Robust Functional Principal Components: a projection-pursuit approach *

Lucas Bali
Departamento de Matemáticas and Instituto de Cálculo, FCEyN, Universidad de Buenos Aires and CONICET
Graciela Boente
Departamento de Matemáticas and Instituto de Cálculo, FCEyN, Universidad de Buenos Aires and CONICET
David E. Tyler
Department of Statistics, Rutgers University
Jane–Ling Wang
Department of Statistics,University of California at Davis

June 19, 2011

Abstract

In many situations, data are recorded over a period of time and may be regarded as realizations of a stochastic process. In this paper, robust estimators for the principal components are considered by adapting the projection pursuit approach to the functional data setting. Our approach combines robust projection–pursuit with different smoothing methods. Consistency of the estimators are shown under mild assumptions. The performance of the classical and robust procedures are compared in a simulation study under different contamination schemes.

Key Words: Fisher–consistency, Functional Data, Principal Components, Outliers, Robust Estimation

AMS Subject Classification: MSC 62G35 MSC 62G20

*This research was partially supported by Grants X018 from Universidad of Buenos Aires, PID 112-200801-00216 from CONICET and PICT 821 from ANPCYT, Argentina (L. Bali and G. Boente), NSF grant DMS-0906773 (D. E. Tyler), and NSF Grant DMS-0906813 (J-L. Wang).

1 Introduction

Analogous to classical principal components analysis (PCA), the projection-pursuit approach to robust PCA is based on finding projections of the data which have maximal dispersion. Instead of using the variance as a measure of dispersion, a robust scale estimator s_n is used for the maximization problem. This approach was introduced by Li and Chen (1985), who proposed estimators based on maximizing (or minimizing) a robust scale. In this way, the first robust principal component vector is defined as

$$\widehat{\mathbf{a}} = \underset{\{\mathbf{a}:\|\mathbf{a}\|=1\}}{\operatorname{argmax}} \; s_n(\mathbf{a}^{\mathrm{T}}\mathbf{x}_1, \cdots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_n),$$

and the subsequent principal component vectors are obtained by imposing orthogonality conditions. In the multivariate setting, Li and Chen (1985) argue that the breakdown point for this projection-pursuit based procedure is the same as that of the scale estimator s_n. Later on, Croux and Ruiz–Gazen (2005) derived the influence functions of the resulting principal components, while their asymptotic distribution was studied in Cui *et al.* (2003). A maximization algorithm for obtaining $\widehat{\mathbf{a}}$ was proposed in Croux and Ruiz–Gazen (1996).

The aim of this paper is to adapt the projection pursuit approach to the functional data setting. We focus on functional data that are recorded over a period of time and regarded as realizations of a stochastic process, often assumed to be in the L^2 space on a real interval. Various choices of robust scales, including the median of the absolute deviation about the median (MAD) and some of its variants which are discussed in Rousseeuw and Croux (1993), will be explored and compared.

Principal components analysis, which was originally developed for multivariate data, has been successfully extended to accommodate functional data, and is usually referred to as functional PCA. It can be described as follows. Let $\{X(t) : t \in \mathcal{I}\}$ be a stochastic process defined in $(\Omega, \mathcal{A}, P)$ with continuous trajectories and finite second moment, where $\mathcal{I} \subset \mathbb{R}$ is a finite interval. Without loss of generality, we may assume that $\mathcal{I} = [0, 1]$. We will denote the covariance function by $\gamma_X(t, s) = \operatorname{cov}(X(t), X(s))$, and the corresponding covariance operator by $\mathbf{\Gamma}_X$. We then have $\gamma_X(t, s) = \sum_{j=1}^{\infty} \lambda_j \phi_j(t)\phi_j(s)$, where $\{\phi_j : j \geq 1\}$ and $\{\lambda_j : j \geq 1\}$ are respectively the eigenfunctions and the eigenvalues of the covariance operator $\mathbf{\Gamma}_X$ with $\lambda_j \geq \lambda_{j+1}$. Moreover, $\sum_{j=1}^{\infty} \lambda_j^2 = \|\mathbf{\Gamma}_X\|_{\mathcal{F}}^2 = \int_0^1 \int_0^1 \gamma_X^2(t, s)dtds < \infty$. Let $Y = \int_0^1 \alpha(t)X(t)dt = \langle \alpha, X\rangle$ be a linear combination of the process $\{X(s)\}$, so that $\operatorname{var}(Y) = \langle \alpha, \mathbf{\Gamma}_X \alpha\rangle$. The first principal component is defined as the random variable $Y_1 = \langle \alpha_1, X\rangle$ such that

$$\operatorname{var}(Y_1) = \sup_{\{\alpha:\|\alpha\|=1\}} \operatorname{var}(\langle \alpha, X\rangle) = \sup_{\{\alpha:\|\alpha\|=1\}} \langle \alpha, \mathbf{\Gamma}_X \alpha\rangle, \tag{1}$$

where $\|\alpha\|^2 = \langle \alpha, \alpha\rangle$. Therefore, if $\lambda_1 > \lambda_2$, the solution of (1) is related to the eigenfunction associated with the largest eigenvalue of the operator $\mathbf{\Gamma}_X$, i.e., $\alpha_1 = \phi_1$ and $\operatorname{var}(Y_0) = \lambda_1$. Dauxois *et al.* (1982) derived the asymptotic properties of the principal components of functional data, which are defined as the eigenfunctions of the sample covariance operator. Rice and Silverman (1991) proposed to smooth the principal components by a roughness penalization method and suggested a leave-one-subject-out cross validation method to select the smoothing parameter. Silverman (1996) and Ramsay and Silverman (2005) introduced smooth principal components for functional data, also based on roughness penalty methods, while Boente and Fraiman (2000) considered a kernel–based approach. More recent work on estimation of the principal components and the covariance function includes Gervini (2006), Hall and Hosseini-Nasab (2006), Hall *et al.* (2006) and Yao and Lee (2006).

The literature on robust principal components in the functional data setting is rather sparse. To our knowledge, the first attempt to provide estimators of the principal components that are less sensitive to anomalous observations was due to Locantore *et al.* (1999), who considered the coefficients of a basis expansion. Their approach, however, is multivariate in nature. Gervini (2008) studied a fully functional approach to robust estimation of the principal components by considering a functional version of the spherical principal components defined in Locantore *et al.* (1999) but assuming a finite and known number of principal components in order to ensure Fisher–consistency. Hyndman and Ullah (2007) proposed a method combining a robust projection–pursuit approach and a smoothing and weighting step to forecast age–specific mortality and fertility rates observed over time. However, they did not study its properties in detail.

In this paper, we introduce several robust estimators of the principal components in the functional data setting and establish their strong consistency. Our approach uses a robust projection–pursuit combined with various smoothing methods and our results hold even if the number of principal components is not finite. In this sense, it provides the first rigorous attempt to tackle the challenging properties of robust functional PCA.

In Section 2, the robust estimators of the principal components, based on both the raw and smoothed approaches, are introduced. Consistency results and the asymptotic robustness of the procedure are given in Section 3, while the selection of the smoothing parameters for the smooth principal components is discussed in Section 4. The results of a Monte Carlo study are reported in Section 5. Section 6 contains some concluding remarks and appendix A provides conditions under which one of the crucial assumptions hold. Most proofs are relegated to Appendix B.

2 The estimators

We consider several robust approaches in this section and define them on a separable Hilbert space $\mathcal{H}$ keeping in mind that the main application will be $\mathcal{H} = L^2(\mathcal{I})$. From now on and throughout the paper, $\{X_i : 1 \le i \le n\}$ denote realizations of the stochastic process $X \sim P$ in a separable Hilbert space $\mathcal{H}$. Thus, $X_i \sim P$ are independent stochastic processes that follow the same law. This independence condition could be relaxed since we only need the strong law of large numbers to hold in order to guarantee that the results established in this paper hold.

2.1 Raw robust projection–pursuit approach

Based on the property (1) of the first principal component and given $\sigma_{\mathrm{R}}(F)$ a robust scale functional, the raw (meaning unsmoothed) robust functional principal component directions are defined as

$$\begin{cases} \phi_{\mathrm{R},1}(P) &= \operatorname*{argmax}_{\|\alpha\|=1} \sigma_{\mathrm{R}}\left(P[\alpha]\right) \\ \phi_{\mathrm{R},m}(P) &= \operatorname*{argmax}_{\|\alpha\|=1, \alpha\in\mathcal{B}_m} \sigma_{\mathrm{R}}\left(P[\alpha]\right), \quad 2 \le m \,, \end{cases} \tag{2}$$

where $P[\alpha]$ stands for the distribution of $\langle \alpha, X\rangle$ when $X \sim P$, and $\mathcal{B}_m = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{\mathrm{R},j}(P)\rangle = 0,\ 1 \le j \le m-1\}$. We will denote the mth largest eigenvalues by

$$\lambda_{\mathrm{R},m}(P) = \sigma^2_{\mathrm{R}}(P[\phi_{\mathrm{R},m}]) = \max_{\|\alpha\|=1, \alpha\in\mathcal{B}_m} \sigma^2_{\mathrm{R}}\left(P[\alpha]\right) \ . \tag{3}$$

Since the unit ball is weakly compact, the maximum above is attained if the scale functional σ_{R} is (weakly) continuous.

Next, denote by $s_n^2 : \mathcal{H} \to \mathbb{R}$ the function $s_n^2(\alpha) = \sigma^2_{\mathrm{R}}\left(P_n[\alpha]\right)$, where $\sigma_{\mathrm{R}}(P_n[\alpha])$ stand for the functional σ_{R} computed at the empirical distribution of $\langle \alpha, X_1\rangle, \dots, \langle \alpha, X_n\rangle$. Analogously, $\sigma : \mathcal{H} \to \mathbb{R}$ will stand for $\sigma(\alpha) = \sigma_{\mathrm{R}}(P[\alpha])$. The components in (2) will be estimated empirically by

$$\begin{cases} \widehat{\phi}_1 &= \operatorname*{argmax}_{\|\alpha\|=1} s_n(\alpha) \\ \widehat{\phi}_m &= \operatorname*{argmax}_{\alpha\in\widehat{\mathcal{B}}_m} s_n(\alpha) \quad 2 \le m, \end{cases} \tag{4}$$

where $\widehat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_j \rangle = 0 \ , \ \forall\, 1 \le j \le m-1\}$. The estimators of the eigenvalues are then computed as

$$\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) \ , \quad 1 \le m \ . \tag{5}$$

2.2 Smoothed robust principal components

Sometimes instead of raw functional principal components, smoothed ones are of interest. The advantages of smoothed functional PCA are well documented, see for instance, Rice and Silverman (1991) and Ramsay and Silverman (2005). One compelling argument is that smoothing is a regularization tool that might reveal more interpretable and interesting feature of the modes of variation for functional data. Rice and Silverman (1991) and Silverman (1996) proposed two smoothing approaches by penalizing the variance and the norm, respectively. To be more specific, Rice and Silverman (1991) estimate the first principal component by maximizing over $\|\alpha\| = 1$, the objective function $\widehat{\text{var}}\,(\langle \alpha, X \rangle) - \tau \lceil \alpha, \alpha \rceil$, where $\widehat{\text{var}}$ stands for the sample variance and $\lceil \alpha, \beta \rceil = \int_0^1 \alpha''(t)\beta''(t)dt$. Silverman (1996) proposed a different way to penalize the roughness by defining the penalized inner product $\langle \alpha, \beta \rangle_\tau = \langle \alpha, \beta \rangle + \tau \lceil \alpha, \beta \rceil$. Then, the smoothed first direction $\widehat{\phi}_1$ is the one that maximizes $\widehat{\text{var}}\,(\langle \alpha, X \rangle)$ over over $\|\alpha\|_\tau = 1$ subject to the condition that $\|\widehat{\phi}_1\|_\tau^2 = \langle \phi_1, \phi_1 \rangle_\tau = 1$.

Silverman (1996) obtained consistency results of the norm–penalized principal components estimators under the assumption that ϕ_j have finite roughness, i.e., $\lceil \phi_j, \phi_j \rceil < \infty$. Clearly the smoothing parameter τ needs to converge to 0 in order to get consistency results.

Let us consider $\mathcal{H}_{\rm S}$, the subset of "smooth elements"of $\mathcal{H}$. In order to obtain consistency results, we need $\phi_{{\rm R},j}(P) \in \mathcal{H}_{\rm S}$, or $\phi_{{\rm R},j}(P)$ belongs to the closure, $\overline{\mathcal{H}_{\rm S}}$, of $\mathcal{H}_{\rm S}$. Let $D : \mathcal{H}_{\rm S} \to \mathcal{H}$, a linear operator that we will call the "differentiator". Using D, we will define the symmetric positive semidefinite bilinear form $\lceil \cdot, \cdot \rceil : \mathcal{H}_{\rm S} \times \mathcal{H}_{\rm S} \to \mathbb{R}$, where $\lceil \alpha, \beta \rceil = \langle D\alpha, D\beta \rangle$. The "penalization operator" is then defined as $\Psi : \mathcal{H}_{\rm S} \to \mathbb{R}$, $\Psi(\alpha) = \lceil \alpha, \alpha \rceil$, and the penalized inner product as $\langle \alpha, \beta \rangle_\tau = \langle \alpha, \beta \rangle + \tau \lceil \alpha, \beta \rceil$. Therefore, $\|\alpha\|_\tau^2 = \|\alpha\|^2 + \tau \Psi(\alpha)$. As in Pezzulli and Silverman (1993), we will assume that the bilinear form is closable.

Remark 2.2.1. The most common setting for functional data is to choose $\mathcal{H} = L^2(\mathcal{I})$, $\mathcal{H}_{\rm S} = \{\alpha \in L^2(\mathcal{I}), \alpha$ is twice differentiable, and $\int_{\mathcal{I}} (\alpha''(t))^2 dt < \infty\}$, $D\alpha = \alpha''$ and $\lceil \alpha, \beta \rceil = \int_{\mathcal{I}} \alpha''(t)\beta''(t)dt$ so that $\Psi(\alpha) = \int_{\mathcal{I}} (\alpha''(t))^2 dt$.

Let $\sigma_{\rm R}(F)$ be a robust scale functional and define $s_n^2(\alpha)$ and $\sigma(\alpha)$ as in Section 2.1. Then, we can adapt the classical procedure by defining the smoothed robust functional principal components estimators either

a) by penalizing the norm as

$$\begin{cases} \widehat{\phi}_{{\rm PN},1} & = \underset{\|\alpha\|_\tau = 1}{\operatorname{argmax}}\, s_n^2(\alpha) = \underset{\beta \ne 0}{\operatorname{argmax}}\, \dfrac{s_n^2(\beta)}{\langle \beta, \beta \rangle + \tau \lceil \beta, \beta \rceil} \\ \widehat{\phi}_{{\rm PN},m} & = \underset{\alpha \in \widehat{\mathcal{B}}_{m,\tau}}{\operatorname{argmax}}\, s_n^2(\alpha) \quad 2 \le m, \end{cases} \tag{6}$$

where $\widehat{\mathcal{B}}_{m,\tau} = \{\alpha \in \mathcal{H} : \|\alpha\|_\tau = 1, \langle \alpha, \widehat{\phi}_{{\rm PN},j} \rangle_\tau = 0 \ , \ \forall\, 1 \le j \le m-1\}$;

b) or by penalizing the scale as

$$\begin{cases} \widehat{\phi}_{{\rm PS},1} & = \underset{\|\alpha\| = 1}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \tau \lceil \alpha, \alpha \rceil \right\} \\ \widehat{\phi}_{{\rm PS},m} & = \underset{\alpha \in \widehat{\mathcal{B}}_{{\rm S},m}}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \tau \lceil \alpha, \alpha \rceil \right\} \quad 2 \le m, \end{cases} \tag{7}$$

where $\widehat{\mathcal{B}}_{{\rm PS},m} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_{{\rm PS},j} \rangle = 0 \ , \ \forall\, 1 \le j \le m-1\}$.

The eigenvalue estimators are thus defined as

$$\widehat{\lambda}_{\mathrm{PS},m} = s_n^2(\widehat{\phi}_{\mathrm{PS},m}) \tag{8}$$
$$\widehat{\lambda}_{\mathrm{PN},m} = s_n^2(\widehat{\phi}_{\mathrm{PN},m}). \tag{9}$$

2.3 Sieve approach for robust functional principal components

A different approach can be defined that is related to $B-$splines, and more generally, the method of sieves. The sieve method involves approximating an infinite–dimensional parameter space Θ by a series of finite–dimensional parameter spaces Θ_n, that depend on the sample size n and estimating the parameter on the spaces Θ_n, not Θ.

Let $\{\delta_i\}_{i\geq 1}$ be a basis of $\mathcal{H}$ and define $\mathcal{H}_{p_n}$ the linear space spanned by $\delta_1,\dots,\delta_{p_n}$ and by $\mathcal{S}_{p_n} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\| = 1\}$, i.e., $\mathcal{H}_{p_n} = \{\alpha \in \mathcal{H} : \ \alpha = \sum_{j=1}^{p_n} a_j\delta_j\}$ and $\mathcal{S}_{p_n} = \{\alpha \in \mathcal{H} : \ \alpha = \sum_{j=1}^{p_n} a_j\delta_j$, $\mathbf{a} = (a_1,\dots,a_{p_n})^{\mathrm{T}}$ such that $\|\alpha\|^2 = \sum_{j=1}^{p_n}\sum_{s=1}^{p_n} a_j a_s\langle\delta_j,\delta_s\rangle = 1\}$. Note that $\mathcal{S}_{p_n}$ approximates the unit sphere $\mathcal{S} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$. Define the robust sieve estimators of the principal components as

$$\begin{cases} \widehat{\phi}_{\mathrm{SI},1} &= \underset{\alpha\in\mathcal{S}_{p_n}}{\operatorname{argmax}}\, s_n(\alpha) \\ \widehat{\phi}_{\mathrm{SI},m} &= \underset{\alpha\in\widehat{\mathcal{B}}_{n,m}}{\operatorname{argmax}}\, s_n(\alpha) \quad 2\leq m, \end{cases} \tag{10}$$

where $\widehat{\mathcal{B}}_{n,m} = \{\alpha \in \mathcal{S}_{p_n} : \ \langle\alpha,\widehat{\phi}_{\mathrm{SI},j}\rangle = 0\ ,\ \forall\ 1\leq j\leq m-1\}$, and define the eigenvalue estimators as

$$\widehat{\lambda}_{\mathrm{SI},m} = s_n^2(\widehat{\phi}_{\mathrm{SI},m})\ . \tag{11}$$

Some of the more frequently used bases in the analysis of functional data are the Fourier, polynomial, spline, or wavelet bases (see, for instance, Ramsay and Silverman, 2005).

3 Consistency results

In this section, we show that under mild conditions the functionals $\phi_{\mathrm{R},m}(P)$ and $\lambda_{\mathrm{R},m}(P)$ are weakly continuous. Moreover, we state conditions that guarantee the consistency of the estimators defined in Section 2. Our results hold in particular for, but are not restricted to, the functional elliptical families defined in Bali and Boente (2009). We recall here their definition for the sake of completeness.

Let X be a random element in a separable Hilbert space $\mathcal{H}$. Let $\mu \in \mathcal{H}$ and $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$ be a self–adjoint, positive semidefinite and compact operator. We will say that X has an elliptical distribution with parameters $(\mu,\mathbf{\Gamma})$, denoted as $X \sim \mathcal{E}(\mu,\mathbf{\Gamma})$, if for any linear and bounded operator $A : \mathcal{H} \to \mathbb{R}^d$, AX has a multivariate elliptical distribution with parameters $A\mu$ and $\mathbf{\Sigma} = A\mathbf{\Gamma}A^*$, i.e., $AX \sim \mathcal{E}_d(A\mu,\mathbf{\Sigma})$, where $A^* : \mathbb{R}^p \to \mathcal{H}$ stands for the adjoint operator of A. As in the finite–dimensional setting, if the covariance operator, $\mathbf{\Gamma}_X$, of X exists then, $\mathbf{\Gamma}_X = a\,\mathbf{\Gamma}$, for some $a \in \mathbb{R}$.

The following transformation can be used to obtain random elliptical elements. Let V_1 be a Gaussian element in $\mathcal{H}$ with zero mean and covariance operator $\mathbf{\Gamma}_{V_1}$, and let Z be a random variable independent of V_1. Given $\mu \in \mathcal{H}$, define $X = \mu + Z\,V_1$. Then, X has an elliptical distribution $\mathcal{E}(\mu,\mathbf{\Gamma})$ with the operator $\mathbf{\Gamma}$ being proportional to $\mathbf{\Gamma}_{V_1}$ and with no moment conditions required. It is worth noting that the converse holds if all the eigenvalues of $\mathbf{\Gamma}$ are positive. Specifically, if $X \sim \mathcal{E}(\mu,\mathbf{\Gamma})$ and the eigenvalues of $\mathbf{\Gamma}$ are positive, then, $X = \mu + ZV$ for some mean zero Gaussian process V and random variable $Z \in \mathbb{R}$, which is independent of V. This result can be obtained as a corollary to the theorem in Kingman (1972), which states that if a random variable can be embedded within a sequence of spherical random vectors of dimension k for any $k = 1,2,\dots$, then the random variable must be distributed as a scale mixture of normals. For random elements which admit a finite Karhunen Loève expansion, i.e., $X(t) = \mu(t) + \sum_{j=1}^{q} \lambda_j^{\frac{1}{2}} U_j \phi_j(t)$, the assumption that X has an elliptical distribution is analogous to assuming that $\mathbf{U} = (U_1, \dots, U_q)^{\mathrm{T}}$ has a spherical distribution. This finite expansion was considered, for instance, by Gervini (2008).

To derive the consistency of the proposed estimators, we need the following assumptions.

S1. $\sup_{\|\alpha\|=1} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \xrightarrow{a.s.} 0$

S2. $\sigma : \mathcal{H} \to \mathbb{R}$ is a weakly continuous function, i.e., continuous with respect to the weak topology in $\mathcal{H}$.

Remark 3.1.

i) Assumption **S1** holds for the classical estimators based on the sample variance since the empirical covariance operator, $\widehat{\mathbf{\Gamma}}$, is consistent in the unit ball. Indeed, as shown in Dauxois *et al.* (1982), $\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}_X\| \xrightarrow{a.s.} 0$, which entails that $\sup_{\|\alpha\|=1} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \le \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}_X\| \xrightarrow{a.s.} 0$. However, this assumption can be harder to verify for other scale functionals since the unit sphere $\mathcal{S} = \{\|\alpha\| = 1\}$ is not compact. The weaker conditions $\sup_{\|\alpha\|_\tau=1} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \xrightarrow{a.s.} 0$ or $\sup_{\alpha \in \mathcal{S}_{p_n}} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \xrightarrow{a.s.} 0$ can be introduced for the smoothed proposals in Section 2.2., since the set $\{\alpha \in \mathcal{S}_{p_n}\}$ is compact. Some more general conditions on the scale functional that guarantee **S1** are stated in Appendix A.

ii) If the scale functional σ_{R} is a continuous functional (with respect to the weak topology), then **S2** follows. This is because if $\alpha_k \to \alpha$, as $k \to \infty$, then $\langle \alpha_k, X\rangle \xrightarrow{\omega} \langle \alpha, X\rangle$ and hence, $\sigma_{\mathrm{R}}(P[\alpha_k]) \to \sigma_{\mathrm{R}}(P[\alpha])$. For the case when the scale functional is taken to be the standard deviation and the underlying probability P has a covariance operator $\mathbf{\Gamma}_X$, we see from the relationship $\sigma^2(\alpha) = \langle \alpha, \mathbf{\Gamma}_X \alpha\rangle$ that condition **S2** holds, even though the standard deviation itself is not a weakly continuous functional.

iii) If X has an elliptical distribution $\mathcal{E}(\mu, \mathbf{\Gamma})$, then there exists a positive constant c such that for any $\alpha \in \mathcal{H}$, $\sigma_{\mathrm{R}}^2(P[\alpha]) = c\langle \alpha, \mathbf{\Gamma}\alpha\rangle$. Furthermore, it immediately follows that the function $\sigma : \mathcal{H} \to \mathbb{R}$ defined as $\sigma(\alpha) = \sigma_R(P[\alpha])$ is weakly continuous. Moreover, since there exists a metric d generating the weak topology in $\mathcal{H}$ and the closed ball $\mathcal{V}_r = \{\alpha : \|\alpha\| \le r\}$ is weakly compact, we see that **S2** implies that $\sigma(\alpha)$ is uniformly continuous with respect to the metric d and hence, with respect to the weak topology, over $\mathcal{V}_r$. Weakly uniform continuity is used in some of the results presented later in this section.

iv) The Fisher-consistency of the functionals defined through (2) follows immediately from the previous result if the underlying distribution is elliptical. More generally, let us consider the following assumption

S3. there exists a constant $c > 0$ and a self–adjoint, positive semidefinite and compact operator $\mathbf{\Gamma}$, such that for any $\alpha \in \mathcal{H}$, we have $\sigma^2(\alpha) = c\langle \alpha, \mathbf{\Gamma}\alpha\rangle$.

*Let $X \sim P$ be a random element such that **S3** holds. Denote by $\lambda_1 \ge \lambda_2 \ge \dots$ the eigenvalues of $\mathbf{\Gamma}$ and by ϕ_j the eigenfunction associated to λ_j. Then, we have that $\phi_{\mathrm{R},j}(P) = \phi_j$ and $\lambda_{\mathrm{R},j}(P) = c\,\lambda_j$.*

As in the finite–dimensional setting, the scale functional σ_{R} can be calibrated to attain Fisher–consistency of the eigenvalues.

v) Assumption **S3** ensures that we are estimating the target directions. It may seem restrictive since it is difficult to verify outside the family of elliptical distributions except when the scale is taken to be the standard deviation. However, even in the finite-dimensional case, asymptotic properties have been derived only under similar restrictions. For instance, both Li and Chen (1985) and Croux and Ruiz–Gazen (2005) assume an underlying elliptical distribution in order to obtain consistency results and influence functions, respectively. Also, in Cui *et al.* (2003) the influence function of the projected data is assumed to be of the form $h(\mathbf{x}, \mathbf{a}) = 2\sigma(F[\mathbf{a}])\mathrm{IF}(\mathbf{x}, \sigma_{\mathbf{a}}; F_0)$, where $F[\mathbf{a}]$ stands for the distribution of $\mathbf{a}^{\mathrm{T}}\mathbf{x}$ when $\mathbf{x} \sim F$. This condition, though, primarily holds when the distribution is elliptical.

Before stating the consistency results, we first establish some notations and then prove the continuity of the eigenfunction and eigenvalue functionals. Denote by $\mathcal{L}_{m-1}$ the linear space spanned by $\{\phi_{\mathrm{R},1}, \ldots, \phi_{\mathrm{R},m-1}\}$ and let $\widehat{\mathcal{L}}_{m-1}$ be the linear space spanned by the first $m-1$ estimated eigenfunctions, i.e., by $\{\widehat{\phi}_1, \ldots, \widehat{\phi}_{m-1}\}$, $\{\widehat{\phi}_{\mathrm{PS},1}, \ldots, \widehat{\phi}_{\mathrm{PS},m-1}\}$, $\{\widehat{\phi}_{\mathrm{PN},1}, \ldots, \widehat{\phi}_{\mathrm{PN},m-1}\}$ or $\{\widehat{\phi}_{\mathrm{SI},1}, \ldots, \widehat{\phi}_{\mathrm{SI},m-1}\}$, where it will be clear in each case which linear space we are considering. Finally, for any linear space $\mathcal{L}$, $\pi_{\mathcal{L}} : \mathcal{H} \to \mathcal{L}$ stands for the orthogonal projection onto the linear space $\mathcal{L}$, which exists if $\mathcal{L}$ is a closed linear space. In particular, $\pi_{\mathcal{L}_{m-1}}$, $\pi_{\widehat{\mathcal{L}}_{m-1}}$ and $\pi_{\mathcal{H}_{p_n}}$ are well defined.

The following Lemma is useful for deriving the consistency and continuity of the eigenfunction estimators. In this lemma and in the subsequent proposition and theorems, it should be noted that $\langle\widehat{\phi}, \phi\rangle^2 \to 1$ implies, under the same mode of convergence, that the sign of $\widehat{\phi}$ can be chosen so that $\widehat{\phi} \to \phi$. Throughout the rest of this section, $\phi_{\mathrm{R},j}(P)$ and $\lambda_{\mathrm{R},j}(P)$ stand for the functionals defined through (2) and (3). For the sake of simplicity, denote by $\lambda_{\mathrm{R},j} = \lambda_{\mathrm{R},j}(P)$ and $\phi_{\mathrm{R},j} = \phi_{\mathrm{R},j}(P)$. Assume that $\lambda_{\mathrm{R},1} > \lambda_{\mathrm{R},2} > \ldots > \lambda_{\mathrm{R},q} > \lambda_{\mathrm{R},q+1}$ for some $q \geq 2$ and that, for $1 \leq m \leq q$, $\phi_{\mathrm{R},j}$ are unique up to changes in their sign.

Lemma 3.1. *Let $\widehat{\phi}_m \in \mathcal{S}$ be such that $\langle\widehat{\phi}_m, \widehat{\phi}_j\rangle = 0$ for $j \neq m$. If **S2** holds, we have that*

a) If $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$, then, $\langle\widehat{\phi}_1, \phi_{\mathrm{R},1}\rangle^2 \xrightarrow{a.s.} 1$.

b) Given $2 \leq m \leq q$, if $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{\mathrm{R},s}$, for $1 \leq s \leq m-1$, we have that for $1 \leq m \leq q$, $\langle\widehat{\phi}_m, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$.

Let $d_{\mathrm{PR}}(P, Q)$ stands for the Prohorov distance between the probability measures P and Q. Thus, $P_n \xrightarrow{\omega} P$ if and only if $d_{\mathrm{PR}}(P_n, P) \to 0$. Proposition 3.1 below establishes the continuity of the functionals defined as (2) and (3) hence, the asymptotic robustness of the estimators derived from them, as defined in Hampel (1971). As it will be shown in Appendix A, the uniform convergence required in assumption ii) is satisfied, for instance, if σ_{R} is a continuous scale functional.

Proposition 3.1. *Assume that **S2** holds and that*

$$\sup_{\|\alpha\|=1} |\sigma_{\mathrm{R}}(P_n[\alpha]) - \sigma_{\mathrm{R}}(P[\alpha])| \to 0 \text{ whenever } P_n \xrightarrow{\omega} P.$$

Then, for any sequence P_n such that $P_n \xrightarrow{\omega} P$, we have that

a) $\lambda_{\mathrm{R},1}(P_n) \to \lambda_{\mathrm{R},1}$ and $\sigma^2(\phi_{\mathrm{R},1}(P_n)) \to \sigma^2(\phi_{\mathrm{R},1})$.

b) $\langle\phi_{\mathrm{R},1}(P_n), \phi_{\mathrm{R},1}\rangle^2 \to 1$.

c) For any $2 \leq m \leq q$, if $\phi_{\mathrm{R},s}(P_n) \to \phi_{\mathrm{R},s}$, for $1 \leq s \leq m-1$, then, $\lambda_{\mathrm{R},m}(P_n) \to \sigma^2(\phi_{\mathrm{R},m}) = \lambda_{\mathrm{R},m}$ and $\sigma^2(\phi_{\mathrm{R},m}(P_n)) \to \sigma^2(\phi_{\mathrm{R},m})$.

d) For $1 \leq m \leq q$, $\langle\phi_{\mathrm{R},m}(P_n), \phi_{\mathrm{R},m}\rangle^2 \to 1$.

3.1 Consistency of the raw robust estimators

Theorem 3.1 establishes the consistency of the raw estimators of the principal components. The proof of the theorem is similar to that of Proposition 3.1.

Theorem 3.1. *Let $\widehat{\phi}_m$ and $\widehat{\lambda}_m$ be the estimators defined in (4) and (5), respectively. Under **S1** and **S2**, we have that*

a) $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ and $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$.

b) $\langle \widehat{\phi}_1, \phi_{\mathrm{R},1}\rangle^2 \xrightarrow{a.s.} 1$.

c) Given $2 \leq m \leq q$, if $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{\mathrm{R},s}$, for $1 \leq s \leq m-1$, then $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$.

d) For $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$.

3.2 Consistency of the smoothed robust approach via penalization of the norm

Recall that $\mathcal{H}_{\mathrm{S}}$ is the subspace of $\mathcal{H}$ of smooth elements α such that $\Psi(\alpha) = \lceil \alpha, \alpha \rceil = \|D\alpha\|^2 < \infty$. To derive the consistency of the proposals given by (6) and (7), we will need one of the following assumptions in **S4**.

S4. a) $\phi_{\mathrm{R},j} \in \mathcal{H}_{\mathrm{S}}, \ \forall j$ or b) $\phi_{\mathrm{R},j} \in \overline{\mathcal{H}_{\mathrm{S}}}, \ \forall j$.

Condition **S4b)** generalizes the assumption of smoothness required in Silverman (1996), and holds, for example, when $\mathcal{H}_{\mathrm{S}}$ is a dense subset of $\mathcal{H}$.

For the sake of simplicity, denote by $\mathcal{T}_k = \mathcal{L}_k^{\perp}$ the linear space orthogonal to $\phi_1, \ldots, \phi_k$ and by $\pi_k = \pi_{\mathcal{T}_k}$ the orthogonal projection with respect to the inner product defined in $\mathcal{H}$. On the other hand, let $\widehat{\pi}_{\tau,k}$ be the projection onto the linear space orthogonal to $\widehat{\phi}_{\mathrm{PN},1}, \ldots, \widehat{\phi}_{\mathrm{PN},k}$ in the space $\mathcal{H}_{\mathrm{S}}$ in the inner product $\langle \cdot, \cdot \rangle_\tau$, i.e., for any $\alpha \in \mathcal{H}_{\mathrm{S}}$, $\widehat{\pi}_{\tau,k}(\alpha) = \alpha - \sum_{j=1}^k \langle \alpha, \widehat{\phi}_{\mathrm{PN},j}\rangle_\tau \widehat{\phi}_{\mathrm{PN},j}$. Moreover, let $\widehat{\mathcal{T}}_{\tau,k}$ be the linear space orthogonal to $\widehat{\mathcal{L}}_k$ with the inner product $\langle \cdot, \cdot \rangle_\tau$. Thus, $\widehat{\pi}_{\tau,k}$ is the orthogonal projection onto $\widehat{\mathcal{T}}_{\tau,k}$ with respect to this inner product.

Theorem 3.2. *Let $\widehat{\phi}_{\mathrm{PN},m}$ and $\widehat{\lambda}_{\mathrm{PN},m}$ be the estimators defined in (6) and (9), respectively. Moreover, assume conditions **S1**, **S2** and **S4b)** holds. If $\tau = \tau_n \to 0$, $\tau_n \geq 0$, then*

a) $\widehat{\lambda}_{\mathrm{PN},,1} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ and $\sigma^2(\widehat{\phi}_{\mathrm{PN},1}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$

b) $\tau \lceil \widehat{\phi}_{\mathrm{PN},1}, \widehat{\phi}_{\mathrm{PN},1} \rceil \xrightarrow{a.s.} 0$, and so, $\|\widehat{\phi}_{\mathrm{PN},1}\| \xrightarrow{a.s.} 1$.

c) $\langle \widehat{\phi}_{\mathrm{PN},1}, \phi_{\mathrm{R},1}\rangle^2 \xrightarrow{a.s.} 1$.

d) Given $2 \leq m \leq q$, if $\widehat{\phi}_{\mathrm{PN},\ell} \xrightarrow{a.s.} \phi_{\mathrm{R},\ell}$ and $\tau \lceil \widehat{\phi}_{\mathrm{PN},\ell}, \widehat{\phi}_{\mathrm{PN},\ell} \rceil \xrightarrow{a.s.} 0$, for $1 \leq \ell \leq m-1$, then $\widehat{\lambda}_{\mathrm{PN},m} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, $\sigma^2(\widehat{\phi}_{\mathrm{PN},m}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, $\tau \lceil \widehat{\phi}_{\mathrm{PN},m}, \widehat{\phi}_{\mathrm{PN},m} \rceil \xrightarrow{a.s.} 0$ and so, $\|\widehat{\phi}_{\mathrm{PN},m}\| \xrightarrow{a.s.} 1$.

e) For $1 \leq m \leq q$, $\langle \widehat{\phi}_{\mathrm{PN},m}, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$.

3.3 Consistency of the smoothed robust approach via penalization of the scale

Consistency of the proposal given by (7) under assumption **S4a)** is given below.

Theorem 3.3. *Let $\widehat{\phi}_{\mathrm{PS},m}$ and $\widehat{\lambda}_{\mathrm{PS},m}$ be the estimators defined in (7) and (8), respectively. Moreover, assume conditions **S1**, **S2** and **S4a)** hold. If $\tau = \tau_n \to 0$, $\tau_n \geq 0$, then*

a) $\widehat{\lambda}_{\mathrm{PS},1} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ *and* $\sigma^2(\widehat{\phi}_{\mathrm{PS},1}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$. *Moreover,* $\tau\lceil \widehat{\phi}_{\mathrm{PN},1}, \widehat{\phi}_{\mathrm{PN},1}\rceil \xrightarrow{a.s.} 0$.

b) $\langle \widehat{\phi}_{\mathrm{PS},1}, \phi_{\mathrm{R},1}\rangle^2 \xrightarrow{a.s.} 1$.

c) Given $2 \leq m \leq q$, *if* $\widehat{\phi}_{\mathrm{PS},\ell} \xrightarrow{a.s.} \phi_{\mathrm{R},\ell}$, *and* $\tau\lceil \widehat{\phi}_{\mathrm{PN},\ell}, \widehat{\phi}_{\mathrm{PN},\ell}\rceil \xrightarrow{a.s.} 0$, *for* $1 \leq \ell \leq m-1$, *then* $\widehat{\lambda}_{\mathrm{PS},m} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, $\sigma^2(\widehat{\phi}_{\mathrm{PS},m}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ *and* $\tau\lceil \widehat{\phi}_{\mathrm{PN},m}, \widehat{\phi}_{\mathrm{PN},m}\rceil \xrightarrow{a.s.} 0$.

d) For $1 \leq m \leq q$, $\langle \widehat{\phi}_{\mathrm{PS},m}, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$.

3.4 Consistency of the robust approach through the method of sieves

The following Theorem establishes the consistency of the estimators of the principal components defined through (10).

Theorem 3.4. *Let $\widehat{\phi}_{\mathrm{SI},m}$ and $\widehat{\lambda}_{\mathrm{SI},m}$ be the estimators defined in (10) and (11), respectively. Under **S1** and **S2**, if $p_n \to \infty$, then*

a) $\widehat{\lambda}_{\mathrm{SI},1} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ *and* $\sigma^2(\widehat{\phi}_{\mathrm{SI},1}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$

b) Given $2 \leq m \leq q$, *if* $\widehat{\phi}_{\mathrm{SI},\ell} \xrightarrow{a.s.} \phi_{\mathrm{R},\ell}$, *for* $1 \leq \ell \leq m-1$, *then* $\widehat{\lambda}_{\mathrm{SI},m} \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ *and* $\sigma^2(\widehat{\phi}_{\mathrm{SI},m}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$

c) For $1 \leq m \leq q$, $\langle \widehat{\phi}_{\mathrm{SI},m}, \phi_m\rangle^2 \xrightarrow{a.s.} 1$.

4 Selection of the smoothing parameters

The selection of the smoothing parameters is an important practical issue. The most popular general approach to address such a selection problem is to use the cross-validation methods. In nonparametric regression, the sensitivity of L^2 cross–validation methods to outliers has been pointed out by Wang and Scott (1994) and by Cantoni and Ronchetti (2001), among others. The latter also proposed more robust alternatives to L^2 cross–validation. The idea of robust cross–validation can be adapted to the present situation. Assume for the moment that we are interested in a fixed number, ℓ, of components. We propose to proceed as follows.

1. Center the data. i.e., define $\widetilde{X}_i = X_i - \widehat{\mu}$ where $\widehat{\mu}$ is a robust location estimator such as the trimmed means proposed by Fraiman and Muniz (2001), the depth–based estimators of Cuevas *et al.* (2007) and López–Pintado and Romo (2007), or the functional median defined in Gervini (2008).

2. For the penalized roughness approaches and for each m in the range $1 \leq m \leq \ell$ and $0 < \tau_n$, let $\widehat{\phi}^{(-j)}_{m,\tau_n}$ denote the robust estimator of the mth principal component computed without the jth observation.

3. Define $X_j^{\perp}(\tau_n) = \widetilde{X}_j - \pi_{\widehat{\mathcal{L}}_\ell^{(-j)}}(\widetilde{X}_j)$, where $\pi_{\mathcal{H}_s}(X)$ stands for the orthogonal projection of X onto the linear (closed) space $\mathcal{H}_s$ and $\widehat{\mathcal{L}}_\ell^{(-j)}$ stands for the linear space spanned by $\widehat{\phi}^{(-j)}_{1,\tau_n}, \ldots, \widehat{\phi}^{(-j)}_{\ell,\tau_n}$.

4. Given a robust scale estimator around zero σ_n, we propose to minimize $RCV_\ell(\tau_n) = \sigma_n^2(\|X_1^\perp(\tau_n)\|, \ldots, \|X_n^\perp(\tau_n)\|)$.

By robust scale estimator around zero, we mean that no location estimator is applied to center the data. For instance, in the classical setting, we will take $\sigma_n^2(z_1, \ldots, z_n) = (1/n)\sum_{i=1}^n z_i^2$ while in the robust situation, one may consider $\sigma_n(z_1, \ldots, z_n) = \text{median}(z_1, \ldots, z_n)$ or the solution of $\sum_{i=1}^n \chi(z_i/\sigma_n) = n/2$. For large sample sizes, it is well understood that cross-validation methods can be computationally prohibitive. In such cases, $K-$fold cross–validation provides a useful alternative. In the following, we briefly describe a robust $K-$fold cross–validation procedure suitable for our proposed estimates.

1. First center the data as above, using $\widetilde{X}_i = X_i - \widehat{\mu}$.
2. Partition the centered data set $\{\widetilde{X}_i\}$ randomly into K disjoint subsets of approximately equal sizes with the jth subset having size $n_j \geq 2$, $\sum_{j=1}^K n_j = n$. Let $\{\widetilde{X}_i^{(j)}\}_{1\leq i\leq n_j}$ be the elements of the jth subset, and $\{\widetilde{X}_i^{(-j)}\}_{1\leq i\leq n-n_j}$ denote the elements in the complement of the jth subset. The set $\{\widetilde{X}_i^{(-j)}\}_{1\leq i\leq n-n_j}$ will be the training set and $\{\widetilde{X}_i^{(j)}\}_{1\leq i\leq n_j}$ the validation set.
3. Similar to Step 2 above but leave the jth validation subset $\{\widetilde{X}_i^{(j)}\}_{1\leq i\leq n_j}$ out instead of the jth observation.
4. Define $X_j^{(j)\perp}(\tau_n)$ the same way as in Step 2 above, using the validation set. For instance, $X_i^{(j)\perp}(\tau_n) = \widetilde{X}_i^{(j)} - \pi_{\widehat{\mathcal{L}}_\ell^{(-j)}}(\widetilde{X}_i^{(j)})$, $1 \leq i \leq n_j$, where $\widehat{\mathcal{L}}_\ell^{(-j)}$ stands for the linear space spanned by $\widehat{\phi}_{1,\tau_n}^{(-j)}, \ldots, \widehat{\phi}_{\ell,\tau_n}^{(-j)}$.
5. Given a scale estimator around zero σ_n, the robust $K-$fold cross-validation method chooses the smoothing parameter which minimizes $RCV_{\ell,\text{KCV}}(\tau_n) = \sum_{j=1}^K \sigma_n^2(\|X_1^{(j)\perp}(\tau_n)\|, \ldots, \|X_{n_j}^{(j)\perp}(\tau_n)\|)$.

A similar approach can be given to choose p_n when considering the sieve estimators.

5 Monte Carlo Study

5.1 Algorithm and notation

All the methods to be considered here are modifications of the basic algorithm proposed by Croux and Ruiz–Gazen (1996) for the computation of principal components using projection-pursuit. The basic algorithm applies to multivariate data, say m-dimensional, and requires a search over projections in $\mathbb{R}^m$. To apply the algorithm to functional data, we discretized the domain of the observed function over $m = 50$ equally spaced points in $\mathcal{I} = [-1, 1]$. We have also adapted the algorithm to allow for smoothed principal components and for different methods of centering. In this sense, there are three main characteristics which distinguish the different computed estimators: the scale function, the method of centering, and the type of smoothing used.

- **Scale function:** Three scale functions are considered here: the classical standard deviation (SD), the Median Absolute Deviation (MAD) and an $M-$estimator of scale ($M-$SCALE). The latter two are robust scale statistics. The $M-$estimator combines both the robustness of the (MAD) with the smoothness of the standard deviation. For the $M-$estimator, we used as score function $\chi_c(y) = \min\left(3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6\,, 1\right)$, introduced by Beaton and Tukey (1974), with tuning constant $c = 1.56$ and breakdown point 1/2. To compute the $M-$scale, the initial estimator of scale was the MAD.

- **Centering:** For the classical procedures, i.e., those based on SD, we used a point–to–point mean as the centering point. For the robust procedures, i.e., those based on MAD or $M-$SCALE, we used either the L^1 median, which is commonly referred to as the spatial median, or the point–to–point median to center the data. This avoids the extra complexity associated with the functional trimmed means or the depth–based estimators. It turned out that the two robust centering methods produced similar results and so, only the results for the L^1 median are reported.

- **Smoothing level τ:** For both the classical and robust procedures defined in Section 2.2, a penalization depending on the L^2 norm of the second derivative is included, multiplied by a smoothing factor. Note that when $\tau = 0$, the raw estimators defined in Section 2.1 are obtained. We also considered smoothing the directional candidates in our algorithm, by using a kernel smoother for the classical procedures and a local median for the robust ones. However, this turned out to be extremely time consuming, without any noticeable difference in the results.

- **Sieve:** Two different sieve bases were considered: the Fourier basis, i.e., taking δ_j to be the Fourier basis functions, and the cubic $B-$spline basis functions. The Fourier basis used in the sieve method is the same basis used to generate the data.

In all Figures and Tables, the estimators corresponding to each scale choice are labeled as SD, MAD, $M-$SCALE. For each scale, we considered four estimators, the raw estimators where no smoothing is used, the estimators obtained by penalizing the scale function defined in (7), those obtained by penalizing the norm defined in (6), and the sieve estimators defined in (10). In all Tables, as in Section 2, the jth principal direction estimators related to each method are labelled as $\widehat{\phi}_j$, $\widehat{\phi}_{\text{PS},j}$, $\widehat{\phi}_{\text{PN},j}$ and $\widehat{\phi}_{\text{SI},j}$, respectively.

When using the penalized estimators, several values for the penalizing parameters τ and ρ were chosen. Since large values of the smoothing parameters make the penalizing term to be the dominant component independently of the amount of contamination considered, we choose τ and ρ equal to $an^{-\alpha}$ for $\alpha = 3$ and 4 and a equal to 0.05, 0.10, 0.15, 0.25, 0.5, 0.75, 1, 1.5 and 2. However, boxplots and density estimators are given only when $\alpha = 3$ and $a = 0.25$, 0.75 and 2.

For the sieve estimators based on the Fourier basis, ordered as $\{1, \cos(\pi x), \sin(\pi x), \ldots, \cos(q_n \pi x), \sin(q_n \pi x), \ldots\}$, the values $p_n = 2q_n$ with $q_n = 5$, 10 and 15 were used, while for the sieve estimators based on the $B-$splines, the dimension of the linear space considered was selected as $p_n = 10$, 20 and 50. The basis for the B-splines is generated from the R function *cSplineDes*, with the knots being equally spaced in the interval [-1,1] and the number of knots equal to $p_n + 1$. The resulting $B-$spline basis, though, is not orthonormal. Since it is easier to apply the algorithm for the sieve estimators when an orthonormal basis is used, a Gram-Schmidt orthogonalization is applied to the $B-$splines basis to obtained a new orthonormal bases spanning the same subspace.

5.2 Simulation settings

The sample was generated using a finite Karhunen-Loève expansion with the functions, $\phi_i : [-1, 1] \to \mathbb{R}$, $i = 1, 2, 3$, where

$$\begin{aligned} \phi_1(x) &= \sin(4\pi x) \\ \phi_2(x) &= \cos(7\pi x) \\ \phi_3(x) &= \cos(15\pi x)\ . \end{aligned}$$

It is worth noticing that, when considering the sieve estimators based on the Fourier basis, the third component cannot be detected when $q_n < 15$, since in this case $\phi_3(x)$ is orthogonal to the estimating space. Likewise, the second component cannot be detected when $q_n < 7$.

We performed $NR = 1000$ replications generating independent samples $\{X_i\}_{i=1}^n$ of size $n = 100$ following the model $X_i = Z_{i1}\phi_1 + Z_{i2}\phi_2 + Z_{i3}\phi_3$, where Z_{ij} are independent random variables whose distribution will depend on the situation to be considered. The central model, denoted C_0, corresponds to Gaussian samples. We have also considered four contaminations of the central model, labelled C_2, $C_{3,a}$, $C_{3,b}$ and C_{23} depending on the components to be contaminated. The central model and the contaminations can be described as follows. For each of the models, we took $\sigma_1 = 4$, $\sigma_2 = 2$ and $\sigma_3 = 1$.

- C_0: $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ and $Z_{i3} \sim N(0, \sigma_3^2)$.
- C_2: Z_{i2} are independent and identically distributed as $0.8\ N(0, \sigma_2^2) + 0.2\ N(10, 0.01)$, while $Z_{i1} \sim N(0, \sigma_1^2)$ and $Z_{i3} \sim N(0, \sigma_3^2)$. This contamination corresponds to a strong contamination on the second component and changes the mean value of the generated data Z_{i2} and also the first principal component. Note that $\text{var}(Z_{i2}) = 19.202$.
- $C_{3,a}$: $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ and $Z_{i3} \sim 0.8\ N(0, \sigma_3^2) + 0.2\ N(15, 0.01)$. This contamination corresponds to a strong contamination on the third component. Note that $\text{var}(Z_{i3}) = 36.802$.
- $C_{3,b}$: $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ and $Z_{i3} \sim 0.8\ N(0, \sigma_3^2) + 0.2\ N(6, 0.01)$. This contamination corresponds to a strong contamination on the third component. Note that $\text{var}(Z_{i3}) = 6.562$.
- C_{23}: Z_{ij} are independent and such that $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim 0.9N(0, \sigma_2^2) + 0.1N(15, 0.01)$ and $Z_{i3} \sim 0.9N(0, \sigma_3^2) + 0.1N(20, 0.01)$. This contamination corresponds to a mild contamination on the two last components. Note that $\text{var}(Z_{i2}) = 23.851$, and $\text{var}(Z_{i3}) = 36.901$.

We also considered a Cauchy situation, labelled C_c, defined by taking $(Z_{i1}, Z_{i2}, Z_{i3}) \sim \mathcal{C}_3(0, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, where $\mathcal{C}_p(0, \mathbf{\Sigma})$ stands for the $p-$dimensional elliptical Cauchy distribution centered at 0 with scatter matrix $\mathbf{\Sigma}$. For this situation, the covariance operator does not exist and thus the classical principal components are not defined.

It is worth noting that the directions ϕ_1, ϕ_2 and ϕ_3 correspond to the classical principal components for the case C_0, but not necessarily for the other cases. For instance, $C_{3,a}$ interchanges the order between ϕ_1 and ϕ_3, i.e., ϕ_3 is now the first classical principal component, i.e., that obtained from the covariance operator, while ϕ_1 is the second and ϕ_2 is the third.

5.3 Simulation results

For each situation, we compute the estimators of the first three principal components and the square distance between the true and the estimated direction (normalized to have L^2 norm 1), i.e.,

$$D_j = \left\| \frac{\widehat{\phi}_j}{\|\widehat{\phi}_j\|} - \phi_j \right\|^2 .$$

Table 7 to 12 give the mean of D_j over replications for the raw and penalized estimators. Table 7 corresponds to the raw and penalized estimators under C_0 for the different choices of the penalizing parameters. This table allows to see that a better performance is achieved in most cases with $\alpha = 3$. Hence, as mentioned above, all the Figures correspond to values of the smoothing parameter equal to $\tau = an^{-3}$. To be more precise, the results in Table 7 show that the best choice for $\widehat{\phi}_{\text{PS},j}$ is $\tau = 2n^{-3}$ for all jNote that $\rho = 1.5n^{-3}$ give quite similar results, when using the $M-$scale, reducing the error in about a half and a third for $j = 2$ and 3, respectively. When penalizing the norm, i.e., when considering $\widehat{\phi}_{\text{PN},j}$ the choice of the penalizing parameter seems to depend both on the component to be estimated and on the estimator to be used. For instance, when using the standard deviation, the best choice is $0.10n^{-3}$, for $j = 1$ and 2 while for $j = 3$ a smaller order is needed to obtain an improvement over the raw estimators. The value $\tau = 0.75n^{-4}$ leads to a small gain over the raw estimators. For the robust procedures, larger values are needed to see the advantage of the penalized approach over the raw estimators. For instance, for $j = 1$, the larger reduction is observed when $\tau = 2n^{-3}$ while for $j = 2$, the best choices correspond to $\tau = 0.5n^{-3}$ and $\tau = 0.25n^{-3}$ when using the MAD and $M-$scale, respectively. For instance, when using the $M-$scale, choosing $\tau = 0.75n^{-3}$ lead to a reduction of about 30% and 50% for the first and second principal directions, respectively. On the other hand, when estimating the third component, again smaller values of τ are needed. Tables 4 and 5 report the mean of D_j over replications for different sizes of the grid under C_0 for some values of the penalizing parameters. The size $m = 50$ selected in our study gives a compromise between the performance of the estimators and the computing time. As it can be seen, some improvement is observed when using $m = 250$ instead of 50 points but at the expense of multiplying by five the computing time.

Besides, Tables 13 to 18 give the mean of D_j over replications for the sieve estimators. Figures 1 to 6 show the density estimates of D_j, for $j = 1, 2$ and 3, respectively when $\alpha = 3$ combined with $a = 1.5$ for the estimators penalizing the scale and $a = 0.75$ for those penalizing the norm. The density estimates were evaluated using the normal kernel with bandwidth 0.6 in all cases. The plots given in black correspond to the densities of D_j evaluated over the $NR = 1000$ normally distributed samples, while those in red, gray, blue and green correspond to C_2, $C_{3,a}$, $C_{3,b}$ and C_{23}, respectively. Finally, Figures 8 to 13 show the boxplots of the ratio $\widehat{\lambda}_m/\lambda_m$ for the different eigenvalue estimators. The classical estimators are labelled SD, while the robust ones MAD and MS. For the norm or scale–penalized estimators the penalization parameter τ is indicated after the estimator type label while for the sieve estimators the parameter p_n follows the name of the scale estimator considered. For the Cauchy distribution, the large values obtained for the classical estimators obscure any differences within the robust procedures and so, separate boxplots for the robust methods only are given at the bottom of Figures 8 to 13.

The simulation confirms the expected inadequate behaviour of the classical estimators, in the presence of outliers. A bias is observed when estimating the eigenvalues. The poorest efficiency of the raw eigenvalue estimates is obtained using the projection–pursuit procedure combined with the MAD estimator. It is also worth noticing that the level of smoothing τ seems to affect the eigenvalue estimators, introducing a bias even for Gaussian samples. Note that for some contaminations, the robust estimators are also biased. However, the order among them is preserved and so, the target eigenfunction is in most cases, recovered.

With respect to the principal direction estimation, under contamination, the classical estimators do not estimate the target eigenfunctions very accurately, which can be seen from the shift in the density of the norm towards 2. Note that when considering the Cauchy distribution, the main effect is observed on the eigenvalue estimators since, even if the covariance operator does not exist, the directions seem to be recovered when using the standard deviation. The robust eigenfunction estimators seem to be affected mainly unaffected by all the contaminations except by $C_{3,a}$. In particular, the projection–pursuit estimators based on an $M-$scale seem to be more affected by this contamination. On the other hand, $C_{3,a}$ affects the estimators of the third eigenfunction when penalizing the norm. With respect to C_{3a}, the robust estimators obtained penalizing the norm $\widehat{\phi}_{\text{PN},j}$ show the lower effect among all the competitors. Note that even if the order of the classical eigenfunctions is modified, as mentioned above, the robust estimators of the first principal direction are not affected by this contamination.

It is worth noting that the classical estimators of the first component are not affected by $C_{3,a}$ for some values of the smoothing parameter, when penalizing the norm since the penalization dominates over the contaminated variances. The same phenomena is observed under $C_{3,b}$ when using the classical estimators for the selected amount of penalization. For the raw estimators, the sensitivity of the classical estimators under this contamination can be observed in Table 6.

As noted in Silverman (1996), for the classical estimators, some degree of smoothing in the procedure based on penalizing the norm will give a better estimation of ϕ_j in the L^2 sense under mild conditions. In particular, both the procedure penalizing the norm and the scale provide some improvement with respect to the raw estimators if $\Psi(\phi_j) < \Psi(\phi_\ell)$, when $j < \ell$. This means that the principal directions are rougher as the eigenvalues decrease (see Pezzulli and Silverman, 1993 and Silverman, 1996), which is also reflected in our simulation study. The advantages of the smooth projection pursuit procedures are most striking when estimating ϕ_2 and ϕ_3 with an $M-$scale and using the penalized scale approach.

As expected, when using the sieve estimators, the Fourier basis gives the best performance over all the methods under C_0, since our data set was generated using this basis (see Table 13). The choice of the $B-$spline basis give results quite similar to those obtained with $\widehat{\phi}_{\text{PS},j}$.

5.4 $K-$th fold simulation

Table 1 reports the computing times in minutes for 1000 replications and for a fixed value of τ. This suggests that the leave-one-out cross–validation may be difficult to perform, and so a $K-$fold approach is adopted instead. A simulation study was performed where the smoothing parameter τ was selected using the procedure described in Section 4 with $K = 4$, $\ell = 1$. We performed 500 replications. The results when penalizing the scale function, i.e., for the estimators defined through (7), are reported in Table 2 and in Figure 7. The classical estimators are sensitive to the considered contaminations and except for contaminations in the third component, the robust counterpart show their advantage. Note that both C_{3a} and C_{3b} affect the robust estimators when the smoothing parameter τ is selected by the robust $K-$fold cross-validation method.

	SD	MAD	$M-$SCALE
Raw	5.62	6.98	17.56
Smoothed	7.75	9.00	20.18
Smoothed Norm	31.87	33.21	44.04

Table 1: Computing times in minutes for 1000 replications and a fixed value of τ.

Model	Scale estimator	$j = 1$	$j = 2$	$j = 3$
			$\widehat{\phi}_{\text{PS},j}$	
C_0	SD	0.0073	0.0094	0.0078
	MAD	0.0662	0.0993	0.0634
	M-scale	0.0225	0.0311	0.0172
C_2	SD	1.2840	1.2837	0.0043
	MAD	0.3731	0.3915	0.0504
	M-scale	0.4261	0.4286	0.0153
C_{3A}	SD	1.7840	1.8901	1.9122
	MAD	0.2271	0.5227	0.5450
	M-scale	0.2176	0.4873	0.5437
C_{3B}	SD	0.0192	0.8350	0.8525
	MAD	0.0986	0.3930	0.3820
	M-scale	0.0404	0.2251	0.2285
C_{23}	SD	1.7645	0.5438	1.6380
	MAD	0.2407	0.3443	0.2064
	M-scale	0.2613	0.3707	0.2174
C_{Cauchy}	SD	0.3580	0.4835	0.2287
	MAD	0.0788	0.1511	0.1082
	M-scale	0.0444	0.0707	0.0434

Table 2: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ when the penalizing parameter is selected using $K-$fold cross–validation.

6 Concluding Remarks

In this paper, we consider robust principal component analysis for functional data based on a projection–pursuit approach. The different procedures correspond to robust versions of the unsmoothed principal component estimators, to the estimators obtained penalizing the scale and to those obtained by penalizing the norm. A sieve approach based on approximating the elements of the unit ball by elements over finite–dimensional spaces is also considered. A robust cross-validation procedure is introduced to select the smoothing parameters. Consistency results are derived for the four type of estimators. Moreover, the functional related to the unsmoothed estimators is shown to be continuous and so, the related estimators are asymptotically robust.

The simulation study confirms the expected inadequate behaviour of the classical estimators in the presence of outliers, with the robust procedures performing significantly better. The proposed robust procedures themselves for the eigenfunctions, however, perform quite similarly to each other under the contaminations studied. A study of the influence functions and the asymptotic distributions of the different robust procedures would be useful for differentiating between them. We leave these important and challenging theoretical problems, though, for future research.

A Appendix A

In this Appendix, we provide conditions under which **S1** hold by requiring continuity to the scale functional. To derive these results, we will first derive some properties regarding the weak convergence of empirical measures that hold not only in $L^2(\mathcal{I})$ but in any complete and separable metric space.

Let $\mathcal{M}$ be a complete and separable metric space (Polish space) and $\mathcal{B}$ the Borel $\sigma-$algebra of $\mathcal{M}$. The Prohorov distance between two probability measures P and Q on $\mathcal{M}$ is defined as: $d_{\rm PR}(P,Q) = \inf\{\epsilon, P(A) \leq Q(A^\epsilon) + \epsilon,\ \forall A \in \mathcal{B}\}$, where $A^\epsilon = \{x \in \mathcal{M}, d(x,A) < \epsilon\}$. Theorem A.1 shows that, analogously to the Glivenko–Cantelli Theorem in finite–dimensional spaces, on a Polish space the empirical measures converge weakly almost surely to the probability measure generating the observations.

Theorem A.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_n : \Omega \to \mathcal{M}$, $n \in \mathbb{N}$, be a sequence of independent and identically distributed random elements such that $X_i \sim P$. Assume that $\mathcal{M}$ is a Polish space and denote by P_n the the empirical probability measure, that is, $P_n(A) = \frac{1}{n}\sum_1^n I_A(X_i)$ with $I_A(X_i) = 1$ if $X_i \in A$ and 0 elsewhere. Then, $P_n \xrightarrow{\omega} P$ almost surely, i.e., $d_{\rm PR}(P_n, P) \xrightarrow{a.s.} 0$.*

Proof. Note that the strong law of large numbers entails that for any borelian set A, $P_n(A) \xrightarrow{a.s.} P(A)$, i.e., $P_n(A) \to P(A)$ except for a set $\mathcal{N}_A \subset \Omega$ of $\mathbb{P}$-measure zero.

Let us show that given $j \in \mathbb{N}$, there exists $\mathcal{N}_j \subset \Omega$ such that $\mathbb{P}(\mathcal{N}_j) = 0$ and, for any $\omega \notin \mathcal{N}_j$, there exists $n_j(\omega) \in \mathbb{N}$ such that if $n \geq n_j(\omega)$, then $d_{\rm PR}(P_n, P) < 1/j$.

The fact that $\mathcal{M}$ is a Polish space entails that there exists a finite class of disjoint sets $\{A_i, 1 \leq i \leq k\}$ with diameter smaller than $\frac{1}{2j}$ such that

$$P\left(\bigcup_{i=1}^k A_i\right) > 1 - \frac{1}{2j}. \tag{A.1}$$

Denote by $\mathcal{A}$ the class of all the sets that are obtained as a finite union of the A_i, i.e., $B \in \mathcal{A}$ if and only if there exists $A_{i_1}, \dots, A_{i_\ell}$ such that $B = \cup_{j=1}^\ell A_{i_j}$. Note that $\mathcal{A}$ has a finite number of elements s. For each $1 \leq i \leq s$, and $B_i \in \mathcal{A}$, let $\mathcal{N}_{B_i} \subset \Omega$ with $\mathbb{P}(\mathcal{N}_{B_i}) = 0$ such that if $\omega \notin \mathcal{N}_{B_i}$, then $|P_n(B_i) - P(B_i)| \to 0$. We define $\mathcal{N}_j = \bigcup_{i=1}^s \mathcal{N}_{B_i}$, then $\mathbb{P}(\mathcal{N}_j) = 0$.

Let $\omega \notin \mathcal{N}_j$, then we have that $|P_n(B_i) - P(B_i)| \to 0$, for $1 \leq i \leq s$. Hence, there exists $n_j(\omega) \in \mathbb{N}$ such that for $n \geq n_j(\omega)$ we have that $|P_n(B) - P(B)| < \frac{1}{2j}$ for any $B \in \mathcal{A}$. We will now show if $n \geq n_j(\omega)$ then $d_{\text{PR}}(P_n, P) < 1/j$.

Consider B a borelian set and let A be the union of all the sets A_i that intersect B. Note that $A \in \mathcal{A}$ and so $|P_n(A) - P(A)| < \frac{1}{2j}$. Therefore, $B \subset A \cup \left(\bigcup_{i=1}^{k} A_i\right)^c$ and $A \subset B^{1/j}$. This last inclusion holds because the sets A_i have diameter smaller than $\frac{1}{2j}$. Thus, using (A.1), we get that

$$P(B) \leq P(A) + P\left[\left(\bigcup_{i=1}^{k} A_i\right)^c\right] = P(A) + 1 - P\left[\left(\bigcup_{i=1}^{k} A_i\right)\right] < P(A) + \frac{1}{2j},$$

which together with the fact that $|P_n(A) - P(A)| < \frac{1}{2j}$ implies that $P(B) \leq P(A) + \frac{1}{2j} < P_n(A) + 1/j$. Using that $A \subset B^{1/j}$, we get that $P_n(A) + 1/j \leq P_n(B^{1/j}) + 1/j$, so $P(B) < P_n(B^{1/j}) + 1/j$ and this holds for every B borelian set. Thus, $d_{\text{PR}}(P_n, P) < 1/j$, as it was desired.

To conclude the proof, we will show that $d_{\text{PR}}(P_n, P) \to 0$ except for a zero $\mathbb{P}$-measure set. Consider all the sets $\mathcal{N}_j$ previously defined and let $\mathcal{N} = \bigcup_{j\in\mathbb{N}} \mathcal{N}_j$. It is clear that $\mathbb{P}(\mathcal{N}) = 0$. Thus, for any $\omega \notin \mathcal{N}$, we will have that for each j there exists $n_j = n_j(\omega)$ such that $d(P_n, P) < 1/j$ if $n \geq n_j$. This concludes the proof. □

Let P be a probability measure in $\mathcal{M}$, a separable Banach space. Then, given $f \in \mathcal{M}^*$, where $\mathcal{M}^*$ stands for the dual space, define $P[f]$ as the real measure of the random variable $f(X)$, with $X \sim P$. Then, we have that

Theorem A.2. *Let $\{P_n\}_{n\in\mathbb{N}}$ and P be probability measures defined on $\mathcal{M}$ such that $P_n \xrightarrow{\omega} P$, i.e., $d_{\text{PR}}(P_n, P) \to 0$. Then, $\sup_{\|f\|_*=1} d_{\text{PR}}(P_n[f], P[f]) \to 0$.*

Proof. Fix $\epsilon > 0$ and let n_0 be such that $d_{\text{PR}}(P_n, P) < \epsilon$, for $n \geq n_0$. We will show that $\sup_{\|f\|_*=1} d_{\text{PR}}(P_n[f], P[f]) < \epsilon$, for $n \geq n_0$. Fix $n \geq n_0$.

Using that $d_{\text{PR}}(P_n, P) < \epsilon$ and Strassen's Theorem, we get that there exists $\{X_n\}_{n\in\mathbb{N}}$ and X in $\mathcal{M}$ such that $X_n \sim P_n$, $X \sim P$ and $\mathbb{P}(\|X_n - X\| \leq \epsilon) > 1 - \epsilon$. Note that for any $f \in \mathcal{M}^*$, with $\|f\|_* = 1$, $f(X_n) \sim P_n[f]$ and $f(X) \sim P[f]$. Using that $|f(X_n) - f(X)| = |f(X_n - X)| \leq \|f\|_* \|X_n - X\| \leq \|X_n - X\|$, we get that for any $f \in M^*$, such that $\|f\|_* = 1$,

$$\{\|X_n - X\| \leq \epsilon\} \subseteq \{|f(X_n) - f(X)| \leq \epsilon\}$$

which entails that

$$1 - \epsilon < \mathbb{P}(\|X_n - X\| \leq \epsilon\ ) \leq \mathbb{P}(|f(X_n) - f(X)| \leq \epsilon),\ \forall f \in M^*,\ \|f\|_* = 1.$$

Thus, $\mathbb{P}(|f(X_n) - f(X)| \leq \epsilon) > 1 - \epsilon$, and so, using again Strassen's Theorem, we get that

$$P_n[f](A) \leq P[f](A^\epsilon) + \epsilon,\ \forall A \in \mathcal{B},\ \forall f \in M^*, \|f\|_* = 1.$$

Therefore, for any $f \in M^*$ such that $\|f\|_* = 1$, we have that $d_{\text{PR}}(P_n[f], P[f]) \leq \epsilon$, i.e., $\sup_{\|f\|_*=1} d_{\text{PR}}(P_n[f], P[f]) \leq \epsilon$ concluding the proof. □

In the particular, when considering a separable Hilbert space $\mathcal{H}$, if $f \in \mathcal{H}^*$ is such that $\|f\|_* = 1$, then $f(X) = \langle \alpha, X\rangle$ with $\|\alpha\| = 1$. The following result states that when σ_{R} is a continuous scale functional, uniform convergence can be attained.

Theorem A.3. *Let $\{P_n\}_{n\in\mathbb{N}}$ and P be probability measures defined on a separable Hilbert space $\mathcal{H}$, such that $P_n \xrightarrow{\omega} P$, i.e., $d_{\text{PR}}(P_n, P) \to 0$. Let σ_{R} be a continuous scale functional. Then, $\sup_{\|\alpha\|=1} |\sigma_{\text{R}}(P_n[\alpha]) - \sigma_{\text{R}}(P[\alpha]))| \longrightarrow 0$.*

PROOF. Denote by $a_n = \sup_{\|\alpha\|=1} |\sigma_{\mathrm{R}}(P_n[\alpha]) - \sigma_{\mathrm{R}}(P[\alpha]))|$, it is enough to show that $L = \limsup_{n\to\infty} a_n = 0$.

First note that since $\mathcal{S} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$ is weakly compact and σ_{R} is a continuous functional, for each fixed n such that $a_n \neq 0$, there exists $\alpha_n \in \mathcal{S}$ such that

$$a_n = |\sigma_{\mathrm{R}}(P_n[\alpha_n]) - \sigma_{\mathrm{R}}(P[\alpha_n])) \ . \tag{A.2}$$

Effectively, let $\gamma_\ell \in \mathcal{S}$ be such that $|\sigma_{\mathrm{R}}(P_n[\gamma_\ell]) - \sigma_{\mathrm{R}}(P[\gamma_\ell]))| \to a_n$, then the weak compactness of $\mathcal{S}$, entails that there exists a subsequence γ_{ℓ_s} such that γ_{ℓ_s} converges weakly to $\gamma \in \mathcal{H}$. It is easy to see that $\|\gamma\| \leq 1$. Besides, using that σ_{R} is continuous we obtain that $|\sigma_{\mathrm{R}}(P_n[\gamma_{\ell_s}]) - \sigma_{\mathrm{R}}(P[\gamma_{\ell_s}]))| \to |\sigma_{\mathrm{R}}(P_n[\gamma]) - \sigma_{\mathrm{R}}(P[\gamma]))|$, as $s \to \infty$. Hence, $|\sigma_{\mathrm{R}}(P_n[\gamma]) - \sigma_{\mathrm{R}}(P[\gamma]))| = a_n$ which entails that $\gamma \neq 0$. Let $\widetilde{\gamma} = \gamma/\|\gamma\|$, then $\widetilde{\gamma} \in \mathcal{S}$ and thus $|\sigma_{\mathrm{R}}(P_n[\widetilde{\gamma}]) - \sigma_{\mathrm{R}}(P[\widetilde{\gamma}]))| \leq a_n$. On the other hand, using that σ_{R} is a scale functional we get that

$$|\sigma_{\mathrm{R}}(P_n[\widetilde{\gamma}]) - \sigma_{\mathrm{R}}(P[\widetilde{\gamma}]))| = \frac{|\sigma_{\mathrm{R}}(P_n[\gamma]) - \sigma_{\mathrm{R}}(P[\gamma]))|}{\|\gamma\|} = \frac{a_n}{\|\gamma\|}$$

which implies that $\|\gamma\| \geq 1$ leading to $\|\gamma\| = 1$ and to the existence of a sequence $\alpha_n \in \mathcal{S}$ satisfying (A.2).

Let a_{n_k} be a subsequence such that $a_{n_k} \to L$, we will assume that $a_{n_k} \neq 0$. Then, using (A.2), we have that $\alpha_{n_k} \in \mathcal{S}$ such that $a_{n_k} = |\sigma_{\mathrm{R}}(P_{n_k}[\alpha_{n_k}]) - \sigma_{\mathrm{R}}(P[\alpha_{n_k}]))| \to L$. Using that $\mathcal{S}$ is weakly compact, we can choose a subsequence $\beta_j = \alpha_{n_{k_j}}$ such that β_j converges weakly to β, i.e., for any $\alpha \in \mathcal{H}$, $\langle \beta_j, \alpha\rangle \to \langle \beta, \alpha\rangle$. Note that since $\|\beta_j\| = 1$, then $\|\beta\| \leq 1$ (β could be 0) and that

$$a_{n_{k_j}} = |\sigma_{\mathrm{R}}(P_{n_{k_j}}[\beta_j]) - \sigma_{\mathrm{R}}(P[\beta_j]))| \to L \tag{A.3}$$

For the sake of simplicity denote $P^{(j)} = P_{n_{k_j}}$. Then, Theorem A.2 entails that

$$d_{\mathrm{PR}}(P^{(j)}[\beta_j], P[\beta_j]) \leq \sup_{\|\alpha\|=1} d_{\mathrm{PR}}(P^{(j)}[\alpha], P[\alpha]) \to 0$$

while the fact that β_j converges weakly to β implies that $d_{\mathrm{PR}}(P[\beta_j], P[\beta]) \to 0$, concluding that $d_{\mathrm{PR}}(P^{(j)}[\beta_j], P[\beta]) \to 0$. The continuity of σ_{R} leads to

$$\sigma_{\mathrm{R}}(P^{(j)}[\beta_j]) \to \sigma_{\mathrm{R}}(P[\beta]) \ . \tag{A.4}$$

Using again that β_j converges weakly to β and the weak continuity of σ_{R} we get that

$$\sigma_{\mathrm{R}}(P[\beta_j]) \to \sigma_{\mathrm{R}}(P[\beta]) \ . \tag{A.5}$$

Thus, (A.4) and (A.5) imply that $\sigma_{\mathrm{R}}(P^{(j)}[\beta_j]) - \sigma_{\mathrm{R}}(P[\beta_j]) \to 0$ and so, from (A.3), $L = 0$, concluding the proof. □

Moreover, using Theorem A.1, we get the following result that shows that **S1** holds if σ_R is a continuous scale functional.

Corollary A.1 *Let P be a probability measure in a separable Hilbert space $\mathcal{H}$, P_n be the empirical measure of a random sample $X_1, \ldots, X_n$ with $X_i \sim P$, and σ_R be a continuous scale functional. Then, we have that*

$$\sup_{\|\alpha\|=1} |\sigma_{\mathrm{R}}(P_n[\alpha]) - \sigma_{\mathrm{R}}(P[\alpha]))| \xrightarrow{a.s.} 0.$$

B Appendix B: Proofs

Proof of Lemma 3.1. a) Let $\mathcal{N} = \{\omega : \sigma^2(\widehat{\phi}_1(\omega)) \not\to \sigma^2(\phi_{\mathrm{R},1})\}$ and fix $\omega \notin \mathcal{N}$, then $\sigma^2(\widehat{\phi}_1(\omega)) \to \sigma^2(\phi_{\mathrm{R},1})$. Using that $\mathcal{S}$ is weakly compact, we have that for any subsequence γ_ℓ of $\widehat{\phi}_1(\omega)$ there exists a subsequence γ_{ℓ_s} such that γ_{ℓ_s} converges weakly to $\gamma \in \mathcal{H}$. It is easy to see that $\|\gamma\| \leq 1$. Besides, using that $\sigma^2(\widehat{\phi}_1(\omega)) \to \sigma^2(\phi_{\mathrm{R},1})$, we get that $\sigma^2(\gamma_{\ell_s}) \to \sigma^2(\phi_{\mathrm{R},1})$ while on the other hand, the weakly continuity of σ entails that $\sigma^2(\gamma_{\ell_s}) \to \sigma^2(\gamma)$, as $s \to \infty$. Hence, $\sigma^2(\gamma) = \sigma^2(\phi_{\mathrm{R},1})$ which entails that $\gamma \neq 0$. Let $\widetilde{\gamma} = \gamma/\|\gamma\|$, then $\widetilde{\gamma} \in \mathcal{S}$ and thus $\sigma^2(\widetilde{\gamma}) \leq \sigma^2(\phi_{\mathrm{R},1})$. On the other hand, using that σ_{R} is a scale functional we get that

$$\sigma(\widetilde{\gamma}) = \frac{\sigma(\gamma)}{\|\gamma\|} = \frac{\sigma(\phi_{\mathrm{R},1})}{\|\gamma\|}$$

which implies that $\|\gamma\| \geq 1$ leading to $\|\gamma\| = 1$ and so, using the uniqueness of $\phi_{\mathrm{R},1}$ we obtain that $\langle \gamma, \phi_{\mathrm{R},1}\rangle^2 = 1$. Therefore, since any subsequence of $\widehat{\phi}_1(\omega)$ will have a limit converging either to $\phi_{\mathrm{R},1}$ or $-\phi_{\mathrm{R},1}$, we obtain a).
b) Write $\widehat{\phi}_m$ as $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{\mathrm{R},j} + \widehat{\gamma}_m$, with $\langle \widehat{\gamma}_m, \phi_{\mathrm{R},j}\rangle = 0$, $1 \leq j \leq m-1$. To obtain b) we only have to show that $\langle \widehat{\gamma}_m, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$. Note that $\langle \widehat{\phi}_m, \widehat{\phi}_j\rangle = 0$, for $j \neq m$, implies that $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{\mathrm{R},j}\rangle = \langle \widehat{\phi}_m, \phi_{\mathrm{R},j} - \widehat{\phi}_j\rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j\rangle = \langle \widehat{\phi}_m, \phi_{\mathrm{R},j} - \widehat{\phi}_j\rangle$. Thus, using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$, $1 \leq j \leq m-1$, and $\|\widehat{\phi}_m\| = 1$, we get that $\widehat{a}_j \xrightarrow{a.s.} 0$ for $1 \leq j \leq m-1$ and so, $\|\widehat{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$. Note that $1 = \|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$, hence, $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$ which implies that $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, where $\widetilde{\gamma}_m = \widehat{\gamma}_m/\|\widehat{\gamma}_m\|$. Using that $\sigma(\alpha)$ is a weakly continuous function and the unit ball is weakly compact, we obtain that

$$\sigma(\widetilde{\gamma}_m) - \sigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0\,. \tag{A.6}$$

Effectively, let $\mathcal{N} = \Omega - \{\omega : \|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0\}$, then $\mathbb{P}(\mathcal{N}) = 1$. Fix $\omega \notin \mathcal{N}$ and let $b_n = \sigma(\widetilde{\gamma}_m) - \sigma(\widehat{\phi}_m) = \sigma(\widetilde{\gamma}_{n,m}) - \sigma(\widehat{\phi}_{n,m})$. It is enough to show that every subsequence of $\{b_n\}$ converges to 0. Denote by $\{b_{n'}\}$ a subsequence, then by the weak compactness of $\mathcal{S}$, there exists a subsequence $\{n_j\} \subset \{n'\}$ such that $\widetilde{\gamma}_{n_j,m})$ and $\widehat{\phi}_{n_j,m})$ converge weakly to γ and ϕ, respectively. The fact that $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \to 0$, we get that $\gamma = \phi$ and so the weak continuity of σ entails that $b_{n_j} \to 0$.

The fact that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and (A.6) imply that $\sigma(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma(\phi_{\mathrm{R},m})$. The proof follows now as in a) using the fact that $\widetilde{\gamma}_m \in \mathcal{C}_m$, with $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{\mathrm{R},j}\rangle = 0,\ 1 \leq j \leq m-1\}$ and $\phi_{\mathrm{R},m}$ is the unique maximizer of $\sigma(\alpha)$ over $\mathcal{C}_m$. □

Proof of Proposition 3.1. For the sake of simplicity denote by $\sigma_n(\alpha) = \sigma_{\mathrm{R}}(P_n[\alpha])$, $\widehat{\phi}_m = \phi_{\mathrm{R},m}(P_n)$ and $\widehat{\lambda}_m = \lambda_{\mathrm{R},m}(P_n)$. Moreover, let $\widehat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_j\rangle = 0\ ,\ \forall\ 1 \leq j \leq m-1\}$ and $\widehat{\mathcal{L}}_{m-1}$ the linear space spanned by $\widehat{\phi}_1, \ldots, \widehat{\phi}_{m-1}$.

a) Using ii), we get that $a_{n,1} = \sigma_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \to 0$ and $b_{n,1} = \sigma_n^2(\phi_{\mathrm{R},1}) - \sigma^2(\phi_{\mathrm{R},1}) \to 0$ which implies that

$$\sigma^2(\phi_{\mathrm{R},1}) = \sigma_n^2(\phi_{\mathrm{R},1}) - b_{n,1} \leq \sigma_n^2(\widehat{\phi}_1) - b_{n,1} = \sigma^2(\widehat{\phi}_1) + a_{n,1} - b_{n,1} \leq \sigma^2(\phi_{\mathrm{R},1}) + a_{n,1} - b_{n,1} = \sigma^2(\phi_{\mathrm{R},1}) + o(1),$$

where $o(1)$ stands for a term converging to 0. Therefore, $\sigma^2(\phi_{\mathrm{R},1}) \leq \sigma^2(\widehat{\phi}_1) + o(1) \leq \sigma^2(\phi_{\mathrm{R},1}) + o(1)$, which entails that $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$, concluding the proof of a).

Note that we have not used the weak continuity of σ as a function of α to derive a).

b) Follows as in Lemma 3.1 a).

c) Let $2 \leq m \leq q$, be fixed and assume that $\widehat{\phi}_s \to \phi_{\mathrm{R},s}$, for $1 \leq s \leq m-1$. We will begin by showing that $\widehat{\lambda}_m \to \lambda_{\mathrm{R},m}$.

$$|\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| \;=\; |\max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma_n^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha)| \leq \max_{\alpha \in \widehat{\mathcal{B}}_m} |\sigma_n^2(\alpha) - \sigma^2(\alpha)| + |\max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha)|$$

$$\leq \max_{\|\alpha\|=1} |\sigma_n^2(\alpha) - \sigma^2(\alpha)| + |\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha\in\mathcal{B}_m} \sigma^2(\alpha)|$$
$$\leq o(1) + |\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha\in\mathcal{B}_m} \sigma^2(\alpha)| = o(1) + |\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) - \sigma^2(\phi_{\mathrm{R},m})| \;.$$

Thus, in order to obtain the desired result, it only remains to show $\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) \to \sigma^2(\phi_{\mathrm{R},m})$. We will show that

$$\sigma^2(\phi_{\mathrm{R},m}) \leq \max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) + o(1) \tag{A.7}$$
$$\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) \leq o(1) + \sigma^2(\phi_{\mathrm{R},m}) \tag{A.8}$$

Using that $\widehat{\phi}_s \to \phi_{\mathrm{R},s}$ for $1 \leq s \leq m-1$, we obtain that $\|\pi_{\widehat{\mathcal{L}}_{m-1}} - \pi_{\mathcal{L}_{m-1}}\| \to 0$. In particular, we have that $\|\pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m} - \pi_{\mathcal{L}_{m-1}}\phi_{\mathrm{R},m}\| \to 0$, which, together with the fact that $\pi_{\mathcal{L}_{m-1}}\phi_{\mathrm{R},m} = 0$, implies that $\pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m} \to 0$ and so, $\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m} \to \phi_{\mathrm{R},m}$. Using that $\phi_{\mathrm{R},m} = \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m} + (\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m})$, we obtain that $\|\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m}\| \to \|\phi_{\mathrm{R},m}\| = 1$. Denote by $\widehat{\alpha}_m = (\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m})/\|\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m}\|$, note that $\widehat{\alpha}_m \in \widehat{\mathcal{B}}_m$. Then, from the fact that $\|\phi_{\mathrm{R},m} - \pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m}\| \to 1$ and $\|\pi_{\widehat{\mathcal{L}}_{m-1}}\phi_{\mathrm{R},m}\| \to 0$, we obtain that $\phi_{\mathrm{R},m} = \widehat{\alpha}_m + o(1)$ which together with the continuity of σ, implies that $\sigma^2(\widehat{\alpha}_m) \to \sigma^2(\phi_{\mathrm{R},m})$. Hence,

$$\sigma^2(\phi_{\mathrm{R},m}) = \sigma^2(\widehat{\alpha}_m) + o(1) \leq \max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) + o(1),$$

where we have used the fact that $\widehat{\alpha}_m$ belongs to $\widehat{\mathcal{B}}_m$, concluding the proof of (A.7).

To derive (A.8) notice that

$$\begin{aligned}
\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) &= \max_{\alpha\in\widehat{\mathcal{B}}_m} \left(\sigma^2(\alpha) - \sigma_n^2(\alpha) + \sigma_n^2(\alpha)\right) \leq \max_{\alpha\in\widehat{\mathcal{B}}_m} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + \sigma_n^2(\widehat{\phi}_m) \\
&\leq \max_{\alpha\in\widehat{\mathcal{B}}_m} (\sigma^2(\alpha) - \sigma_n^2(\alpha)) + \sigma_n^2(\widehat{\phi}_m) - \sigma^2(\widehat{\phi}_m) + \sigma^2(\widehat{\phi}_m) \\
&\leq 2\max_{\|\alpha\|=1} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + \sigma^2(\widehat{\phi}_m) = o(1) + \sigma^2(\widehat{\phi}_m) \;.
\end{aligned}$$

Using that $\pi_{\widehat{\mathcal{L}}_{m-1}}\widehat{\phi}_m = 0$ and $\|\pi_{\widehat{\mathcal{L}}_{m-1}}\widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}}\widehat{\phi}_m\| \to 0$ (since $\|\widehat{\phi}_m\| = 1$) we get that

$$\widehat{\phi}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}}\widehat{\phi}_m + (\pi_{\mathcal{L}_{m-1}} - \pi_{\widehat{\mathcal{L}}_{m-1}})\widehat{\phi}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}}\widehat{\phi}_m + o(1) \;.$$

Denote by $\widehat{b}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}}\widehat{\phi}_m$, then we have that $\widehat{\phi}_m = \widehat{b}_m + o(1)$, which entails that $\|\widehat{b}_m\| \to 1$. Let $\widehat{\beta}_m$ stand for $\widehat{\beta}_m = \widehat{b}_m/\|\widehat{b}_m\|$. Note that $\widehat{\beta}_m \in \mathcal{B}_m$, then $\sigma(\widehat{\beta}_m) \leq \sigma(\phi_{\mathrm{R},m})$. On the other hand, using that $\widehat{\phi}_m - \widehat{\beta}_m = o(1)$ and the fact that σ is weakly continuous and $\mathcal{S}$ is weakly compact, we obtain, as in Lemma 3.1, that $\sigma(\widehat{\phi}_m) - \sigma(\widehat{\beta}_m) = o(1)$. Then,

$$\max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma^2(\alpha) \leq o(1) + \sigma^2(\widehat{\phi}_m) = o(1) + \sigma^2(\widehat{\beta}_m) \leq o(1) + \sigma^2(\phi_{\mathrm{R},m}) \;,$$

concluding the proof of (A.8) and so, $\widehat{\lambda}_m = \max_{\alpha\in\widehat{\mathcal{B}}_m} \sigma_n^2(\alpha) \to \sigma^2(\phi_{\mathrm{R},m}) = \lambda_{\mathrm{R},m}$.

Let us show that $\sigma^2(\widehat{\phi}_m) \to \sigma^2(\phi_{\mathrm{R},m})$.

$$\begin{aligned}
|\sigma^2(\widehat{\phi}_m) - \sigma^2(\phi_{\mathrm{R},m})| &\leq |\sigma^2(\widehat{\phi}_m) - \sigma_n^2(\widehat{\phi}_m)| + |\sigma_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{\mathrm{R},m})| \\
&\leq |\sigma^2(\widehat{\phi}_m) - \sigma_n^2(\widehat{\phi}_m)| + |\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| \\
&\leq \sup_{\|\alpha\|=1} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + |\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| \;.
\end{aligned}$$

and the proof follows now using ii) and the fact that $\widehat{\lambda}_m \to \sigma^2(\phi_{\mathrm{R},m})$.

d) We have already proved that when $m = 1$ the result holds. We will proceed by induction, we will assume that $\langle \widehat{\phi}_j, \phi_{\mathrm{R},j}\rangle^2 \to 1$ for $1 \le j \le m-1$ and we will show that $\langle \widehat{\phi}_m, \phi_{\mathrm{R},m}\rangle^2 \to 1$. Using c) we have that $\sigma^2(\widehat{\phi}_m) \to \sigma^2(\phi_{\mathrm{R},m})$ and so, as in Lemma 3.1 b) we conclude the proof. □

Proof of Theorem 3.2. To avoid burden notation, we will denote $\widehat{\phi}_j = \widehat{\phi}_{\mathrm{PN},j}$ and $\widehat{\lambda}_j = \widehat{\lambda}_{\mathrm{PN},,j}$.
a) We will prove that

$$\sigma^2(\phi_{\mathrm{R},1}) \quad \ge \quad \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1) \tag{A.9}$$

and that under **S4a)**

$$\sigma^2(\phi_{\mathrm{R},1}) \quad \le \quad \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1)\ , \tag{A.10}$$

holds. A weaker inequality than (A.10) will be obtained under **S4b)**.

Let us prove the first inequality. Using that σ is a scale functional, and that $\|\widehat{\phi}_1\| \le 1$, we get easily that

$$\sigma^2(\phi_{\mathrm{R},1}) = \sup_{\alpha\in\mathcal{S}} \sigma^2(\alpha) \ge \sigma^2\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) = \frac{\sigma^2\left(\widehat{\phi}_1\right)}{\|\widehat{\phi}_1\|^2} \ge \sigma^2(\widehat{\phi}_1).$$

On the other hand, **S1** entails that $\widehat{a}_{n,1} = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ and so, $\sigma^2(\phi_{\mathrm{R},1}) \ge \sigma^2(\widehat{\phi}_1) = s_n^2(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1) = \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1)$, concluding the proof of (A.9).

We will derive a). Since clearly, **S4b)** implies **S4a)**, we begin by showing the result under **S4a)** to have an idea of the arguments to be used. The extension to **S4b)** can be made using some technical arguments.

i) Assume that **S4a)** holds, then $\phi_{\mathrm{R},1} \in \mathcal{H}_{\mathrm{S}}$, so that $\|\phi_{\mathrm{R},1}\|_\tau < \infty$. Note that $\|\phi_{\mathrm{R},1}\|_\tau \ge \|\phi_{\mathrm{R},1}\| = 1$, then, defining $\beta_1 = \phi_{\mathrm{R},1}/\|\phi_{\mathrm{R},1}\|_\tau$, we have that $\|\beta_1\|_\tau = 1$, which implies that $\widehat{\lambda}_1 = s_n^2(\widehat{\phi}_1) \ge s_n^2(\beta_1)$. Again, using **S1** we get that $\widehat{b}_{n,1} = s_n^2(\beta_1) - \sigma^2(\beta_1) \xrightarrow{a.s.} 0$, hence,

$$\widehat{\lambda}_1 \ge s_n^2(\beta_1) = \sigma^2(\phi_{\mathrm{R},1}/\|\phi_{\mathrm{R},1}\|_\tau) + o_{\mathrm{a.s.}}(1) = \frac{\sigma^2(\phi_{\mathrm{R},1})}{\|\phi_{\mathrm{R},1}\|_\tau^2} + o_{\mathrm{a.s.}}(1) = \sigma^2(\phi_{\mathrm{R},1}) + o_{\mathrm{a.s.}}(1)$$

where, in the last inequality, we have used that $\|\phi_{\mathrm{R},1}\|_\tau \to \|\phi_{\mathrm{R},1}\| = 1$ since $\tau \to 0$, concluding the proof of a) in this case.

ii) Assume that **S4b)** holds. In this case, we cannot consider $\|\phi_{\mathrm{R},1}\|_\tau$ since $\phi_{\mathrm{R},1}$ does not belong to $\mathcal{H}_{\mathrm{S}}$, otherwise we argue as in i). Since, $\phi_{\mathrm{R},1} \in \overline{\mathcal{H}_{\mathrm{S}}}$, we can choose a sequence $\widetilde{\phi}_{1,k} \in \mathcal{H}_{\mathrm{S}}$ such that $\widetilde{\phi}_{1,k} \to \phi_{\mathrm{R},1}$, $\|\widetilde{\phi}_{1,k}\| = 1$ and $|\sigma^2(\widetilde{\phi}_{1,k}) - \sigma^2(\phi_{\mathrm{R},1})| < 1/k$. Note that for any fixed k, $\|\widetilde{\phi}_{1,k}\|_\tau \ge \|\widetilde{\phi}_{1,k}\| = 1$ and $\|\widetilde{\phi}_{1,k}\|_\tau \longrightarrow \|\widetilde{\phi}_{1,k}\| = 1$ since $\tau_n \longrightarrow 0$. Thus, using that $\widehat{\lambda}_1 = \max_{\|\alpha\|_\tau=1} s_n^2(\alpha)$ and defining $\beta_{1,k} = \widetilde{\phi}_{1,k}/\|\widetilde{\phi}_{1,k}\|_\tau$, we obtain that $\|\beta_{1,k}\|_\tau = 1$ and $\widehat{\lambda}_1 = s_n^2(\widehat{\phi}_1) \ge s_n^2(\beta_{1,k})$.

Note that **S1** entails that $\widehat{b}_{n,1} = s_n^2(\beta_{1,k}) - \sigma^2(\beta_{1,k}) \xrightarrow{a.s.} 0$, hence,

$$\widehat{\lambda}_1 \ge s_n^2(\beta_{1,k}) = \sigma^2(\beta_{1,k}) + o_{\mathrm{a.s.}}(1) = \frac{\sigma^2(\widetilde{\phi}_{1,k})}{\|\widetilde{\phi}_{1,k}\|_\tau^2} + o_{\mathrm{a.s.}}(1) = \frac{\sigma^2(\phi_{\mathrm{R},1}) - 1/k}{\|\widetilde{\phi}_{1,k}\|_\tau} + o_{\mathrm{a.s.}}(1)\ .$$

Therefore, using (A.9) and the fact that $\|\widetilde{\phi}_{1,k}\|_\tau \ge 1$, we have that

$$\sigma^2(\phi_{\mathrm{R},1}) \ge \widehat{\lambda}_1 + O_{1,n} \ge \frac{\sigma^2(\phi_{\mathrm{R},1}) - 1/k}{\|\widetilde{\phi}_{1,k}\|_\tau} + o_{\mathrm{a.s.}}(1) \ge \sigma^2(\phi_{\mathrm{R},1}) - \left(1 - \frac{1}{\|\widetilde{\phi}_{1,k}\|_\tau}\right)\sigma^2(\phi_{\mathrm{R},1}) - \frac{1}{k} + O_{2,n}\ ,$$

where $O_{i,n} = o_{\text{a.s.}}(1)$, $i = 1, 2$. Let $\mathcal{N} = \cup_{i=1,2}\{\omega : O_{i,n}(\omega) \not\to 0\}$ and fix $\omega \notin \mathcal{N}$. Given $\epsilon > 0$, fix k_0 such that $1/k_0 < \epsilon$. Let n_0 be such that, for $n \geq n_0$, $O_{i,n}(\omega) < \epsilon$, $i = 1, 2$ and

$$0 \leq \left(1 - \frac{1}{\|\widetilde{\phi}_{1,k_0}\|_\tau}\right) \sigma^2(\phi_{\text{R},1}) < \epsilon ,$$

where we have used that $\tau_n \to 0$ and thus $\|\widetilde{\phi}_{1,k_0}\|_\tau \to 1$. Then, using

$$|\widehat{\lambda}_1(\omega) - \sigma^2(\phi_{\text{R},1})| \leq \max\{|O_{1,n}|, |O_{2,n}| + 2\epsilon\} \leq 3\epsilon$$

which entails that $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1})$, as desired.

Using **S1**, we get that $\widehat{\lambda}_1 - \sigma^2(\widehat{\phi}_1) = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, using that $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1})$, we obtain that $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1})$, concluding the proof of a).

It is worth noticing that as a consequence of the above results, we get that the following inequalities converge to equalities

$$\sigma^2(\phi_{\text{R},1}) \geq \sigma^2\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \geq \sigma^2(\widehat{\phi}_1) = \widehat{\lambda}_1 + o_{\text{a.s.}}(1) ,$$

in particular,

$$\sigma^2\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1}) \qquad \text{and} \qquad \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1}) . \tag{A.11}$$

b) Note that

$$\tau\lceil\widehat{\phi}_1, \widehat{\phi}_1\rceil = 1 - \|\widehat{\phi}_1\|^2 = 1 - \frac{\sigma^2(\widehat{\phi}_1)}{\sigma^2(\widehat{\phi}_1/\|\widehat{\phi}_1\|)}.$$

Thus, using (A.11) we have that the second term is $1 + o_{\text{a.s.}}(1)$, concluding the proof of b).

c) Note that since $\|\widehat{\phi}_1\|_\tau = 1$, we have that $\|\widehat{\phi}_1\| \leq 1$. Moreover, from b) $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$. Let $\widetilde{\phi}_1 = \widehat{\phi}_1/\|\widehat{\phi}_1\|$. Then, we have that $\widetilde{\phi}_1 \in \mathcal{S}$ and $\sigma(\widetilde{\phi}_1) = \sigma(\widehat{\phi}_1)/\|\widehat{\phi}_1\|$. Using that $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1})$ and $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$, we obtain that $\sigma^2(\widetilde{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},1})$ and thus, the proof follows using Lemma 3.1.

d) Let us show that $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\text{R},m})$. We begin by proving the following extension of **S1**

$$\sup_{\|\alpha\|_\tau \leq 1} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\widehat{\pi}_{\tau,m-1}\alpha)| \xrightarrow{a.s.} 0 . \tag{A.12}$$

Using **S1** and the fact that s_n is a scale estimator and so, $s_n(\alpha) = \|\alpha\|_\tau \, s_n(\alpha/\|\alpha\|_\tau)$, we get that

$$\sup_{\|\alpha\|_\tau \leq 1} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \xrightarrow{a.s.} 0 \tag{A.13}$$

Note that

$$\sup_{\|\alpha\|_\tau \leq 1} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\widehat{\pi}_{\tau,m-1}\alpha)| \leq \sup_{\|\alpha\|_\tau \leq 1} |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{\tau,m-1}\alpha)| + \sup_{\|\alpha\|_\tau \leq 1} |\sigma^2(\widehat{\pi}_{\tau,m-1}\alpha) - s_n^2(\widehat{\pi}_{\tau,m-1}\alpha)|$$

Using (A.13) and the fact that if $\|\alpha\|_\tau \leq 1$ then $\|\widehat{\pi}_{\tau,m-1}\alpha\|_\tau \leq 1$, we get that the second term on the right hand side converges to 0 almost surely.
To conclude the proof of (A.12), it remains to show that

$$\sup_{\|\alpha\|_\tau \leq 1} |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{\tau,m-1}\alpha)| \xrightarrow{a.s.} 0 . \tag{A.14}$$

As in Silverman (1996), using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$ and that $\tau\Psi(\widehat{\phi}_j) = \tau\lceil\widehat{\phi}_j, \widehat{\phi}_j\rceil \xrightarrow{a.s.} 0$, for $1 \le j \le m-1$, we get that

$$\sup_{\|\alpha\|_\tau \le 1} \|\langle\alpha, \phi_{\mathrm{R},j}\rangle\phi_{\mathrm{R},j} - \langle\alpha, \widehat{\phi}_j\rangle_\tau\widehat{\phi}_j\| \xrightarrow{a.s.} 0 \quad \text{for} \quad 1 \le j \le m-1 \tag{A.15}$$

Effectively, for any $\alpha \in \mathcal{H}_{\mathrm{S}}$ such that $\|\alpha\|_\tau^2 = \|\alpha\|^2 + \tau\Psi(\alpha) \le 1$, we have that

$$\begin{aligned}
\|\langle\alpha, \phi_{\mathrm{R},j}\rangle\phi_{\mathrm{R},j} - \langle\alpha, \widehat{\phi}_j\rangle_\tau\widehat{\phi}_j\| &\le \|\alpha\|\|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + \|\widehat{\phi}_j\| \left|\langle\alpha, \phi_{\mathrm{R},j}\rangle - \langle\alpha, \widehat{\phi}_j\rangle_\tau\right| \\
&\le \|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + \left|\langle\alpha, \phi_{\mathrm{R},j} - \widehat{\phi}_j\rangle + \tau\lceil\alpha, \widehat{\phi}_j\rceil\right| \\
&\le \|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + \left\{\|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + (\tau\Psi(\alpha))^{\frac{1}{2}} \left(\tau\Psi(\widehat{\phi}_j)\right)^{\frac{1}{2}}\right\} \\
&\le \|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + \left\{\|\phi_{\mathrm{R},j} - \widehat{\phi}_j\| + \left(\tau\Psi(\widehat{\phi}_j)\right)^{\frac{1}{2}}\right\}
\end{aligned}$$

and so, (A.15) holds entailing that $\sup_{\|\alpha\|_\tau \le 1} \|\widehat{\pi}_{\tau,m-1}\alpha - \pi_{m-1}\alpha\| \xrightarrow{a.s.} 0$. Therefore, using that σ is weakly continuous and the unit ball is weakly compact, we get easily that (A.14) holds, concluding the proof of (A.12).

As in a), we will show that

$$\sigma^2(\phi_{\mathrm{R},m}) \ge \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1) \tag{A.16}$$

and that when **S4a)** holds

$$\sigma^2(\phi_{\mathrm{R},m}) \le \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1)\ . \tag{A.17}$$

holds. A weaker inequality than (A.17) will be obtained under **S4b)**.

Using again that σ is a scale functional, we get easily that $\sup_{\alpha\in\mathcal{S}\cap\mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha\in\mathcal{S}} \sigma^2(\pi_{m-1}\alpha)$ and so,

$$\sigma^2(\phi_{\mathrm{R},m}) = \sup_{\alpha\in\mathcal{S}\cap\mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha\in\mathcal{S}} \sigma^2(\pi_{m-1}\alpha) \ge \sigma^2\left(\pi_{m-1}\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\ .$$

From (A.12) we get that $\widehat{b}_m = \sigma^2\left(\pi_{m-1}\widehat{\phi}_m\right) - s_n^2\left(\widehat{\pi}_{\tau,m-1}\widehat{\phi}_m\right) \xrightarrow{a.s.} 0$ and so, since $\widehat{\pi}_{\tau,m-1}\widehat{\phi}_m = \widehat{\phi}_m$ and $\|\widehat{\phi}_m\| \le 1$, we get that

$$\begin{aligned}
\sigma^2(\phi_{\mathrm{R},m}) &\ge \sigma^2\left(\pi_{m-1}\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right) = \frac{\sigma^2\left(\pi_{m-1}\widehat{\phi}_m\right)}{\|\widehat{\phi}_m\|^2} \\
&\ge \sigma^2\left(\pi_{m-1}\widehat{\phi}_m\right) = s_n^2\left(\widehat{\pi}_{\tau,m-1}\widehat{\phi}_m\right) + o_{\mathrm{a.s.}}(1) = s_n^2(\widehat{\phi}_m) + o_{\mathrm{a.s.}}(1) = \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1)\ ,
\end{aligned}$$

concluding the proof of (A.16).

Let us show that (A.17) holds if **S4a)** holds.

i) If **S4a)** holds, $\phi_{\mathrm{R},m} \in \mathcal{H}_{\mathrm{S}}$, so that $\|\phi_{\mathrm{R},m}\|_\tau < \infty$ and $\|\phi_{\mathrm{R},m}\|_\tau \to \|\phi_{\mathrm{R},m}\| = 1$. Using that s_n is a scale estimator and the fact that for any $\alpha \in \mathcal{H}_{\mathrm{S}}$ such that $\|\alpha\|_\tau = 1$ we have that $\|\widehat{\pi}_{\tau,m-1}\alpha\|_\tau \le 1$, we get easily that

$$\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) = \sup_{\|\alpha\|_\tau=1, \alpha\in\widehat{\mathcal{T}}_{\tau,m-1}} s_n^2(\alpha) = \sup_{\|\alpha\|_\tau=1} s_n^2(\widehat{\pi}_{\tau,m-1}\alpha) \ge s_n^2\left(\frac{\widehat{\pi}_{\tau,m-1}\phi_{\mathrm{R},m}}{\|\phi_{\mathrm{R},m}\|_\tau}\right)$$

which together with (A.12) and the fact that $\|\phi_{\mathrm{R},m}\|_\tau \to \|\phi_{\mathrm{R},m}\| = 1$, entails that

$$\begin{aligned}
\widehat{\lambda}_m &\ge s_n^2\left(\frac{\widehat{\pi}_{\tau,m-1}\phi_{\mathrm{R},m}}{\|\phi_{\mathrm{R},m}\|_\tau}\right) = \sigma^2\left(\frac{\pi_{m-1}\phi_{\mathrm{R},m}}{\|\phi_{\mathrm{R},m}\|_\tau}\right) + o_{\mathrm{a.s.}}(1) \\
&\ge \sigma^2\left(\frac{\phi_{\mathrm{R},m}}{\|\phi_{\mathrm{R},m}\|_\tau}\right) + o_{\mathrm{a.s.}}(1) = \frac{\sigma^2\left(\phi_{\mathrm{R},m}\right)}{\|\phi_{\mathrm{R},m}\|_\tau^2} + o_{\mathrm{a.s.}}(1) \ge \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1)
\end{aligned}$$

concluding the proof of (A.17) in this case and so, when **S4a)** holds $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$.

ii) Assume that **S4b)** holds. As in a), let us consider a sequence $\widetilde{\phi}_{m,k} \in \mathcal{H}_{\mathrm{S}}$ such that $\widetilde{\phi}_{m,k} \to \phi_{\mathrm{R},m}$, as $k \to \infty$, $\|\widetilde{\phi}_{m,k}\| = 1$ and $|\sigma^2(\pi_{m-1}\widetilde{\phi}_{m,k}) - \sigma^2(\phi_{\mathrm{R},m})| < 1/k$, since $\pi_{m-1}\phi_{\mathrm{R},m} = \phi_{\mathrm{R},m}$. Then, for each fixed k, we have that $\|\widetilde{\phi}_{m,k}\|_\tau \longrightarrow \|\widetilde{\phi}_{m,k}\| = 1$ since $\tau \to 0$.

Using that s_n is a scale estimator and the fact that for any $\alpha \in \mathcal{H}_{\mathrm{S}}$ such that $\|\alpha\|_\tau = 1$ we have that $\|\widehat{\pi}_{\tau,m-1}\alpha\|_\tau \leq 1$, we get that

$$\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) = \sup_{\|\alpha\|_\tau=1, \alpha\in\widehat{\mathcal{T}}_{\tau,m-1}} s_n^2(\alpha) = \sup_{\|\alpha\|_\tau=1} s_n^2(\widehat{\pi}_{\tau,m-1}\alpha) \geq s_n^2\left(\frac{\widehat{\pi}_{\tau,m-1}\widetilde{\phi}_{m,k}}{\|\widetilde{\phi}_{m,k}\|_\tau}\right)$$

Using (A.12) and the fact that $|\sigma^2(\pi_{m-1}\widetilde{\phi}_{m,k}) - \sigma^2(\phi_{\mathrm{R},m})| < 1/k$ and $\|\widetilde{\phi}_{m,k}\|_\tau \geq 1$, we get that

$$\begin{aligned}
\widehat{\lambda}_m &\geq s_n^2\left(\frac{\widehat{\pi}_{\tau,m-1}\widetilde{\phi}_{m,k}}{\|\widetilde{\phi}_{m,k}\|_\tau}\right) = \sigma^2\left(\frac{\pi_{m-1}\widetilde{\phi}_{m,k}}{\|\widetilde{\phi}_{m,k}\|_\tau}\right) + o_{\mathrm{a.s.}}(1) \\
&\geq \frac{\sigma^2\left(\phi_{\mathrm{R},m}\right) - 1/k}{\|\widetilde{\phi}_{m,k}\|_\tau^2} + o_{\mathrm{a.s.}}(1) \\
&\geq \sigma^2\left(\phi_{\mathrm{R},m}\right) - \sigma^2\left(\phi_{\mathrm{R},m}\right)\left(1 - \frac{1}{\|\widetilde{\phi}_{m,k}\|_\tau^2}\right) - \frac{1}{k} + o_{\mathrm{a.s.}}(1)
\end{aligned}$$

Therefore, arguing as in a) we obtain that $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2\left(\phi_{\mathrm{R},m}\right)$.

On the other hand, as in a), using **S1**, we get that $\widehat{\lambda}_m - \sigma^2(\widehat{\phi}_m) = s_n^2(\widehat{\phi}_m) - \sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, using that $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, we obtain that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$.

Thus, it remains to show that $\tau\lceil\widehat{\phi}_m, \widehat{\phi}_m\rceil \xrightarrow{a.s.} 0$. As in a), we have that the following inequalities converge to equalities

$$\sigma^2(\phi_{\mathrm{R},m}) \geq \sigma^2\left(\pi_{m-1}\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right) \geq \sigma^2\left(\pi_{m-1}\widehat{\phi}_m\right) \geq \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1)\,. \tag{A.18}$$

Note that since σ is a scale estimator, we have that

$$\tau\lceil\widehat{\phi}_m, \widehat{\phi}_m\rceil = 1 - \|\widehat{\phi}_m\|^2 = 1 - \frac{\sigma^2(\pi_{m-1}\widehat{\phi}_m)}{\sigma^2(\pi_{m-1}\widehat{\phi}_m/\|\widehat{\phi}_m\|)}\,,$$

which together with (A.18) entails that the second term on the right hand side is $1 + o_{\mathrm{a.s.}}(1)$, concluding the proof of d).

e) For $m = 1$ the result was derived in c). Let us assume that for $1 \leq j \leq m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$ and that $\tau\lceil\widehat{\phi}_j, \widehat{\phi}_j\rceil \xrightarrow{a.s.} 0$, we will show that $\langle\widehat{\phi}_m, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$, i.e., we will use an induction argument. By d) we already know that $\tau\lceil\widehat{\phi}_m, \widehat{\phi}_m\rceil \xrightarrow{a.s.} 0$ which entails that $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$. Denote by $\widetilde{\phi}_j = \widehat{\phi}_j/\|\widehat{\phi}_j\|$, it is enough to show that $\langle\phi_{\mathrm{R},m}, \widetilde{\phi}_m\rangle^2 \xrightarrow{a.s.} 1$. We have that, $\langle\widehat{\phi}_m, \widehat{\phi}_j\rangle_\tau = 0$. Using that $\tau\lceil\widehat{\phi}_j, \widehat{\phi}_j\rceil \xrightarrow{a.s.} 0$, for $1 \leq j \leq m-1$, we get that $\tau\lceil\widehat{\phi}_j, \widehat{\phi}_m\rceil \xrightarrow{a.s.} 0$ for $1 \leq j \leq m-1$ and so, $\langle\widetilde{\phi}_m, \widetilde{\phi}_j\rangle \xrightarrow{a.s.} 0$. Therefore, arguing as in Lemma 3.1, we can write $\widetilde{\phi}_m$ as $\widetilde{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j\phi_{\mathrm{R},j} + \widehat{\gamma}_m$, with $\langle\widehat{\gamma}_m, \phi_{\mathrm{R},j}\rangle = 0$, $1 \leq j \leq m-1$. To obtain e) it remains to show that $\langle\widehat{\gamma}_m, \phi_{\mathrm{R},m}\rangle^2 \xrightarrow{a.s.} 1$. Note that $\langle\widetilde{\phi}_m, \widetilde{\phi}_j\rangle \xrightarrow{a.s.} 0$, for $j \neq m$, implies that $\widehat{a}_j = \langle\widetilde{\phi}_m, \phi_{\mathrm{R},j}\rangle = \langle\widetilde{\phi}_m, \phi_{\mathrm{R},j} - \widetilde{\phi}_j\rangle + \langle\widetilde{\phi}_m, \widetilde{\phi}_j\rangle = \langle\widetilde{\phi}_m, \phi_{\mathrm{R},j} - \widetilde{\phi}_j\rangle + o_{\mathrm{a.s.}}(1)$. Thus, using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$, $1 \leq j \leq m-1$, and $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, we get that $\widehat{a}_j \xrightarrow{a.s.} 0$ for $1 \leq j \leq m-1$ and so, $\|\widetilde{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$. Note that $1 = \|\widetilde{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$, hence, $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$ which implies that $\|\widetilde{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, where $\widetilde{\gamma}_m = \widehat{\gamma}_m/\|\widehat{\gamma}_m\|$. Using that $\sigma(\alpha)$ is a weakly continuous and $\mathcal{S}$ is weakly compact, we obtain that $\sigma(\widetilde{\gamma}_m) - \sigma(\widetilde{\phi}_m) \xrightarrow{a.s.} 0$ which together with the fact that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, implies that $\sigma(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma(\phi_{\mathrm{R},m})$. The proof follows now as in Lemma 3.1 using the fact that $\widetilde{\gamma}_m \in \mathcal{C}_m$, with $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{\mathrm{R},j}\rangle = 0,\ 1 \le j \le m-1\}$ and $\phi_{\mathrm{R},m}$ is the unique maximizer of $\sigma(\alpha)$ over $\mathcal{C}_m$. □

Proof of Theorem 3.3. To avoid burden notation, we will denote $\widehat{\phi}_j = \widehat{\phi}_{\mathrm{PS},j}$ and $\widehat{\lambda}_j = \widehat{\lambda}_{\mathrm{PS},j}$.
a) We will prove that

$$\sigma^2(\phi_{\mathrm{R},1}) \quad \ge \quad \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1) \tag{A.19}$$

and that under **S4a)**

$$\sigma^2(\phi_{\mathrm{R},1}) \quad \le \quad \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1)\,, \tag{A.20}$$

holds.

Let us prove the first inequality. We easily get that

$$\sigma^2(\phi_{\mathrm{R},1}) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\alpha) \ge \sigma^2(\widehat{\phi}_1).$$

On the other hand, **S1** entails that $\widehat{a}_{n,1} = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ and so, $\sigma^2(\phi_{\mathrm{R},1}) \ge \sigma^2(\widehat{\phi}_1) = s_n^2(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1) = \widehat{\lambda}_1 + o_{\mathrm{a.s.}}(1)$, concluding the proof of (A.19).

We will derive (A.20). Since **S4a)** holds, we have that $\phi_{\mathrm{R},1} \in \mathcal{H}_{\mathrm{S}}$, so that $\|\phi_{\mathrm{R},1}\|_\tau < \infty$. Note that

$$\widehat{\lambda}_1 = s_n^2(\widehat{\phi}_1) \ge s_n^2(\widehat{\phi}_1) - \tau\lceil \widehat{\phi}_1, \widehat{\phi}_1 \rceil = \sup_{\alpha \in \mathcal{S}} \{s_n^2(\alpha) - \tau\lceil \alpha, \alpha \rceil\} \ge s_n^2(\phi_{\mathrm{R},1}) - \tau\lceil \phi_{\mathrm{R},1}, \phi_{\mathrm{R},1} \rceil. \tag{A.21}$$

Using that $\tau \to 0$, we obtain that $\tau\lceil \phi_{\mathrm{R},1}, \phi_{\mathrm{R},1} \rceil \to 0$. Also, using **S1** we get that $s_n^2(\phi_{\mathrm{R},1}) = \sigma^2(\phi_{\mathrm{R},1}) + o_{\mathrm{a.s.}}(1)$. Therefore, (A.21) can be written as

$$\widehat{\lambda}_1 \ge s_n^2(\phi_{\mathrm{R},1}) - \tau\lceil \phi_{\mathrm{R},1}, \phi_{\mathrm{R},1} \rceil = \sigma^2(\phi_{\mathrm{R},1}) + o_{\mathrm{a.s.}}(1)$$

Hence, (A.20) follows which together with (A.19) implies that $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$. Using **S1** we have that

$$\widehat{\lambda}_1 - \sigma^2(\widehat{\phi}_1) = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$$

therefore we also get that $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1}))$. From the fact that $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$, $s_n^2(\phi_{\mathrm{R},1}) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$, $\tau \to 0$ and since

$$\widehat{\lambda}_1 \ge s_n^2(\widehat{\phi}_1) - \tau\lceil \widehat{\phi}_1, \widehat{\phi}_1 \rceil \ge s_n^2(\phi_{\mathrm{R},1}) - \tau\lceil \phi_{\mathrm{R},1}, \phi_{\mathrm{R},1} \rceil$$

we get that $\tau\lceil \widehat{\phi}_1, \widehat{\phi}_1 \rceil \xrightarrow{a.s.} 0$, concluding the proof of a).
b) Follows easily using Lemma 3.1 a), the fact that $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ and $\|\widehat{\phi}_1\| = 1$.

c) We will prove that

$$\sigma^2(\phi_{\mathrm{R},m}) \quad \ge \quad \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1) \tag{A.22}$$

and that under **S4a)**

$$\sigma^2(\phi_{\mathrm{R},m}) \quad \le \quad \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1)\,, \tag{A.23}$$

holds. In order to derive (A.22), note that

$$\sigma^2(\phi_{\mathrm{R},1}) = \sup_{\alpha\in\mathcal{S}\cap\mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha\in\mathcal{S}} \sigma^2(\pi_{m-1}\alpha) \geq \sigma^2(\pi_{m-1}\widehat{\phi}_m). \tag{A.24}$$

Let us show that $\sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{m-1}\widehat{\phi}_m) + o_{\mathrm{a.s.}}(1)$. Indeed, if $\alpha \in \mathcal{S}$, then

$$|\sigma^2(\pi_{m-1}\alpha) - s_n^2(\widehat{\pi}_{m-1}\alpha)| \leq |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{m-1}\alpha)| + |\sigma^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\widehat{\pi}_{m-1}\alpha)|. \tag{A.25}$$

The second term on the right hand side of (A.25) will be $o_{\mathrm{a.s.}}(1)$ since **S1** holds. Let us show that the first one will also be $o_{\mathrm{a.s.}}(1)$. Using that $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$, we get that $\widehat{\pi}_{m-1}\alpha \xrightarrow{a.s.} \pi_{m-1}\alpha$. Finally, from **S2**, i.e., the continuity of σ, we get that the first term on the right hand side of (A.25) will be $o_{\mathrm{a.s.}}(1)$. Therefore, $\sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{m-1}\widehat{\phi}_m) + o_{\mathrm{a.s.}}(1)$. Therefore, (A.24) entails that

$$\sigma^2(\phi_{\mathrm{R},1}) \geq \sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{m-1}\widehat{\phi}_m) + o_{\mathrm{a.s.}}(1) = s_n^2(\widehat{\phi}_m) + o_{\mathrm{a.s.}}(1) = \widehat{\lambda}_m + o_{\mathrm{a.s.}}(1)$$

concluding the proof of (A.22).

Let us now proof that, under **S4a**, (A.23) holds.

$$\begin{aligned}
\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) &\geq s_n^2(\widehat{\phi}_m) - \tau\lceil\widehat{\phi}_m, \widehat{\phi}_m\rceil = \sup_{\alpha\in\mathcal{S}\cap\widehat{\mathcal{T}}_{m-1}} \{s_n^2(\alpha) - \tau\lceil\alpha,\alpha\rceil\} \\
&\geq \sup_{\alpha\in\mathcal{S}}\{s_n^2(\widehat{\pi}_{m-1}\alpha) - \tau\lceil\widehat{\pi}_{m-1}\alpha, \widehat{\pi}_{m-1}\alpha\rceil\} && \text{(A.26)} \\
&\geq s_n^2(\widehat{\pi}_{m-1}\phi_{\mathrm{R},m}) - \tau\lceil\widehat{\pi}_{m-1}\phi_{\mathrm{R},m}, \widehat{\pi}_{m-1}\phi_{\mathrm{R},m}\rceil. && \text{(A.27)}
\end{aligned}$$

Let us show that $\sup_{\alpha\in\mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \xrightarrow{a.s.} 0$. Effectively,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \\
\leq\ &\sup_{\alpha\in\mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{m-1}\alpha)| + \sup_{\alpha\in\mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| + \sup_{\alpha\in\mathcal{S}} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \\
\leq\ &\sup_{\alpha\in\mathcal{S}} |s_n^2(\alpha) - \sigma^2(\alpha)| + \sup_{\alpha\in\mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| + \sup_{\alpha\in\mathcal{S}} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)|
\end{aligned}$$

The first and third terms of the last inequality converge to 0 almost surely since **S1** holds. Thus we only have to show that $\sup_{\alpha\in\mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| = o_{\mathrm{a.s.}}(1)$. Using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$ for $1 \leq j \leq m-1$, it is easy to show that $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$ since it reduces to a difference of finite dimensional projections. Therefore, we have uniform convergence in the set $\{\alpha, \|\alpha\| \leq 1\}$. Using that σ is weakly continuous in $\mathcal{S} = \{\alpha, \|\alpha\| \leq 1\}$ which is weakly compact we obtain that $\sup_{\alpha\in\mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)|$, converges to 0 almost surely. In conclusion,

$$\sup_{\alpha\in\mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| = o_{\mathrm{a.s.}}(1).$$

Using that $\tau\lceil\widehat{\phi}_\ell, \widehat{\phi}_\ell\rceil \xrightarrow{a.s.} 0$, $1 \leq \ell \leq m-1$, analogous arguments to those considered in Pezzulli and Silverman (1993) and the fact that $\tau \to 0$ implies that $\tau\lceil\phi_{\mathrm{R},m}, \phi_{\mathrm{R},m}\rceil = o(1)$, it is not hard to see that

$$\tau_n\lceil\widehat{\pi}_{m-1}\phi_{\mathrm{R},m}, \widehat{\pi}_{m-1}\phi_{\mathrm{R},m}\rceil \xrightarrow{a.s.} 0.$$

Those two results essentially allow us to replace $\widehat{\pi}_{m-1}\alpha$ by $\pi_{m-1}\alpha$ in (A.27). Therefore,

$$\begin{aligned}
\widehat{\lambda}_m &\geq s_n^2(\pi_{m-1}\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1) = \sup_{\alpha\in\mathcal{S}\cap\mathcal{T}_{m-1}} \{s_n^2(\alpha) - \tau\lceil\alpha,\alpha\rceil\} + o_{\mathrm{a.s.}}(1) \\
&\geq s_n^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1) = \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1) + o_{\mathrm{a.s.}}(1) = \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1)
\end{aligned}$$

where we have used **S1**. Using that

$$\widehat{\lambda}_m \geq s_n^2(\widehat{\phi}_m) - \tau\lceil \widehat{\phi}_m, \widehat{\phi}_m \rceil \geq \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1)$$

and the fact that $\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ imply that $\tau\lceil \widehat{\phi}_m, \widehat{\phi}_m \rceil \xrightarrow{a.s.} 0$, concluding the proof of c).

d) We have already proved that when $m = 1$ the result holds. We will proceed by induction, we will assume that $\langle \widehat{\phi}_j, \phi_{\mathrm{R},j} \rangle^2 \xrightarrow{a.s.} 1$ for $1 \leq j \leq m-1$ and we will show that $\langle \widehat{\phi}_m, \phi_{\mathrm{R},m} \rangle^2 \xrightarrow{a.s.} 1$. By definition $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$, for $j \neq m$ and $\widehat{\phi}_m \in \mathcal{S}$ thus, using the fact that c) entails that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and Lemma 3.1 b), the proof follows. □

Proof of Theorem 3.4. For the sake of simplicity, we will avoid the subscript SI and we will denote $\widehat{\phi}_j = \widehat{\phi}_{\mathrm{SI},j}$ and $\widehat{\lambda}_j = \widehat{\lambda}_{\mathrm{SI},j}$.
a) The proof follows using similar arguments as those considered in the proof of Proposition 3.1. Using **S1** we get that

$$\widehat{a}_{n,1} = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0\ . \tag{A.28}$$

Let $\widetilde{\phi}_{1,p_n} = \pi_{\mathcal{H}_{p_n}} \phi_{\mathrm{R},1} / \|\pi_{\mathcal{H}_{p_n}} \phi_{\mathrm{R},1}\|$, then, $\widetilde{\phi}_{1,p_n} \in \mathcal{S}_{p_n}$ and $\widetilde{\phi}_{1,p_n} \to \phi_{\mathrm{R},1}$. Hence, **S2** entails that $\sigma(\widetilde{\phi}_{1,p_n}) \to \sigma(\phi_1)$ while using **S1**, we get that $s_n^2(\widetilde{\phi}_{1,p_n}) - \sigma^2(\widetilde{\phi}_{1,p_n}) \xrightarrow{a.s.} 0$. Thus, $\widehat{b}_{n,1} = s_n^2(\widetilde{\phi}_{1,p_n}) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$. Note that

$$\sigma^2(\phi_{\mathrm{R},1}) = s_n^2(\widetilde{\phi}_{1,p_n}) - \widehat{b}_{n,1} \leq s_n^2(\widehat{\phi}_1) - \widehat{b}_{n,1} = \sigma^2(\widehat{\phi}_1) + \widehat{a}_{n,1} - \widehat{b}_{n,1} \leq \sigma^2(\phi_{\mathrm{R},1}) + \widehat{a}_{n,1} - \widehat{b}_{n,1}\ ,$$

that is, $\sigma^2(\phi_{\mathrm{R},1}) - \widehat{a}_{n,1} + \widehat{b}_{n,1} \leq \sigma^2(\widehat{\phi}_1) \leq \sigma^2(\phi_{\mathrm{R},1})$ and so,

$$\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})\ , \tag{A.29}$$

which together with (A.28) implies that $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$.
b) We have that

$$\begin{aligned} |\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| &= |s_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{\mathrm{R},m})| = |\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} s_n^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha)| \\ &\leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |s_n^2(\alpha) - \sigma^2(\alpha)| + |\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha)| \\ &\leq \max_{\alpha \in \mathcal{S}_{p_n}} |s_n^2(\alpha) - \sigma^2(\alpha)| + |\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha)|\ . \end{aligned} \tag{A.30}$$

Using **S1**, we get that the first term on the right hand side of (A.30) converges to 0 almost surely. Thus, it will be enough to show that $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, since $\max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) = \sigma^2(\phi_{\mathrm{R},m})$. Using that $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{\mathrm{R},s}$, for $1 \leq s \leq m-1$, we get that

$$\|\pi_{\widehat{\mathcal{L}}_{m-1}} - \pi_{\mathcal{L}_{m-1}}\| \xrightarrow{a.s.} 0\ , \tag{A.31}$$

which implies that

$$\|\pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widetilde{\phi}_m\| \xrightarrow{a.s.} 0\ , \tag{A.32}$$

where $\widetilde{\phi}_m = \pi_{\mathcal{H}_{p_n}} \phi_{\mathrm{R},m}$. On the other hand, using that $p_n \to \infty$, we get that $\widetilde{\phi}_m \to \phi_{\mathrm{R},m}$, thus $\|\pi_{\mathcal{L}_{m-1}}(\widetilde{\phi}_m - \phi_{\mathrm{R},m})\| \to 0$ which together with (A.32) and the fact that $\pi_{\mathcal{L}_{m-1}} \phi_{\mathrm{R},m} = 0$, entails that $\pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m \xrightarrow{a.s.} 0$ and $\widetilde{\beta}_m = \widetilde{\phi}_m - \pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m \xrightarrow{a.s.} \phi_{\mathrm{R},m}$. Denoting by $a_m = \|\widetilde{\beta}_m\|$, $a_m \xrightarrow{a.s.} 1$, we obtain that $\widehat{\alpha}_m = \widetilde{\beta}_m / a_m \xrightarrow{a.s.} \phi_{\mathrm{R},m}$. Moreover, $\widehat{\alpha}_m \in \widehat{\mathcal{B}}_{n,m}$ since $\widetilde{\phi}_m \in \mathcal{H}_{p_n}$ and $\widehat{\phi}_j \in \mathcal{H}_{p_n}$,

$1 \leq j \leq m-1$, implying that $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \geq \sigma^2(\widehat{\alpha}_m)$. Using the weak continuity of σ, we get that $\sigma(\widehat{\alpha}_m) \xrightarrow{a.s.} \sigma(\phi_{\mathrm{R},m})$, i.e.,

$$\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \geq \sigma^2(\widehat{\alpha}_m) = \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1) \, .$$

On the other hand,

$$\begin{aligned} \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) &\leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |\sigma^2(\alpha) - s_n^2(\alpha)| + s_n^2(\widehat{\phi}_m) \\ &\leq 2 \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |\sigma^2(\alpha) - s_n^2(\alpha)| + \sigma^2(\widehat{\phi}_m) = o_{\mathrm{a.s.}}(1) + \sigma^2(\widehat{\phi}_m) \end{aligned}$$

Using (A.31) and the fact that $\widehat{\phi}_m \in \widehat{\mathcal{B}}_{n,m}$, $\|\widehat{\phi}_m\| = 1$, we get that $\widehat{\phi}_m = \widehat{\phi}_m - \pi_{\widehat{\mathcal{L}}_{m-1}} \widehat{\phi}_m = \widehat{b}_m + o_{\mathrm{a.s.}}(1)$, where $\widehat{b}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m$. Thus $\|\widehat{b}_m\| \xrightarrow{a.s.} 1$. Denote $\widehat{\beta}_m = \widehat{b}_m / \|\widehat{b}_m\|$, then $\widehat{\beta}_m \in \mathcal{B}_m$ which implies that $\sigma(\widehat{\beta}_m) \leq \sigma(\phi_{\mathrm{R},m})$. Besides, the weak continuity of σ and the weak compactness of the unit ball entail, as in Lemma 3.1, that $\sigma(\widehat{\beta}_m) - \sigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, since $\widehat{\phi}_m - \widehat{\beta}_m = o_{\mathrm{a.s.}}(1)$. Summarizing,

$$\begin{aligned} \sigma^2(\phi_{\mathrm{R},m}) + o_{\mathrm{a.s.}}(1) = \sigma^2(\widehat{\alpha}_m) \leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) &\leq o_{\mathrm{a.s.}}(1) + \sigma^2(\widehat{\phi}_m) = o_{\mathrm{a.s.}}(1) + \sigma^2(\widehat{\beta}_m) \\ &\leq o_{\mathrm{a.s.}}(1) + \sigma^2(\phi_{\mathrm{R},m}) \end{aligned}$$

concluding that $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ and thus the proof that $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$. Moreover, it is easy to see that **S1** and the fact that $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$ entail that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$, since

$$\begin{aligned} |\sigma^2(\widehat{\phi}_m) - \sigma^2(\phi_{\mathrm{R},m})| &\leq |\sigma^2(\widehat{\phi}_m) - s_n^2(\widehat{\phi}_m)| + |s_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{\mathrm{R},m})| \\ &= |\sigma^2(\widehat{\phi}_m) - s_n^2(\widehat{\phi}_m)| + |\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| \\ &\leq \sup_{\alpha \in \mathcal{S}_{p_n}} |\sigma^2(\alpha) - s_n^2(\alpha)| + |\widehat{\lambda}_m - \sigma^2(\phi_{\mathrm{R},m})| \, . \end{aligned}$$

c) We begin by proving that $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{\mathrm{R},1}$. Using $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},1})$ and Lemma 3.1a), the result follows easily since $\widehat{\phi}_1 \in \mathcal{S}$. Let us show that if $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$ for $1 \leq j \leq m-1$ then $\langle \widehat{\phi}_m, \phi_{\mathrm{R},m} \rangle^2 \xrightarrow{a.s.} 1$, which will lead to c). Since $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{\mathrm{R},j}$ for $1 \leq j \leq m-1$, we have that $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{\mathrm{R},m})$. Besides, using that $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$, $j \neq m$ and $\widehat{\phi}_m \in \mathcal{S}$, Lemma 3.1b) concludes the proof. □

References

[1] Bali, J. and Boente, G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statist. Probab. Lett.*, **79**, 1858-1865.

[2] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.*, **48**, 335-345.

[3] Cantoni, E. and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141-146.

[4] Croux, C. and Ruiz–Gazen, A. (1996). A fast algorithm for robust principal components based on projection pursuit. In *Compstat: Proceedings in Computational Statistics*, ed. A. Prat, Heidelberg: Physica–Verlag, 211-217.

[5] Croux, C. and Ruiz–Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection–Pursuit Approach Revisited. *J. Multivar. Anal.*, **95**, 206-226.

[6] Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comp. Statist.*, **22**, 48196.

[7] Cui, H., He, X. and Ng, K. W. (2003). Asymptotic Distribution of Principal Components Based on Robust Dispersions. *Biometrika*, **90**, 953-966.

[8] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivar. Anal.*, **12**, 136-154.

[9] Fraiman, R. and Muñiz, G. (2001). Trimmed means for functional data. *Test*, **10**, 41940.

[10] Gervini, D. (2006). Free-knot spline smoothing for functional data. *J. Roy. Statist. Soc. Ser. B*, **68**, 67187.

[11] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.

[12] Hall, P. and Hosseini–Nasab, M. (2006). On properties of functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, **68**, 10926.

[13] Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493517.

[14] Hyndman, R. J. and S. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comp. Statist. Data Anal.*, **51**, 4942-4956.

[15] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with Discussion). *Test*, **8**, 173.

[16] Li, G. and Chen, Z. (1985). Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *J. Amer. Statist. Assoc.*, **80**, 759-766.

[17] López–Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Comp. Statist. Data Anal.*, **51**, 495768.

[18] Pezzulli, S. D. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist.*, **8**, 1-16.

[19] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.

[20] Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, **53**, 233-243.

[21] Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the Median Absolute Deviation. *J. Amer. Statist. Assoc.*, **88**, 1273-1283.

[22] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24**, 1-24.

[23] van der Geer, S. (2000). *Empirical Processes in $M-$Estimation*. Cambridge University Press.

[24] Wang, F. and Scott, D. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Stat. Assoc.*, **89**, 65-76.

[25] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Statist. Soc. Ser. B*, **68**, 325.

Scale estimator	m	$j=1$	$j=2$	$j=3$
SD	50	0.0080	0.0117	0.0100
MAD	50	0.0744	0.1288	0.0879
M-scale	50	0.0243	0.0424	0.0295
SD	100	0.0078	0.0113	0.0079
MAD	100	0.0700	0.1212	0.0827
M-scale	100	0.0237	0.0416	0.0271
SD	150	0.0077	0.0112	0.0075
MAD	150	0.0703	0.1216	0.0824
M-scale	150	0.0234	0.0414	0.0268
SD	200	0.0077	0.0112	0.0073
MAD	200	0.0705	0.1223	0.0825
M-scale	200	0.0233	0.0416	0.0269
SD	250	0.0077	0.0112	0.0073
MAD	250	0.0701	0.1212	0.0815
M-scale	250	0.0233	0.0414	0.0267

Table 3: Mean values of $\|\widehat{\phi}_j - \phi_j\|^2$ under C_0, for the raw estimators, for different sizes m of the grid.

Method	Scale estimator	m	$a = 0.15$			$a = 0.5$			$a = 1$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_{\mathrm{PS},j}$	SD	50	0.0078	0.0106	0.0090	0.0077	0.0090	0.0074	0.0077	0.0081	0.0064
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	50	0.0737	0.1187	0.0780	0.0720	0.1061	0.0663	0.0702	0.0929	0.0531
$\widehat{\phi}_{\mathrm{PS},j}$	M-scale	50	0.0240	0.0377	0.0249	0.0239	0.0317	0.0187	0.0232	0.0270	0.0136
$\widehat{\phi}_{\mathrm{PS},j}$	SD	100	0.0076	0.0095	0.0061	0.0075	0.0079	0.0043	0.0073	0.0069	0.0032
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	100	0.0698	0.1092	0.0697	0.0687	0.0927	0.0529	0.0668	0.0782	0.0380
$\widehat{\phi}_{\mathrm{PS},j}$	M-scale	100	0.0234	0.0345	0.0198	0.0226	0.0269	0.0121	0.0221	0.0231	0.0078
$\widehat{\phi}_{\mathrm{PS},j}$	SD	150	0.0076	0.0094	0.0057	0.0075	0.0077	0.0038	0.0072	0.0068	0.0027
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	150	0.0695	0.1088	0.0692	0.0678	0.0883	0.0483	0.0663	0.0758	0.0346
$\widehat{\phi}_{\mathrm{PS},j}$	M-scale	150	0.0231	0.0340	0.0190	0.0224	0.0262	0.0111	0.0218	0.0223	0.0068
$\widehat{\phi}_{\mathrm{PS},j}$	SD	200	0.0075	0.0093	0.0054	0.0074	0.0076	0.0036	0.0071	0.0067	0.0025
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	200	0.0699	0.1080	0.0678	0.0680	0.0880	0.0475	0.0663	0.0751	0.0337
$\widehat{\phi}_{\mathrm{PS},j}$	M-scale	200	0.0230	0.0336	0.0186	0.0223	0.0259	0.0106	0.0217	0.0221	0.0065
$\widehat{\phi}_{\mathrm{PS},j}$	SD	250	0.0075	0.0093	0.0054	0.0074	0.0075	0.0035	0.0071	0.0066	0.0024
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	250	0.0695	0.1080	0.0679	0.0679	0.0881	0.0474	0.0661	0.0750	0.0333
$\widehat{\phi}_{\mathrm{PS},j}$	M-scale	250	0.0228	0.0333	0.0184	0.0223	0.0258	0.0105	0.0216	0.0219	0.0063
$\widehat{\phi}_{\mathrm{PN},j}$	SD	50	0.0075	0.0075	0.0161	0.0087	0.0095	0.0490	0.0093	0.0113	0.1197
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	50	0.0619	0.0731	0.1465	0.0552	0.0650	0.2687	0.0511	0.0658	0.4073
$\widehat{\phi}_{\mathrm{PN},j}$	M-scale	50	0.0203	0.0216	0.0310	0.0193	0.0213	0.0715	0.0192	0.0233	0.1470
$\widehat{\phi}_{\mathrm{PN},j}$	SD	100	0.0075	0.0086	0.0078	0.0073	0.0073	0.0099	0.0075	0.0074	0.0151
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	100	0.0675	0.1012	0.0988	0.0617	0.0799	0.1226	0.0603	0.0706	0.1499
$\widehat{\phi}_{\mathrm{PN},j}$	M-scale	100	0.0220	0.0293	0.0250	0.0205	0.0227	0.0257	0.0198	0.0206	0.0297
$\widehat{\phi}_{\mathrm{PN},j}$	SD	150	0.0075	0.0100	0.0073	0.0074	0.0087	0.0073	0.0073	0.0079	0.0077
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	150	0.0694	0.1153	0.0864	0.0676	0.1072	0.0994	0.0652	0.0914	0.0997
$\widehat{\phi}_{\mathrm{PN},j}$	M-scale	150	0.0229	0.0371	0.0263	0.0221	0.0306	0.0246	0.0213	0.0264	0.0245
$\widehat{\phi}_{\mathrm{PN},j}$	SD	200	0.0076	0.0107	0.0072	0.0075	0.0099	0.0072	0.0075	0.0091	0.0072
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	200	0.0699	0.1183	0.0821	0.0689	0.1138	0.0865	0.0677	0.1119	0.0953
$\widehat{\phi}_{\mathrm{PN},j}$	M-scale	200	0.0232	0.0396	0.0262	0.0226	0.0360	0.0253	0.0222	0.0327	0.0250
$\widehat{\phi}_{\mathrm{PN},j}$	SD	250	0.0076	0.0109	0.0072	0.0075	0.0105	0.0072	0.0075	0.0101	0.0071
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	250	0.0700	0.1208	0.0829	0.0695	0.1176	0.0831	0.0690	0.1153	0.0859
$\widehat{\phi}_{\mathrm{PN},j}$	M-scale	250	0.0231	0.0404	0.0263	0.0229	0.0387	0.0259	0.0228	0.0368	0.0255

Table 4: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ under C_0 when $\rho = an^{-3}$ and $\tau = an^{-3}$ for different sizes m of the grid.

Method	Scale estimator	m	$a = 0.15$			$a = 0.5$			$a = 1$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_{\text{PS},j}$	SD	50	0.0080	0.0117	0.0100	0.0080	0.0116	0.0100	0.0080	0.0116	0.0100
$\widehat{\phi}_{\text{PS},j}$	MAD	50	0.0744	0.1288	0.0879	0.0743	0.1281	0.0872	0.0743	0.1278	0.0869
$\widehat{\phi}_{\text{PS},j}$	M-scale	50	0.0243	0.0423	0.0294	0.0243	0.0422	0.0294	0.0243	0.0420	0.0291
$\widehat{\phi}_{\text{PS},j}$	SD	100	0.0078	0.0113	0.0079	0.0078	0.0113	0.0079	0.0077	0.0112	0.0078
$\widehat{\phi}_{\text{PS},j}$	MAD	100	0.0700	0.1211	0.0825	0.0700	0.1211	0.0825	0.0703	0.1212	0.0821
$\widehat{\phi}_{\text{PS},j}$	M-scale	100	0.0237	0.0415	0.0271	0.0237	0.0413	0.0268	0.0237	0.0412	0.0267
$\widehat{\phi}_{\text{PS},j}$	SD	150	0.0077	0.0112	0.0075	0.0077	0.0111	0.0074	0.0077	0.0111	0.0073
$\widehat{\phi}_{\text{PS},j}$	MAD	150	0.0703	0.1214	0.0822	0.0704	0.1213	0.0821	0.0703	0.1210	0.0819
$\widehat{\phi}_{\text{PS},j}$	M-scale	150	0.0234	0.0413	0.0267	0.0234	0.0411	0.0265	0.0234	0.0408	0.0261
$\widehat{\phi}_{\text{PS},j}$	SD	200	0.0077	0.0112	0.0073	0.0077	0.0111	0.0073	0.0076	0.0110	0.0072
$\widehat{\phi}_{\text{PS},j}$	MAD	200	0.0705	0.1221	0.0823	0.0705	0.1221	0.0823	0.0705	0.1221	0.0823
$\widehat{\phi}_{\text{PS},j}$	M-scale	200	0.0233	0.0415	0.0268	0.0233	0.0410	0.0263	0.0233	0.0405	0.0258
$\widehat{\phi}_{\text{PS},j}$	SD	250	0.0076	0.0111	0.0072	0.0076	0.0111	0.0072	0.0076	0.0109	0.0070
$\widehat{\phi}_{\text{PS},j}$	MAD	250	0.0701	0.1210	0.0812	0.0701	0.1205	0.0807	0.0701	0.1204	0.0807
$\widehat{\phi}_{\text{PS},j}$	M-scale	250	0.0233	0.0413	0.0265	0.0233	0.0410	0.0263	0.0232	0.0402	0.0255
$\widehat{\phi}_{\text{PN},j}$	SD	50	0.0079	0.0113	0.0100	0.0078	0.0106	0.0098	0.0078	0.0098	0.0096
$\widehat{\phi}_{\text{PN},j}$	MAD	50	0.0737	0.1262	0.0885	0.0732	0.1234	0.0927	0.0720	0.1176	0.0965
$\widehat{\phi}_{\text{PN},j}$	M-scale	50	0.0240	0.0407	0.0291	0.0239	0.0384	0.0288	0.0233	0.0350	0.0281
$\widehat{\phi}_{\text{PN},j}$	SD	100	0.0077	0.0113	0.0079	0.0077	0.0111	0.0079	0.0077	0.0109	0.0078
$\widehat{\phi}_{\text{PN},j}$	MAD	100	0.0702	0.1215	0.0829	0.0701	0.1212	0.0839	0.0699	0.1195	0.0838
$\widehat{\phi}_{\text{PN},j}$	M-scale	100	0.0237	0.0414	0.0271	0.0236	0.0409	0.0271	0.0235	0.0402	0.0270
$\widehat{\phi}_{\text{PN},j}$	SD	150	0.0077	0.0112	0.0075	0.0077	0.0111	0.0075	0.0077	0.0111	0.0075
$\widehat{\phi}_{\text{PN},j}$	MAD	150	0.0704	0.1213	0.0822	0.0704	0.1213	0.0825	0.0703	0.1214	0.0827
$\widehat{\phi}_{\text{PN},j}$	M-scale	150	0.0234	0.0414	0.0268	0.0234	0.0413	0.0268	0.0234	0.0412	0.0268
$\widehat{\phi}_{\text{PN},j}$	SD	200	0.0077	0.0112	0.0073	0.0077	0.0111	0.0073	0.0077	0.0111	0.0073
$\widehat{\phi}_{\text{PN},j}$	MAD	200	0.0705	0.1223	0.0826	0.0705	0.1224	0.0827	0.0705	0.1226	0.0830
$\widehat{\phi}_{\text{PN},j}$	M-scale	200	0.0233	0.0416	0.0269	0.0233	0.0416	0.0269	0.0233	0.0415	0.0269
$\widehat{\phi}_{\text{PN},j}$	SD	250	0.0077	0.0112	0.0073	0.0076	0.0111	0.0073	0.0076	0.0111	0.0073
$\widehat{\phi}_{\text{PN},j}$	MAD	250	0.0701	0.1212	0.0814	0.0701	0.1212	0.0815	0.0701	0.1211	0.0815
$\widehat{\phi}_{\text{PN},j}$	M-scale	250	0.0233	0.0414	0.0267	0.0233	0.0414	0.0267	0.0233	0.0414	0.0267

Table 5: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ under C_0 when $\rho = an^{-4}$ and $\tau = an^{-4}$, for different sizes m of the grid.

Scale estimator	Model	$j=1$	$j=2$	$j=3$	Model	$j=1$	$j=2$	$j=3$
SD		0.0080	0.0117	0.0100		0.0254	1.6314	1.6554
MAD	C_0	0.0744	0.1288	0.0879	$C_{3,b}$	0.1183	0.6177	0.5971
$M-$scale		0.0243	0.0424	0.0295		0.0730	0.6274	0.6346
SD		1.2308	1.2307	0.0040		1.7825	0.3857	1.7563
MAD	C_2	0.3730	0.4016	0.0638	C_{23}	0.2590	0.4221	0.2847
$M-$scale		0.4231	0.4271	0.0173		0.2879	0.4655	0.3053
SD		1.7977	1.8885	1.9139		0.3071	0.4659	0.2331
MAD	$C_{3,a}$	0.2729	0.8004	0.7922	C_{Cauchy}	0.0854	0.1592	0.1100
$M-$scale		0.3014	0.9660	0.9849		0.0502	0.0850	0.0542

Table 6: Mean values of $\|\widehat{\phi}_j - \phi_j\|^2$ for the raw estimators.

Method	Scale estimator	a	$\alpha=3$			$\alpha=4$		
			$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widehat{\phi}_j$	SD		0.0080	0.0117	0.0100	0.0080	0.0117	0.0100
$\widehat{\phi}_j$	MAD		0.0744	0.1288	0.0879	0.0744	0.1288	0.0879
$\widehat{\phi}_j$	$M-$scale		0.0243	0.0424	0.0295	0.0243	0.0424	0.0295
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.05	0.0079	0.0113	0.0097	0.0080	0.0117	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.05	0.0739	0.1259	0.0848	0.0744	0.1288	0.0879
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.05	0.0242	0.0405	0.0277	0.0243	0.0424	0.0295
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.10	0.0079	0.0109	0.0093	0.0080	0.0117	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.10	0.0739	0.1233	0.0823	0.0744	0.1288	0.0879
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.10	0.0241	0.0390	0.0262	0.0243	0.0423	0.0295
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.15	0.0078	0.0106	0.0090	0.0080	0.0117	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.15	0.0737	0.1187	0.0780	0.0744	0.1288	0.0879
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.15	0.0240	0.0377	0.0249	0.0243	0.0423	0.0294
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.25	0.0078	0.0099	0.0082	0.0080	0.0117	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.25	0.0730	0.1145	0.0744	0.0744	0.1288	0.0879
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.25	0.0239	0.0353	0.0224	0.0243	0.0423	0.0294
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.5	0.0077	0.0090	0.0074	0.0080	0.0116	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.5	0.0720	0.1061	0.0663	0.0743	0.1281	0.0872
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.5	0.0239	0.0317	0.0187	0.0243	0.0422	0.0294
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.75	0.0077	0.0084	0.0067	0.0080	0.0116	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.75	0.0710	0.0982	0.0588	0.0743	0.1280	0.0871
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.75	0.0234	0.0287	0.0155	0.0243	0.0422	0.0293
$\widehat{\phi}_{\mathrm{PS},j}$	SD	1	0.0077	0.0081	0.0064	0.0080	0.0116	0.0100
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	1	0.0702	0.0929	0.0531	0.0743	0.1278	0.0869
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	1	0.0232	0.0270	0.0136	0.0243	0.0420	0.0291
$\widehat{\phi}_{\mathrm{PS},j}$	SD	1.5	0.0075	0.0075	0.0058	0.0080	0.0116	0.0099
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	1.5	0.0689	0.0845	0.0449	0.0742	0.1275	0.0866
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	1.5	0.0228	0.0246	0.0112	0.0243	0.0416	0.0287
$\widehat{\phi}_{\mathrm{PS},j}$	SD	2	0.0074	0.0071	0.0053	0.0080	0.0115	0.0099
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	2	0.0688	0.0797	0.0391	0.0742	0.1272	0.0863
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	2	0.0222	0.0229	0.0094	0.0243	0.0413	0.0284
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.05	0.0076	0.0080	0.0103	0.0079	0.0116	0.0100
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.05	0.0660	0.0911	0.1130	0.0741	0.1273	0.0875
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.05	0.0214	0.0254	0.0265	0.0242	0.0417	0.0293
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.10	0.0074	0.0074	0.0128	0.0079	0.0115	0.0100
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.10	0.0644	0.0801	0.1321	0.0739	0.1267	0.0880
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.10	0.0209	0.0228	0.0285	0.0242	0.0412	0.0291
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.15	0.0075	0.0075	0.0161	0.0079	0.0113	0.0100
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.15	0.0619	0.0731	0.1465	0.0737	0.1262	0.0885
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.15	0.0203	0.0216	0.0310	0.0240	0.0407	0.0291
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.25	0.0080	0.0081	0.0239	0.0078	0.0111	0.0100
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.25	0.0583	0.0678	0.1848	0.0735	0.1249	0.0892
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.25	0.0198	0.0210	0.0410	0.0240	0.0401	0.0290
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.5	0.0087	0.0095	0.0490	0.0078	0.0106	0.0098
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.5	0.0552	0.0650	0.2687	0.0732	0.1234	0.0927
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.5	0.0193	0.0213	0.0715	0.0239	0.0384	0.0288
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.75	0.0089	0.0103	0.0834	0.0078	0.0101	0.0096
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.75	0.0526	0.0651	0.3458	0.0728	0.1200	0.0939
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.75	0.0190	0.0220	0.1081	0.0235	0.0366	0.0284
$\widehat{\phi}_{\mathrm{PN},j}$	SD	1	0.0093	0.0113	0.1197	0.0078	0.0098	0.0096
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	1	0.0511	0.0658	0.4073	0.0720	0.1176	0.0965
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	1	0.0192	0.0233	0.1470	0.0233	0.0350	0.0281
$\widehat{\phi}_{\mathrm{PN},j}$	SD	1.5	0.0100	0.0134	0.1905	0.0078	0.0094	0.0097
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	1.5	0.0462	0.0648	0.4897	0.0714	0.1107	0.0960
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	1.5	0.0190	0.0255	0.2241	0.0230	0.0327	0.0275
$\widehat{\phi}_{\mathrm{PN},j}$	SD	2	0.0109	0.0160	0.2608	0.0076	0.0089	0.0096
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	2	0.0440	0.0677	0.5736	0.0704	0.1053	0.0971
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	2	0.0184	0.0271	0.2990	0.0228	0.0312	0.0273

Table 7: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when $\tau = an^{-\alpha}$.

Method	Scale estimator	a	$\alpha=3$			$\alpha=4$		
			$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widehat{\phi}_j$	SD		1.2308	1.2307	0.0040	1.2308	1.2307	0.0040
$\widehat{\phi}_j$	MAD		0.3730	0.4016	0.0638	0.3730	0.4016	0.0638
$\widehat{\phi}_j$	$M-$scale		0.4231	0.4271	0.0173	0.4231	0.4271	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.05	1.2301	1.2300	0.0039	1.2308	1.2307	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.05	0.3730	0.3998	0.0619	0.3730	0.4016	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.05	0.4229	0.4266	0.0169	0.4231	0.4271	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.10	1.2296	1.2295	0.0039	1.2308	1.2307	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.10	0.3725	0.3992	0.0616	0.3730	0.4016	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.10	0.4228	0.4264	0.0168	0.4231	0.4271	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.15	1.2289	1.2288	0.0039	1.2308	1.2307	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.15	0.3722	0.3981	0.0608	0.3730	0.4016	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.15	0.4219	0.4254	0.0166	0.4231	0.4271	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.25	1.2271	1.2271	0.0039	1.2308	1.2307	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.25	0.3720	0.3963	0.0589	0.3730	0.4016	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.25	0.4219	0.4249	0.0162	0.4231	0.4271	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.5	1.2228	1.2227	0.0038	1.2308	1.2307	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.5	0.3706	0.3906	0.0541	0.3730	0.4016	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.5	0.4212	0.4233	0.0151	0.4231	0.4270	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	0.75	1.2177	1.2175	0.0037	1.2306	1.2305	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	0.75	0.3687	0.3867	0.0507	0.3730	0.4015	0.0638
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.75	0.4201	0.4219	0.0143	0.4231	0.4270	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	1	1.2122	1.2120	0.0037	1.2303	1.2303	0.0040
$\widehat{\phi}_{\text{PS},j}$	MAD	1	0.3684	0.3852	0.0490	0.3730	0.4014	0.0637
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1	0.4190	0.4204	0.0137	0.4231	0.4270	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	1.5	1.2017	1.2015	0.0036	1.2302	1.2301	0.0039
$\widehat{\phi}_{\text{PS},j}$	MAD	1.5	0.3662	0.3804	0.0445	0.3730	0.4013	0.0636
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1.5	0.4154	0.4161	0.0125	0.4231	0.4270	0.0173
$\widehat{\phi}_{\text{PS},j}$	SD	2	1.1921	1.1919	0.0036	1.2302	1.2301	0.0039
$\widehat{\phi}_{\text{PS},j}$	MAD	2	0.3631	0.3739	0.0403	0.3730	0.4013	0.0636
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	2	0.4131	0.4136	0.0117	0.4230	0.4269	0.0172
$\widehat{\phi}_{\text{PN},j}$	SD	0.05	1.2177	1.2198	0.0056	1.2303	1.2303	0.0040
$\widehat{\phi}_{\text{PN},j}$	MAD	0.05	0.3666	0.3820	0.0837	0.3730	0.4000	0.0629
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.05	0.4170	0.4206	0.0226	0.4229	0.4269	0.0174
$\widehat{\phi}_{\text{PN},j}$	SD	0.10	1.2078	1.2123	0.0074	1.2303	1.2303	0.0040
$\widehat{\phi}_{\text{PN},j}$	MAD	0.10	0.3637	0.3789	0.1088	0.3726	0.3990	0.0629
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.10	0.4128	0.4184	0.0293	0.4228	0.4265	0.0173
$\widehat{\phi}_{\text{PN},j}$	SD	0.15	1.1934	1.2002	0.0095	1.2303	1.2303	0.0040
$\widehat{\phi}_{\text{PN},j}$	MAD	0.15	0.3605	0.3752	0.1297	0.3723	0.3983	0.0632
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.15	0.4092	0.4173	0.0368	0.4226	0.4262	0.0174
$\widehat{\phi}_{\text{PN},j}$	SD	0.25	1.1736	1.1850	0.0148	1.2301	1.2302	0.0040
$\widehat{\phi}_{\text{PN},j}$	MAD	0.25	0.3560	0.3736	0.1657	0.3720	0.3966	0.0625
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.25	0.4042	0.4174	0.0513	0.4226	0.4260	0.0175
$\widehat{\phi}_{\text{PN},j}$	SD	0.5	1.1092	1.1319	0.0307	1.2288	1.2289	0.0041
$\widehat{\phi}_{\text{PN},j}$	MAD	0.5	0.3379	0.3654	0.2499	0.3719	0.3959	0.0637
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.5	0.3865	0.4121	0.0979	0.4220	0.4248	0.0174
$\widehat{\phi}_{\text{PN},j}$	SD	0.75	1.0430	1.0767	0.0470	1.2277	1.2279	0.0041
$\widehat{\phi}_{\text{PN},j}$	MAD	0.75	0.3254	0.3633	0.3172	0.3724	0.3947	0.0645
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.75	0.3707	0.4073	0.1436	0.4220	0.4249	0.0181
$\widehat{\phi}_{\text{PN},j}$	SD	1	0.9727	1.0174	0.0653	1.2279	1.2282	0.0042
$\widehat{\phi}_{\text{PN},j}$	MAD	1	0.3111	0.3586	0.3807	0.3726	0.3950	0.0674
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1	0.3531	0.3996	0.1832	0.4218	0.4246	0.0185
$\widehat{\phi}_{\text{PN},j}$	SD	1.5	0.8598	0.9248	0.1084	1.2263	1.2269	0.0044
$\widehat{\phi}_{\text{PN},j}$	MAD	1.5	0.2876	0.3521	0.4924	0.3723	0.3925	0.0690
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1.5	0.3253	0.3895	0.2613	0.4227	0.4254	0.0186
$\widehat{\phi}_{\text{PN},j}$	SD	2	0.7401	0.8205	0.1525	1.2264	1.2271	0.0045
$\widehat{\phi}_{\text{PN},j}$	MAD	2	0.2664	0.3454	0.5727	0.3703	0.3922	0.0736
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	2	0.2987	0.3780	0.3363	0.4205	0.4234	0.0192

Table 8: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_2 when $\tau = an^{-\alpha}$.

Method	Scale estimator	a	$\alpha=3$			$\alpha=4$		
			$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widehat{\phi}_j$	SD		1.7977	1.8885	1.9139	1.7977	1.8885	1.9139
$\widehat{\phi}_j$	MAD		0.2729	0.8004	0.7922	0.2729	0.8004	0.7922
$\widehat{\phi}_j$	$M-$scale		0.3014	0.9660	0.9849	0.3014	0.9660	0.9849
$\widehat{\phi}_{\text{PS},j}$	SD	0.05	1.7976	1.8886	1.9136	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.05	0.2714	0.7840	0.7761	0.2729	0.8004	0.7922
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.05	0.3000	0.9555	0.9746	0.3014	0.9659	0.9849
$\widehat{\phi}_{\text{PS},j}$	SD	0.10	1.7967	1.8888	1.9135	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.10	0.2707	0.7763	0.7697	0.2729	0.8004	0.7922
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.10	0.2990	0.9444	0.9656	0.3014	0.9657	0.9847
$\widehat{\phi}_{\text{PS},j}$	SD	0.15	1.7956	1.8890	1.9136	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.15	0.2689	0.7668	0.7609	0.2729	0.8004	0.7922
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.15	0.2982	0.9384	0.9598	0.3014	0.9653	0.9844
$\widehat{\phi}_{\text{PS},j}$	SD	0.25	1.7930	1.8889	1.9128	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.25	0.2660	0.7449	0.7400	0.2729	0.8004	0.7922
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.25	0.2964	0.9235	0.9461	0.3014	0.9653	0.9844
$\widehat{\phi}_{\text{PS},j}$	SD	0.5	1.7872	1.8896	1.9117	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.5	0.2622	0.7069	0.7063	0.2729	0.7999	0.7913
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.5	0.2912	0.8651	0.8933	0.3014	0.9652	0.9843
$\widehat{\phi}_{\text{PS},j}$	SD	0.75	1.7823	1.8900	1.9112	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	0.75	0.2577	0.6645	0.6678	0.2721	0.7962	0.7884
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.75	0.2851	0.8120	0.8457	0.3014	0.9652	0.9843
$\widehat{\phi}_{\text{PS},j}$	SD	1	1.7752	1.8904	1.9096	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	1	0.2555	0.6402	0.6456	0.2721	0.7941	0.7867
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1	0.2806	0.7728	0.8132	0.3011	0.9632	0.9823
$\widehat{\phi}_{\text{PS},j}$	SD	1.5	1.7612	1.8906	1.9071	1.7977	1.8886	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	1.5	0.2469	0.5819	0.5936	0.2721	0.7928	0.7854
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1.5	0.2701	0.6999	0.7479	0.3009	0.9613	0.9806
$\widehat{\phi}_{\text{PS},j}$	SD	2	1.7394	1.8903	1.9067	1.7977	1.8885	1.9139
$\widehat{\phi}_{\text{PS},j}$	MAD	2	0.2365	0.5259	0.5408	0.2721	0.7925	0.7848
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	2	0.2570	0.6314	0.6856	0.3005	0.9587	0.9778
$\widehat{\phi}_{\text{PN},j}$	SD	0.05	1.1991	1.8687	1.9247	1.7961	1.8884	1.9144
$\widehat{\phi}_{\text{PN},j}$	MAD	0.05	0.1654	0.6677	0.9259	0.2712	0.7955	0.7911
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.05	0.1720	0.8166	1.0945	0.2997	0.9664	0.9886
$\widehat{\phi}_{\text{PN},j}$	SD	0.10	0.3089	1.7563	1.9172	1.7929	1.8881	1.9146
$\widehat{\phi}_{\text{PN},j}$	MAD	0.10	0.1164	0.5668	0.9717	0.2691	0.7926	0.7920
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.10	0.0986	0.5954	1.0570	0.2988	0.9702	0.9959
$\widehat{\phi}_{\text{PN},j}$	SD	0.15	0.0959	1.6179	1.9006	1.7917	1.8882	1.9151
$\widehat{\phi}_{\text{PN},j}$	MAD	0.15	0.0884	0.4555	0.9694	0.2677	0.7908	0.7944
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.15	0.0619	0.4484	1.0327	0.2974	0.9675	0.9972
$\widehat{\phi}_{\text{PN},j}$	SD	0.25	0.0740	1.3619	1.8378	1.7863	1.8886	1.9148
$\widehat{\phi}_{\text{PN},j}$	MAD	0.25	0.0648	0.3332	0.9896	0.2645	0.7836	0.7952
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.25	0.0364	0.2517	0.9864	0.2946	0.9643	1.0008
$\widehat{\phi}_{\text{PN},j}$	SD	0.5	0.0812	1.0008	1.6349	1.7720	1.8890	1.9165
$\widehat{\phi}_{\text{PN},j}$	MAD	0.5	0.0526	0.2080	0.9771	0.2593	0.7855	0.8131
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.5	0.0276	0.1046	0.9265	0.2873	0.9660	1.0190
$\widehat{\phi}_{\text{PN},j}$	SD	0.75	0.0882	0.9942	1.6186	1.7586	1.8885	1.9184
$\widehat{\phi}_{\text{PN},j}$	MAD	0.75	0.0468	0.1606	0.9668	0.2533	0.7831	0.8254
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.75	0.0271	0.0631	0.8130	0.2783	0.9566	1.0271
$\widehat{\phi}_{\text{PN},j}$	SD	1	0.0921	1.0504	1.6565	1.7354	1.8887	1.9200
$\widehat{\phi}_{\text{PN},j}$	MAD	1	0.0457	0.1375	0.9365	0.2467	0.7826	0.8418
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1	0.0267	0.0527	0.7608	0.2721	0.9635	1.0493
$\widehat{\phi}_{\text{PN},j}$	SD	1.5	0.0980	1.1618	1.7271	1.6922	1.8870	1.9218
$\widehat{\phi}_{\text{PN},j}$	MAD	1.5	0.0456	0.1120	0.9501	0.2338	0.7712	0.8603
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1.5	0.0266	0.0507	0.7417	0.2553	0.9394	1.0591
$\widehat{\phi}_{\text{PN},j}$	SD	2	0.1031	1.2519	1.7678	1.6482	1.8871	1.9239
$\widehat{\phi}_{\text{PN},j}$	MAD	2	0.0446	0.1152	0.9982	0.2216	0.7533	0.8699
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	2	0.0266	0.0552	0.7756	0.2401	0.9237	1.0732

Table 9: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under $C_{3,a}$ when $\tau = an^{-\alpha}$.

Method	Scale estimator	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_j$	SD		0.0254	1.6314	1.6554	0.0254	1.6314	1.6554
$\widehat{\phi}_j$	MAD		0.1183	0.6177	0.5971	0.1183	0.6177	0.5971
$\widehat{\phi}_j$	$M-$scale		0.0730	0.6274	0.6346	0.0730	0.6274	0.6346
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.05	0.0251	1.6117	1.6360	0.0254	1.6314	1.6553
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.05	0.1180	0.6116	0.5903	0.1183	0.6177	0.5971
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.05	0.0717	0.6097	0.6167	0.0730	0.6274	0.6346
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.10	0.0247	1.5937	1.6182	0.0254	1.6313	1.6553
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.10	0.1177	0.6022	0.5808	0.1183	0.6177	0.5971
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.10	0.0703	0.5947	0.6019	0.0730	0.6270	0.6342
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.15	0.0244	1.5743	1.5990	0.0254	1.6312	1.6551
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.15	0.1171	0.5936	0.5724	0.1183	0.6177	0.5971
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.15	0.0685	0.5785	0.5867	0.0730	0.6268	0.6340
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.25	0.0235	1.5300	1.5559	0.0254	1.6305	1.6545
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.25	0.1156	0.5622	0.5418	0.1183	0.6177	0.5971
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.25	0.0667	0.5488	0.5580	0.0730	0.6266	0.6338
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.5	0.0222	1.3781	1.4045	0.0254	1.6296	1.6537
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.5	0.1109	0.5045	0.4840	0.1183	0.6177	0.5971
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.5	0.0603	0.4819	0.4909	0.0730	0.6257	0.6330
$\widehat{\phi}_{\mathrm{PS},j}$	SD	0.75	0.0209	1.1843	1.2106	0.0254	1.6288	1.6529
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	0.75	0.1083	0.4522	0.4347	0.1183	0.6174	0.5969
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	0.75	0.0570	0.4196	0.4296	0.0729	0.6248	0.6320
$\widehat{\phi}_{\mathrm{PS},j}$	SD	1	0.0198	0.9222	0.9475	0.0254	1.6276	1.6516
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	1	0.1033	0.4005	0.3844	0.1183	0.6168	0.5962
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	1	0.0532	0.3641	0.3742	0.0729	0.6244	0.6316
$\widehat{\phi}_{\mathrm{PS},j}$	SD	1.5	0.0183	0.4274	0.4473	0.0253	1.6253	1.6493
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	1.5	0.0950	0.3119	0.2947	0.1183	0.6150	0.5945
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	1.5	0.0452	0.2371	0.2462	0.0729	0.6233	0.6304
$\widehat{\phi}_{\mathrm{PS},j}$	SD	2	0.0173	0.1491	0.1640	0.0252	1.6231	1.6472
$\widehat{\phi}_{\mathrm{PS},j}$	MAD	2	0.0891	0.2514	0.2368	0.1183	0.6150	0.5945
$\widehat{\phi}_{\mathrm{PS},j}$	$M-$scale	2	0.0413	0.1650	0.1714	0.0729	0.6231	0.6302
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.05	0.0141	0.3471	0.5910	0.0250	1.6211	1.6480
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.05	0.0750	0.3075	0.5051	0.1180	0.6150	0.5974
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.05	0.0318	0.2159	0.4374	0.0717	0.6201	0.6321
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.10	0.0131	0.0397	0.1720	0.0246	1.6136	1.6429
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.10	0.0644	0.1890	0.4423	0.1177	0.6148	0.6011
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.10	0.0264	0.0926	0.3101	0.0709	0.6125	0.6291
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.15	0.0135	0.0206	0.1199	0.0244	1.6049	1.6370
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.15	0.0617	0.1394	0.4084	0.1169	0.6121	0.6025
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.15	0.0254	0.0516	0.2415	0.0686	0.6048	0.6273
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.25	0.0150	0.0177	0.1176	0.0232	1.5875	1.6256
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.25	0.0576	0.0988	0.3934	0.1146	0.5981	0.5954
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.25	0.0254	0.0350	0.2000	0.0667	0.5938	0.6255
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.5	0.0170	0.0207	0.1981	0.0220	1.5442	1.5978
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.5	0.0542	0.0748	0.4397	0.1101	0.5758	0.5913
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.5	0.0254	0.0304	0.1958	0.0610	0.5627	0.6147
$\widehat{\phi}_{\mathrm{PN},j}$	SD	0.75	0.0191	0.0246	0.2803	0.0206	1.4916	1.5622
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	0.75	0.0532	0.0707	0.4945	0.1078	0.5644	0.5967
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	0.75	0.0247	0.0302	0.2358	0.0571	0.5427	0.6149
$\widehat{\phi}_{\mathrm{PN},j}$	SD	1	0.0205	0.0284	0.3638	0.0197	1.4343	1.5216
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	1	0.0520	0.0737	0.5518	0.1030	0.5395	0.5857
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	1	0.0242	0.0308	0.2833	0.0534	0.5158	0.6054
$\widehat{\phi}_{\mathrm{PN},j}$	SD	1.5	0.0229	0.0399	0.5143	0.0180	1.3033	1.4270
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	1.5	0.0493	0.0745	0.6525	0.0954	0.4867	0.5626
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	1.5	0.0241	0.0339	0.3798	0.0471	0.4621	0.5835
$\widehat{\phi}_{\mathrm{PN},j}$	SD	2	0.0240	0.0486	0.6355	0.0172	1.1662	1.3250
$\widehat{\phi}_{\mathrm{PN},j}$	MAD	2	0.0473	0.0822	0.7511	0.0892	0.4435	0.5449
$\widehat{\phi}_{\mathrm{PN},j}$	$M-$scale	2	0.0235	0.0369	0.4719	0.0428	0.4154	0.5639

Table 10: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under $C_{3,b}$ when $\tau = an^{-\alpha}$.

Method	Scale estimator	a	$\alpha=3$			$\alpha=4$		
			$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widehat{\phi}_j$	SD		1.7825	0.3857	1.7563	1.7825	0.3857	1.7563
$\widehat{\phi}_j$	MAD		0.2590	0.4221	0.2847	0.2590	0.4221	0.2847
$\widehat{\phi}_j$	$M-$scale		0.2879	0.4655	0.3053	0.2879	0.4655	0.3053
$\widehat{\phi}_{\text{PS},j}$	SD	0.05	1.7824	0.3887	1.7550	1.7825	0.3857	1.7563
$\widehat{\phi}_{\text{PS},j}$	MAD	0.05	0.2587	0.4195	0.2823	0.2590	0.4220	0.2845
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.05	0.2875	0.4628	0.3029	0.2879	0.4655	0.3052
$\widehat{\phi}_{\text{PS},j}$	SD	0.10	1.7804	0.3911	1.7529	1.7825	0.3857	1.7563
$\widehat{\phi}_{\text{PS},j}$	MAD	0.10	0.2577	0.4131	0.2770	0.2590	0.4220	0.2845
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.10	0.2872	0.4592	0.2998	0.2879	0.4655	0.3052
$\widehat{\phi}_{\text{PS},j}$	SD	0.15	1.7791	0.3953	1.7518	1.7824	0.3874	1.7564
$\widehat{\phi}_{\text{PS},j}$	MAD	0.15	0.2563	0.4092	0.2735	0.2591	0.4223	0.2844
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.15	0.2860	0.4555	0.2966	0.2879	0.4655	0.3052
$\widehat{\phi}_{\text{PS},j}$	SD	0.25	1.7779	0.4008	1.7490	1.7824	0.3874	1.7563
$\widehat{\phi}_{\text{PS},j}$	MAD	0.25	0.2553	0.4037	0.2694	0.2591	0.4223	0.2844
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.25	0.2843	0.4485	0.2905	0.2879	0.4655	0.3052
$\widehat{\phi}_{\text{PS},j}$	SD	0.5	1.7723	0.4196	1.7363	1.7824	0.3875	1.7563
$\widehat{\phi}_{\text{PS},j}$	MAD	0.5	0.2530	0.3872	0.2546	0.2591	0.4223	0.2844
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.5	0.2823	0.4378	0.2813	0.2879	0.4649	0.3047
$\widehat{\phi}_{\text{PS},j}$	SD	0.75	1.7663	0.4342	1.7215	1.7825	0.3877	1.7563
$\widehat{\phi}_{\text{PS},j}$	MAD	0.75	0.2517	0.3775	0.2454	0.2591	0.4223	0.2844
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.75	0.2797	0.4242	0.2694	0.2879	0.4648	0.3046
$\widehat{\phi}_{\text{PS},j}$	SD	1	1.7632	0.4571	1.7112	1.7825	0.3878	1.7562
$\widehat{\phi}_{\text{PS},j}$	MAD	1	0.2496	0.3649	0.2342	0.2591	0.4222	0.2843
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1	0.2771	0.4138	0.2611	0.2879	0.4646	0.3044
$\widehat{\phi}_{\text{PS},j}$	SD	1.5	1.7526	0.4907	1.6835	1.7825	0.3878	1.7562
$\widehat{\phi}_{\text{PS},j}$	MAD	1.5	0.2435	0.3419	0.2137	0.2591	0.4222	0.2843
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1.5	0.2728	0.3959	0.2455	0.2879	0.4644	0.3042
$\widehat{\phi}_{\text{PS},j}$	SD	2	1.7408	0.5360	1.6523	1.7823	0.3881	1.7559
$\widehat{\phi}_{\text{PS},j}$	MAD	2	0.2394	0.3232	0.1963	0.2591	0.4216	0.2837
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	2	0.2669	0.3777	0.2299	0.2878	0.4641	0.3039
$\widehat{\phi}_{\text{PN},j}$	SD	0.05	1.5841	1.0675	1.3787	1.7794	0.3924	1.7554
$\widehat{\phi}_{\text{PN},j}$	MAD	0.05	0.2120	0.3628	0.4227	0.2583	0.4222	0.2872
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.05	0.2361	0.4120	0.4700	0.2875	0.4657	0.3085
$\widehat{\phi}_{\text{PN},j}$	SD	0.10	1.4791	1.3522	0.8009	1.7785	0.3951	1.7542
$\widehat{\phi}_{\text{PN},j}$	MAD	0.10	0.1927	0.3281	0.5028	0.2579	0.4224	0.2892
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.10	0.2114	0.3694	0.5512	0.2864	0.4653	0.3111
$\widehat{\phi}_{\text{PN},j}$	SD	0.15	1.4571	1.4101	0.4440	1.7745	0.4018	1.7516
$\widehat{\phi}_{\text{PN},j}$	MAD	0.15	0.1813	0.3015	0.5521	0.2572	0.4227	0.2928
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.15	0.1932	0.3242	0.5842	0.2861	0.4645	0.3126
$\widehat{\phi}_{\text{PN},j}$	SD	0.25	1.4344	1.4216	0.3038	1.7701	0.4189	1.7453
$\widehat{\phi}_{\text{PN},j}$	MAD	0.25	0.1642	0.2648	0.6089	0.2550	0.4167	0.2926
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.25	0.1713	0.2642	0.6139	0.2839	0.4628	0.3172
$\widehat{\phi}_{\text{PN},j}$	SD	0.5	1.3906	1.3926	0.4061	1.7568	0.4502	1.7320
$\widehat{\phi}_{\text{PN},j}$	MAD	0.5	0.1523	0.2298	0.6664	0.2518	0.4160	0.3055
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.5	0.1540	0.2113	0.6139	0.2816	0.4595	0.3275
$\widehat{\phi}_{\text{PN},j}$	SD	0.75	1.3716	1.3846	0.5333	1.7446	0.4832	1.7208
$\widehat{\phi}_{\text{PN},j}$	MAD	0.75	0.1429	0.2126	0.7109	0.2496	0.4077	0.3094
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.75	0.1463	0.1951	0.5917	0.2773	0.4576	0.3400
$\widehat{\phi}_{\text{PN},j}$	SD	1	1.3303	1.3475	0.6386	1.7347	0.5234	1.7066
$\widehat{\phi}_{\text{PN},j}$	MAD	1	0.1354	0.2025	0.7465	0.2462	0.4032	0.3156
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1	0.1387	0.1837	0.5833	0.2751	0.4572	0.3516
$\widehat{\phi}_{\text{PN},j}$	SD	1.5	1.2728	1.3010	0.8004	1.7113	0.5988	1.6797
$\widehat{\phi}_{\text{PN},j}$	MAD	1.5	0.1261	0.1933	0.7833	0.2396	0.3920	0.3269
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1.5	0.1299	0.1827	0.6111	0.2691	0.4540	0.3747
$\widehat{\phi}_{\text{PN},j}$	SD	2	1.2048	1.2419	0.9162	1.6962	0.6816	1.6488
$\widehat{\phi}_{\text{PN},j}$	MAD	2	0.1194	0.1938	0.8525	0.2365	0.3959	0.3493
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	2	0.1233	0.1845	0.6364	0.2620	0.4498	0.3917

Table 11: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_{23} when $\tau = an^{-\alpha}$.

Method	Scale estimator	a	$\alpha=3$			$\alpha=4$		
			$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
$\widehat{\phi}_j$	SD		0.3071	0.4659	0.2331	0.3071	0.4659	0.2331
$\widehat{\phi}_j$	MAD		0.0854	0.1592	0.1100	0.0854	0.1592	0.1100
$\widehat{\phi}_j$	$M-$scale		0.0502	0.0850	0.0542	0.0502	0.0850	0.0542
$\widehat{\phi}_{\text{PS},j}$	SD	0.05	0.3071	0.4658	0.2329	0.3071	0.4659	0.2331
$\widehat{\phi}_{\text{PS},j}$	MAD	0.05	0.0855	0.1590	0.1092	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.05	0.0501	0.0837	0.0529	0.0502	0.0849	0.0540
$\widehat{\phi}_{\text{PS},j}$	SD	0.10	0.3071	0.4656	0.2328	0.3071	0.4659	0.2331
$\widehat{\phi}_{\text{PS},j}$	MAD	0.10	0.0850	0.1577	0.1082	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.10	0.0500	0.0818	0.0512	0.0502	0.0849	0.0540
$\widehat{\phi}_{\text{PS},j}$	SD	0.15	0.3071	0.4654	0.2326	0.3071	0.4659	0.2331
$\widehat{\phi}_{\text{PS},j}$	MAD	0.15	0.0850	0.1561	0.1067	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.15	0.0497	0.0806	0.0502	0.0502	0.0849	0.0540
$\widehat{\phi}_{\text{PS},j}$	SD	0.25	0.3071	0.4655	0.2323	0.3071	0.4659	0.2331
$\widehat{\phi}_{\text{PS},j}$	MAD	0.25	0.0849	0.1542	0.1049	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.25	0.0497	0.0795	0.0490	0.0502	0.0849	0.0540
$\widehat{\phi}_{\text{PS},j}$	SD	0.5	0.3069	0.4645	0.2314	0.3071	0.4659	0.2330
$\widehat{\phi}_{\text{PS},j}$	MAD	0.5	0.0846	0.1469	0.0976	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.5	0.0491	0.0749	0.0444	0.0502	0.0848	0.0540
$\widehat{\phi}_{\text{PS},j}$	SD	0.75	0.3069	0.4637	0.2306	0.3071	0.4659	0.2330
$\widehat{\phi}_{\text{PS},j}$	MAD	0.75	0.0837	0.1391	0.0900	0.0854	0.1590	0.1098
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	0.75	0.0489	0.0724	0.0415	0.0502	0.0848	0.0539
$\widehat{\phi}_{\text{PS},j}$	SD	1	0.3069	0.4629	0.2298	0.3071	0.4659	0.2330
$\widehat{\phi}_{\text{PS},j}$	MAD	1	0.0829	0.1332	0.0841	0.0854	0.1592	0.1097
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1	0.0487	0.0694	0.0384	0.0501	0.0846	0.0538
$\widehat{\phi}_{\text{PS},j}$	SD	1.5	0.3069	0.4610	0.2280	0.3071	0.4659	0.2330
$\widehat{\phi}_{\text{PS},j}$	MAD	1.5	0.0824	0.1216	0.0728	0.0854	0.1590	0.1096
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	1.5	0.0481	0.0639	0.0333	0.0501	0.0845	0.0537
$\widehat{\phi}_{\text{PS},j}$	SD	2	0.3069	0.4593	0.2262	0.3071	0.4659	0.2330
$\widehat{\phi}_{\text{PS},j}$	MAD	2	0.0815	0.1134	0.0655	0.0855	0.1591	0.1093
$\widehat{\phi}_{\text{PS},j}$	$M-$scale	2	0.0474	0.0606	0.0300	0.0501	0.0841	0.0533
$\widehat{\phi}_{\text{PN},j}$	SD	0.05	0.2750	0.3631	0.2238	0.3066	0.4644	0.2331
$\widehat{\phi}_{\text{PN},j}$	MAD	0.05	0.0723	0.1034	0.1273	0.0857	0.1589	0.1103
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.05	0.0436	0.0550	0.0553	0.0499	0.0833	0.0532
$\widehat{\phi}_{\text{PN},j}$	SD	0.10	0.2670	0.3332	0.2298	0.3052	0.4631	0.2339
$\widehat{\phi}_{\text{PN},j}$	MAD	0.10	0.0701	0.0891	0.1474	0.0849	0.1574	0.1106
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.10	0.0420	0.0488	0.0627	0.0497	0.0828	0.0535
$\widehat{\phi}_{\text{PN},j}$	SD	0.15	0.2649	0.3258	0.2418	0.3047	0.4592	0.2321
$\widehat{\phi}_{\text{PN},j}$	MAD	0.15	0.0684	0.0837	0.1693	0.0848	0.1569	0.1111
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.15	0.0415	0.0468	0.0711	0.0493	0.0814	0.0530
$\widehat{\phi}_{\text{PN},j}$	SD	0.25	0.2544	0.3096	0.2708	0.3042	0.4575	0.2323
$\widehat{\phi}_{\text{PN},j}$	MAD	0.25	0.0677	0.0808	0.2166	0.0844	0.1549	0.1118
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.25	0.0408	0.0453	0.0841	0.0491	0.0804	0.0529
$\widehat{\phi}_{\text{PN},j}$	SD	0.5	0.2387	0.2927	0.3304	0.3020	0.4486	0.2310
$\widehat{\phi}_{\text{PN},j}$	MAD	0.5	0.0665	0.0798	0.3079	0.0830	0.1503	0.1127
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.5	0.0390	0.0445	0.1245	0.0488	0.0779	0.0535
$\widehat{\phi}_{\text{PN},j}$	SD	0.75	0.2317	0.2941	0.3921	0.2986	0.4404	0.2306
$\widehat{\phi}_{\text{PN},j}$	MAD	0.75	0.0643	0.0791	0.3714	0.0830	0.1468	0.1140
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	0.75	0.0356	0.0426	0.1745	0.0482	0.0755	0.0535
$\widehat{\phi}_{\text{PN},j}$	SD	1	0.2266	0.2903	0.4480	0.2962	0.4374	0.2337
$\widehat{\phi}_{\text{PN},j}$	MAD	1	0.0623	0.0801	0.4391	0.0826	0.1401	0.1125
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1	0.0338	0.0421	0.2202	0.0475	0.0729	0.0533
$\widehat{\phi}_{\text{PN},j}$	SD	1.5	0.2147	0.2980	0.5370	0.2919	0.4197	0.2259
$\widehat{\phi}_{\text{PN},j}$	MAD	1.5	0.0574	0.0810	0.5517	0.0802	0.1306	0.1116
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	1.5	0.0313	0.0430	0.3212	0.0468	0.0692	0.0536
$\widehat{\phi}_{\text{PN},j}$	SD	2	0.1992	0.2883	0.6208	0.2880	0.4057	0.2242
$\widehat{\phi}_{\text{PN},j}$	MAD	2	0.0526	0.0826	0.6515	0.0789	0.1263	0.1156
$\widehat{\phi}_{\text{PN},j}$	$M-$scale	2	0.0298	0.0449	0.4050	0.0462	0.0660	0.0533

Table 12: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_c when $\tau = an^{-\alpha}$.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
			Fourier Basis				$B-$splines Basis		
$\widehat{\phi}_{\mathrm{SI},j}$	SD	10	0	2	2	10	0.0380	1.9325	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	10	0	2	2	10	0.0385	1.9335	1.9747
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	10	0	2	2	10	0.0380	1.9326	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	SD	20	0.0046	0.0046	2	20	0.0076	0.0117	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	20	0.0594	0.0594	2	20	0.0588	0.0658	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	20	0.0178	0.0178	2	20	0.0176	0.0246	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	SD	30	0.0053	0.0097	0.0059	50	0.0076	0.0111	0.0071
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	30	0.0703	0.1237	0.0836	50	0.0703	0.1237	0.0836
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	30	0.0230	0.0410	0.0262	50	0.0230	0.0410	0.0262

Table 13: Mean values of $\|\widehat{\phi}_{j,\mathrm{SI}} - \phi_j\|^2$, under C_0.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
			Fourier Basis				$B-$splines Basis		
$\widehat{\phi}_{\mathrm{SI},j}$	SD	10	0	2	2	10	0.0380	1.9325	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	10	0	2	2	10	0.0404	1.9330	1.9745
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	10	0	2	2	10	0.0381	1.9326	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	SD	20	1.1477	1.1477	2	20	1.1472	1.1336	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	20	0.3536	0.3536	2	20	0.3497	0.3556	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	20	0.4076	0.4076	2	20	0.4053	0.4109	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	SD	30	1.1477	1.1479	0.0013	50	1.1472	1.1469	0.0022
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	30	0.3611	0.3890	0.0583	50	0.3611	0.3890	0.0583
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	30	0.4081	0.4125	0.0150	50	0.4081	0.4125	0.0150

Table 14: Mean values of $\|\widehat{\phi}_{j,\mathrm{SI}} - \phi_j\|^2$, under C_2.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
			Fourier Basis				$B-$splines Basis		
$\widehat{\phi}_{\mathrm{SI},j}$	SD	10	0	2	2	10	0.0380	1.9803	1.9925
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	10	0	2	2	10	0.0391	1.9519	1.9817
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	10	0	2	2	10	0.0380	1.9435	1.9786
$\widehat{\phi}_{\mathrm{SI},j}$	SD	20	0.0044	0.0044	2	20	1.7884	0.0117	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	20	0.0588	0.0588	2	20	0.0589	0.0691	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	20	0.0204	0.0204	2	20	0.0202	0.0272	1.9744
$\widehat{\phi}_{\mathrm{SI},j}$	SD	30	1.8028	1.8942	1.9412	50	1.7884	1.8932	1.9142
$\widehat{\phi}_{\mathrm{SI},j}$	MAD	30	0.2716	0.8199	0.8083	50	0.2716	0.8199	0.8083
$\widehat{\phi}_{\mathrm{SI},j}$	$M-$scale	30	0.2977	0.9922	1.0013	50	0.2977	0.9922	1.0013

Table 15: Mean values of $\|\widehat{\phi}_{j,\mathrm{SI}} - \phi_j\|^2$, under $C_{3,a}$.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
				Fourier Basis				$B-$splines Basis	
$\widehat{\phi}_{\text{SI},j}$	SD	10	0	2	2	10	0.0380	1.9327	1.9744
$\widehat{\phi}_{\text{SI},j}$	MAD	10	0	2	2	10	0.0386	1.9382	1.9765
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	10	0	2	2	10	0.0380	1.9332	1.9746
$\widehat{\phi}_{\text{SI},j}$	SD	20	0.0044	0.0044	2	20	0.0235	0.0116	1.9744
$\widehat{\phi}_{\text{SI},j}$	MAD	20	0.0588	0.0588	2	20	0.0597	0.0673	1.9744
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	20	0.0204	0.0204	2	20	0.0202	0.0272	1.9744
$\widehat{\phi}_{\text{SI},j}$	SD	30	0.0171	1.6657	1.6668	50	0.0235	1.6650	1.6648
$\widehat{\phi}_{\text{SI},j}$	MAD	30	0.1154	0.6330	0.6033	50	0.1154	0.6330	0.6033
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	30	0.0679	0.6231	0.6191	50	0.0679	0.6231	0.6191

Table 16: Mean values of $\|\widehat{\phi}_{j,\text{SI}} - \phi_j\|^2$, under $C_{3,b}$.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
				Fourier Basis				$B-$splines Basis	
$\widehat{\phi}_{\text{SI},j}$	SD	10	0	2	2	10	0.0380	1.9327	1.9744
$\widehat{\phi}_{\text{SI},j}$	MAD	10	0	2	2	10	0.0412	1.9411	1.9776
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	10	0	2	2	10	0.0381	1.9371	1.9761
$\widehat{\phi}_{\text{SI},j}$	SD	20	1.4572	1.4572	2	20	1.7707	1.4505	1.9744
$\widehat{\phi}_{\text{SI},j}$	MAD	20	0.1516	0.1516	2	20	0.1500	0.1585	1.9744
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	20	0.1620	0.1620	2	20	0.1622	0.1687	1.9744
$\widehat{\phi}_{\text{SI},j}$	SD	30	1.7910	0.3910	1.7569	50	1.7707	0.4060	1.7448
$\widehat{\phi}_{\text{SI},j}$	MAD	30	0.2440	0.4146	0.2751	50	0.2440	0.4146	0.2751
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	30	0.2754	0.4644	0.2923	50	0.2754	0.4644	0.2923

Table 17: Mean values of $\|\widehat{\phi}_{j,\text{SI}} - \phi_j\|^2$, under C_{23}.

Method	Scale estimator	p_n	$j=1$	$j=2$	$j=3$	p_n	$j=1$	$j=2$	$j=3$
				Fourier Basis				$B-$splines Basis	
$\widehat{\phi}_{\text{SI},j}$	SD	10	0	2	2	10	0.0404	1.9347	1.9752
$\widehat{\phi}_{\text{SI},j}$	MAD	10	0	2	2	10	0.0401	1.9347	1.9752
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	10	0	2	2	10	0.0382	1.9330	1.9745
$\widehat{\phi}_{\text{SI},j}$	SD	20	0.4380	0.4380	2	20	0.5770	0.4446	1.9744
$\widehat{\phi}_{\text{SI},j}$	MAD	20	0.2538	0.2538	2	20	0.1641	0.1711	1.9744
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	20	0.1964	0.1964	2	20	0.1166	0.1233	1.9744
$\widehat{\phi}_{\text{SI},j}$	SD	30	0.5558	0.7519	0.4991	50	0.5770	0.7806	0.4942
$\widehat{\phi}_{\text{SI},j}$	MAD	30	0.2960	0.5189	0.3549	50	0.2024	0.3566	0.2401
$\widehat{\phi}_{\text{SI},j}$	$M-$scale	30	0.2339	0.4166	0.2766	50	0.1401	0.2549	0.1715

Table 18: Mean values of $\|\widehat{\phi}_{j,\text{SI}} - \phi_j\|^2$, under C_c.

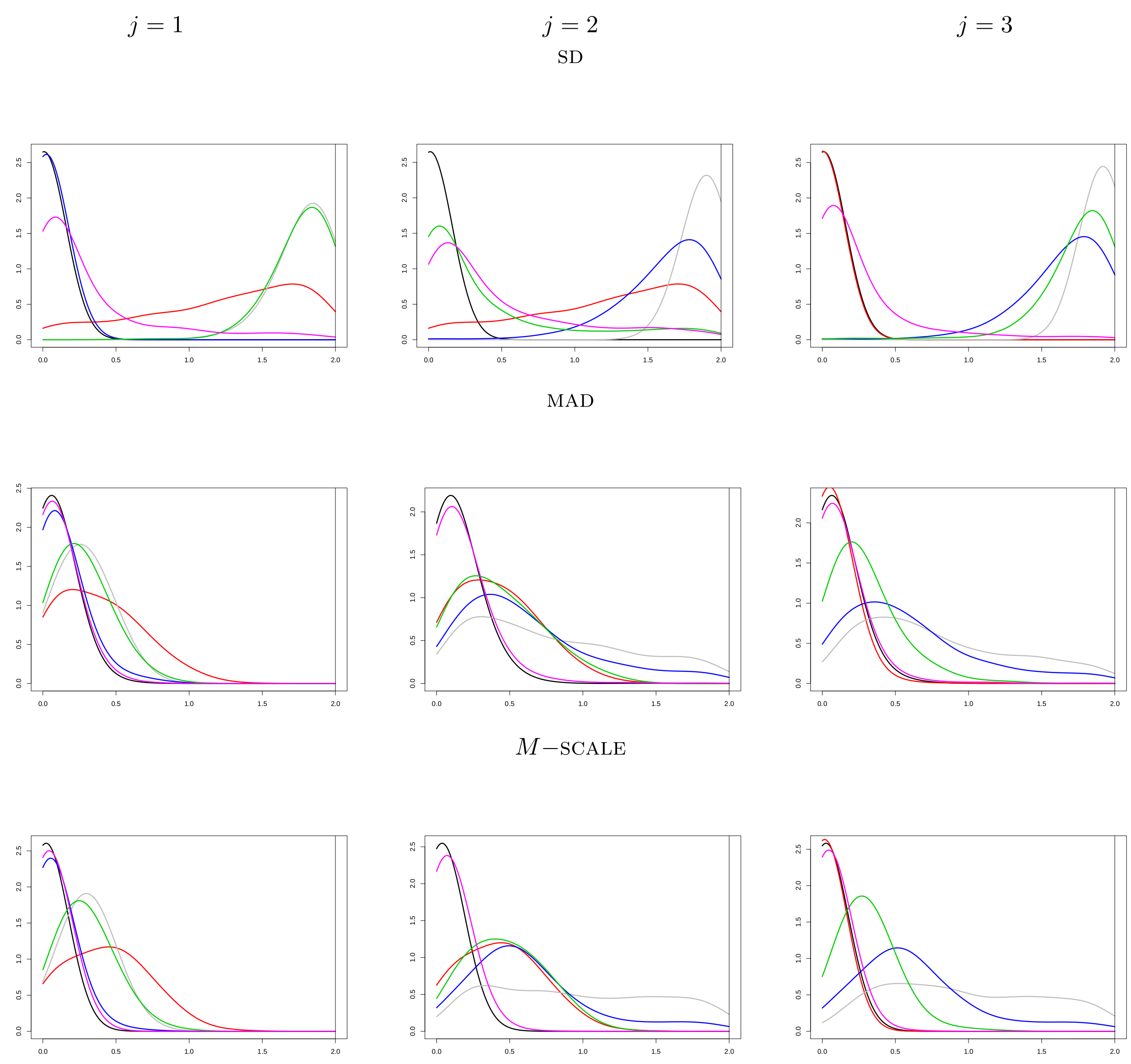

Figure 1: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 when using the raw estimators. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

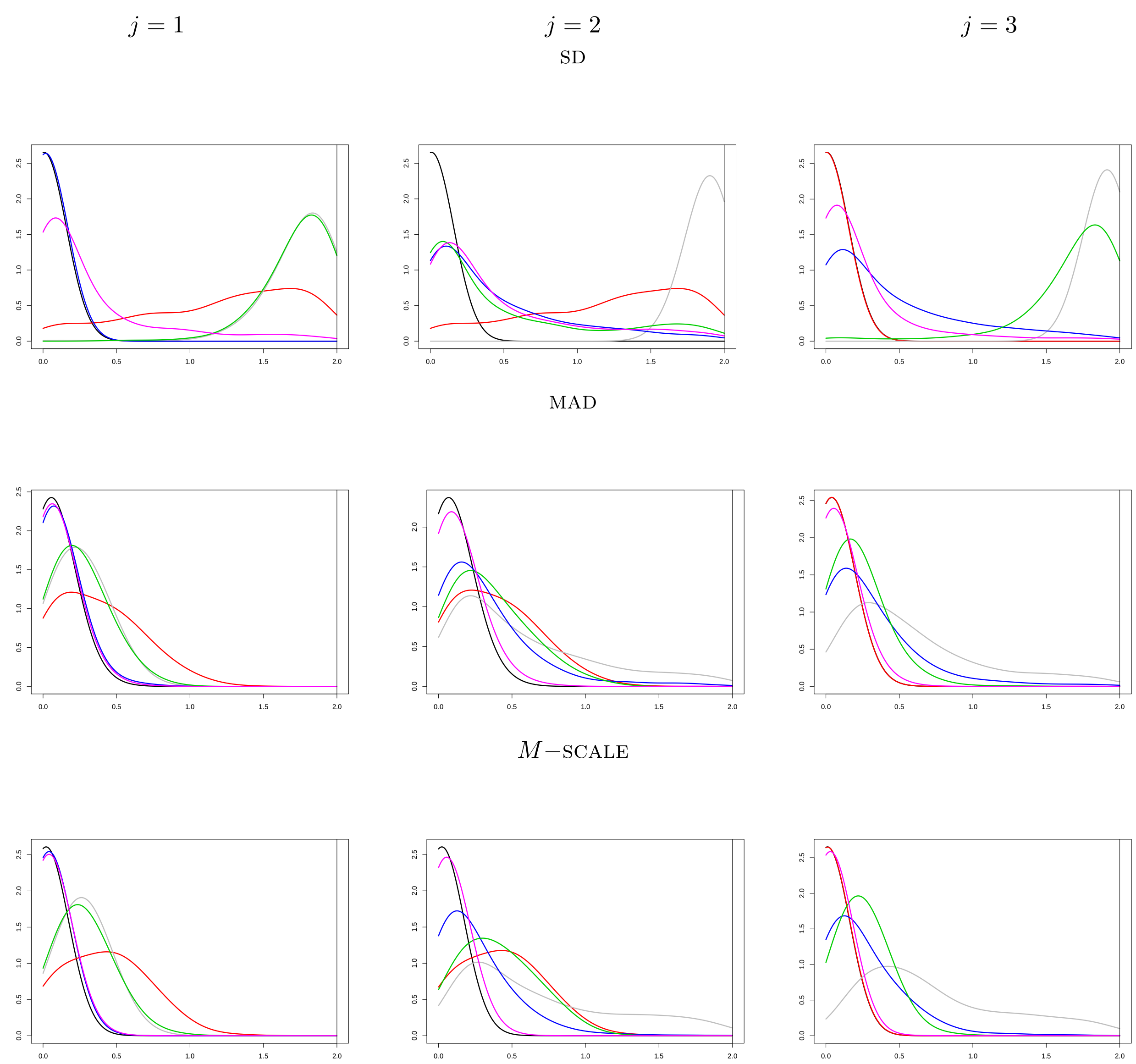

Figure 2: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 when using the scale–penalized estimators, $\widehat{\phi}_{\mathrm{PS},j}$, with a penalization of $\lambda = 1.50n^{-3}$. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

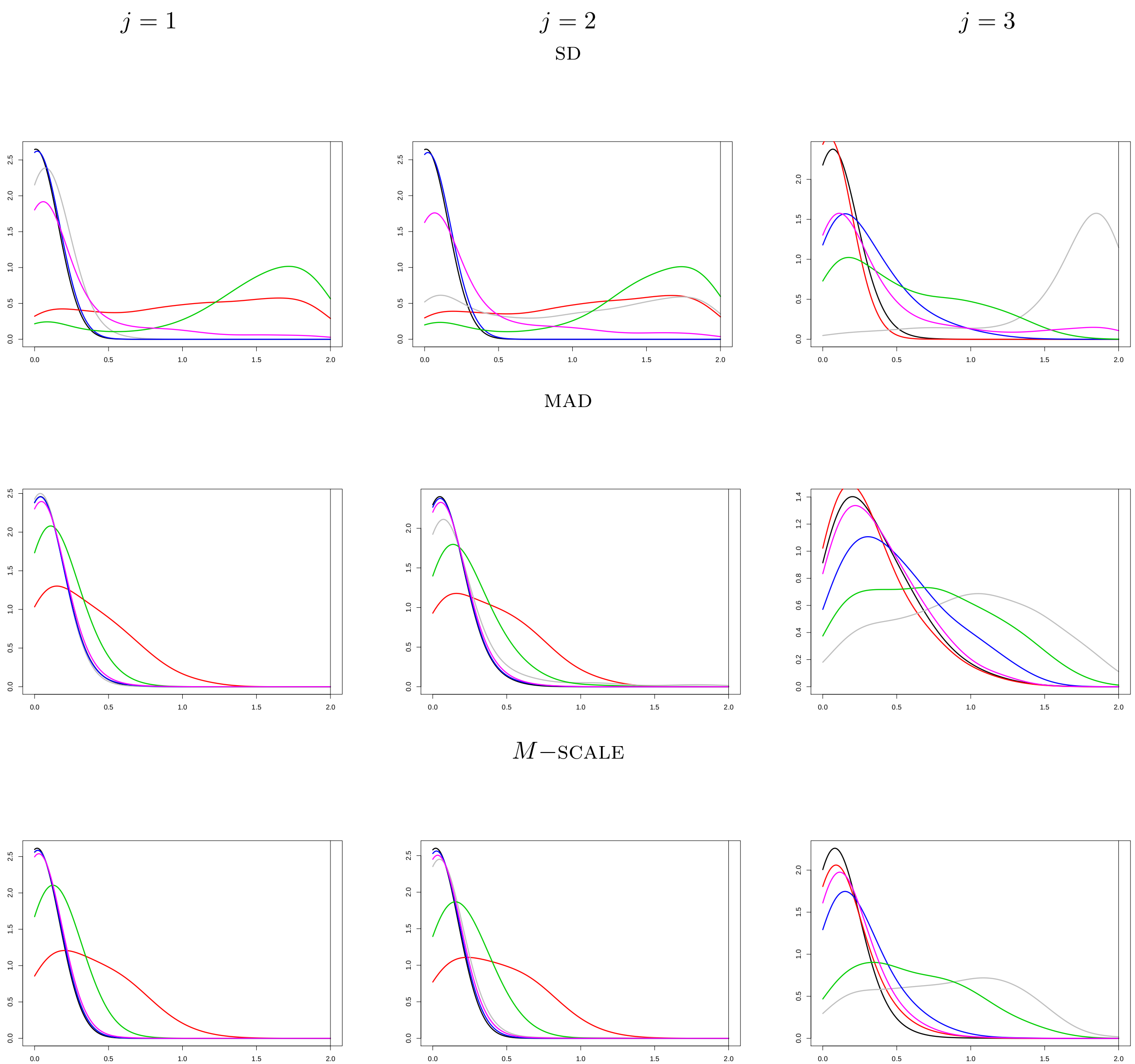

Figure 3: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 when using the norm–penalized estimators, $\widehat{\phi}_{\mathrm{PN},j}$, with a penalization of $\lambda = 0.75n^{-3}$. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

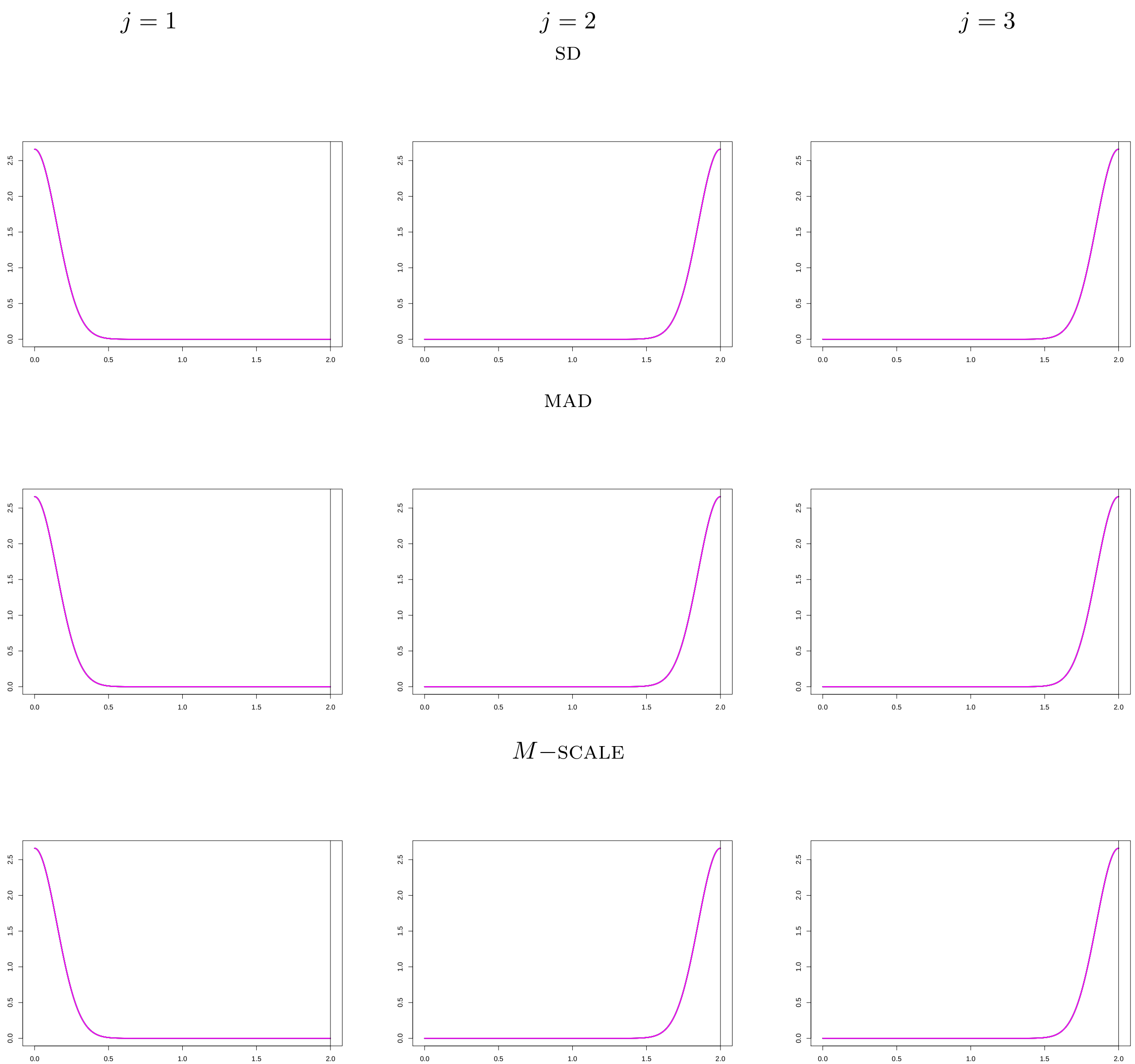

Figure 4: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 for the estimators $\widehat{\phi}_{\mathrm{SI},j}$ defined in (10), when $p_n = 10$ using the Fourier Basis. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

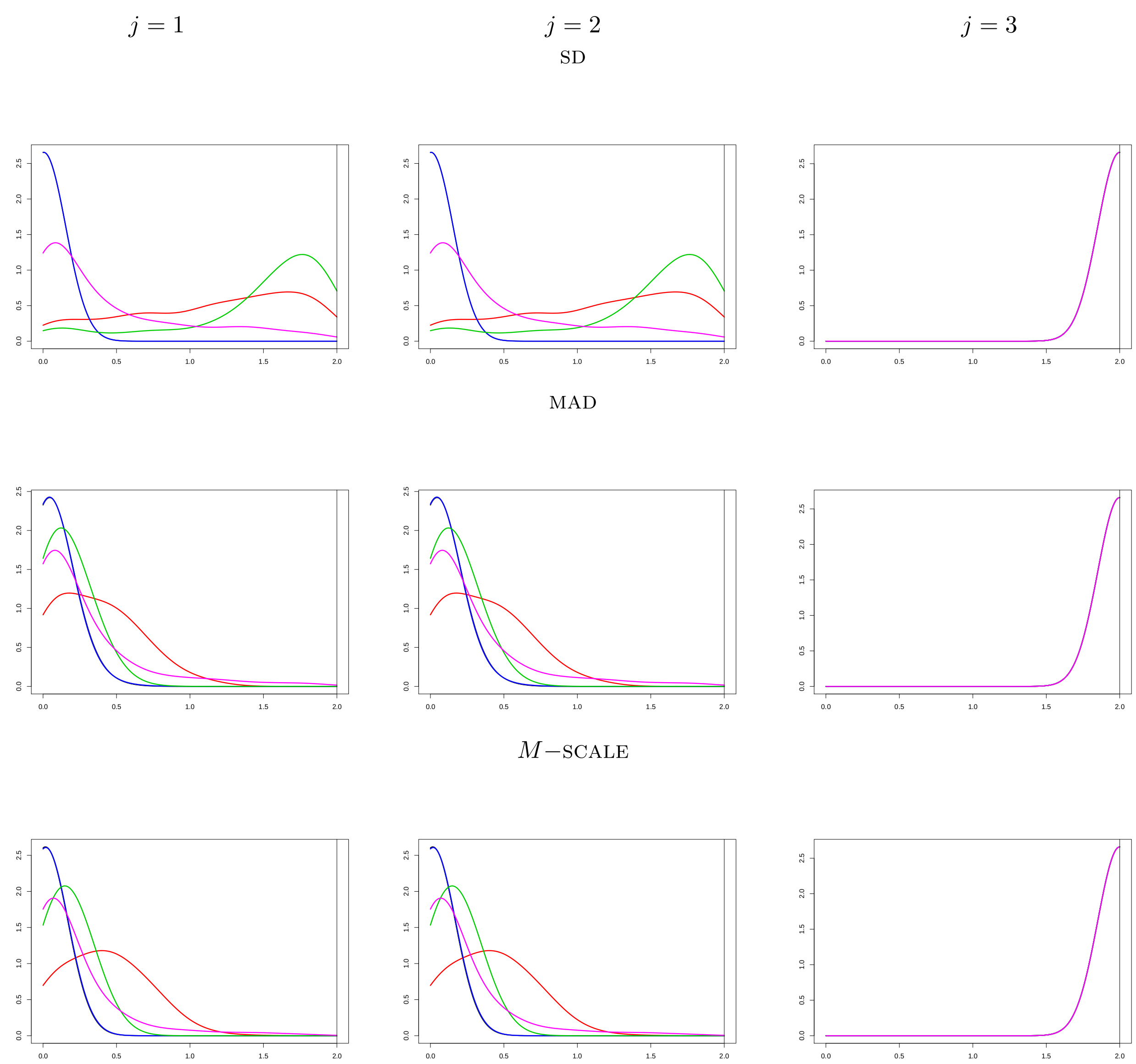

Figure 5: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 for the estimators $\widehat{\phi}_{\mathrm{SI},j}$ defined in (10), when $p_n = 20$ using the Fourier Basis. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

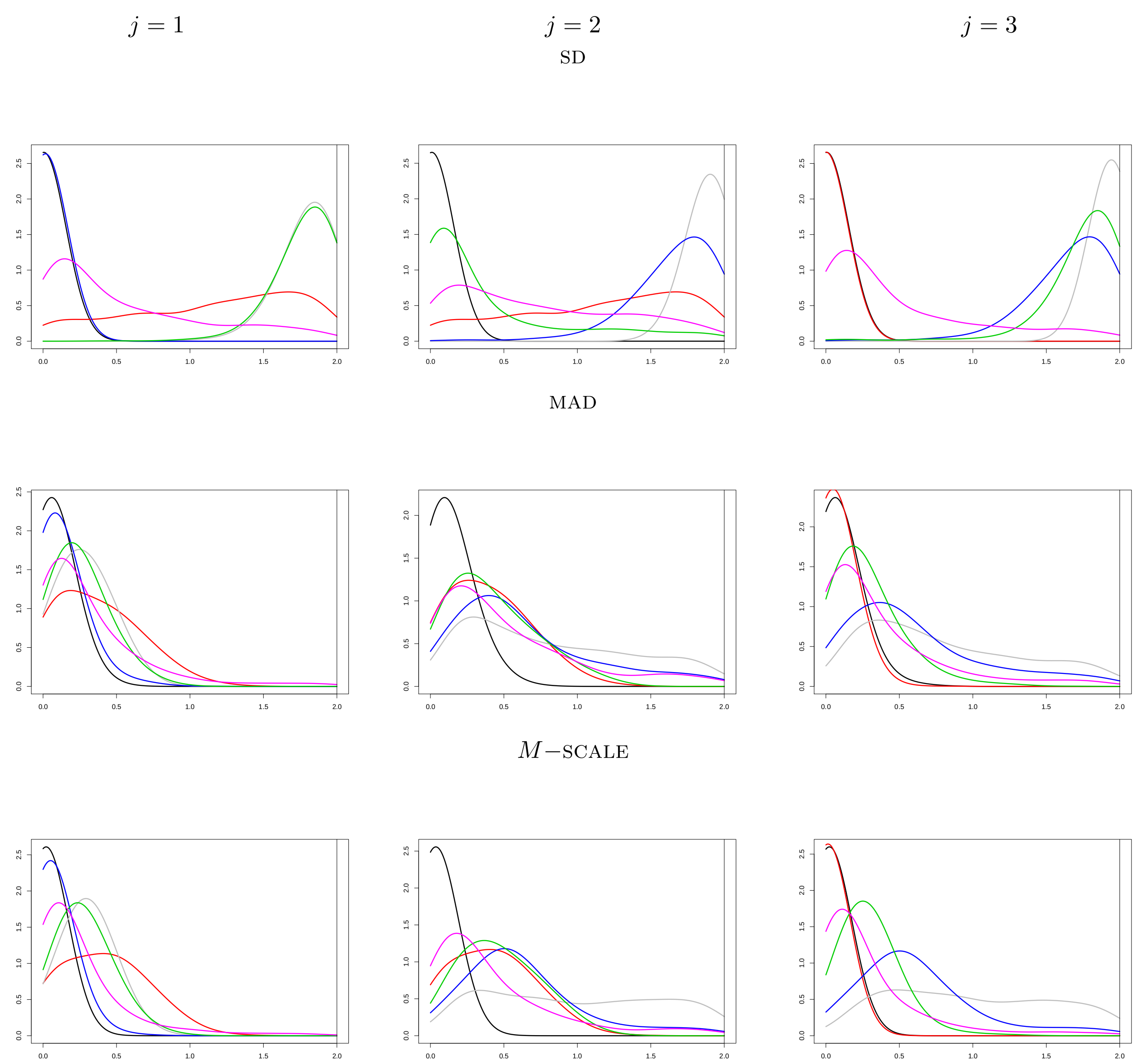

Figure 6: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ with bandwidth 0.6 for the estimators $\widehat{\phi}_{\mathrm{SI},j}$ defined in (10), when $p_n = 30$ using the Fourier Basis. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

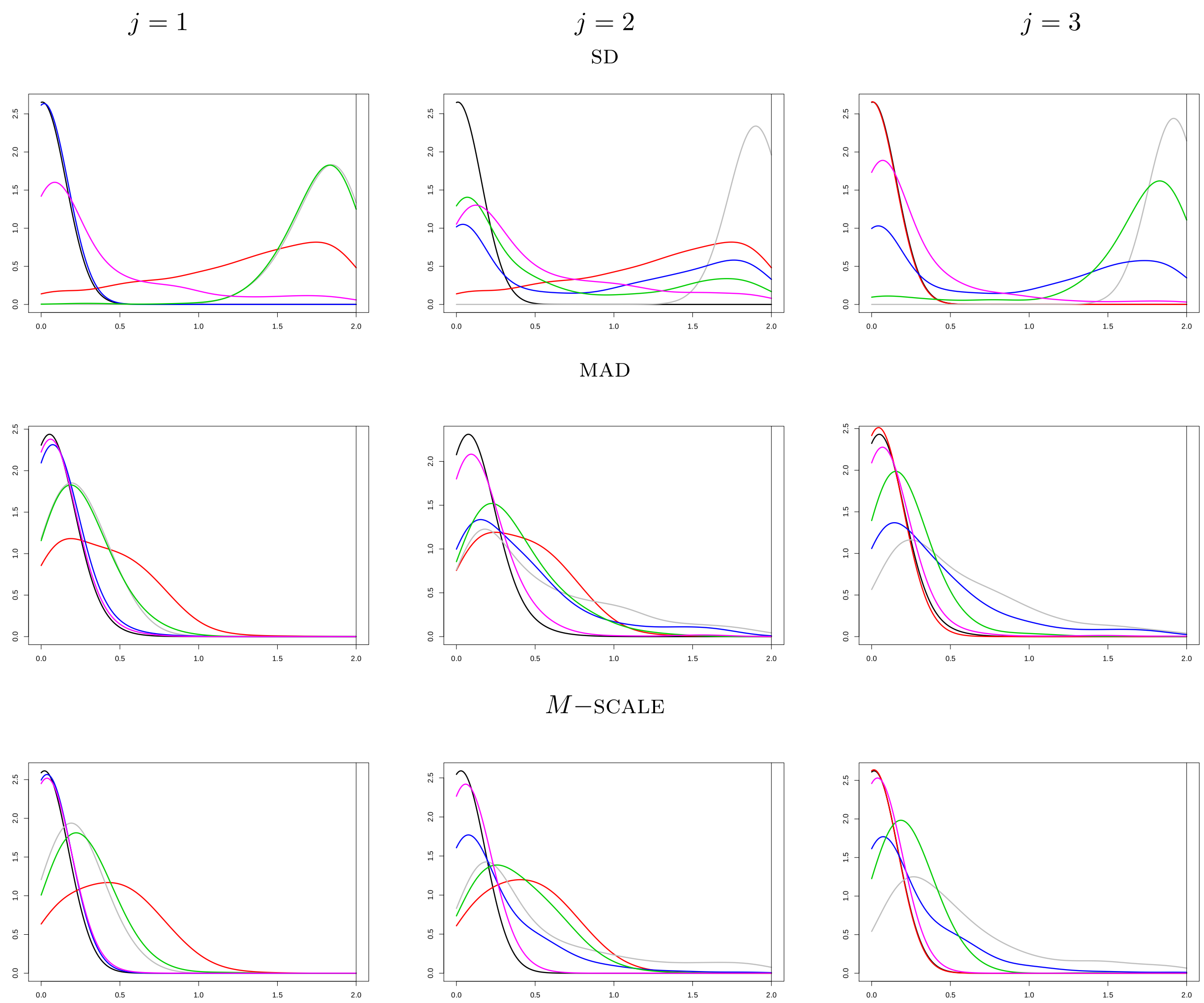

Figure 7: Density estimates of $C(\widehat{\phi}_j, \phi_j)$ when the penalization τ is selected via $K-$fold Cross–validation and the scale–penalized estimators, $\widehat{\phi}_{\text{PS},j}$, are used. The black lines correspond to C_0, while those in red, gray, blue, green and pink correspond to C_2, $C_{3,a}$, $C_{3,b}$, C_{23} and C_c, respectively.

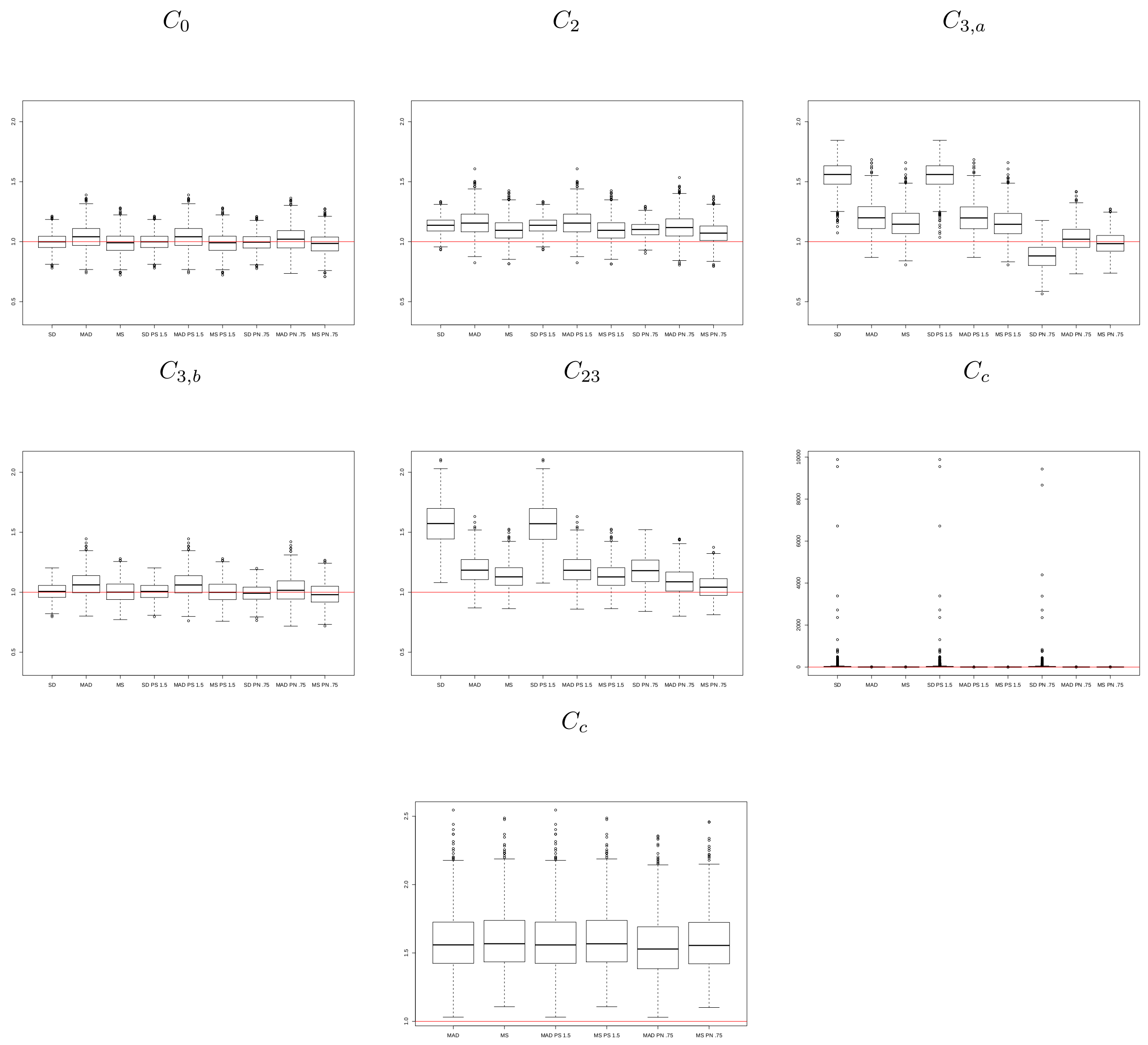

Figure 8: Boxplots of the ratio $\widehat{\lambda}_1/\lambda_1$ for the raw and penalized estimators.

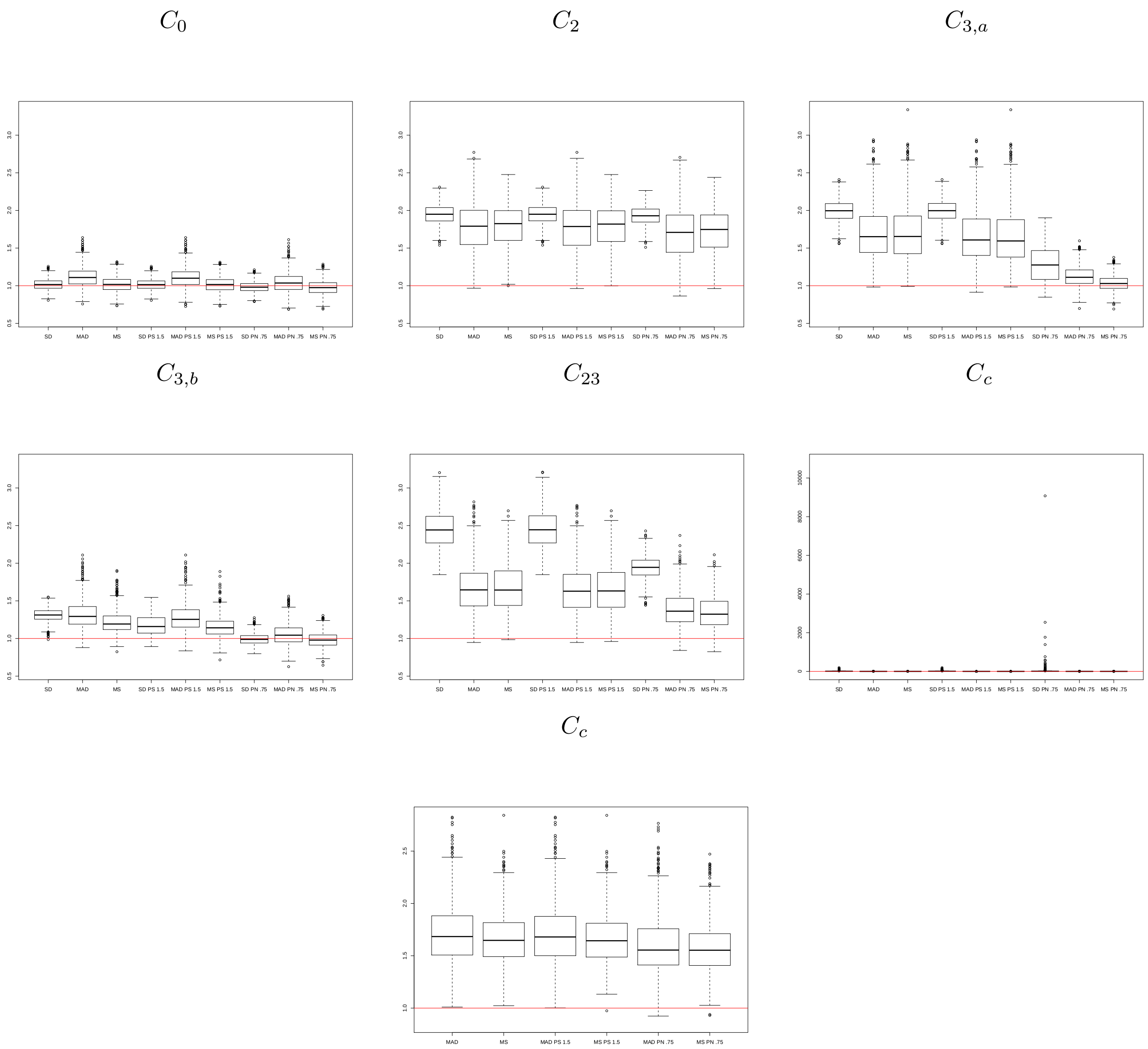

Figure 9: Boxplots of the ratio $\widehat{\lambda}_2/\lambda_2$ for the raw and penalized estimators.

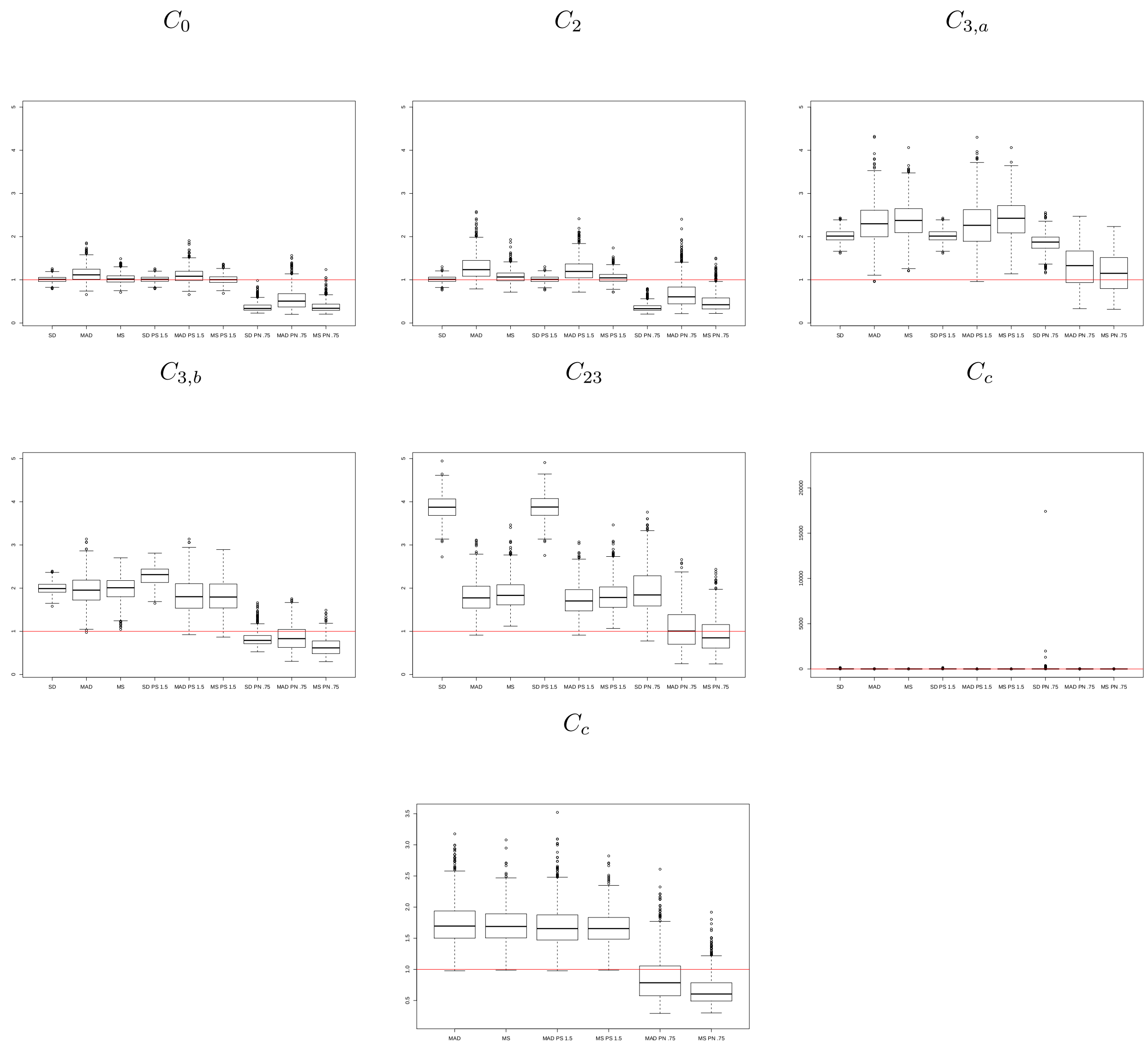

Figure 10: Boxplots of the ratio $\widehat{\lambda}_3/\lambda_3$ for the raw and penalized estimators.

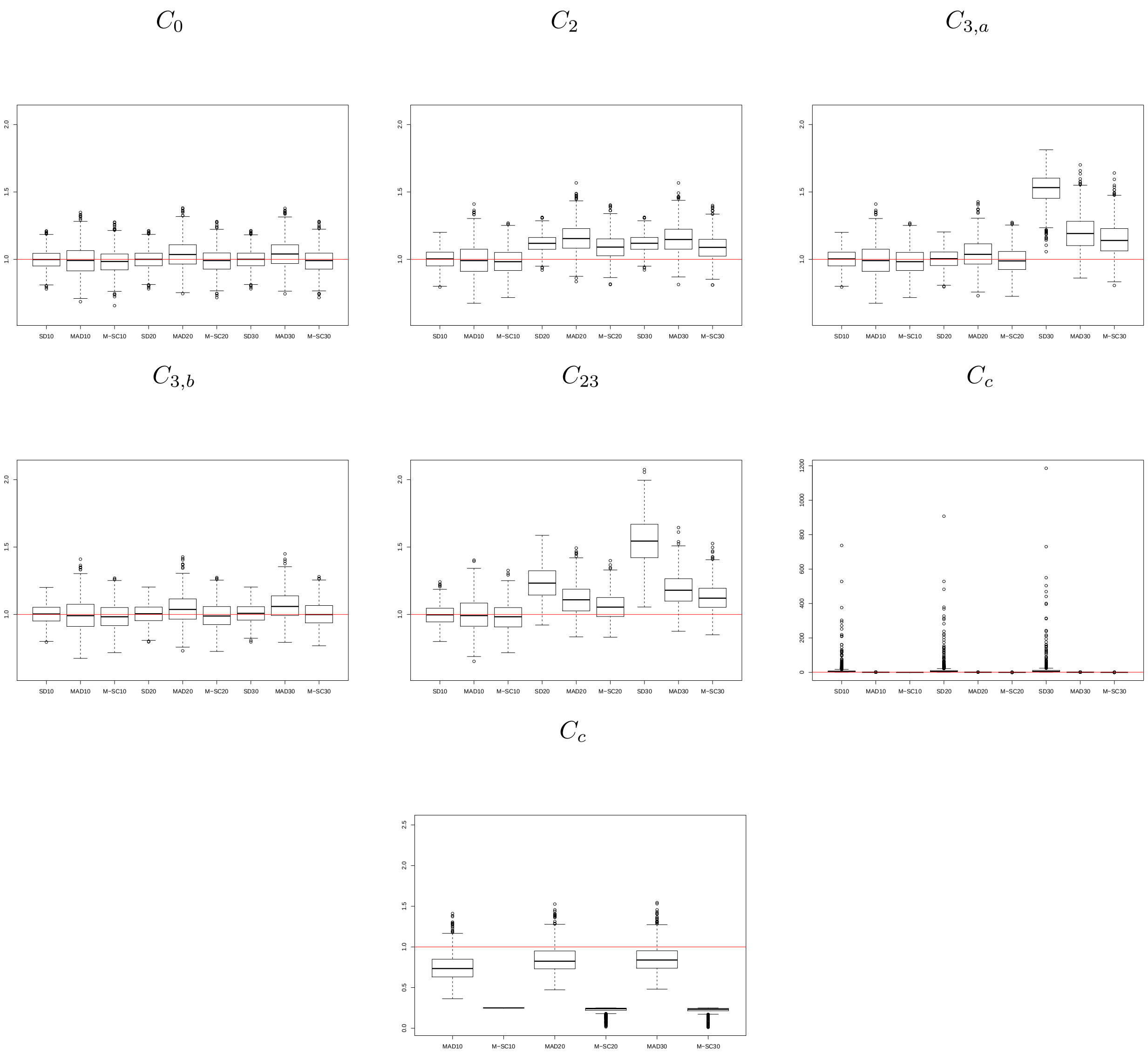

Figure 11: Boxplots of the ratio $\widehat{\lambda}_1/\lambda_1$ for the sieve estimators when using the Fourier basis.

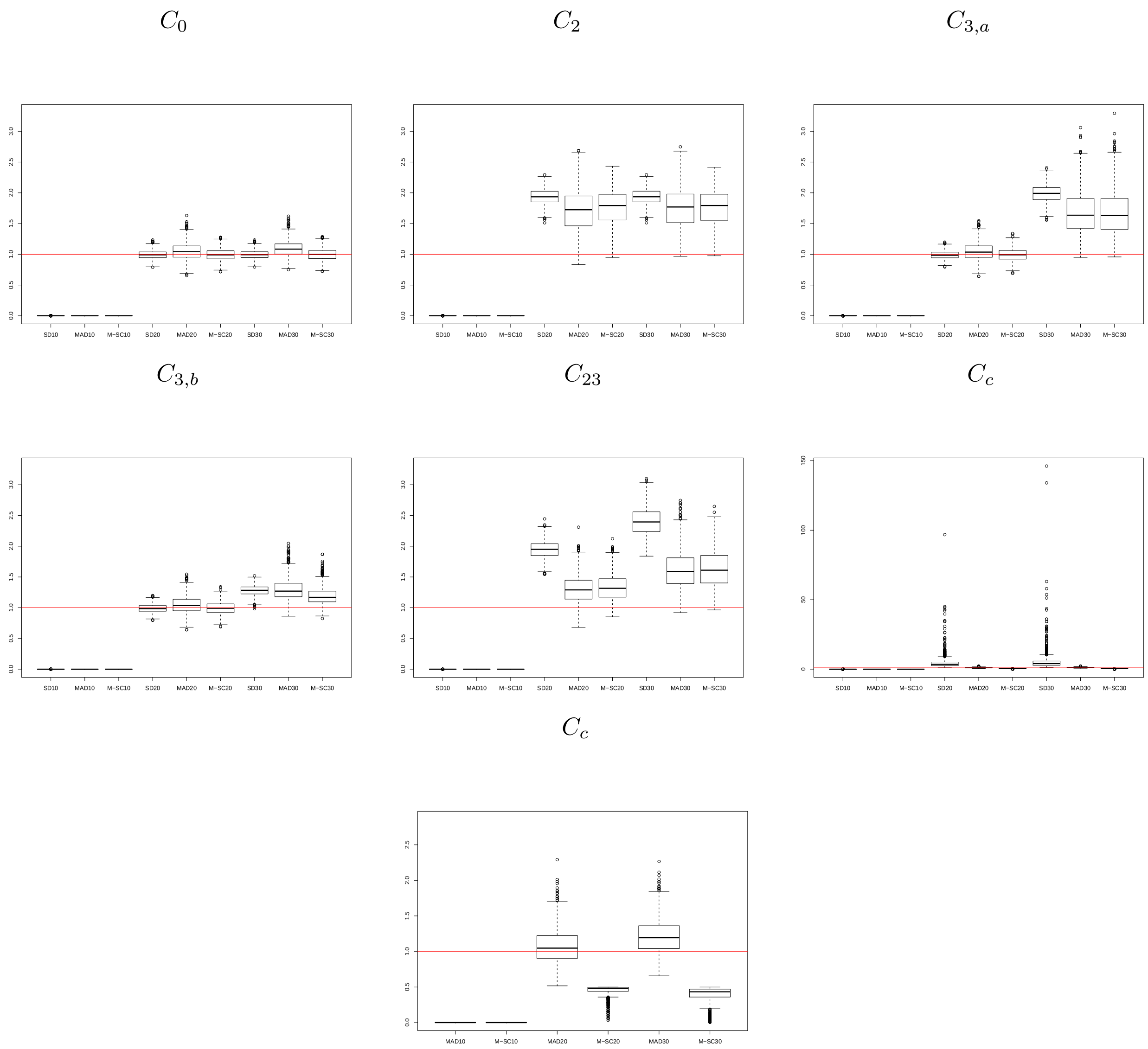

Figure 12: Boxplots of the ratio $\widehat{\lambda}_2/\lambda_2$ for the sieve estimators when using the Fourier basis.

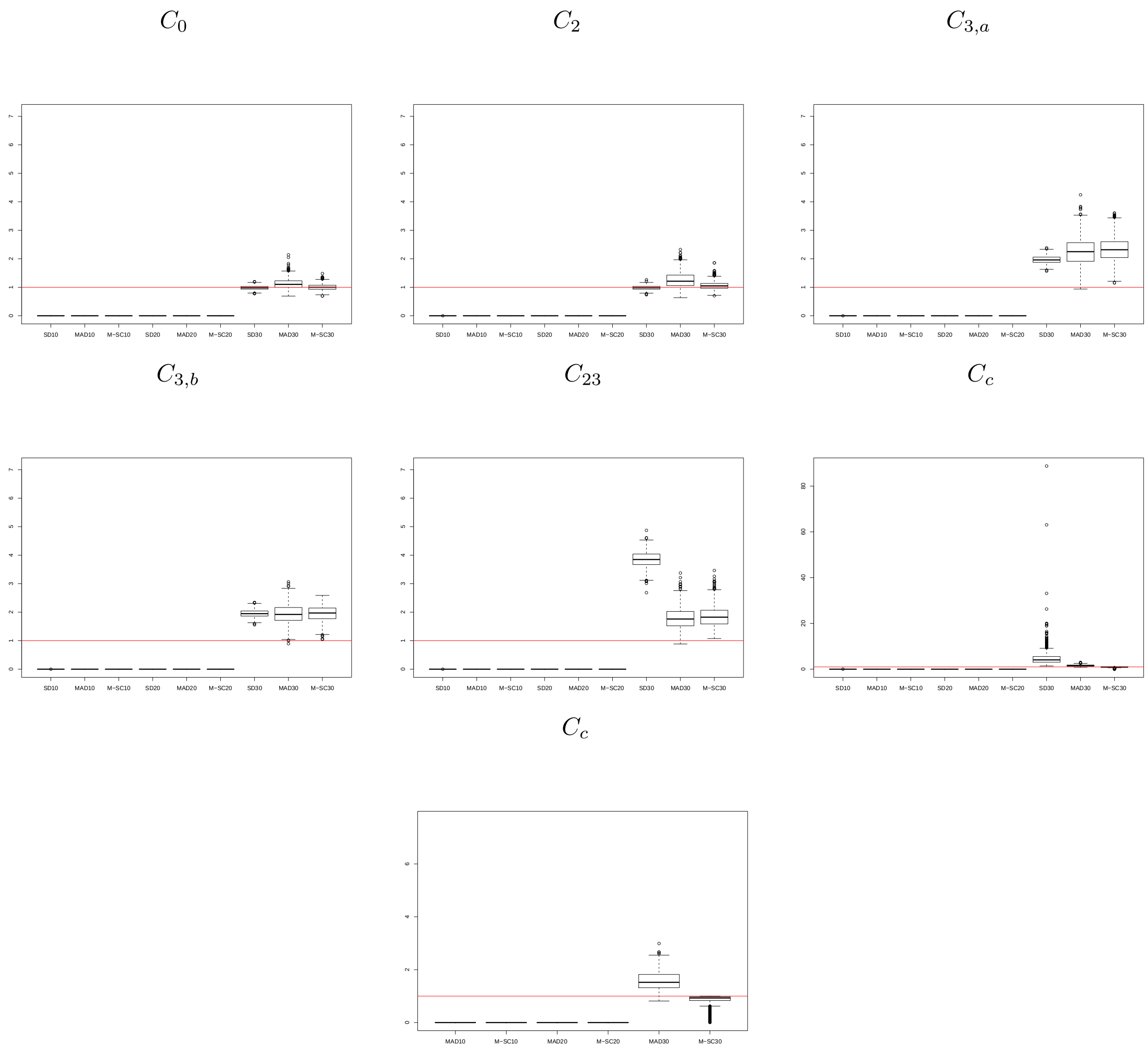

Figure 13: Boxplots of the ratio $\widehat{\lambda}_3/\lambda_3$ for the sieve estimators when using the Fourier basis.