_R = \sum_{k=1}^{k_n} \widehat{b}_{k,R} \widehat{\psi}_k = \sum_{k=1}^{k_n} \frac{\widehat{c}_{y,s_k}}{\widehat{a}_k + r_n} \widehat{\psi}_k, \tag{15}$$

where r_n is a certain positive number called regularization parameter.

Once β has been estimated through (15), fitted values of the responses and residuals can be obtained similarly to the case of (11). First, the vector of fitted responses is $\widehat{Y}_R = (\widehat{y}_{R,1}, \ldots, \widehat{y}_{R,n})'$, where $\widehat{y}_{R,i} = \widehat{m}_y - \left\langle \chi_i - \widehat{\mu}_\chi, \widehat{\beta}_R \right\rangle$, that, in matrix form, is given by $\widehat{Y}_R = H_R Y$, where H_R is the $n \times n$ hat matrix:

$$H_R = \frac{1}{n} \left(1_n 1_n' + \sum_{k=1}^{k_n} \frac{\widehat{s}_k \widehat{s}_k'}{\widehat{a}_k + r_n} \right).$$

Second, the vector of residuals is given by $\widehat{e}_R = Y - \widehat{Y}_R = (I_n - H_R) Y$ that allows the error variance σ_e^2 to be estimated through the functional residual variance as in (12). Finally, the coefficient of determination in (8) can be estimated with:

$$\widehat{R}_R^2 = \sum_{k=1}^{k_n} \widehat{a}_k \widehat{b}_{k,R}^2 = \sum_{k=1}^{k_n} \frac{\widehat{a}_k}{(\widehat{a}_k + r_n)^2} \widehat{c}_{y,s_k}^2.$$

Regarding finite sample properties, the conditional mean square error of $\widehat{\beta}_R$ in (15) was derived by Ferraty et al. (2012), and is given by:

$$E\left[\left\| \beta - \widehat{\beta}_R \right\|^2 | \chi_1, \ldots, \chi_n \right] = \frac{\sigma_e^2}{n} \sum_{k=1}^{k_n} \frac{\widehat{a}_k}{(\widehat{a}_k + r_n)^2} + \left\| \sum_{k=1}^{k_n} \frac{r_n}{\widehat{a}_k + r_n} \left\langle \beta, \widehat{\psi}_k \right\rangle \widehat{\psi}_k \right\|^2 + \left\| \widehat{R}_{k_n} \right\|^2, \tag{16}$$

where the first term in the right hand side of (16) is the contribution of the variability of $\widehat{\beta}_R$ while the second and third terms are the contribution of the square bias of $\widehat{\beta}_R$. Thus, $\widehat{\beta}_R$ is a biased estimator of β, as $\widehat{\beta}_S$ is. Moreover, given r_n, the larger k_n, the smaller the bias and the larger the variability, as in the case of $\widehat{\beta}_S$, while given k_n, the larger r_n, the larger the bias and the smaller the variability. The main point here is that an appropriate selection of k_n and r_n may lead (16) to be smaller than (13). The difference between both MSEs is given by:

$$
\begin{aligned}
&E\left[\left\|\beta-\widehat{\beta}_S\right\|^2|\chi_1,\ldots,\chi_n\right]-E\left[\left\|\beta-\widehat{\beta}_R\right\|^2|\chi_1,\ldots,\chi_n\right]=\\
&=\frac{\sigma_e^2}{n}\sum_{k=1}^{k_n}\left(\frac{1}{\widehat{a}_k}-\frac{\widehat{a}_k}{(\widehat{a}_k+r_n)^2}\right)-\left\|\sum_{k=1}^{k_n}\frac{r_n}{\widehat{a}_k+r_n}\left\langle\beta,\widehat{\psi}_k\right\rangle\widehat{\psi}_k\right\|
\end{aligned}
$$

that is positive when σ_e^2 is large and/or n is small enough, see Ferraty et al. (2012) for a deeper analysis of this difference. On the other hand, the conditional mean square prediction error of the fitted model is given by:

$$
\begin{aligned}
&E\left[(y-\widehat{y}_{R,n+1})^2|\chi_1,\ldots,\chi_{n+1}\right]=\\
&=\sigma_e^2+\frac{\sigma_e^2}{n}\left(1+\sum_{k=1}^{k_n}\frac{\widehat{a}_k}{(\widehat{a}_k+r_n)^2}\widehat{s}_{n+1,k}^2\right)+\left\langle\chi_{n+1}-\widehat{\mu}_\chi,\widehat{R}_{k_n}+\sum_{k=1}^{k_n}\frac{\widehat{a}_k}{\widehat{a}_k+r_n}\left\langle\beta,\widehat{\psi}_k\right\rangle\widehat{\psi}_k\right\rangle^2,
\end{aligned}
\tag{17}
$$

that is slightly different than the one that appears in Theorem 3 in Ferraty et al (2012) because (17) includes the variability due to the estimation of the means of both the response and the predictor variables. The behavior of (17) is much more complicated than the one of (14). Comparison of both MSPEs will be done using simulations in Section 5.

The asymptotic behavior of the estimator $\widehat{\beta}_R$ in (15) has been analyzed in Hall and Horowitz (2007) and Ferraty et al. (2012). On the one hand, Hall and Horowitz (2007) proved similar results of $\widehat{\beta}_R$ to those of $\widehat{\beta}_S$ in the same paper, under similar conditions on k_n, the spacings of $a_1, a_2, \ldots$ and the coefficients $b_1, b_2, \ldots$ On the other hand, Ferraty et al. (2012) established the consistency of $\widehat{\beta}_R$ under conditions close to the ones given in Cardot, Ferraty and Sarda (1999, 2003) to establish the consistency of $\widehat{\beta}_S$.

3.4 Best subset functional principal component regression estimation

Both the standard and regularized functional principal component regression estimates of β in (11) and (15), respectively, are based on regressing the response variable on the functional principal component scores linked with the largest eigenvalues of the predictor covariance operator. Thus, both estimates includes the scores without regard to how well they predict the response. However, as shown in (9), the contribution to the fit of each score is measured in terms of the squared correlation between the corresponding score and the response. Particularly, the proportion of total variance of the scalar response y explained by the score s_k is given by r^2_{y,s_k}. Therefore, the scores ordered accordingly with the magnitudes of their associated eigenvalues do not contribute monotonically to explain the total variance of the scalar response. This result suggests that selecting the scores with larger squared correlations with the scalar response may be more appropriate than selecting the scores linked with the largest eigenvalues, an idea hinted by Aguilera, Ocaña and Valderrama (1997) in the problem of predicting continuous-time stochastic processes. Hence, we also consider an estimate of β that takes into account the response when choosing the scores to include in the regression. The resulting estimator can be seen as a functional version of the best subset selection strategy usual in multiple linear regression and it is given by:

$$\widehat{\beta}_{BS} = \sum_{k \in K_n} \widehat{b}_{k,S} \widehat{\psi}_k = \sum_{k \in K_n} \frac{\widehat{c}_{y,s_k}}{\widehat{a}_k} \widehat{\psi}_k, \tag{18}$$

where K_n is a subset of $\{1, 2, \ldots, k_{\max}\}$ that includes those components with largest sample square correlations with the responses sorted in descending order in terms of square correlations r^2_{y,s_k}.

The best subset estimate of β in (18) is constructed using the responses. This makes the computation of the conditional mean square error of $\widehat{\beta}_{BS}$ and the mean square prediction error of the fitted model are intractable. Nevertheless, one might expect that taking a principal component score correlated with the response but with a small associated eigenvalue may increase the variability of (18). To see why, note that as $\sum_{k=1}^{\infty} b_k^2 < \infty$, from (6), this quantity can be rewritten as:

$$\sigma_y^2 \sum_{k=1}^{\infty} \frac{r^2_{y,s_k}}{a_k} < \infty.$$

Therefore, the square correlations r^2_{y,s_k} should converge to 0 at a faster rate than the eigenvalues a_k, which indeed converges very rapidly to 0. Therefore, the functional linear model with scalar response in (2) consider cases in which the large square correlations are mostly associated with large eigenvalues. Consequently, a score linked with a small eigenvalue should not enter in estimation unless it is highly correlated with the scalar response.

Finally, as for $\widehat{\beta}_S$ and $\widehat{\beta}_R$, once that β has been estimated through (18), one can obtain fitted values of the responses and residuals. Expressions of fitted values, hat matrix, residuals, functional residual variance and estimated coefficient of determination for (18) are similar to those for (11) but replacing the quantities associated with the first eigenvalues with the quantities associated with the eigenvalues belonging to the set K_n.

4 Estimation through functional partial least squares

As shown in (5), the functional linear model with scalar response can be written in terms of an infinite linear combination of functional principal component scores. Alternatively, Preda and Saporta (2005) developed functional partial least squares as a method to write the functional linear model in (2) in terms of an infinite linear combination of uncorrelated random variables relevant to predict the real response. The idea is to decompose the functional predictor χ and the real response y in terms of zero mean uncorrelated random variables, denoted by $p_1, p_2, \ldots$, with maximum predictive performance. These are obtained in an iterative fashion as follows:

1. Define $y_0 = y - m_y$ and $\chi_0 = \chi - \mu_\chi$ and let $l = 0$.

2. Let $p_{l+1} = \langle \chi_l, \varphi_{l+1} \rangle$, where φ_{l+1} is a function in $L^2(T)$, such that $c^2_{y_l, p_{l+1}}$ is maximal, with $c_{y_l, p_{l+1}} = Cov\left[y_l, p_{l+1}\right]$, which is given by:

$$\varphi_{l+1} = \frac{\Delta_{y_l, \chi_l}}{\|\Delta_{y_l, \chi_l}\|},$$

   where $\Delta_{y_l, \chi_l} = Cov\left[y_l, \chi_l\right]$.

3. Define $y_{l+1} = y_l - v_{l+1} p_{l+1}$, where v_{l+1} is a constant given by:

$$v_{l+1} = \frac{c_{y_l, p_{l+1}}}{\sigma^2_{p_{l+1}}},$$

   where $\sigma^2_{p_{l+1}} = Var\left(p_{l+1}\right)$, and define $\chi_{l+1} = \chi_l - \varrho_{l+1} p_{l+1}$, where ϱ_{l+1} is a function in $L^2(T)$ given by:

$$\varrho_{l+1} = \frac{\Delta_{\chi_l, p_{l+1}}}{\sigma^2_{p_{l+1}}},$$

   where $\Delta_{\chi_l, p_{l+1}} = Cov\left[\chi_l, p_{l+1}\right]$.

4. Let $l = l + 1$ and back to step 2.

Preda and Saporta (2005) showed that $p_1, p_2, \ldots$ obtained in this way form an orthogonal basis such that:

$$\chi = \mu_\chi + \sum_{l=1}^{\infty} p_l \varrho_l,$$

and,

$$y = m_y + \sum_{l=1}^{\infty} v_l p_l + e, \tag{19}$$

respectively. The functional slope β can be written in terms of $p_1, p_2, \ldots$ as well. For that, note that $p_l = \langle \chi_{l-1}, \varphi_l \rangle = \langle \chi - \mu_\chi, \phi_l \rangle$, where $\phi_1 = \varphi_1$ and $\phi_l = \varphi_l - \langle \varrho_1, \varphi_l \rangle \phi_1 - \cdots - \langle \varrho_{l-1}, \varphi_l \rangle \phi_{l-1}$, for $l \geq 2$. Therefore, (19) leads to:

$$y = m_y + \sum_{l=1}^{\infty} v_l \langle \chi_{l-1}, \varphi_l \rangle + e = m_y + \left\langle \chi - \mu_\chi, \sum_{l=1}^{\infty} v_l \phi_l \right\rangle + e,$$

which, for uniqueness of β, shows that β can be written as:

$$\beta = \sum_{l=1}^{\infty} v_l \phi_l.$$

Also, from (19), and using Proposition 3 in Preda and Saporta (2005), it is not difficult to see that:

$$R^2 = \sum_{l=1}^{\infty} r_{y,p_l}^2,$$

where $r_{y,p_l} = Cor\left[y, p_l\right]$, for $l = 1, 2, \ldots$ Therefore, the proportion of the total variance of the scalar response y explained by the functional predictor χ, is measured in terms of the squared correlations between the scalar response and the functional partial least squares components p_l, for $l = 1, 2, \ldots$ In particular, the proportion of the total variance of y explained by the random variable p_l is given by r_{y,p_l}^2. Moreover, the square correlations r_{y,p_l}^2 are naturally sorted in descending order. Therefore, the variables p_l contribute monotonically to explain the total variance of the scalar response.

Assume now that we have $\{(\chi_i, y_i)\,,\ i = 1, \ldots, n\}$ a sample of independent and identically distributed random variables drawn from the pair (χ, y). In order to estimate the model parameters through functional partial least squares, one has to apply the algorithm given before but replacing the unknown quantities with their sample counterparts, as done in the methods based on functional principal components. Particularly, the estimate of the functional slope β provided by the functional partial least squares components is given by:

$$\widehat{\beta}_{PLS} = \sum_{l=1}^{k_n} \widehat{v}_l \widehat{\phi}_l, \tag{20}$$

where k_n is a threshold, similar to the threshold used in the standard and regularized functional principal component regression estimation methods, and $\widehat{v}_l$ and $\widehat{\varphi}_l$ are the estimates of v_l and φ_l, respectively, given by the procedure.

As mentioned previously, the estimate $\widehat{\beta}_{PLS}$, as the best subset functional principal component regression estimate $\widehat{\beta}_{BS}$, is constructed using the responses. Thus, computation of the conditional mean square error of $\widehat{\beta}_{PLS}$ and the mean square prediction error of the fitted model are intractable. Regarding the asymptotic behavior of the estimator $\widehat{\beta}_{PLS}$ in (20), Preda and Saporta (2005) stated the consistency of $\widehat{\beta}_{PLS}$ to estimate β, although, as noted by Delaigle and Hall (2012), without a proof and without regularity conditions or convergence rates. Delaigle and Hall (2012) presented an alternative formulation of the functional partial least squares problem that allows these authors to show the consistency of $\widehat{\beta}_{PLS}$, as well as rates of convergence, under minor regularity conditions.

Finally, as for $\widehat{\beta}_S$, $\widehat{\beta}_R$ and $\widehat{\beta}_{BS}$, one may obtain fitted values and residuals and, afterwards, estimate σ_e^2 and R^2 using the sample correlations, $\widehat{r}_{y,p_l}$, similarly to the case of (11), so we do not show the details.

5 A performance evaluation of the estimation methods

This section is devoted to compare the finite sample performance of the estimation methods described in Sections 3 and 4 by means of an extensive simulation study. For that, we estimate the mean square error of the functional slope estimates as well as the mean square prediction error of the fitted models. Several configurations of models and sample sizes are considered to provide researchers and practitioners with a summary of the performance of the methods. The Monte Carlo experiments in this section have been carried out by means of various routines written by the authors in R (http://www.r-project.org/).

As described in Section 3, the key point in estimating the parameters of the model in (2) is, for

$\widehat{\beta}_S$ in (11) and $\widehat{\beta}_{BS}$ in (18), the selection of the functional principal components, for $\widehat{\beta}_R$ in (15), the selection of the functional principal components as well as the selection of the regularization parameter, and for $\widehat{\beta}_{PLS}$ in (20), the selection of the functional partial least squares components. In asymptotic theory, the threshold k_n and the regularization parameter r_n have been considered as parameters that depend on the sample size n and which allows the convergence of the estimates of β to the true slope β to be shown. However, in real-world applications, the previous quantities have to be fixed based on the information provided by the data and rules based on asymptotic theory are of limited help. Here, selection of functional principal components, functional partial least squares and regularization parameter is done using predictive cross-validation (PCV). The idea behind PCV is to fit the model parameters with all but one observation, which is posteriorly predicted. Therefore, PCV chooses the model that better predict the discarded observations. In terms of the standard functional principal component regression estimation, PCV chooses the model that minimizes:

$$PCV\,(k) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{m}_{y,-i} - \left\langle \chi_i - \widehat{\mu}_{\chi,-i}, \widehat{\beta}_{S,-i}^{k}\right\rangle\right)^2,$$

for $k = 1, \ldots, k_{\max}$, where $\widehat{m}_{y,-i}$ and $\widehat{\mu}_{\chi,-i}$, respectively, are the sample estimates of m_y and μ_χ excluding the i-th pair (χ_i, y_i), and $\widehat{\beta}_{S,-i}^{k}$ is the estimate (11) with k functional principal components and excluding (χ_i, y_i). The PCV for the other three estimators can be obtained straightforwardly. Note that we do not consider other alternatives, such as generalized cross-validation (GCV), the Akaike Information Criterion (AIC), the corrected Akaike Information Criterion (AICc) or the Bayesian Information Criterion (BIC) that have been used in some papers dealing with the standard functional principal component regression estimator, see Cardot, Ferraty and Sarda (2003), Hall and Hosseini-Nasab (2006), Chiou and Müller (2007) and Febrero-Bande, Galeano and González-Manteiga (2010), among others. This is because, these selection methods depend on the degrees of freedom parameter that is only properly defined for the standard and the regularized functional principal component regression estimation methods. The usual definition of the degrees of freedom parameter in regression (see, Krämer and Sugiyama, 2011) can not be applied for the best subset functional principal component regression and functional partial least squares regression estimation methods because the corresponding hat matrices depends on the responses. The proper definition of the degrees of freedom parameter in these two cases is an open question outside the scope of this paper.

For the simulation study, we analyze 5 different scenarios for the model $y = \langle \chi, \beta \rangle + e$, i.e., both the response and the functional predictor are centered random variables. In all the scenarios, the functional predictor is defined in the interval $T = [0, 1]$ and is given by:

$$\chi = \sum_{k=1}^{\infty} g_k z_k \psi_k$$

where $z_1, z_2, \ldots$ are i.i.d. random variables with mean 0 and variance 1, $\psi_1, \psi_2, \ldots$ are the eigenfunctions of the predictor covariance operator Γ_χ and $g_1, g_2, \ldots$ are such that $g_k^2 = a_k$, for $k = 1, 2, \ldots$, where $a_1, a_2, \ldots$ are the eigenvalues associated to $\psi_1, \psi_2, \ldots$ These quantities, as well as the functional slope β, depend on the scenario considered. Particularly, the random variables $z_1, z_2, \ldots$ are either standard Gaussian or standardized exponential with rate 1. On the other hand, the errors are either Gaussian or centered exponential with variances given by σ_e^2 which is chosen such that $R^2 = 0.75$.

The considered scenarios are the following:

1. We take $\psi_k = \sqrt{2} \sin\left((k - 0.5)\pi t\right)$ and $g_k = 1/\left((k - 0.5)\pi\right)$, for $k = 1, 2, \ldots$ and $\beta(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t)$. This model has been used in Cardot, Ferraty and Sarda (2003) and represents a case in which β can be written exactly in terms of the first three eigenfunctions.

2. We take ψ_k and g_k, for $k = 1, 2, \ldots$ as in scenario 1, and $\beta(t) = \log\left(1.5t^2 + 10\right) + \cos(4\pi t)$. This model has been also used in Cardot, Ferraty and Sarda (2003) and represents a case in which β is an infinite linear combination of the eigenfunctions.

3. We take $\psi_k = \sqrt{2} \cos(k\pi t)$ and $g_k = k^{-1}$, for $k = 1, 2, \ldots$ and $\beta(t) = \pi^2\left(t^2 - 1/3\right)$. This model has been used in Hall and Housseini-Nasab (2006) and represents a case in which β is an infinite linear combination of the eigenfunctions with decreasing weights.

4. We take $\psi_1 = 1$, $\psi_k = \sqrt{2}\cos\left((k-1)\pi t\right)$, for $k = 2, 3, \ldots$, and $g_k = (-1)^{k+1} k^{-1}$, for $k = 1, 2, \ldots$, and $\beta = \sum_{k=1}^{\infty} b_k \psi_k$, where $b_1 = 0.3$ and $b_k = 4(-1)^{k+1} k^{-2}$, for $k = 2, 3, \ldots$ This model has been used in Hall and Horowitz (2007) and represents a case in which β is an

infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues $a_1, a_2, \ldots$ are well spaced.

5. We take ψ as in scenario 4, $g_1 = 1$, $g_k = 0.2\,(-1)^{k+1}\,(1 - 0.0001k)$, for $k = 2, 3, 4$, $g_{5k+j} = 0.2\,(-1)^{5k+j+1}\,\{(5k)^{-1} - 0.0001j\}$, for $k = 1, 2, \ldots$ and $j = 0, 1, \ldots, 4$, and β as in scenario 4. This model has been used in Hall and Horowitz (2007) and represents a case in which β is an infinite linear combination of the eigenfunctions with decreasing weights and the eigenvalues $a_1, a_2, \ldots$ are closely spaced. Hall and Horowitz (2007) pointed out that this is a complex scenario for estimators based on functional principal components.

Figure 1 shows the slope functions used in the 5 scenarios. Particularly, the last row shows the slope used in scenarios 4 and 5. Additionally, Figure 1 also shows the approximations of the slope function using the first five eigenfunctions of the predictor covariance operator. Note that in the first scenario, β can be written exactly as a function of the first three eigenfunctions. In the other scenarios, the number of functions needed to provide with a good approximation of β is larger than in the first scenario. Particularly, in the second scenario, more than 5 eigenfunctions are needed to obtain a good approximation of the corresponding β.

For numerical calculations the infinite series are truncated at $k = 50$. The sample functions are discretized by 100 equidistant points in the interval $T = [0, 1]$. Gaussian errors with mean 0 and variance 0.01 are added to each generated point. Then, the discrete functions are converted to functional observations using a B-spline basis of order 6 with 20 basis functions which seem enough to fit the data well. We considered different orders and number of basis functions but the results were very similar to those that are going to be presented next. The considered sample sizes are $n = 50, 100$ and 200. For each scenario and sample size, we generate 500 sets of $2n$ pairs, $(\chi_1, y_1), \ldots, (\chi_{2n}, y_{2n})$. Then, for each set of $2n$ pairs, we estimate the model parameters by means of the four estimation methods described in Section 3 using the first n pairs of observations. Cross-validation is used to select the adequate model with $k_{max} = 8$. Additionally, to apply the regularized estimator, we consider a grid of possible values for the regularization parameter r_n ranging from 10^{-4} to 10^{-2} in steps of size 10^{-4}. Then, a number of 100 possible values of the regularization parameter is accounted. To compare the behavior of the different methods, we use the mean square error of the estimates of β, given by:

$$\int_T \left(\beta(t) - \widehat{\beta}(t)\right)^2 dt,$$

where $\widehat{\beta}$ can be any of the considered estimators, and the mean square prediction error of the fitted model given by:

$$\frac{1}{n}\sum_{i=n+1}^{2n}(y_i - \widehat{y}_i)^2$$

where $\widehat{y}_{n+1}, \ldots, \widehat{y}_{2n}$ are the predictions of the last n observation of the response variable from the generated data set using the parameter estimates obtained with the first n observations.

Tables 1–5 show the mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models from the 500 pairs, obtained with the four estimation methods described in Sections 3 and 4. Additionally, Tables 6–10 contain the mean, median, standard deviation and mean absolute deviation of the number of components used in estimation. Several comments regarding these tables are in order.

Regarding the mean square error of estimation of β, note that the means are much larger than the medians in almost all the situations. This is because very bad estimates are made for some of the generated data sets. Nevertheless, large errors are reduced when the sample size increases, but even with $n = 200$, means are sometimes still much larger than medians. This is in agreement with the number of components shown in Tables 6–10. Note that the largest mean square errors are associated with the largest number of components selected. These extreme errors make comparison between methods more difficult. Indeed, different results are found in different scenarios. For instance, it is reasonable to conclude that in scenarios 1 and 3, the behavior of the standard and regularized functional principal component regression estimation method is the best one, although, as expected, the regularized estimator es less affected by bad estimates than the standard estimator. Scenario 2 is more complex because the number of components selected by all the methods is, in general, smaller than needed. In this case, if we focus on the mean values, the regularized estimator has the best performance. However, if we focus on the median values, the functional partial least squares estimation method appears to be the best method. Thus, this last method can be largely affected by extreme errors maybe because the inclusion of additional components. In scenario 4, the best method is clearly the regularized estimator. Finally, the behavior of the estimation methods in scenario 5 is close to the corresponding for scenario 2. Particularly, the functional partial least squares has better performance than the other methods in terms of median values, but not in terms of mean values.

Regarding the mean square prediction error, the differences between estimators are very small. This suggest that, even with a bad selection of the number of components to include in estimation, the predictions are not very affected, except in the small sample size. In this situation, a good choice of the number of components appears to slightly improve prediction. This is in accordance with the results derived in Hall and Horowitz (2007).

Finally, regarding the number of components included in estimation, the results appear to confirm our initial expectations taken from first part of Tables 1–5. In general, the number of components considered by the functional partial least squares estimator is smaller than the number considered by the methods based of functional principal components. In other words, for a fixed number of components, FPLS fits closer than FPCR, a result proved by de Jong (1993) and Phatak and de Hoog (2001) in the multivariate framework and already outlined in the functional framework in Preda and Saporta (2005) and Delaigle and Hall (2012). However, note that this is not always an advantage if we want to obtain good estimates of the functional slope, as can be seen in scenarios 4 and 5.

In summary, there is not a dominant method in all the scenarios. Nevertheless, the regularized functional principal component estimator and the functional partial least squares methods appear to have the best performances. Comparing these two estimators, the regularized estimator appears to be more stable than the functional partial least squares estimator but requires to fix the regularization parameter with the cost of adding computational burden. Finally, the best subset estimator does not appear to have an outstanding performance in none of the considered scenarios. This is somehow in agreement with the comments made in Section 3.

6 Conclusions

This paper have provided an overview of estimation methods based on the functional principal component and functional partial least squares techniques to estimate the parameters of the functional linear model with scalar response. In particular, we have illustrated that the role played by the functional principal components and the functional partial least squares components that are used in estimation appears to be very important in estimating the functional slope of the model. However the importance of this selection appears to be less relevant in terms of out of sample prediction. A functional version of the best subset selection strategy usual in multiple linear regression is also analyzed. Nevertheless, the behavior of this estimator in the simulations is not as good as the other alternatives.

Other aspects of the functional linear model with scalar response have been extensively analyzed. For instance, Cardot, Ferraty, Mas and Sarda (2003) and Cardot, Goia and Sarda (2004) developed and analyzed tests for the nullity of the functional slope while García-Portugués, González-Manteiga and Febrero-Bande (2014) proposed a goodness-of-fit test for the null hypothesis of a functional linear model with scalar response. Chiou and Müller (2007) proposed diagnostics for the model via residual processes. Febrero-Bande, Galeano and González-Manteiga (2010) proposed measures of influence for pairs of observations generated from the model. Horváth and Reeder (2012) proposed a method to test if an integral operator that connect two sequences of curves generated by a functional linear model changes during the observation period. Fremdt et al. (2014) analyzed the role played by the number of functional principal components in the Karhunen-Loève expansion of χ and applied their results to derive two inferential procedures for the mean function: a change-point test and a two-sample test. See also Horváth and Rice (2015) for a recent survey on functional data analysis and the two-sample problem using functional principal component analysis.

There are a number of alternative estimation methods of the parameters of the functional linear model with scalar response that are of interest. For instance, methods based on the use of basis functions can be found in Hastie and Mallows (1993), Marx and Eilers (1996), Ramsay and Silverman (2006), Li and Hsing (2007), and Crambes, Kneip and Sarda (2009), among others. Also, Cardot and Johannes (2010) proposed projection estimators which combine dimension reduction and thresholding. Finally, James, Wang and Zhu (2009) introduces a method called FLiRTI which uses variable selection ideas to produce accurate estimates of the functional slope β.

Finally, other interesting models not covered in this paper are linear models in which the response is functional as in James (2002) and Yao, Müller and Wang (2005), generalized functional linear models as in Müller and Stadtmüller (2005), additive functional linear models as in Müller and Yao (2008) and Fan, James and Radchenko (2015) and generalized additive functional models as in Febrero-Bande and González-Manteiga (2013). Also, Yao and Müller (2010) and Horváth and Reeder (2013) considered a functional regression model in which the scalar response depends on the functional predictor in a quadratic way.

Acknowledgments

The first and third author acknowledges financial support from Ministerio de Economía y Competitividad grant MTM2013-41383-P. The second author acknowledges financial support from Ministerio de Economía y Competitividad grant ECO2012-38442.

References

Aguilera, A. M., Escabias, M., Preda, C. and Saporta, G. (2010) Using basis expansions for estimating functional PLS regression applications with chemometric data. Chemometrics and Intelligent Laboratory Systems, 104, 289-305.

Aguilera, A., Ocaña, F. A. and Valderrama, M. J. (1997) An approximated principal component prediction model for continuous-time stochastic processes. Applied Stochastic Models and Data Analysis, 13, 61-72.

Cai, T. T. and Hall, P. (2006) Prediction in functional linear regression. Annals of Statistics, 34, 2159-2179.

Crambes, C., Kneip, A. and Sarda, P. (2009) Smoothing splines estimators for functional linear regression. Annals of Statistics, 37, 35-72.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003) Testing hypotheses in the functional linear model. Scandinavian Journal of Statistics, 30, 241-255.

Cardot, H., Ferraty, F. and Sarda, P. (1999) Functional linear model. Statistics and Probability Letters, 45, 11-22.

Cardot, H., Ferraty, F. and Sarda, P. (2003) Spline estimators for the functional linear model. Statistica Sinica, 13, 571-591.

Cardot, H., Goia, A. and Sarda, P. (2004) Testing for no effects in functional linear regression models, some computational approaches. Communications in Statistics - Simulation and Computation, 33, 179-199.

Cardot, H. and Johannes, J. (2010) Thresholding projection estimators in functional linear models. Journal of Multivariate Analysis, 101, 395-408.

Cardot, H., Mas, A. and Sarda, P. (2007) CLT in functional linear regression models. Probability Theory and Related Fields, 138, 325-361.

Cardot, H. and Sarda, P. (2011) Functional linear regression. In The Oxford Handbook of Functional Data Analysis (F. Ferraty and Y. Romain, eds.) 21–46. Oxford: Oxford Univ. Press.

Chiou, J. M. and Müller, H. G. (2007) Diagnostics for functional regression via residual processes. Computational Statistics and Data Analysis, 51, 4849-4863.

de Jong, S. (1993) PLS fits closer than PCR. Journal of Chemometrics, 7, 551–557.

Delaigle, A. and Hall, P. (2012) Methodology and theory for partial least squares applied to functional data. Annals of Statistis, 40, 322-352.

Fan, Y., James, G. M. and Radchenko, P. (2015) Functional additive regression. Working paper, University of Southern California.

Febrero-Bande, M., Galeano, P. and González-Manteiga, W. (2010) Measures of influence for the functional linear model with scalar response. Journal of Multivariate Analysis, 101, 327-339.

Febrero-Bande, M. and González-Manteiga, W. (2013) Generalized Additive Models for Functional Data. Test, 22, 278-292.

Ferraty, F. and Vieu, P. (2006) Nonparametric functional data analysis. Springer-Verlag, New York.

Ferraty, F., González-Manteiga, W., Martínez-Calvo, A. and Vieu, P. (2012) Presmoothing in functional linear regression. Statistica Sinica, 22, 69-94.

Fremdt, S., Horváth, L., Kokoszka, P. and Steinebach, J. G. (2014) Functional data analysis with increasing number of projections. Journal of Multivariate Analysis, 124, 313-332.

García-Portugués, E., González-Manteiga, W. and Febrero-Bande, M. (2014) A goodness-of-fit test for the functional linear model with scalar response. Journal of Computational and Graphical Statistics, 23, 761-778.

Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis. Journal of the Royal Statistical Society, Series B, 68, 109-126.

Hall, P. and Horowitz, J. L. (2007) Methodology and convergence rates for functional linear regression. Annals of Statistics, 35, 70-91.

Hastie, T. and Mallows, C. (1993) A discussion of "A statistical view of some chemometrics regression tools" by i. E. Frank and J. H. Friedman. Technometrics, 35, 140-143.

Horváth, L. and Kokoszka, P. (2012) Inference for functional data with applications. Springer, New York.

Horváth, L. and Reeder, R. (2012) Detecting changes in functional linear models. Journal of Multivariate Analysis, 111, 310-334.

Horváth, L. and Reeder, R. (2013) A test of significance in functional quadratic regression. Bernoulli, 19, 2120-2151.

Horváth, L. and Rice, G. (2015) An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. Revista Matemática Complutense, 28, 505-548.

James, G. M. (2002) Generalized linear models with functional predictors. Journal of the Royal Statistical Society, Series B 64, 411–432.

James, G., Wang, J. and Zhu, J. (2009) Functional linear regression that's interpretable. Annals of Statistics, 37, 2083-2108.

Krämer, N. and Sugiyama, M. (2011) The Degrees of Freedom of Partial Least Squares Regression. Journal of the American Statistical Association, 106, 697-705.

Li, Y. and Hsing, T. (2007) On rates of convergence in functional linear regression. Journal of Multivariate Analysis, 98, 1782-1804.

Marx, B. D. and Eilers, P. H. (1996) Generalized linear regression on sampled signals with penalized likelihood. In: Forcina, A., Marchetti, G. M., Hatzinger, R., Galmacci, G. (Eds.), Statistical Modelling. Proceedings of the Eleventh International Workshop On Statistical Modelling, Orvietto.

Mas, A. and Pumo, B. (2011) Linear processes for functional data. In The Oxford Handbook of Functional Data Analysis (F. Ferraty and Y. Romain, eds.) 47–71. Oxford: Oxford Univ. Press.

Müller, H-G. and Yao, F. (2008) Functional additive models. Journal of the American Statistical Association, 103, 426-437.

Müller, H-G. and Stadtmüller, U. (2005) Generalized functional linear models. Annals of statistics, 33, 774-805.

Phatak, A. and De Hoog, F. (2001) PLSR, Lanczos, and conjugate gradients. CSIRO Mathematical & Information Sciences, Report No. CMIS 01/122, Canberra.

Preda, C. and Saporta, G. (2005) PLS regression on a stochastic process. Computational statistics and data analysis, 48, 149-158.

Ramsay, J. O. and Silverman, B. W. (2002) Applied Functional Data Analysis. Springer-Verlag, New York.

Ramsay, J. O. and Silverman, B. W. (2006) Functional data analysis, 2nd edition. Springer-Verlag, New York.

Reiss, P. T. and Ogden, R. T. (2007) Functional principal component regression and functional partial least squares. Journal of the American Statistical Association, 102, 984-996.

Yao, F. and Müller, H-G. (2010) Functional quadratic regression. Biometrika, 97, 49–64.

Yao, F., Müller, H-G., and Wang, J-L. (2005) Functional linear regression analysis for longitudinal data. Annals of Statistics, 33, 2873-2903.

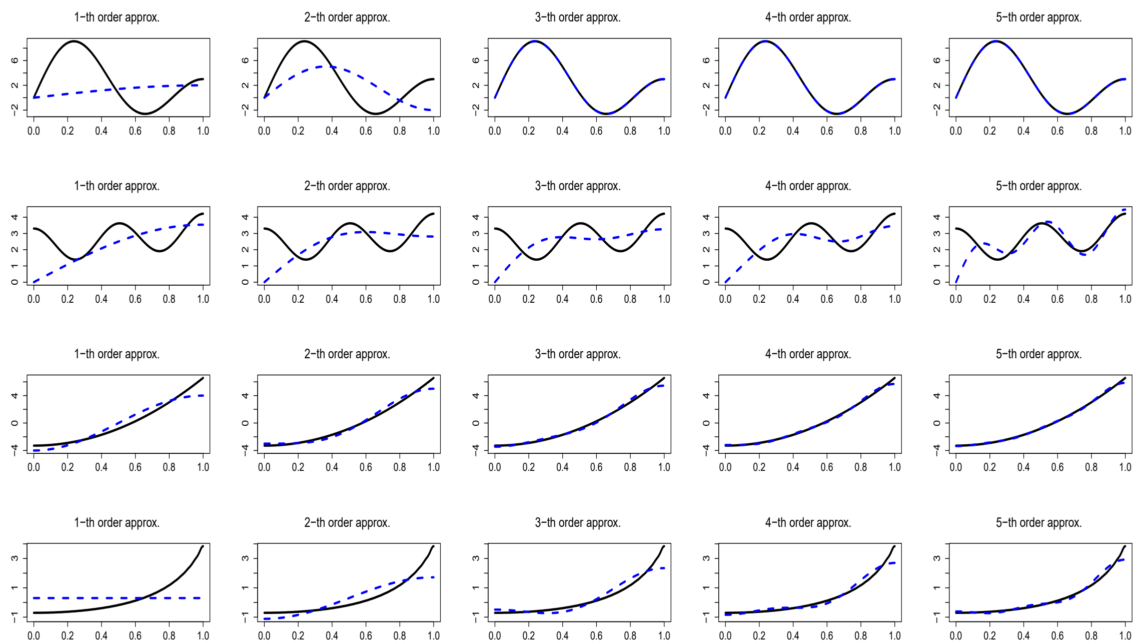

Figure 1: Approximations of the slope function for the five scenarios using the corresponding eigenfunctions. Last row corresponds to scenarios 4 and 5.

Table 1: Mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models for scenario 1.

	n	*Distrib.*	$\widehat{\beta}_S$ *Mean* (*sd*)	$\widehat{\beta}_S$ *Median* (*mad*)	$\widehat{\beta}_R$ *Mean* (*sd*)	$\widehat{\beta}_R$ *Median* (*mad*)	$\widehat{\beta}_{BS}$ *Mean* (*sd*)	$\widehat{\beta}_{BS}$ *Median* (*mad*)	$\widehat{\beta}_{PLS}$ *Mean* (*sd*)	$\widehat{\beta}_{PLS}$ *Median* (*mad*)
MSE	50	(G,G)	10.9350 (17.4411)	3.7792 (3.8559)	8.0407 (10.9323)	4.0410 (3.9542)	16.0573 (21.3070)	6.4107 (7.6911)	15.1115 (54.0051)	6.2377 (3.2917)
		(G,E)	8.5239 (13.8583)	2.9505 (2.5813)	6.7921 (8.7475)	3.5490 (3.1807)	11.3576 (15.2924)	4.0717 (4.2632)	15.0989 (66.5026)	5.3112 (3.5430)
		(E,G)	11.1059 (16.3963)	4.1351 (4.1058)	7.9138 (10.1527)	4.3901 (4.1877)	15.1885 (19.2481)	6.4650 (7.7188)	31.4991 (136.4391)	6.8399 (4.9921)
		(E,E)	9.3384 (16.4213)	3.6072 (3.5574)	6.9065 (8.9136)	4.1937 (3.9711)	14.9923 (21.8997)	5.6575 (7.1170)	43.5654 (196.8475)	5.6691 (4.5662)
	100	(G,G)	4.4080 (7.5004)	1.5291 (1.3652)	3.5100 (4.3662)	1.6075 (1.4497)	6.9786 (8.7883)	2.5420 (2.9450)	17.3113 (91.4501)	2.4769 (1.3499)
		(G,E)	4.5033 (7.1838)	1.4202 (1.3520)	3.7265 (4.4449)	1.6582 (1.6504)	7.3498 (8.8390)	3.3134 (4.0724)	4.0193 (12.2754)	2.3734 (1.2533)
		(E,G)	3.4980 (6.7913)	1.3060 (1.0807)	2.9178 (4.2349)	1.4658 (1.3067)	6.0536 (8.2881)	1.7825 (1.8976)	10.3521 (64.7839)	2.3146 (1.2955)
		(E,E)	4.2333 (7.7906)	1.5190 (1.2681)	3.2584 (4.0781)	1.6729 (1.4648)	5.4734 (8.6664)	1.6175 (1.5909)	6.4269 (30.8827)	2.4319 (1.3419)
	200	(G,G)	2.2981 (4.0647)	0.7264 (0.6739)	1.7671 (2.3276)	0.7609 (0.7301)	3.7188 (4.4173)	1.6584 (2.1458)	7.7156 (43.6583)	1.2045 (0.5208)
		(G,E)	2.0382 (3.5984)	0.7395 (0.6657)	1.6282 (2.0545)	0.8022 (0.7407)	3.0408 (3.9297)	1.1135 (1.2534)	7.5144 (36.2329)	1.0664 (0.5163)
		(E,G)	1.9365 (3.0896)	0.7660 (0.6924)	1.6339 (2.0763)	0.7421 (0.6677)	3.1216 (3.9363)	1.3810 (1.4977)	6.1538 (42.1460)	1.1838 (0.5983)
		(E,E)	1.6470 (2.9465)	0.6945 (0.6097)	1.4721 (1.9918)	0.7120 (0.6299)	2.9920 (3.9108)	0.9838 (1.0945)	7.1104 (35.4659)	1.0288 (0.5829)
MSPE	50	(G,G)	0.8017 (0.1753)	0.7941 (0.1573)	0.7909 (0.1638)	0.7843 (0.1515)	0.8164 (0.1746)	0.8058 (0.1699)	0.8037 (0.1795)	0.7835 (0.1585)
		(G,E)	0.8343 (0.3323)	0.7778 (0.2964)	0.8287 (0.3331)	0.7693 (0.3022)	0.8480 (0.3393)	0.7903 (0.3365)	0.8403 (0.3232)	0.7901 (0.2936)
		(E,G)	0.8134 (0.1911)	0.8013 (0.1749)	0.8022 (0.1825)	0.7802 (0.1811)	0.8292 (0.2006)	0.8092 (0.1864)	0.8368 (0.2326)	0.8088 (0.1935)
		(E,E)	0.7996 (0.2866)	0.7684 (0.2760)	0.7889 (0.2796)	0.7555 (0.2771)	0.8201 (0.3017)	0.7744 (0.2785)	0.8246 (0.2977)	0.7885 (0.2782)
	100	(G,G)	0.7294 (0.1048)	0.7315 (0.1101)	0.7272 (0.1022)	0.7274 (0.0996)	0.7367 (0.1037)	0.7416 (0.0999)	0.7298 (0.1116)	0.7267 (0.1027)
		(G,E)	0.7205 (0.1813)	0.6905 (0.1899)	0.7163 (0.1805)	0.6840 (0.1912)	0.7269 (0.18014)	0.6965 (0.1824)	0.7140 (0.1839)	0.6867 (0.1935)
		(E,G)	0.7274 (0.1030)	0.7262 (0.1062)	0.7257 (0.1007)	0.7240 (0.1032)	0.7342 (0.1028)	0.7327 (0.1083)	0.7266 (0.1031)	0.7311 (0.1027)
		(E,E)	0.7305 (0.1978)	0.7013 (0.1609)	0.7294 (0.1962)	0.7009 (0.1565)	0.7341 (0.1986)	0.7037 (0.1664)	0.7283 (0.1973)	0.7004 (0.1731)
	200	(G,G)	0.7084 (0.0756)	0.6987 (0.0758)	0.7069 (0.0747)	0.6977 (0.0728)	0.7119 (0.0765)	0.6999 (0.0711)	0.7091 (0.0812)	0.6966 (0.0745)
		(G,E)	0.7198 (0.1315)	0.7124 (0.1346)	0.7185 (0.1319)	0.7111 (0.1339)	0.7211 (0.1327)	0.7131 (0.1378)	0.7193 (0.1332)	0.7077 (0.1316)
		(E,G)	0.7151 (0.0700)	0.7156 (0.0653)	0.7143 (0.0705)	0.7134 (0.0671)	0.7174 (0.0708)	0.7153 (0.0678)	0.7144 (0.0766)	0.7103 (0.0641)
		(E,E)	0.7162 (0.1447)	0.7045 (0.1520)	0.7161 (0.1448)	0.7043 (0.1515)	0.7192 (0.1449)	0.7032 (0.1457)	0.7170 (0.1479)	0.7011 (0.1483)

Table 2: Mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models for scenario 2.

			$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
	n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
MSE	50	(G,G)	12.5631 (31.0663)	1.7957 (0.6140)	7.4753 (17.1116)	1.8202 (0.6489)	21.2754 (36.2221)	1.9089 (0.8604)	16.3419 (85.6385)	1.6065 (0.3879)
		(G,E)	10.2913 (27.9715)	1.6519 (0.3709)	6.1865 (15.4516)	1.6307 (0.3711)	17.1814 (35.1586)	1.6649 (0.3940)	28.2881 (200.4855)	1.5755 (0.2668)
		(E,G)	7.2109 (16.9109)	1.6492 (0.3539)	5.9528 (14.1241)	1.7040 (0.5161)	10.8468 (20.6223)	1.6631 (0.3709)	20.4145 (106.3146)	1.5415 (0.2574)
		(E,E)	8.7419 (22.3767)	1.6641 (0.3122)	6.7485 (15.4095)	1.7252 (0.5258)	14.4438 (27.4198)	1.6968 (0.3412)	23.3655 (116.6716)	1.5776 (0.2414)
	100	(G,G)	6.0747 (13.2985)	1.5947 (0.2403)	3.7737 (6.0522)	1.5820 (0.2704)	11.9158 (16.0118)	2.7762 (2.1132)	21.1219 (158.0015)	1.5152 (0.2086)
		(G,E)	4.8944 (11.6491)	1.5486 (0.2004)	3.3874 (6.4207)	1.5697 (0.2929)	9.3568 (15.2482)	1.6116 (0.3277)	9.7837 (41.4416)	1.5030 (0.2000)
		(E,G)	5.6262 (11.5907)	1.5687 (0.2582)	3.9346 (7.0257)	1.5823 (0.3203)	11.2172 (15.8704)	1.7025 (0.5157)	9.4729 (51.0925)	1.5175 (0.2275)
		(E,E)	7.1533 (13.5811)	1.6074 (0.2777)	4.6411 (7.5212)	1.6114 (0.3212)	8.7055 (13.8030)	1.6428 (0.2803)	21.5798 (117.5768)	1.5514 (0.2244)
	200	(G,G)	3.9023 (6.7801)	1.4868 (0.2340)	2.6628 (3.4654)	1.4925 (0.3376)	7.3297 (8.1286)	3.8498 (3.6396)	16.1000 (98.0895)	1.4793 (0.1754)
		(G,E)	3.0686 (3.9270)	1.5188 (0.2836)	2.1591 (1.5141)	1.5052 (0.3184)	5.4543 (6.0808)	2.2056 (1.2992)	7.0585 (37.7174)	1.4605 (0.1721)
		(E,G)	3.6129 (6.2210)	1.5076 (0.2215)	2.4552 (3.3607)	1.4978 (0.2209)	6.4997 (8.0530)	1.7170 (0.6742)	8.0990 (31.4560)	1.4578 (0.1634)
		(E,E)	4.2164 (7.5395)	1.5067 (0.2803)	2.7439 (3.4840)	1.5029 (0.4108)	6.6854 (8.2867)	2.5907 (1.7762)	3.6122 (9.4751)	1.4636 (0.1738)
MSPE	50	(G,G)	1.4004 (0.3122)	1.3537 (0.2790)	1.3689 (0.2824)	1.3318 (0.2695)	1.4157 (0.3100)	1.3697 (0.2941)	1.3805 (0.3127)	1.3287 (0.2831)
		(G,E)	1.3791 (0.5818)	1.2219 (0.4337)	1.3551 (0.5769)	1.2081 (0.4279)	1.3955 (0.5733)	1.2711 (0.4160)	1.3674 (0.5847)	1.2067 (0.4477)
		(E,G)	1.3902 (0.3164)	1.3610 (0.3176)	1.3750 (0.3055)	1.3420 (0.3074)	1.3898 (0.3111)	1.3547 (0.2906)	1.3861 (0.3241)	1.3534 (0.3001)
		(E,E)	1.4446 (0.5251)	1.3506 (0.4721)	1.4292 (0.5166)	1.3223 (0.4426)	1.4539 (0.5331)	1.3582 (0.4529)	1.4488 (0.5337)	1.3363 (0.4573)
	100	(G,G)	1.3589 (0.2076)	1.3341 (0.1773)	1.3492 (0.2015)	1.3282 (0.1900)	1.3782 (0.2170)	1.3522 (0.1884)	1.3637 (0.2253)	1.3224 (0.1948)
		(G,E)	1.3582 (0.3818)	1.2844 (0.3297)	1.3511 (0.3812)	1.2769 (0.3227)	1.3722 (0.3815)	1.2948 (0.3169)	1.3596 (0.3772)	1.2939 (0.3238)
		(E,G)	1.3624 (0.2014)	1.3568 (0.2029)	1.3527 (0.1976)	1.3464 (0.1899)	1.3807 (0.2116)	1.3636 (0.2151)	1.3607 (0.2109)	1.3425 (0.2032)
		(E,E)	1.3978 (0.3988)	1.3265 (0.3916)	1.3811 (0.3910)	1.3253 (0.3716)	1.3918 (0.3903)	1.3265 (0.3916)	1.4016 (0.4095)	1.3241 (0.4020)
	200	(G,G)	1.3240 (0.1370)	1.3276 (0.1241)	1.3182 (0.1332)	1.3171 (0.1260)	1.3344 (0.1364)	1.3368 (0.1299)	1.3276 (0.1456)	1.3277 (0.1408)
		(G,E)	1.3184 (0.2508)	1.3099 (0.2247)	1.3107 (0.2490)	1.3009 (0.2239)	1.3247 (0.2493)	1.3182 (0.2290)	1.3215 (0.2506)	1.3059 (0.2319)
		(E,G)	1.3252 (0.1303)	1.3182 (0.1257)	1.3195 (0.1272)	1.3127 (0.1207)	1.3346 (0.1322)	1.3254 (0.1254)	1.3303 (0.1381)	1.3207 (0.1268)
		(E,E)	1.2961 (0.2477)	1.2737 (0.2400)	1.2865 (0.2452)	1.2687 (0.2383)	1.3049 (0.2478)	1.2837 (0.2476)	1.2911 (0.2451)	1.2669 (0.2216)

Table 3: Mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models for scenario 3.

			$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
	n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
MSE	50	(G, G)	5.5032 (12.8841)	0.9282 (0.7782)	4.2855 (8.2634)	0.9464 (0.8502)	9.4978 (15.7871)	1.1604 (1.0935)	8.2390 (43.1345)	0.8266 (0.5809)
		(G, E)	3.5702 (9.6489)	0.9194 (0.6622)	2.8430 (5.3877)	0.9379 (0.7580)	7.7462 (14.3459)	1.1499 (0.9081)	12.5546 (75.0546)	0.8484 (0.4713)
		(E, G)	4.7239 (13.0502)	0.9359 (0.7151)	3.7919 (9.4054)	0.9794 (0.7998)	7.9973 (16.2793)	1.0397 (0.9131)	21.6167 (86.1103)	0.8518 (0.6277)
		(E, E)	4.1740 (10.3708)	1.0587 (0.8261)	3.5392 (8.2450)	1.0031 (0.8255)	7.3195 (13.2568)	1.1142 (1.0328)	10.1602 (45.7856)	0.8098 (0.5372)
	100	(G, G)	2.2436 (5.0962)	0.6159 (0.4985)	1.7057 (2.8897)	0.6150 (0.5027)	4.2934 (5.9488)	1.2631 (1.4519)	7.9987 (45.9459)	0.7509 (0.3823)
		(G, E)	1.4831 (3.3170)	0.6240 (0.4324)	1.4048 (2.4984)	0.6397 (0.4580)	3.6652 (5.6828)	0.9006 (0.8071)	6.3790 (39.9878)	0.6366 (0.3283)
		(E, G)	1.7267 (3.5561)	0.5619 (0.4125)	1.5514 (2.3973)	0.5593 (0.4268)	3.5538 (5.2723)	0.8759 (0.7956)	12.7350 (62.8530)	0.6526 (0.3661)
		(E, E)	2.5981 (6.2464)	0.6186 (0.5579)	2.0386 (4.1165)	0.6609 (0.6124)	4.4372 (7.2369)	0.9690 (1.0003)	3.7705 (18.3182)	0.6205 (0.3353)
	200	(G, G)	1.2659 (2.5272)	0.3779 (0.2963)	0.9329 (1.3134)	0.3603 (0.2910)	2.3085 (2.9253)	0.9197 (1.0860)	4.1722 (23.4223)	0.4629 (0.1913)
		(G, E)	1.0495 (2.0835)	0.3578 (0.2423)	0.8828 (1.2099)	0.4063 (0.2929)	2.1452 (2.8148)	0.7120 (0.7651)	2.8988 (23.6653)	0.4424 (0.2262)
		(E, G)	1.2517 (2.5531)	0.4266 (0.3317)	1.0930 (1.6394)	0.4244 (0.3476)	2.4753 (3.2175)	0.7574 (0.8431)	3.6152 (16.7311)	0.4799 (0.2602)
		(E, E)	0.9228 (1.6918)	0.3452 (0.2391)	0.7681 (1.0163)	0.3441 (0.2394)	2.1328 (2.8027)	0.7192 (0.7802)	4.3834 (36.9197)	0.4282 (0.2280)
MSPE	50	(G, G)	4.5595 (1.0698)	4.4977 (0.9717)	4.5166 (1.0302)	4.4444 (0.9924)	4.6366 (1.0782)	4.4926 (0.9811)	4.4875 (1.0701)	4.2749 (1.0185)
		(G, E)	4.6008 (1.6102)	4.2411 (1.4468)	4.5659 (1.5881)	4.2185 (1.4212)	4.7361 (1.6754)	4.3000 (1.5234)	4.6331 (1.7109)	4.2299 (1.5678)
		(E, G)	4.5952 (1.1064)	4.5706 (1.0596)	4.5532 (1.0657)	4.5513 (1.0775)	4.6600 (1.1449)	4.5735 (1.1015)	4.6566 (1.2902)	4.5418 (1.0683)
		(E, E)	4.5526 (1.9603)	4.2033 (1.5716)	4.5149 (1.9411)	4.1588 (1.5880)	4.6301 (1.9882)	4.3026 (1.6002)	4.6011 (2.0854)	4.1996 (1.5712)
	100	(G, G)	4.3444 (0.6799)	4.2975 (0.7327)	4.3150 (0.6698)	4.2544 (0.6765)	4.4041 (0.6999)	4.3585 (0.7533)	4.3686 (0.7371)	4.3098 (0.7403)
		(G, E)	4.2336 (1.1937)	4.0773 (1.0401)	4.2184 (1.1868)	4.0628 (1.0322)	4.2946 (1.1960)	4.1292 (0.9992)	4.2367 (1.2163)	4.1332 (0.9442)
		(E, G)	4.3375 (0.5996)	4.3365 (0.5754)	4.3181 (0.5912)	4.2999 (0.5919)	4.3971 (0.6071)	4.3572 (0.6099)	4.3913 (0.7021)	4.3491 (0.6308)
		(E, E)	4.3322 (1.1965)	4.1338 (1.1444)	4.3030 (1.1856)	4.1224 (1.1354)	4.3829 (1.2125)	4.2029 (1.1886)	4.3226 (1.1893)	4.1437 (1.1857)
	200	(G, G)	4.1711 (0.4364)	4.2060 (0.3950)	4.1555 (0.4314)	4.1916 (0.4028)	4.1973 (0.4480)	4.2381 (0.4074)	4.1810 (0.4643)	4.2138 (0.4143)
		(G, E)	4.1407 (0.7775)	4.0725 (0.7313)	4.1304 (0.7795)	4.0808 (0.7171)	4.1849 (0.7833)	4.1470 (0.6926)	4.1367 (0.7814)	4.0805 (0.7114)
		(E, G)	4.1898 (0.4271)	4.1350 (0.4288)	4.1764 (0.4268)	4.1151 (0.3906)	4.2239 (0.4390)	4.1637 (0.4182)	4.2084 (0.4630)	4.1467 (0.4306)
		(E, E)	4.1481 (0.8100)	4.0232 (0.7277)	4.1400 (0.8112)	4.0279 (0.7105)	4.1808 (0.8091)	4.0622 (0.7385)	4.1586 (0.8147)	4.0452 (0.7189)

Table 4: Mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models for scenario 4.

			$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
	n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
MSE	50	(G, G)	0.3497 (0.3707)	0.2156 (0.1390)	0.2683 (0.2086)	0.2152 (0.1259)	0.4341 (0.4116)	0.2612 (0.1901)	0.6045 (2.6028)	0.2258 (0.1238)
		(G, E)	0.3507 (0.3647)	0.2290 (0.1369)	0.2697 (0.2337)	0.1982 (0.1185)	0.4092 (0.3760)	0.2695 (0.1875)	0.5958 (1.8827)	0.2057 (0.1197)
		(E, G)	0.3833 (0.4548)	0.2467 (0.1559)	0.3121 (0.3215)	0.2350 (0.1338)	0.4966 (0.5138)	0.2928 (0.2181)	0.4023 (0.7827)	0.2426 (0.1171)
		(E, E)	0.3790 (0.4226)	0.2313 (0.1532)	0.3050 (0.2754)	0.2318 (0.1325)	0.4950 (0.4894)	0.3151 (0.2250)	1.1835 (4.7920)	0.2495 (0.1213)
	100	(G, G)	0.2107 (0.1752)	0.1511 (0.0834)	0.1632 (0.1168)	0.1292 (0.0662)	0.2675 (0.1994)	0.1923 (0.1360)	0.4393 (1.6481)	0.1537 (0.0709)
		(G, E)	0.1867 (0.1462)	0.1494 (0.0809)	0.1494 (0.0842)	0.1308 (0.0584)	0.2284 (0.1667)	0.1733 (0.1048)	0.4166 (2.0087)	0.1484 (0.0605)
		(E, G)	0.1955 (0.1440)	0.1534 (0.0761)	0.1589 (0.0972)	0.1385 (0.0661)	0.2512 (0.1738)	0.1916 (0.1299)	0.5505 (2.1265)	0.1479 (0.0690)
		(E, E)	0.2050 (0.1947)	0.1375 (0.0773)	0.1614 (0.1220)	0.1305 (0.0700)	0.2439 (0.2141)	0.1641 (0.1084)	0.4210 (1.6700)	0.1484 (0.0659)
	200	(G, G)	0.1292 (0.0930)	0.1000 (0.0446)	0.0980 (0.0536)	0.0843 (0.0370)	0.1586 (0.1043)	0.1190 (0.0678)	0.2850 (1.2996)	0.0826 (0.0315)
		(G, E)	0.1211 (0.0842)	0.0996 (0.0440)	0.0903 (0.0429)	0.0784 (0.0402)	0.1388 (0.0914)	0.1128 (0.0574)	0.1581 (0.5443)	0.0834 (0.0294)
		(E, G)	0.1359 (0.1086)	0.0976 (0.0498)	0.1005 (0.0687)	0.0817 (0.0351)	0.1537 (0.1133)	0.1197 (0.0708)	0.1827 (0.5812)	0.0889 (0.0391)
		(E, E)	0.1249 (0.0954)	0.0946 (0.0435)	0.0939 (0.0468)	0.0784 (0.0339)	0.1445 (0.1038)	0.1115 (0.0648)	0.1693 (0.7341)	0.0847 (0.0320)
MSPE	50	(G, G)	0.1635 (0.0342)	0.1606 (0.0317)	0.1597 (0.0326)	0.1552 (0.0305)	0.1663 (0.0353)	0.1627 (0.0332)	0.1626 (0.0349)	0.1584 (0.0300)
		(G, E)	0.1653 (0.0627)	0.1545 (0.0540)	0.1620 (0.0625)	0.1510 (0.0546)	0.1661 (0.0622)	0.1574 (0.0571)	0.1639 (0.0620)	0.1510 (0.0553)
		(E, G)	0.1576 (0.0326)	0.1568 (0.0320)	0.1550 (0.0319)	0.1544 (0.0330)	0.1600 (0.0343)	0.1561 (0.0325)	0.1565 (0.0317)	0.1533 (0.0306)
		(E, E)	0.1637 (0.0660)	0.1491 (0.0588)	0.1605 (0.0631)	0.1464 (0.0545)	0.1690 (0.0669)	0.1573 (0.0594)	0.1658 (0.0679)	0.1511 (0.0529)
	100	(G, G)	0.1512 (0.0222)	0.1501 (0.0232)	0.1489 (0.0210)	0.1473 (0.0225)	0.1527 (0.0220)	0.1512 (0.0236)	0.1512 (0.0231)	0.1487 (0.0225)
		(G, E)	0.1518 (0.0403)	0.1447 (0.0399)	0.1501 (0.0404)	0.1433 (0.0397)	0.1531 (0.0403)	0.1466 (0.0401)	0.1518 (0.0413)	0.1452 (0.0429)
		(E, G)	0.1523 (0.0224)	0.1514 (0.0237)	0.1506 (0.0221)	0.1515 (0.0221)	0.1535 (0.0229)	0.1534 (0.0229)	0.1528 (0.0245)	0.1529 (0.0236)
		(E, E)	0.1543 (0.0392)	0.1477 (0.0367)	0.1523 (0.0390)	0.1461 (0.0366)	0.1556 (0.0391)	0.1497 (0.0370)	0.1538 (0.0393)	0.1491 (0.0374)
	200	(G, G)	0.1424 (0.0144)	0.1419 (0.0141)	0.1415 (0.0142)	0.1410 (0.0147)	0.1433 (0.0148)	0.1432 (0.0141)	0.1424 (0.0150)	0.1420 (0.0149)
		(G, E)	0.1447 (0.0268)	0.1425 (0.0250)	0.1436 (0.0267)	0.1416 (0.0235)	0.1451 (0.0268)	0.1433 (0.0249)	0.1440 (0.0269)	0.1425 (0.0251)
		(E, G)	0.1461 (0.0155)	0.1475 (0.0159)	0.1447 (0.0152)	0.1456 (0.0152)	0.1467 (0.0157)	0.1479 (0.0161)	0.1455 (0.0161)	0.1451 (0.0157)
		(E, E)	0.1412 (0.0310)	0.1350 (0.0285)	0.1400 (0.0306)	0.1329 (0.0291)	0.1418 (0.0312)	0.1349 (0.0295)	0.1404 (0.0306)	0.1333 (0.0283)

Table 5: Mean, median, standard deviation and mean absolute deviation of the estimated mean square errors of the estimates of β and mean square prediction errors of the fitted models for scenario 5.

	n	*Distrib.*	$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
			Mean (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
MSE	50	(G,G)	0.5727 (1.0130)	0.1269 (0.0620)	0.3365 (0.5083)	0.1464 (0.0876)	0.7629 (1.0728)	0.1574 (0.1089)	1.5850 (6.4757)	0.1288 (0.0577)
		(G,E)	0.5972 (1.0176)	0.1424 (0.0854)	0.3397 (0.5191)	0.1500 (0.0937)	0.7201 (1.1048)	0.1528 (0.0979)	0.5534 (1.7322)	0.1218 (0.0541)
		(E,G)	0.4596 (0.7159)	0.1459 (0.0785)	0.3248 (0.4174)	0.1767 (0.1187)	0.6832 (0.9168)	0.1818 (0.1245)	0.7056 (2.3657)	0.1459 (0.0746)
		(E,E)	0.5864 (1.0092)	0.1423 (0.0880)	0.3912 (0.6076)	0.1493 (0.0976)	0.8356 (1.4358)	0.1582 (0.1156)	1.6909 (8.1471)	0.1372 (0.0706)
	100	(G,G)	0.3046 (0.4629)	0.0960 (0.0370)	0.1917 (0.2038)	0.1073 (0.0507)	0.4015 (0.4962)	0.1202 (0.0770)	0.9971 (4.4376)	0.0909 (0.0285)
		(G,E)	0.2581 (0.3677)	0.0921 (0.0262)	0.1644 (0.1459)	0.1010 (0.0432)	0.3255 (0.4195)	0.1017 (0.0423)	1.2093 (7.5413)	0.0846 (0.0261)
		(E,G)	0.3105 (0.5348)	0.0973 (0.0361)	0.1954 (0.2433)	0.1031 (0.0456)	0.4146 (0.5741)	0.1145 (0.0624)	0.9877 (4.4533)	0.0920 (0.0304)
		(E,E)	0.3997 (0.7414)	0.0974 (0.0400)	0.2331 (0.3922)	0.0982 (0.0396)	0.4566 (0.7202)	0.1038 (0.0558)	0.5010 (2.4823)	0.0908 (0.0270)
	200	(G,G)	0.1481 (0.1901)	0.0779 (0.0161)	0.1112 (0.0868)	0.0791 (0.0237)	0.2074 (0.2410)	0.0824 (0.0271)	0.4806 (2.3055)	0.0707 (0.0132)
		(G,E)	0.1799 (0.1958)	0.0810 (0.0209)	0.1213 (0.0927)	0.0825 (0.0258)	0.2237 (0.2417)	0.0879 (0.0325)	0.3850 (2.6277)	0.0739 (0.0186)
		(E,G)	0.1805 (0.2174)	0.0833 (0.0246)	0.1256 (0.1005)	0.0855 (0.0292)	0.2416 (0.2647)	0.0876 (0.0369)	0.5209 (3.5476)	0.0736 (0.0191)
		(E,E)	0.1734 (0.1928)	0.0802 (0.0222)	0.1208 (0.0989)	0.0828 (0.0309)	0.2027 (0.2089)	0.0860 (0.0311)	0.4177 (2.7155)	0.0701 (0.0142)
MSPE	50	(G,G)	0.0296 (0.0058)	0.0292 (0.0063)	0.0291 (0.0054)	0.0286 (0.0054)	0.0304 (0.0060)	0.0302 (0.0063)	0.0295 (0.0069)	0.0286 (0.0061)
		(G,E)	0.0290 (0.0099)	0.0277 (0.0095)	0.0281 (0.0096)	0.0267 (0.0087)	0.0295 (0.0101)	0.0281 (0.0093)	0.0283 (0.0097)	0.0270 (0.0085)
		(E,G)	0.0297 (0.0065)	0.0291 (0.0064)	0.0293 (0.0061)	0.0287 (0.0063)	0.0302 (0.0065)	0.0291 (0.0063)	0.0295 (0.0066)	0.0287 (0.0065)
		(E,E)	0.0289 (0.0106)	0.0269 (0.0088)	0.0284 (0.0105)	0.0264 (0.0090)	0.0298 (0.0112)	0.0273 (0.0093)	0.0288 (0.0109)	0.0268 (0.0093)
	100	(G,G)	0.0273 (0.0040)	0.0272 (0.0043)	0.0270 (0.0040)	0.0269 (0.0041)	0.0276 (0.0040)	0.0275 (0.0043)	0.0274 (0.0041)	0.0275 (0.0042)
		(G,E)	0.0274 (0.0072)	0.0268 (0.0070)	0.0271 (0.0072)	0.0259 (0.0066)	0.0277 (0.0074)	0.0268 (0.0068)	0.0273 (0.0073)	0.0264 (0.0067)
		(E,G)	0.0270 (0.0042)	0.0270 (0.0042)	0.0267 (0.0041)	0.0265 (0.0042)	0.0272 (0.0042)	0.0271 (0.0042)	0.0269 (0.0043)	0.0267 (0.0042)
		(E,E)	0.0278 (0.0079)	0.0271 (0.0074)	0.0274 (0.0077)	0.0262 (0.0074)	0.0279 (0.0080)	0.0271 (0.0073)	0.0276 (0.0078)	0.0268 (0.0070)
	200	(G,G)	0.0259 (0.0027)	0.0258 (0.0027)	0.0258 (0.0027)	0.0258 (0.0028)	0.0260 (0.0027)	0.0259 (0.0029)	0.0259 (0.0028)	0.0256 (0.0029)
		(G,E)	0.0266 (0.0057)	0.0258 (0.0051)	0.0264 (0.0056)	0.0256 (0.0052)	0.0267 (0.0057)	0.0258 (0.0053)	0.0265 (0.0058)	0.0256 (0.0050)
		(E,G)	0.0266 (0.0026)	0.0266 (0.0026)	0.0264 (0.0025)	0.0264 (0.0027)	0.0267 (0.0026)	0.0267 (0.0026)	0.0265 (0.0026)	0.0266 (0.0028)
		(E,E)	0.0259 (0.0046)	0.0258 (0.0046)	0.0258 (0.0046)	0.0257 (0.0046)	0.0260 (0.0046)	0.0259 (0.0046)	0.0258 (0.0047)	0.0257 (0.0046)

Table 6: Mean, median, standard deviation and mean absolute deviation of the number of components of the estimates of β for scenario 1.

		$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
50	(G, G)	3.97 (1.46)	3 (0)	4.26 (1.65)	3.5 (0.74)	3.91 (1.33)	3 (1.48)	2.86 (0.57)	3 (0)
	(G, E)	3.74 (1.28)	3 (0)	4.28 (1.71)	3 (1.48)	3.55 (1.15)	3 (0)	2.88 (0.60)	3 (0)
	(E, G)	4.17 (1.54)	4 (1.48)	4.47 (1.72)	4 (1.48)	3.8 (1.37)	3 (1.48)	2.89 (0.81)	3 (0)
	(E, E)	3.97 (1.39)	3 (1.48)	4.36 (1.64)	4 (1.48)	3.77 (1.41)	3 (1.48)	2.99 (0.93)	3 (0)
100	(G, G)	3.78 (1.31)	3 (0)	4.06 (1.56)	3 (0)	3.84 (1.13)	3 (0.74)	3.10 (0.67)	3 (0)
	(G, E)	3.88 (1.26)	3 (0)	4.24 (1.61)	4 (1.48)	3.91 (1.10)	4 (1.48)	3 (0.24)	3 (0)
	(E, G)	3.68 (1.09)	3 (0)	3.92 (1.30)	3 (0)	3.67 (0.95)	3 (0)	2.98 (0.57)	3 (0)
	(E, E)	3.84 (1.31)	3 (0)	4.16 (1.58)	3 (0)	3.65 (1.04)	3 (0)	3.02 (0.38)	3 (0)
200	(G, G)	3.80 (1.36)	3 (0)	3.97 (1.55)	3 (0)	3.86 (1.08)	4 (1.48)	3.08 (0.50)	3 (0)
	(G, E)	3.79 (1.26)	3 (0)	3.94 (1.38)	3 (0)	3.79 (1.00)	4 (1.48)	3.11 (0.57)	3 (0)
	(E, G)	3.77 (1.16)	3 (0)	3.90 (1.33)	3 (0)	3.79 (0.94)	4 (1.48)	3.06 (0.49)	3 (0)
	(E, E)	3.66 (1.16)	3 (0)	3.85 (1.39)	3 (0)	3.75 (0.98)	3 (0)	3.1 (0.49)	3 (0)

Table 7: Mean, median, standard deviation and mean absolute deviation of the number of components of the estimates of β for scenario 2.

		$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
50	(G, G)	2.25 (1.79)	1 (0)	2.67 (2.13)	2 (1.48)	1.96 (1.35)	1 (0)	1.44 (0.83)	1 (0)
	(G, E)	2.03 (1.66)	1 (0)	2.43 (2.03)	2 (1.48)	1.78 (1.42)	1 (0)	1.42 (0.96)	1 (0)
	(E, G)	1.88 (1.40)	1 (0)	2.44 (1.99)	1 (0)	1.56 (1.00)	1 (0)	1.38 (0.91)	1 (0)
	(E, E)	1.84 (1.53)	1 (0)	2.49 (2.04)	1 (0)	1.67 (1.23)	1 (0)	1.43 (0.95)	1 (0)
100	(G, G)	2.09 (1.77)	1 (0)	2.64 (2.22)	2 (1.48)	2.13 (1.31)	2 (1.48)	1.43 (0.93)	1 (0)
	(G, E)	1.95 (1.55)	1 (0)	2.60 (2.10)	2 (1.48)	1.89 (1.35)	1 (0)	1.48 (0.85)	1 (0)
	(E, G)	2.05 (1.64)	1 (0)	2.45 (2.03)	2 (1.48)	2.02 (1.47)	1 (0)	1.42 (0.76)	1 (0)
	(E, E)	2.33 (1.82)	1 (0)	2.80 (2.18)	2 (1.48)	1.82 (1.19)	1 (0)	1.58 (1.02)	1 (0)
200	(G, G)	2.36 (1.76)	2 (1.48)	3.00 (2.18)	2 (1.48)	2.36 (1.25)	2 (1.48)	1.65 (1.05)	1 (0)
	(G, E)	2.36 (1.61)	2 (1.48)	2.99 (2.13)	2 (1.48)	2.26 (1.27)	2 (1.48)	1.65 (0.84)	1 (0)
	(E, G)	2.34 (1.74)	2 (1.48)	2.82 (2.18)	2 (1.48)	2.37 (1.44)	2 (1.48)	1.64 (0.84)	1 (0)
	(E, E)	2.50 (1.84)	2 (1.48)	3.33 (2.33)	2 (1.48)	2.44 (1.51)	2 (1.48)	1.60 (0.72)	1 (0)

Table 8: Mean, median, standard deviation and mean absolute deviation of the number of components of the estimates of β for scenario 3.

		$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
50	(G,G)	2.35 (1.79)	2 (1.48)	2.58 (2.01)	2 (1.48)	2.09 (1.58)	1 (0)	1.37 (0.82)	1 (0)
	(G,E)	2.13 (1.57)	2 (1.48)	2.35 (1.81)	2 (1.48)	1.97 (1.38)	1 (0)	1.47 (1.03)	1 (0)
	(E,G)	2.26 (1.66)	2 (1.48)	2.46 (1.83)	2 (1.48)	2.01 (1.47)	1 (0)	1.62 (1.17)	1 (0)
	(E,E)	2.48 (1.73)	2 (1.48)	2.73 (2.06)	2 (1.48)	2.10 (1.48)	2 (1.48)	1.50 (0.99)	1 (0)
100	(G,G)	2.42 (1.56)	2 (1.48)	2.63 (1.76)	2 (1.48)	2.46 (1.32)	2 (1.48)	1.62 (0.89)	2 (1.48)
	(G,E)	2.15 (1.30)	2 (1.48)	2.40 (1.62)	2 (1.48)	2.20 (1.37)	2 (1.48)	1.52 (0.92)	1 (0)
	(E,G)	2.37 (1.35)	2 (1.48)	2.66 (1.68)	2 (1.48)	2.25 (1.22)	2 (1.48)	1.74 (1.07)	2 (1.48)
	(E,E)	2.59 (1.61)	2 (1.48)	2.88 (1.81)	2 (1.48)	2.43 (1.58)	2 (1.48)	1.56 (0.77)	1 (0)
200	(G,G)	2.87 (1.51)	2 (1.48)	3.10 (1.70)	2 (1.48)	2.88 (1.16)	3 (1.48)	1.82 (0.79)	2 (0)
	(G,E)	2.63 (1.36)	2 (0)	2.93 (1.64)	2 (1.48)	2.72 (1.33)	2 (1.48)	1.75 (0.63)	2 (0)
	(E,G)	2.63 (1.49)	2 (1.48)	3.07 (1.89)	2 (1.48)	2.71 (1.34)	2 (1.48)	1.81 (0.80)	2 (0)
	(E,E)	2.74 (1.33)	2 (0)	2.96 (1.54)	2 (1.48)	2.84 (1.30)	3 (1.48)	1.88 (0.76)	2 (0)

Table 9: Mean, median, standard deviation and mean absolute deviation of the number of components of the estimates of β for scenario 4.

		$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
50	(G,G)	3.96 (1.52)	4 (1.48)	4.27 (1.73)	4 (1.48)	3.60 (1.47)	3 (1.48)	2.56 (0.82)	2 (1.48)
	(G,E)	4.24 (1.51)	4 (1.48)	4.55 (1.71)	4 (1.48)	3.80 (1.52)	3 (1.48)	2.66 (0.81)	3 (1.48)
	(E,G)	4.26 (1.53)	4 (1.48)	4.59 (1.70)	4 (1.48)	3.92 (1.51)	4 (1.48)	2.53 (0.63)	2 (0)
	(E,E)	4.33 (1.58)	4 (1.48)	4.70 (1.79)	4 (1.48)	4.00 (1.81)	4 (1.48)	2.78 (1.04)	3 (1.48)
100	(G,G)	4.46 (1.35)	4 (1.48)	4.77 (1.50)	4 (1.48)	4.27 (1.24)	4 (1.48)	2.90 (0.74)	3 (1.48)
	(G,E)	4.30 (1.42)	4 (1.48)	4.77 (1.62)	5 (1.48)	4.32 (1.39)	4 (1.48)	2.87 (0.73)	3 (0)
	(E,G)	4.52 (1.46)	4 (1.48)	4.91 (1.68)	5 (1.48)	4.42 (1.46)	4 (1.48)	2.81 (0.90)	3 (0)
	(E,E)	4.58 (1.43)	4 (1.48)	5.00 (1.64)	5 (1.48)	4.19 (1.36)	4 (1.48)	2.91 (0.79)	3 (0)
200	(G,G)	4.87 (1.34)	5 (1.48)	5.27 (1.51)	5 (1.48)	4.81 (1.11)	5 (1.48)	3.15 (0.61)	3 (0)
	(G,E)	4.88 (1.30)	5 (1.48)	5.19 (1.48)	5 (1.48)	4.75 (1.12)	5 (1.48)	3.08 (0.43)	3 (0)
	(E,G)	4.91 (1.40)	5 (1.48)	5.36 (1.53)	5 (1.48)	4.75 (1.16)	5 (1.48)	3.13 (0.52)	3 (0)
	(E,E)	4.89 (1.33)	4 (1.48)	5.39 (1.50)	5 (1.48)	4.70 (1.20)	4 (1.48)	3.05 (0.49)	3 (0)

Table 10: Mean, median, standard deviation and mean absolute deviation of the number of components of the estimates of β for scenario 5.

		$\widehat{\beta}_S$		$\widehat{\beta}_R$		$\widehat{\beta}_{BS}$		$\widehat{\beta}_{PLS}$	
n	*Distrib.*	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)	*Mean* (*sd*)	*Median* (*mad*)
50	(G, G)	4.54 (1.31)	4 (0)	5.12 (1.63)	4 (0)	4.27 (1.34)	4 (1.48)	3.23 (0.89)	3 (0)
	(G, E)	4.52 (1.18)	4 (0)	5.10 (1.51)	5 (1.48)	4.11 (1.28)	4 (1.48)	3.08 (0.70)	3 (0)
	(E, G)	4.40 (1.12)	4 (0)	5.04 (1.53)	4 (1.48)	4.12 (1.25)	4 (1.48)	3.27 (0.77)	3 (0)
	(E, E)	4.58 (1.28)	4 (0)	5.15 (1.62)	4 (1.48)	4.22 (1.46)	4 (1.48)	3.34 (0.94)	3 (0)
100	(G, G)	4.45 (1.13)	4 (0)	5.08 (1.55)	4 (1.48)	4.24 (1.10)	4 (1.48)	3.14 (0.88)	3 (0)
	(G, E)	4.40 (0.99)	4 (0)	5.15 (1.49)	5 (1.48)	4.19 (1.08)	4 (1.48)	3.08 (0.88)	3 (0)
	(E, G)	4.49 (1.11)	4 (0)	5.18 (1.58)	4 (0)	4.32 (1.05)	4 (0)	3.33 (0.87)	3 (0)
	(E, E)	4.66 (1.17)	4 (0)	5.10 (1.49)	4 (0)	4.39 (1.12)	4 (0)	3.24 (0.72)	3 (0)
200	(G, G)	4.33 (0.95)	4 (0)	5.08 (1.53)	4 (0)	4.26 (1.00)	4 (0)	3.07 (0.77)	3 (0)
	(G, E)	4.46 (1.02)	4 (0)	5.17 (1.52)	4 (1.48)	4.42 (1.11)	4 (0)	3.07 (0.69)	3 (0)
	(E, G)	4.55 (1.06)	4 (0)	5.24 (1.50)	5 (1.48)	4.57 (1.06)	4 (0)	3.24 (0.73)	3 (0)
	(E, E)	4.48 (1.04)	4 (0)	5.07 (1.47)	4 (1.48)	4.32 (1.02)	4 (0)	3.10 (070)	3 (0)