How can we discern whether the covariance operator of a stochastic process is of reduced rank, and if so, what its precise rank is? And how can we do so at a given level of confidence? This question is central to a great deal of methods for functional data, which require low-dimensional representations whether by functional PCA or other methods. The difficulty is that the determination is to be made on the basis of i.i.d. replications of the process observed discretely and with measurement error contamination. This adds a ridge to the empirical covariance, obfuscating the underlying dimension. We build a matrix-completion inspired test statistic that circumvents this issue by measuring the best possible least square fit of the empirical covariance's off-diagonal elements, optimised over covariances of given finite rank. For a fixed grid of sufficiently large size, we determine the statistic's asymptotic null distribution as the number of replications grows. We then use it to construct a bootstrap implementation of a stepwise testing procedure controlling the family-wise error rate corresponding to the collection of hypotheses formalising the question at hand. Under minimal regularity assumptions we prove that the procedure is consistent and that its bootstrap implementation is valid. The procedure circumvents smoothing and associated smoothing parameters, is indifferent to measurement error heteroskedasticity, and does not assume a low-noise regime. An extensive simulation study reveals an excellent practical performance, stably across a wide range of settings, and the procedure is further illustrated by means of two data analyses.

CONTENTS

1 Introduction .......................................................... 2
2 Problem Statement and Background .................................. 4
3 Identifiability .......................................................... 5
4 The Testing Procedure .................................................. 6
5 Asymptotic Theory ...................................................... 8
6 Bootstrap Calibration ................................................... 9
7 Practical Implementation ............................................... 12
  7.1 Hypothesis Boundary, Grid Size, Bootstrap Parameters ...... 12
  7.2 Computation ......................................................... 13
8 Simulation study ........................................................ 13
  8.1 Homoskedastic errors .............................................. 14
  8.2 Heteroskedastic errors ............................................. 18
  8.3 Spiked functional data ............................................. 18
  8.4 Infinite dimensional models ..................................... 21
9 Data Analysis .......................................................... 22
  9.1 Analysis of the Tecator data ..................................... 23
  9.2 Analysis of the fruit-fly data ................................... 25
10 Extension to the Case of Irregular and/or Sparse Measurements . 25
11 Concluding remarks ................................................... 27
References ................................................................. 28

 AMS 2000 subject classifications: Primary 62G05, 62M40; secondary 15A99

 Keywords and phrases: bootstrap, functional data analysis, functional PCA, Karhunen-Loève expansion, matrix completion, measurement error, scree-plot
1. Introduction. Principal component analysis (PCA) plays a fundamental role in statistics due to its ability to focus on a parsimonious data subspace that is most relevant for many practical purposes. In the case of functional data, it assumes an even more prominent role because a reduction in the data dimension maps the statistical problem back to a more familiar multivariate setting. Furthermore, regularization techniques that are necessary for functional regression, testing, prediction, and classification typically hinge on the identification of the most prominent sources of variation in the data. A subtlety in principal component analysis (for any kind of data) is the estimation/selection of the number of components to retain. This choice is often made in exploratory manner, either inspecting the scree plot or retaining enough components to explain (say) 85% of the total variation (see, e.g., Jolliffe (2002)). Various types of confirmatory procedures exist (see, e.g., Horn (1965), Velicer (1976) and Peres-Neto et al. (2005)) often depending on an interpretation of what the “dimension of the data” means. To cite from the introduction of Choi et al. (2017), who propose a formal hypothesis testing approach in a multivariate setting: “Jolliffe (2002) provides an excellent summary of existing approaches to determining the number of principal components, grouping them into three branches: subjective methods (e.g., the scree plot), distribution-based test tools (e.g., Bartlett’s test) and computational procedures (e.g., cross-validation). Each branch has advantages as well as disadvantages, and no single method has emerged as the community standard”. In this paper, we view the problem from the perspective of hypothesis testing. This perspective is particularly fitting when the latent dimension of the signal might be a primary object of inference in itself, or when dimension reduction is an intermediate step in an inferential task that ultimately reports a significance statement. The former is a classical problem in the spectral analysis of random signals (Stoica et al., 2005, Ch. 4) whereas the latter occurs widely in inferential methods for functional data (Horváth and Kokoszka, 2012, Ch. 5-12).

When the data are random functions, one deals with a covariance kernel $k_X$ (rather than a covariance matrix), which a priori is an infinite dimensional object. One would then wish to test $\{H_0 : \text{rank}(k_X) \leq d\}$ versus $\{H_1 : \text{rank}(k_X) > d\}$, for some finite integer $d$ where the rank of $k_X$ is at most $d$ if and only if its Mercer expansion has no more than $d$ terms. This is the global null. Thus, we want to test whether the data gives us enough evidence to conclude that its intrinsic variation is at most of dimension $d$. If this null hypothesis is not rejected based on the observed data, one can also consider a more detailed analysis and test the local hypotheses $\{H_{0,q} : \text{rank}(k_X) = q\}$ versus $\{H_{1,q} : \text{rank}(k_X) > q\}$ for $1 \leq q \leq d$.

At first sight, this problem seems simple: if the number of observations exceeds $d$, then perfect inference will be feasible. However, functional data are most often measured discretely and additively corrupted by unobservable measurement errors that are usually modelled as independent random variables. This additional noise adds a “ridge” to the true covariance: if each curve is measured on a common grid of $L$ nodes, the $L \times L$ covariance matrix of the observed erroneous data will be of full rank. Clearly, this gives rise to a problem of the true rank being confounded by the additive noise. One way of removing the effect of the errors is to use some smoothing procedure on the data (see e.g., Ramsay and Silverman (2005)). But this smoothing step obfuscates the problem since the relationship between the rank of $k_X$ and the rank of the smoothed data is unknown, and further depends on the choice of tuning parameter(s) used for smoothing. At this stage, it would seem that the problem is “almost insolubly difficult” as pointed out by Hall and Vial (2006), who further concluded that “conventional approaches based on formal hypothesis testing will not be effective”. As a workaround, Hall and Vial (2006) considered a “low noise” setting (assuming the noise variance vanishes as the number of observations increases) and used an unconventional rank selection procedure based on the amount of unconfounded noise variance. A weakness of the procedure was that it required the analyst to provide acceptable values of the noise variance for the procedure to
be implemented in practice, and these bounds are to be selected in an ad hoc manner. Bathia et al. (2010) do produce a hypothesis test, but in a temporal context where the functional observations are serially correlated but the measurement errors remain independent – therefore by considering the lag 1 cross-covariance (instead of the lag-0 covariance) denoises the problem, removing the ridge effect. Tests developed in the context of multivariate factor models (e.g. Onatski (2010), Lam and Yao (2012), Luo and Li (2016)) generally do not apply either, as key components of their framework are usually incompatible with functional data.

An alternative approach altogether is to view the problem not as one of testing, but rather as one of model selection. For instance, as part of their PACE method, and assuming Gaussian data, Yao et al. (2005) offer a solution based on a pseudo-AIC criterion applied to a smoothed covariance whose diagonal has been removed. Later work by Li et al. (2013) provides estimates of the effective dimension based on a BIC criterion employing the estimate of the error variance obtained using the PACE approach with the difference being that they used an adaptive penalty term in place of that used in the classical BIC technique. For densely observed functional data, Li et al. (2013) also studied a modification of the AIC technique in Yao et al. (2005) by assuming a Gaussian likelihood for the data. Li et al. (2013) finally considered versions of information theoretic criteria studied earlier by Bai and Ng (2002) in the context of factor models in econometrics, where the latter method is used to choose the number of factors. For all of the procedures studied by Yao et al. (2005) and Li et al. (2013), the main drawback is the involvement of smoothing parameters which enter due to the use of smoothing prior to dimension estimation. The asymptotic consistency of these procedures also requires the true rank to actually be finite, specific decay rates for the smoothing parameters, as well as regularity conditions of the true mean and covariance functions. In early work, not explicitly framed in the context of FDA, Kneip (1994) used several smooth versions of the data constructed using a progression of smoothing parameters, to select a dimension based on a sum of residual estimated eigenvalues. Here too, the method’s performance and asymptotics depend on regularity assumptions and decay rates for smoothing parameters.

In this paper, we steer back to a formal hypothesis testing perspective for the dimension problem. We demonstrate that it is possible to construct a valid test that circumvents the smoothing step entirely, by means of matrix completion. The proposed test statistic measures the best possible least square fit of the empirical covariance’s off-diagonal elements by non-negative matrices of a given finite rank, exploiting the fact that the corruption affects only the diagonal. Compared to smoothing based alternatives, our approach presents the following advantages:

- It provides a genuine testing procedure, inferring the rank with confidence guarantees.
- It does not rely on pre-smoothing and consequently on the choice of smoothing parameters.
- It rests on minimal regularity, in particular continuity of the covariance and sample paths.
- It can handle heteroskedastic measurement errors, which are detrimental to smoothing.
- It does not require a “low noise” regime, indeed the noise variances can be arbitrary.
- It exhibits excellent finite sample performance, stably across a wide range of scenarios.
- It remains valid if selecting $H_0$ from the data, so can be combined with model selection.

The paper is organized as follows. In subsection 2, we discuss the problem statement and setup in detail. We then develop a key identifiability result in subsection 3, which elucidates how the rank can be identified. Exploiting this result, Section 4 describes the testing procedure. The asymptotic distribution of the test statistic, and a valid bootstrap-based calibration approach are introduced in 5 and 6. Practical and computational aspects of its implementation are discussed in Section 7. An extensive simulation study is presented in section 8, where we benchmark the performance of our procedure relative to those studied by Yao et al. (2005) and Li et al. (2013). Two illustrative data analyses are presented in 9. Proofs of formal statements are collected in section 1 of the Supplement, and further technical details are given in sections 2 and 3 of the Supplement.
2. Problem Statement and Background. Let $X = \{X(t) : t \in [0,1]\}$ be the stochastic process in question, assumed zero mean and with continuous covariance kernel on $[0,1]^2$, 

$$k_X(s,t) = \mathbb{E}[X(s)X(t)], \quad (s,t) \in [0,1]^2.$$ 

Continuity of $k_X$ implies that it admits the Mercer expansion,

$$k_X(s,t) = \sum_{m \geq 1} \lambda_m \varphi_m(s)\varphi_m(t) \quad (2.1)$$

with the series on the right hand side converging uniformly and absolutely. Consequently, $X$ is mean square continuous and admits a Karhunen-Loève expansion

$$X(t) = \sum_{m \geq 1} Y_m \varphi_m(t), \quad (2.2)$$

where $\{Y_m\}$ is a sequence of uncorrelated zero-mean random variables with variances $\lambda_m$, respectively. Convergence of the series is in the mean square sense, uniformly in $t$. Given $n$ i.i.d. replications $\{X_1,\ldots,X_n\}$ of $X$, we observe the noise-corrupted discrete measurements

$$W_{ij} = X_i(t_j) + \epsilon_{ij}, \quad i = 1,\ldots,n, \ j = 1,\ldots,L, \quad (2.3)$$

for a grid of $L$ points

$$0 \leq t_1 < t_2 < \ldots < t_L \leq 1.$$

We will assume that the grid nodes are regularly spaced, i.e. $(j-1)/L \leq t_j < j/L$, but this can be considerably relaxed: we show later in Section 10 that the case of irregular and/or sparse measurements can be easily transformed to fit this setting. We assume that the $n \times L$ random variables $\epsilon_{ij}$’s are continuous random variables, independent of the $X_i$’s and themselves independent across both indices, with moments up to second order given by

$$\mathbb{E}[\epsilon_{ij}] = 0 \quad \& \quad \text{var}[\epsilon_{ij}] = \sigma_j^2 < \infty, \quad i = 1,\ldots,n, \ j = 1,\ldots,L.$$ 

Note, in particular that the $\epsilon_{ij}$ are allowed to be heteroskedastic in $j$, i.e. the measurement precision may vary over the grid points. The measured vectors $\{(W_{i1},\ldots,W_{iL})^\top\}_{i=1}^n$ are now i.i.d. random vectors in $\mathbb{R}^L$ with $L \times L$ covariance matrix

$$K_{W,L} = \text{cov}\{(X_1(t_1) + \epsilon_{11},X_1(t_2) + \epsilon_{12},\ldots,X_1(t_L) + \epsilon_{1L})^\top\} = K_{X,L} + D,$$

where:

- $K_{X,L} := \{k_X(t_p,t_q)\}_{p,q=1}^L$ is the $L \times L$ matrix obtained by pointwise evaluation of $k_X(\cdot,\cdot)$ on the pairs $(t_i,t_j)$, and
- $D = \text{diag}\{\sigma_1^2,\sigma_2^2,\ldots,\sigma_L^2\}$ is the $L \times L$ covariance matrix of the $L$-vector $(\epsilon_{i1},\ldots,\epsilon_{iL})^\top$.

In this setup, we wish to use the observations $\{W_{ij} : i \leq n, j \leq L\}$ in order to infer whether the stochastic process $X$ is, in fact, low dimensional, and if so what its dimension might be. We use the term infer in its formal sense, i.e. we wish to be able to make statements in the form of hypothesis tests with a given level of significance. Concretely, the question posed pertains to whether the covariance $k_X$ is of reduced rank, in the sense of a finite Mercer expansion (2.1), and if so of what rank.

Formally, for some dimension $d < \infty$, we wish to test the hypothesis pair

$$\{H_0 : \text{rank}(k_X) \leq d\} \quad \& \quad \{H_1 : \text{rank}(k_X) > d\} \quad (2.4)$$
Notice that we can never actually choose \( d = \infty \), since we have finite data, which is why we have to settle with a \( d < L \wedge n \). Typically \( n \gg L \) so that \( L \wedge n = L \). This global hypothesis pair is related to the sequence of local hypotheses

\[
\begin{align*}
H_{0,q} & : \operatorname{rank}(k_X) = q \\
H_{1,q} & : \operatorname{rank}(k_X) > q,
\end{align*}
\]

In particular, if we can sequentially test all \( d \) local hypotheses with a controlled family-wise error rate, then we will have a test for the global hypothesis, and a means to infer what the rank is, when \( H_0 \) is valid (more details in the next section).

As noted in the introduction, while this question is of clear intrinsic theoretical interest, it also arises very prominently when carrying out a functional PCA as a first step for further analysis, in particular when evaluating a scree plot to choose a truncation level: the choice of a truncation dimension \( q \) can be translated into testing whether the rank of \( k_X \) is equal to \( q \).

The frustrating tradeoff faced by the statistician in the context of this problem is that:

1. Without any smoothing, the noise covariance \( D \) confounds the problem by the addition of a ridge to the empirical covariance, leading to an inflation of the underlying dimensionality. Specifically, the rank of \( K_{W,L} = K_{X,L} + D \) is at most \( n \wedge L \), with probability 1.

2. Attempts to denoise \( K_{W,L} \) and approximate \( K_{X,L} \) by means of smoothing will obfuscate the problem, since the choice of smoothing/tuning parameters will interfere with the problem of rank selection.

It is this tradeoff that Hall and Vial (2006) presumably had in mind when referring to this problem of rank inference as “almost insolubly difficult”. Despite the apparent difficulty, we wish to challenge their statement that “conventional approaches based on formal hypothesis testing will not be effective”, demonstrating that this can be achieved via matrix completion.

The crucial observation is that the corrupted diagonal can be entirely disregarded, while still being able to identify the rank, owing to the continuity of the problem. How precisely is described in the next section.

3. Identifiability. The main idea we wish to put forward here is that it is feasible to make inferences about the rank without resorting to smoothing or low noise assumptions, simply by focussing on the off-diagonal elements of the matrix \( K_{W,L} \) for any sufficiently large but finite grid size \( L \). The point is that we have no information whatsoever on the diagonal matrix \( D \), and cannot attempt to annihilate it by means of smoothing without biasing inference on the rank. Still, we have

\[
K_{X,L}(i,j) = K_{W,L}(i,j), \quad \forall i \neq j
\]

i.e. the matrices are equal off the diagonal, even if their relationship on the diagonal is completely unknown. So the rank may still be identifiable from its off-diagonal entries. The first of our main results shows that this is indeed the case, owing to the continuity of \( k_X \).

**Theorem 1 (Identifiability).** Assume that the kernel \( k_X \) is continuous on \([0,1]^2\), let \( d \geq 1 \), and let \( q \in \{1, \ldots, d\} \). Then, there exists a critical \( L_1 = L_1(d) < \infty \) such that, for all \( L > L_1 \), the functional

\[
\Theta \mapsto \sum_{i \neq j} \left( K_{W,L}(i,j) - \Theta(i,j) \right)^2
\]

restricted on the set \( \mathcal{M}_q = \{ \Theta \in \mathbb{R}^{L \times L} : \operatorname{rank}(\Theta) \leq q \} \) of matrices of rank at most \( q \).
1. Vanishes uniquely at \( K_{X,L} \) when \( \text{rank}(k_X) = q \).
2. Is bounded below by a positive constant when \( \text{rank}(k_X) > q \).

**Remark 1 (Notation).** The sum-of-squares term \( \sum_{i \neq j} (K_{W,L}(i,j) - \Theta(i,j))^2 \) is simply the squared Frobenius distance between \( \Theta \) and \( K_{W,L} \) when disregarding their diagonal entries. We can re-write it more compactly as \( \| P_L \circ (K_{W,L} - \Theta) \|_F^2 \), where \( P_L = \{1 \{i \neq j\}\}_{i,j=1}^L \), \( \|A\|_F = \sqrt{\text{trace}(A^T A)} \) is the Frobenius matrix norm, and ‘\( \circ \)’ denotes the Hadamard (element-wise) product.

**Remark 2 (Critical Grid Size).** The precise critical value \( L_\dagger < \infty \) in Theorem 1 will generally depend on the boundary value \( d \) in the global hypothesis pair (10.3), and the spectrum of \( k_X \). For most scenarios encountered in functional data analysis, the value

\[
L_\dagger = 2d + 1
\]

suffices. This includes polynomial or trigonometric eigensystems and warped versions thereof, systems comprised of splines or other piecewise (non-vanishing) analytic basis elements, and more generally systems with eigenfunctions that are linearly independent over sets of positive Lebesgue measure. Note that it is not the regularity of the eigenfunctions that is elemental here – for instance, the last class described can include eigenfunctions that are nowhere differentiable. See Section 2 of the Supplement for a detailed discussion.

The theorem affirms that sequentially checking whether the rank of \( k_X \) is equal to \( q \) or exceeds \( q \), for \( q \in \{1, \ldots, d\} \), is feasible by means of the off-diagonal entries of \( K_{X,L} \) alone, and indeed for any finite grid \( L > L_\dagger \). That is, the collection of local hypothesis pairs \( \{H_{0,q}, H_{1,q}\}_{q=1}^d \) is identifiable non-asymptotically in the grid size, even when observation is discrete and noisy. Consequently, we will henceforth be working in a framework where \( L \) is assumed fixed but sufficiently large relative to \( d \) (i.e. \( L > L_\dagger \), where \( L_\dagger \) is as in Theorem 1).

Indeed, the identifiability is constructive, in that if we had access to the true matrix \( K_{W,L} \), starting with \( q = 1 \) and proceeding sequentially, we could discern all \( d \) hypothesis pairs as follows:

1. For any candidate rank \( q \leq d \), we check whether

\[
\min_{\Theta: \text{rank}(\Theta) \leq q} \| P_L \circ (K_{W,L} - \Theta) \|_F^2 = 0.
\]

2. If the minimum is positive we are certain that \( \text{rank}(K_{X,L}) > q \).

### 4. The Testing Procedure.

This constructive identifiability can be leveraged to construct a testing procedure. Of course, in practice the matrix \( K_{W,L} \) is unobservable and we must rely on \( \hat{K}_{W,L} \), the empirical covariance of the observed vector \((W_{i1}, \ldots, W_{iL})^\top\),

\[
\hat{K}_{W,L} := \frac{1}{n} \sum_{i=1}^n (W_{i1}, \ldots, W_{iL})(W_{i1}, \ldots, W_{iL})^\top.
\]

This motivates testing the local hypothesis pair \( \{H_{0,q}, H_{1,q}\} \) by means of the test statistic

\[
T_q = \min_{\Theta: \text{rank}(\Theta) \leq q} \| P_L \circ (\hat{K}_{W,L} - \Theta) \|_F^2,
\]

rejecting \( H_{0,q} \) in favour of \( H_{1,q} \) for large values of \( T_q \). Note the interpretation of the test statistic: to test whether the rank is \( q \), we measure the best possible fit of the off-diagonal elements of the empirical covariance \( \hat{K}_{W,L} \) by a matrix of rank \( q \). We reject when this fit
is poor, and the calibration of \( T_q \) is considered in the next two sections, via an asymptotic analysis based on \( M \)-estimation, and hinging on Theorem 1.

For the moment, though, assume that we can obtain a \( p \)-value \( p_q \) for \( T_q \) (or some appropriately re-scaled version, e.g. \( n \times T_q \)) under the hypothesis \( H_{0,q} \). In order to be able to test the global pair \( \{H_0, H_1\} \) \eqref{eq:10.3}, and infer the rank when the global null \( \{H_0 : \text{rank}(k_X) \leq d\} \) is valid, we consider a stepwise procedure, for a given significance level \( \alpha \):

**Step 1:** Test \( H_{0,1} : \text{rank}(K_{X,L}) = 1 \) vs \( H_1 : \text{rank}(K_{X,L}) > 1 \) by means of \( T^{(1)} \).

Stop if the corresponding \( p \)-value, \( p_1 \), exceeds \( \alpha \); otherwise continue to Step 2.

**Step 2:** Test \( H_{0,2} : \text{rank}(K_{X,L}) = 2 \) vs \( H_1 : \text{rank}(K_{X,L}) > 2 \) by means of \( T^{(2)} \).

Stop if the corresponding \( p \)-value, \( p_2 \), exceeds \( \alpha \); otherwise continue similarly.

\[ \vdots \]

We reject the global null \( \{H_0 : \text{rank}(k_X) \leq d\} \) in \eqref{eq:10.3} if and only if the sequential procedure terminates with the rejection of all local hypotheses up to and including the \( d \)-th one. If the procedure terminates earlier, the global null is not rejected, and we subsequently declare the rank of the functional data to be the value

\[ \hat{r} := \min \{q \geq 1 : p_q > \alpha\}, \]

i.e. the smallest \( q \) for which we fail to reject \( H_{0,q} \). This stepwise procedure strongly controls the Family Wise Error Rate (FWER) at level \( \alpha \) (see Maurer et al. (1995) and Lynch et al. (2017)). Indeed, observe that at most one of the hypotheses \( \{H_{0,q}\}_{q=1}^{d} \) can be true, and suppose it corresponds to \( q = q_0 \). Then, if \( V \) denotes the number of false discoveries among the number of rejections, one has

\[ \{V > 0\} \Leftrightarrow \{V = 1\} \Leftrightarrow \{H_{0,q_0} \text{ has been rejected}\} \Leftrightarrow \{p_{q_0} \leq \alpha\}. \]
So, FWER = \( P(V > 0) = P(p_{r_{\text{true}}} > \alpha) \leq \alpha \), where the probabilities are calculated under the given configuration of true and false null hypotheses, equivalently, under the assumption that \( \{H_{0,q} : \text{rank}(K_{X,L}) = q_0\} \) is true (which automatically ensures that the other hypotheses are false). Since \( q_0 \) is arbitrary, the FWER is controlled at level \( \alpha \).

Finally, if \( \text{rank}(k_X) < d \), we have

\[
P(\tilde{r} > r_{\text{true}}) \leq P(r_{\text{true}} \leq \alpha) = P(V > 0) \leq \alpha.
\]

Thus, the control over the FWER translates into a control over the probability of over-estimating the true rank.

To implement the procedure, we will require the \( p \)-values \( \{p_q\} \) corresponding to (an appropriately re-scaled version of) the test statistic \( T_q \) under \( H_{0,q} \). To this aim, the next two sections determine the large-\( n \) sampling distribution of \( n \times T_q \) under \( H_{0,q} \) and describe a valid bootstrap procedure for approximating \( p \)-values \( \{p^*_q\} \) under \( H_{0,q} \) in practice. En route, they also establish the consistency of the resulting test (and bootstrap procedure) as \( n \to \infty \) under \( H_{1,q} \), for all \( L \) sufficiently large.

5. Asymptotic Theory. To justify the use of the test statistic \( T_q \) for testing \( \{H_{0,q} \text{ vs } H_{1,q}\} \) (for some given \( q \leq d \)), we will derive its asymptotic distribution under the null \( H_{0,q} \) and the alternative \( H_{1,q} \) as \( n \to \infty \) for any \( q \leq d \) and \( L > L_\dag \), after appropriate re-scaling (by \( n \), in particular). To this aim, we introduce the functional,

\[
\Psi: \mathbb{R}^{L \times q} \to [0, \infty), \quad \Psi(C) = \|P_L \circ (K_{W,L} - CC^T)\|_F^2.
\]

Furthermore, we collect the following assumptions:

**Assumption (C):** The covariance kernel \( k_X(\cdot, \cdot) \) is continuous on \([0, 1]^2\), the grid nodes \( \{t_1, \ldots, t_L\} \) are regularly spaced, and \( \text{var}[\varepsilon_{ij}] = \sigma^2_j \in [0, \infty) \).

**Assumption (H):** Under \( H_{0,q} \), there exists a factor \( C_0 \in \mathbb{R}^{L \times q} \) of \( K_{X,L} \), i.e. \( K_{X,L} = C_0 C_0^T \), such that the Hessian \( \nabla^2 \Psi(C_0) \) is non-singular.

**Remark 3 (On The Hessian Condition).** A sufficient condition for (H) to hold true is

**Assumption (E):** The \( q \) leading eigenvectors of \( K_{X,L} \) have non-zero entries.

In particular, if (E) is valid, then \( C_0 \) can be taken to be equal to \( V \Lambda^{1/2} \) where \( K_{X,L} = V \Lambda V^T \) is the eigendecomposition of \( K_{X,L} \), and the Hessian \( \nabla^2 \Psi(V \Lambda^{1/2}) \) is provably non-singular. Condition (E), and hence Assumption (H), is automatically satisfied in all the settings listed in Remark 2. See Section 3 of the Supplement for more details.

We can now state our second main result:

**Theorem 2 (Asymptotic Distribution of the Test Statistic).** Suppose that Assumptions (C) and (H) hold and let \( q \leq d \leq L_\dag < \infty \) be as in Theorem 1. Denote the weak (centered Gaussian) limit of \( \sqrt{n}(\tilde{K}_{W,L} - K_{W,L}) \) by the random matrix \( Z \). Then, for any \( L > L_\dag \),

- When \( H_{0,q} \) is valid, we have as \( n \to \infty \)

\[
nT_q \xrightarrow{d} \|P_L \circ Z\|_F^2 - 8(\text{vec}(P_L \circ Z))^\top \{(C_0 \otimes I_L)(\nabla^2 \Psi(C_0))^{-1}(C_0^\top \otimes I_L)\}\text{vec}(P_L \circ Z)
\]

- When \( H_{1,q} \) is valid, \( nT_q \) diverges to infinity as \( n \to \infty \).

The theorem justifies the use of \( nT_q \) as a test statistic: though \( T_q \) will not be precisely zero even when the true rank is \( q \), the test statistic will converge to zero under \( H_{0,q} \), with an asymptotic variance of the order of \( n^{-2} \). The diffuse limiting law of \( nT_q \) under \( H_{0,q} \) in principle allows for calibration (though it does depend on unknown quantities, see the next Section). That \( nT_q \) diverges under \( H_{1,q} \) establishes the consistency of a test based on \( nT_q \).
Remark 4 (On the rate of convergence). Observe that the $n$-rate of convergence in Theorem 2 is exact (non-asymptotic) in the grid size $L$, i.e. valid pointwise for any given $L$. Indeed, one could also renormalise a priori all the curves by $L$, and then the $n$-rate would also be uniform in $L$ (said differently, $n T_q / L^2$ is tight with respect to $n$ uniformly in $L$, under $H_{0,q}$). Such rescaling has no bearing on our methods/results, since these are non-asymptotic in $L$. It would effectively rescale the test statistic $T_q$ and its critical values by a factor of $L^{-2}$ in all instances, but of course all the $p$-values would remain the same, as they are invariant to rescaling the data by a fixed constant.

6. Bootstrap Calibration. Since the limiting null distribution of $n T_q$ established in Theorem 2 depends on unknown quantities, we consider a bootstrap strategy in order to generate approximate $p$-values of $n T_q$ for testing the pair $\{H_{0,q}, H_{1,q}\}$. If $H_{0,q}$ is truly valid, then a naïve bootstrap would suffice. But if $H_{1,q}$ is actually valid instead, a naïve bootstrap will fail to correctly approximate the sought $p$-values under $H_{0,q}$. In effect, we need a re-centering (or rather, re-ranking) scheme in order to generate bootstrap replications “conforming” to $H_{0,q}$, even when $H_{1,q}$ holds true in reality. The purpose of this section is to present such a scheme and establish its validity. The proposed bootstrap scheme is:
(1) Find a minimiser $\hat{\Theta}_q$ of $\|P_L \circ (\hat{K}_{W,L} - \Theta)\|_F$ over nonnegative definite matrices $\Theta$ satisfying $\text{rank}(\Theta) \leq q$.

(2) For each $1 \leq i \leq n$, define

$$\tilde{m}(W_i) = \tilde{W} + \hat{\Theta}_q \hat{K}_{W,L}^{-1}(W_i - \tilde{W}),$$

where $\tilde{W} = n^{-1} \sum_{i=1}^n W_i$. Under the null hypothesis $\{H_{0,q} : \text{rank}(K_{X,L}) = q\}$, this is an estimator of the best linear predictor of the discretely sampled curve $X$, given the noise-corrupted version $W_i$, i.e., $m(W_i) = \tilde{W} + K_{X,L}^{(q)} \hat{K}_{W,L}^{-1}(W_i - \tilde{W})$.

(3) Estimate $D$ by $\hat{D}$, defined as the diagonal matrix with $j$th diagonal element defined as

$$\hat{D}(j,j) = \max\{\hat{K}_{W,L}(j,j) - \hat{\Theta}_M(j,j), 0\},$$

where $\hat{\Theta}_M$ is a minimiser of $\|P_L \circ (\hat{K}_{W,L} - \Theta)\|_F$ over nonnegative definite matrices $\Theta$ satisfying $\text{rank}(\Theta) \leq M$, and

$$M = m_n 1\{m_n < d\} + d 1\{m_n \geq d\}$$

with

$$m_n = \min \left\{ m : T_m \leq \epsilon \frac{\log n}{n} \right\},$$

and $0 < \epsilon \leq 1$ an arbitrary constant.

(4a) Draw $n$ bootstrap observations $U_1^*, U_2^*, \ldots, U_n^*$ from $\{\tilde{m}(W_i) : 1 \leq i \leq n\}$.

(4b) Draw $n$ i.i.d. observations $V_1^*, V_2^*, \ldots, V_n^*$ from an $L$-dimensional centered Gaussian distribution with covariance matrix $\hat{D} + \hat{A}$, where $\hat{A} \equiv \hat{\Theta}_q - \hat{\Theta}_q \hat{K}_{W,L}^{(q)} \hat{\Theta}_q$.

(5) Define the $L$-vectors $\zeta_j = U_j^* + V_j^*$ for $j = 1, 2, \ldots, n$.

(6) Let $F^*_q$ be the law of

$$T^*_q = \min_{\Theta : \text{rank}(\Theta) \leq q} \left\| P_L \circ \left( \frac{1}{n} \sum_{j=1}^n \zeta_j \zeta_j^T - \Theta \right) \right\|_F^2.$$

(7) To test the pair $\{H_{0,q}, H_{1,q}\}$ use the bootstrap $p$-value $p^*_q = F^*_q(T_q)$.

Of course, in practice we use $B < \infty$ random samples $\{\zeta_{1,b}, \ldots, \zeta_{n,b}\}_{b=1}^B$ to approximate the $p$-value $p^*$ in Step (7) by

$$\hat{p}^*_{q,B} = \frac{1}{B} \sum_{b=1}^B 1\{T^*_{q,b} \leq T_q\} = F^*_{q,B}(T_q)$$

where

$$T^*_{q,b} = \min_{\Theta : \text{rank}(\Theta) \leq q} \left\| P_L \circ \left( \frac{1}{n} \sum_{j=1}^n \zeta_{j,b} \zeta_{j,b}^T - \Theta \right) \right\|_F^2.$$

If one is willing to assume that the measurement errors are homoskedastic, one can replace $\hat{D}$ in Step 3 by its diagonally averaged version,

$$\hat{D} = \text{diag}\left\{ L^{-1} \sum_{j=1}^L \hat{D}(j,j), \ldots, L^{-1} \sum_{j=1}^L \hat{D}(j,j) \right\}.$$
The next remark explains the heuristic behind the bootstrap procedure, and the theorem succeeding it establishes the bootstrap procedure’s validity. The procedure’s finite sample performance is investigated thoroughly in the next Section.

Remark 5 (Bootstrap Heuristic). Assume that the errors \( \varepsilon_{ij} \) in (10.2) are Gaussian. Let \( X_i^{(q)}(u) = \sum_{m=1}^{q}(X_i, \varphi_m)_L \varphi_m(u) \) be the q-truncated Karhunen-Loève expansion of the curve \( X_i \) and \( X_i^{(q)} = \{X_i^{(q)}(t_j)\}_{j=1}^{L} \) its discrete version when evaluated at the \( \{t_j\}_{j=1}^{L} \). If we had access to the \( L \)-vectors \( \{X_i^{(q)}\}_{i=1}^{n} \) and \( \{\varepsilon_i\}_{i=1}^{n} \), then we would generate a bootstrap sample conforming to \( H_{0,q} \) by means of constructing \( n \) random \( L \)-vectors

\[
W_i^{(q)} = X_i^{(q)} + \delta_i, \quad \delta_i \text{ sampled randomly with replacement from } \{\varepsilon_1, \ldots, \varepsilon_n\}.
\]

These bootstrapped vectors would have covariance matrix \( K_{X,L}^{(q)} + D \), where

\[
K_{X,L}^{(q)}(i,j) = \sum_{m=1}^{q} \lambda_m \varphi_m(t_i) \varphi_m(t_j).
\]

If instead of observing \( \{X_i^{(q)}\}_{i=1}^{n} \) and \( \{\varepsilon_i\}_{i=1}^{n} \), we only had access to their covariance \( K_{X,L}^{(q)} \) and \( D \), then we would do the “next best thing”, i.e. replace \( X_i^{(q)} \) by its best linear predictor given the actual observations,

\[
m(W_i) = W + K_{X,L}^{(q)} K_{W,L}^{-1}(W_i - W)
\]

and replace \( \delta_i \) by

\[
V_i \sim N(L, 0 + K_{X,L}^{(q)} - K_{X,L}^{(q)} K_{W,L}^{-1} K_{X,L}^{(q)})
\]

The reason this is the “next best thing” is that the resulting \( m(W_i) + V_i \) has zero mean and covariance matrix

\[
\text{Cov}\{m(W_i)\} + D + K_{X,L}^{(q)} - K_{X,L}^{(q)} K_{W,L}^{-1} K_{X,L}^{(q)} = K_{X,L}^{(q)} + D = \text{Cov}\{W_i^{(q)}\}.
\]

In other words, \( \zeta_i = m(W_i) + V_i \) is a “rank q proxy version of \( X_i + \text{Gaussian measurement error} \)”.

whose first and second moments match those of the ideal (but unobservable) bootstrap samples \( W_i^{(q)} = X_i^{(q)} + \delta_i \) (and thus when \( X \) and \( \varepsilon \) are Gaussian, their laws match, too).

The idea of the bootstrap procedure is to materialise this heuristic, replacing the unknown matrices \( \{K_{W,L}^{(q)}, K_{X,L}^{(q)}, D\} \) by their “hat counterparts” \( \{\hat{K}_{W,L}^{(q)}, \hat{\Theta}_{q}, \hat{D}\} \). In particular, as part of the next theorem, the informal statement that the bootstrap scheme generates samples conforming to \( H_{0,q} \) even when \( H_{1,q} \) is true will be made rigorous, by means of establishing validity of the bootstrap.

Theorem 3 (Bootstrap Validity). Let \( q \leq d \leq L_1 < \infty \) be as in Theorem 1 and assume that (C) and (H) hold true. Let \( p_q^* = 1 - F_q^*(T_q) \) be the bootstrapped p-value as defined in Step (7) of the bootstrap procedure above. Then, for all \( L > L_1 \),

- When \( H_{0,q} \) holds true, one has

\[
\mathbb{P}\{p_q^* \leq u\} \xrightarrow{n \rightarrow \infty} u, \quad \forall u \in [0,1],
\]

provided the underlying processes \( \{X_i\} \) and errors \( \{\varepsilon_{ij}\} \) are Gaussian.
• When \( H_{1,q} \) holds true, one has 
\[
\mathbb{P}\{p^*_q \leq u \text{ eventually as } n \to \infty\} = 1, \quad \forall u \in [0, 1].
\]

Remark 6. Regardless of whether or not the \( \{W_i\} \) are Gaussian, as part of the proof of the theorem we establish that under \( H_{0,q} \) the (random) bootstrap law \( F_q^* \) converges pointwise almost surely to the distribution function of the random variable 
\[
\|P_L \circ Z_t\|_F^2 - 8 \left( \text{vec}(P_L \circ Z_t) \right)^\top \{(C_0 \otimes I_L)(\nabla^2 \Psi(C_0))^{-1}(C_0^\top \otimes I_L)\}\text{vec}(P_L \circ Z_t),
\]
where \( C_0 \) is as in Assumption (H), and the random vector \( Z_t \) is the (centred Gaussian) weak limit of \( \sqrt{n}\left\{ \frac{1}{n} \sum_{j=1}^n \xi_j \xi_j^\top - (\tilde{\Theta}_q + \tilde{D}) \right\} \) under \( H_{0,q} \). When the \( \{W_i\} \) are Gaussian, the covariance of \( Z_t \) coincides with that of the centred Gaussian \( Z \) encountered in Theorem 2, and so the bootstrap distribution asymptotically coincides with the limiting law of \( nT_q \) under \( H_{0,q} \) as given by Theorem 2.

When the \( \{W_i\} \) are not Gaussian, it is not guaranteed the centred Gaussians \( Z_t \) and \( Z \) will share the same covariance. Hence the large \( n \) limit of \( p^*_q = 1-F_q^*(T_q) \) (given by \( 1-G(T_q) \)) may not behave as a uniform random variable under \( H_{0,q} \), leading to a significance level different than the nominal one. We investigate the potential effect of non-Gaussianity on calibration of the bootstrap in our simulation study (Section 8), and find that this effect is negligible (in fact undetectable). We expect that Gaussianity can be weakened to higher-order moment conditions, at the expense of an even lengthier proof.

7. Practical Implementation. We now discuss practical aspects related to the implementation of our procedure.

7.1. Hypothesis Boundary, Grid Size, Bootstrap Parameters. Recall that the global hypothesis pair (10.3) to be tested is given by \( \{H_0: \text{rank}(k_X) \leq d\} \) versus \( \{H_1: \text{rank}(k_X) > d\} \) for some prescribed \( d < \infty \). Notice, furthermore, that the bottom-up nature of our iterative testing procedure (Section 4) translates to the FWER remaining invariant to the choice of boundary value \( d \) in the global hypothesis pair (10.3). This means that as far as FWER control is concerned, we may choose \( d \) as we wish. Indeed, we are even free to “data snoop” when choosing \( d \) to set up the global hypothesis pair, i.e. formulate our hypothesis boundary by looking at the data.

The only constraint on the choice of \( d \) is the need to ensure that the grid size \( L \) is sufficiently large relative to \( d \) for our identifiability result (Theorem 1) to hold true. As per Remark 2, it suffices to have grid size \( L \geq 2d + 1 \) for virtually any type of covariance operator encountered in FDA practice, so it is prudent to always respect the constraint \( d \leq \lfloor (L - 1)/2 \rfloor \). Of course, one can always choose \( d \) to be smaller if an inspection of the data suggests so: for instance we can set \( d \) to be a value near an elbow of the off-diagonal scree plot\(^1\)
\[
\quad j \mapsto T^{(j)} - T^{(j-1)}, \quad j = 1, \ldots, L,
\]
provided this choice not exceed \( \lfloor (L - 1)/2 \rfloor \).

The value \( M \) in Step (3) of the bootstrap procedure can similarly be chosen by inspection of the off-diagonal scree plot, as its formal definition suggests: it should represent an elbow of the graph, but can be taken no larger than our choice of \( d \).

These observations motivate the following practical recommendations:

\(^1\)use of the off-diagonal rather than the classical scree plot is recommended, since the former is immune to the presence of measurement errors when \( n \) is large
The boundary $d$ should be no larger than $\lfloor (L - 1)/2 \rfloor$.

In particular, $d$ can be chosen empirically, for instance as a value near an elbow of the off-diagonal scree-plot $j \mapsto T^{(j)} - T^{(j-1)}$.

If the empirical choice is equivocal or exceeds $\lfloor (L - 1)/2 \rfloor$, we simply recommend fixing $d = \lfloor (L - 1)/2 \rfloor$.

Either way, we recommend setting $M$ in Step (3) of the bootstrap procedure as the minimum of $d$ or a value slightly above an elbow of the off-diagonal scree-plot.

In our simulations, we set $d = \lfloor (L - 1)/2 \rfloor$ throughout for reasons of automation. As for $M$, we inspected the off-diagonal scree plots from a sample simulation run in each scenario, and fixed the value of $M$ as a value distinctly above an apparent elbow in that runs’ plot, to accommodate potential variation in other realisations of the plot (unless this exceeded $d$, in which case we took $M = d$). This yielded excellent results irrespectively of the simulation setting.

7.2. Computation. Recall that evaluation of the test statistic $T^{(j)}$ requires the solution of the optimisation problem

$$\min_{\Theta \in \mathbb{R}^{L \times k}, \text{rank}(\Theta) \leq j} \|P_L \circ (\hat{K}_{W, L} - \Theta)\|^2 = \min_{C \in \mathbb{R}^{j \times j}} \|P_L \circ (\hat{K}_{W, L} - CC^T)\|^2.$$ 

This being a non-convex optimization problem, we cannot ensure that standard techniques like gradient descent will converge to a global minimum (note that there are infinitely many minima when using the parametrisation $CC^T$ due to the fact that if $C_1$ is a minimum, so is $C_1 V$ for any $j \times j$ orthogonal matrix $V$).

However, recent work by Chen and Wainwright (2015) shows that projected gradient descent methods with a suitable starting point have a high probability of returning a “good” local optimum in factorised matrix completion problems. For our simulation study, we used the in-built solver `optim` in the R software with starting point $C_1 = U_j \Sigma_j^{1/2}$, where $U \Sigma U^T$ is the spectral decomposition of $\hat{K}_{W, L}$, $U_j$ is the matrix obtained by retaining the first $j$ columns of $U$, and $\Sigma_j$ is the matrix obtained by keeping the first $j$ rows and columns of $\Sigma$. Although we do not exactly use the approach by Chen and Wainwright (2015), it is seen in the simulations that our chosen method of optimisation converges reasonably quickly and yields stable results.

Although our procedure bootstrapping a statistic whose value is the solution of a non-convex problem, its implementation was feasible within very reasonable computational time in all the simulations that we carried out. The following table shows the typical run times for the proposed method (from the initial hypothesis testing till the point when the p-value $p_n$ exceeds the significance level $\alpha$). It is observed that the computation is fast, across different choices of true rank and grid size. We should also mention that, as expected, these values did not change much when the sample size was increased to $n = 250$, so we do not report them.

8. Simulation study. We will now investigate the finite sample performance of our procedure. Recall that in our notation

$$X(t) = \mu(t) + \sum_{j=1}^{r_{\text{true}}} Y_j \varphi_j(t), \quad t \in [0, 1]$$
where \( \{ \varphi_j, \lambda_j \} \) are the eigenfunction/eigenvalue pairs of \( k_X \) and the principal component scores \( Y_j = \int_0^1 X(u) \varphi_j(u) \, du \) satisfy \( E(Y_j) = 0 \) and \( \text{Var}(Y_j) = \lambda_j \) for all \( 1 \leq j \leq r_{\text{true}} \). We observe \( W_{ij} = X_i(t_j) + \varepsilon_{ij} \) for \( 1 \leq i \leq n \) and \( 1 \leq j \leq L \), where \( 0 < t_1 < t_2 < \ldots < t_L < 1 \) are equispaced grid points. For the purposes of the simulation, the errors \( \{ \varepsilon_{ij} \} \) are taken to be independent and normally distributed, potentially heteroskedastic in the grid index, \( \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_j^2) \) for each \( 1 \leq j \leq L \). We will initially define our simulation scenarios with homoskedastic errors, and in a later section switch to heteroskedastic regimes.

8.1. Homoskedastic errors. In the case of homoskedastic measurement errors, we consider the following models (and we comment on their features as we define them):

**Model A1** \( r_{\text{true}} = 3, \; \mu(t) = 5(t - 0.6)^2, \; (\lambda_1, \lambda_2, \lambda_3) = (0.6, 0.3, 0.1), \; Y_j \sim N(0, \lambda_j), \; \varphi_1(t) = 1, \; \varphi_2(t) = \sqrt{2} \sin(2\pi t), \; \varphi_3(t) = \sqrt{2} \cos(2\pi t), \) and \( \sigma_j^2 = 1 \) for all \( j \).

**Model A2** Same as Model A1 except that we now set \( \varphi_3(t) = \sqrt{2} \cos(4\pi t) \), and \( Y_j \) now has a mixture distribution that is \( N(2\sqrt{\lambda_j}/3, \lambda_j/3) \) with probability \( 1/3 \) and \( N(-\sqrt{\lambda_j}/3, \lambda_j/3) \) with probability \( 2/3 \). Thus, the \( X \)-paths are somewhat “curvier” and the principal component scores follow skewed Gaussian mixture models. The latter is chosen to investigate the behaviour of the bootstrap procedure for non-Gaussian processes (see Remark 6).

**Model A3** \( r_{\text{true}} = 3, \; \mu(t) = 12.5(t - 0.5)^2 - 1.25, \; (\lambda_1, \lambda_2, \lambda_3) = (4, 2, 1), \; Y_j \sim N(0, \lambda_j), \; \varphi_1(t) = 1, \; \varphi_2(t) = \sqrt{2} \cos(2\pi t), \; \varphi_3(t) = \sqrt{2} \sin(4\pi t), \) and \( \sigma_j^2 = 2 \) for all \( j \).

**Model A4** Same Model A3 but with principal component scores having a skewed Gaussian mixture law as in Model A2.

**Model A5** \( r_{\text{true}} = 6, \; \mu(t) = 0, \; (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (4, 3.5, 3, 2.5, 2, 1.5), \; Y_j \sim N(0, \lambda_j), \; \varphi_1(t) = 1, \; \varphi_{2k}(t) = \sqrt{2} \sin(2k\pi t) \) for \( k = 1, 2, 3 \), \( \varphi_{2k+1}(t) = \sqrt{2} \cos(2k\pi t) \) for \( k = 1, 2, \) and \( \sigma_j^2 = 3 \) for all \( j \).

Models (A1)-(A3) are similar to those considered in Li et al. (2013). To go beyond globally defined eigenfunctions, the next set of models feature piecewise polynomial eigenfunctions.

**Model S1** \( r_{\text{true}} = 6, \; \mu(t) = 5(t - 0.6)^2, \; (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (2, 1.7, 1.4, 1.1, 0.8, 0.5), \; Y_j \sim N(0, \lambda_j), \) the eigenfunctions \( \varphi_k \) are orthonormalised functions obtained from the basis of cubic splines with knots at \( (0.3, 0.5, 0.7) \), and \( \sigma_j^2 = 3 \) for all \( j \).

**Model S2** The model parameters are the same as in Model S1, with the only difference that being that the principal component scores are now distributed according to the skewed Gaussian mixture form in Model A2.

**Model S3** \( r_{\text{true}} = 4, \; \mu(t) = 5(t - 0.6)^2, \; (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (1.4, 1.1, 0.8, 0.5), \; Y_j \sim N(0, \lambda_j), \) the \( \varphi_k \)’s are orthonormalised functions obtained from the basis of quadratic splines with knots at \( (0.2, 0.6) \), and \( \sigma_j^2 = 2 \) for all \( j \).
**Model S4** The model parameters are the same as in Model S3 except that now the principal component scores now have a skewed Gaussian mixture distribution as in Model A2.

**Model S5** \( r_{\text{true}} = 3, \mu(t) = 5(t - 0.6)^2, (\lambda_1, \lambda_2, \lambda_3) = (1.1, 0.8, 0.5) \), the \( \varphi_j \)'s are orthonormalised functions obtained from the basis of linear splines with knots at \((0.2, 0.6)\), and \( \sigma_j^2 = 1 \) for all \( j \). The principal component scores now have the same skewed Gaussian mixture form as in Model A2.

For each of these models, we have considered two combinations of sample size \( n \) and grid size \( L \), namely \((n, L) = (150, 25)\) and \((150, 50)\), to emulate more sparsely/densely observed settings. The parameter \( M \) described in the bootstrap algorithm in the previous section is set to \( M = 10 \) for all the simulations in this and the next sub-section. As discussed in the previous section, we can choose \( M \) by visual inspection of the off-diagonal scree plot \( j \to T^{(j)} - T^{(j-1)} \). When using this approach in a trial runs from each scenario, the plot was suggestive of \( M = 9 \) for models A5, S1 and S2, and \( M = 6 \) for the other models. The fixed value of \( M = 10 \) was thus chosen for use across the simulation scenarios.

To probe the performance of the bootstrap procedure, we set the number of bootstrap samples to \( B = 500 \) and set the significance level to \( \alpha = 0.05 \). For each model, we carried out 100 independent replications to report the empirical distribution of the estimated rank.

We benchmark the performance of our procedure by comparing to well-known techniques for selecting the rank of a functional data, namely: the AIC based criterion \((AIC_{\text{yao}})\) in Sec. 2.5 of Yao et al. (2005); the modified AIC \((AIC_m)\) and modified BIC \((BIC_m)\) criteria proposed in equations (16) and (6), respectively, in Li et al. (2013); the modified information theoretic criteria \(PC_{p1} \) and \( IC_{p1} \) given in equation (20) in Li et al. (2013). On the suggestion of a referee, we also compare to the Ladle estimator considered in Luo and Li (2016) and the estimator developed in Onatski (2010) (hereafter denoted by ED as in that paper), though with the caveat that these are intrinsically multivariate/high-dimensional procedures, which translated into a functional setting would only work for a finite rank that is strictly less than the grid size. The information theoretic criteria are inspired by similar techniques in Bai and Ng (2002) who used them to estimate the number of factors in an approximate factor model. We underline that comparison to these procedures is purely for the purpose of benchmarking, since these are procedures whose purpose is model selection and thus are geared toward inducing parsimony, though some come with theoretical guarantees of consistently selecting the true rank asymptotically, if the rank is truly finite. The results are tabulated in Tables 2–5.

It is observed from Tables 2–5 that the proposed method selects the true rank in at least 90% of the iterations for all of the chosen models, irrespective of whether the true rank is large/small, the observation grid is sparse/dense, the distribution is Gaussian or not, the signal is smooth/rough, and the noise is large/small compared to the signal. In fact, the when \((n, L) = (150, 50)\), the bootstrap procedure chooses the true rank in all the 100 iterations under all of the above simulation models. Moreover, the evidence (as seen from the magnitude of the \( p \)-values) is quite strong. In cases where the detection of the true rank is not perfect, we found that on making the test procedure more conservative (by choosing a smaller \( \alpha \), e.g., \( \alpha = 0.01 \) or 0.001), the rate of correct identification of the rank surged to 100%. It is observed from the results shown in Tables 2–5 that \( AIC_{\text{yao}} \) estimates the true rank accurately if the rank is large (equal to 6 as in Models (A5), (S1) and (S2)). When the rank is small, the performance of \( AIC_{\text{yao}} \) varies depending on the model. Investigating a bit more, it may be observed that it overestimates the rank under Models A1 and A2 (rank = 3), where the error dominates the leading eigenvalue of the signal. On the other hand, for Models (S3) and (S4) (rank = 4), \( AIC_{\text{yao}} \) accurately selects the true rank. When the rank is small (equal to 3 or 4) but the grid is dense \((L = 50)\), it is seen that \( AIC_{\text{yao}} \) grossly over-estimates the true rank in almost all models. This over-estimation is exacerbated when the eigenfunctions are
Table 2
Table showing the true rank (in bold) and the empirical distribution of the estimated rank under Models A1–A5 with homoskedastic errors for $(n, L) = (150, 25)$

| Selected rank | Model A1 | Model A2 | Model A3 |
|---------------|----------|----------|----------|
|               | 1 2 3 4 5 | 1 2 3 4 5 | 1 2 3 4 5 |
| Proposed test | 0 2 97 0 1 | 0 1 98 1 0 | 0 0 100 0 0 |
| $AIC_{yao}$   | 0 0 13 59 26 | 0 0 16 64 20 | 0 0 74 25 1 |
| $AIC_{m}$     | 34 54 12 0 0 | 41 53 6 0 0 | 80 20 0 0 0 |
| $BIC_{m}$     | 0 0 100 0 0 | 0 0 100 0 0 | 0 0 100 0 0 |
| $PC_{p1}$     | 37 33 0 0 0 | 70 30 0 0 0 | 99 1 0 0 0 |
| $IC_{p1}$     | 53 44 3 0 0 | 55 45 0 0 0 | 92 8 0 0 0 |
| Ladle         | 0 0 100 0 0 | 0 0 100 0 0 | 0 0 100 0 0 |

| Model A4 | Model A5 |
|----------|----------|
| Selected rank | 1 2 3 4 5 | 1 2 3 4 5 6 7 8 |
| Proposed test | 0 0 100 0 0 | 0 0 100 0 0 0 97 2 1 |
| $AIC_{yao}$   | 0 0 0 0 100 | 0 0 0 0 0 92 8 0 |
| $AIC_{m}$     | 37 49 14 0 0 | 24 28 40 6 2 0 0 0 |
| $BIC_{m}$     | 0 0 100 0 0 | 0 0 100 0 0 0 100 0 0 |
| $PC_{p1}$     | 78 22 0 0 0 | 40 40 19 1 0 0 0 0 |
| $IC_{p1}$     | 62 36 2 0 0 | 46 37 16 1 0 0 0 0 |
| Ladle         | 0 0 100 0 0 | 0 0 100 0 0 0 0 0 0 |
| ED            | 0 0 100 0 0 | 0 0 100 0 0 0 0 0 0 |

Table 3
Table showing the true rank (in bold) and the empirical distribution of the estimated rank under Models A1–A5 with homoskedastic errors for $(n, L) = (150, 50)$

| Selected rank | Model A1 | Model A2 | Model A3 |
|---------------|----------|----------|----------|
|               | 1 2 3 4 5 | 1 2 3 4 5 | 1 2 3 4 5 |
| Proposed test | 0 2 97 0 1 | 0 1 98 1 0 | 0 0 100 0 0 |
| $AIC_{yao}$   | 0 0 13 59 26 | 0 0 16 64 20 | 0 0 74 25 1 |
| $AIC_{m}$     | 34 54 12 0 0 | 41 53 6 0 0 | 80 20 0 0 0 |
| $BIC_{m}$     | 0 0 100 0 0 | 0 0 100 0 0 | 0 0 100 0 0 |
| $PC_{p1}$     | 37 33 0 0 0 | 70 30 0 0 0 | 99 1 0 0 0 |
| $IC_{p1}$     | 53 44 3 0 0 | 55 45 0 0 0 | 92 8 0 0 0 |
| Ladle         | 0 0 100 0 0 | 0 0 100 0 0 | 0 0 100 0 0 |

trigonometric, which is surprising since one would expect this to be an easier setting than in (S3), (S4) and (S5). The over-estimation of the rank by $AIC_{yao}$ was also observed by Li et al. (2013).

The $AIC_{m}$, $PC_{p1}$ and $IC_{p1}$ criteria do not perform well in general and mostly underestimate the rank irrespective of the sample size and the sparse/dense regime. The $BIC_{m}$ procedure, on the other hand, yields the same perfect estimation results as our procedure when the grid is dense $(L = 50)$. It does so also when the grid is sparse provided that the true rank is small (equal to 3 or 4). However, for Models (S1) and (S2) with $L = 25$, where the rank is large (equal to 6) and the grid is sparse $(L = 25)$, the $BIC_{m}$ criterion mostly selects a smaller rank. This is different from its performance under Model (A5) with $L = 25$ (which is also of rank 6), where it selects the true rank in 67% of iterations. The difference in this
Table 4

| Model S1 | Model S2 |
|----------|----------|
| Selected rank | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | $\geq 8$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | $\geq 8$ |
| Proposed test | | | | | | | | | | | | | | | | |
| $AIC_{ga}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $AIC_m$ | 31 | 31 | 38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $BIC_m$ | 33 | 36 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $PC_{pi}$ | 65 | 31 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ladle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ED | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Prop. test | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |

Table 5

| Model S3 | Model S4 | Model S5 |
|----------|----------|----------|
| Selected rank | | | |
| 1 | 2 | 3 | 4 | 5 | $\geq 6$ | 1 | 2 | 3 | 4 | 5 | $\geq 6$ | 1 | 2 | 3 | 4 | 5 | $\geq 6$ |
| Proposed test | | | | | | | | | | | | | | | | |
| $AIC_{ga}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| $AIC_m$ | 1 | 8 | 56 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| $BIC_m$ | 1 | 10 | 60 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| $PC_{pi}$ | 7 | 14 | 64 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Ladle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| ED | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Prop. test | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |

The behaviour of $BIC_m$ may be attributed to the fact that for Model (A5), the eigenfunctions are smooth, while they are only twice continuously differentiable for Models (S1) and (S2) due to the presence of knots. The Ladle estimator selects the true rank always in all of the above models. The ED method has its best performance for models A1–A5 under both error regimes where it perfectly estimates the rank with the exception of A2, where it always overestimates. This overestimation of the rank is also seen in all of the non-analytic spline models (S1–S5).

Summarising the observations from Tables 2–5, it may be concluded that the $BIC_m$ and the $AIC_{ga}$ criteria are most appropriate among the competing information-based procedures. Some tentative conclusions on the two methods are as follows. While the latter works
well when the rank is large (irrespective of the sparsity/denseness of the grid), the former is suited when the grid is dense (irrespective of the magnitude of the rank). The \( BIC_m \) procedure also works very well when the grid is sparse, provided that the rank is small. However, both procedures appear to be quite sensitive to departures from the above situations — \( AIC_{yao} \) grossly over-estimates, while \( BIC_m \) mildly under-estimates. Note that the difference in performance is observed between \( L = 25 \) and \( L = 50 \). This change in number of observations is not so stark so as to be classified immediately as sparse versus dense, and the fact that the performance of these two procedures vary in such a moderate change of grid size is concerning. We also mention in passing that the performance of the \( AIC_{yao} \) and the \( BIC_m \) procedures crucially depend on the choice of the smoothing parameters. Indeed, Li et al. (2013) considered models similar to Models (A1)-(A5) but worked with an undersmoothing choice of the bandwidth parameter, and the relative performance of the above two procedures differs from that observed in our simulation results.

On the other hand, Table 2–5 shows that our proposed procedure always selects the true rank in at least 90% of the iterations (the percentage being much higher in most cases), irrespective of the magnitude of the rank and the sparsity/denseness of the grid. Thus, the proposed method seems to provide an effective and stable alternative. Beyond this advantage, our method also comes with a probabilistic guarantee on overestimation, and hence provides an automatic quantification of uncertainty about the true rank, while not relying on smoothing.

8.2. Heteroskedastic errors. Our theory suggests that our testing procedure automatically adapts to a heteroskedastic variance structure for the measurement errors. We therefore use the same model scenarios as before, but this time with heteroskedastic errors in order to gauge how this translates into practical performance. All else being the same, the measurement error variances are now given by

\[
\sigma^2_{(p-1)U+k} = U^{-1} \sum_{l=(p-1)U+1}^{pU} k_X(t_l, t_l)/1.5,
\]

where \( U = L/5 \), \( k = 1, 2, \ldots, U \) and \( p = 1, 2, \ldots, 5 \). This specific error structure may be viewed from the perspective of a local averaging of the signal along with a downscaling by a factor of 3/2. For these simulation models, the results obtained are provided in Tables 6 to 9. It is observed that the performance of the proposed procedure remains invariant to the presence of homoskedasticity, as our theory predicts.

8.3. Spiked functional data. One may also consider a spiked covariance model, in analogy to high-dimensional statistics (see, e.g., Johnstone (2001), Paul (2007)) where some of the eigenvalues are considerably larger than the rest (Amini and Wainwright, 2012). One instance of the latter setting is the Tecator data set considered in Section 9. This is a particularly challenging setting: heuristically, a prominent bend is expected to appear in the scree plot, well before the index value of the true rank (see Figure 2 which shows the scree plots for the spiked models considered immediately below). To probe the performance of our method in this setting, we also consider the following spiked scenarios:

Model SF1 Model (A1) is modified to now have \((\lambda_1, \lambda_2, \lambda_3) = (4, 0.2, 0.1)\) and \(\sigma^2_j = 1\) for all \(j\). Here the first eigenvalue explains about 93% of the total variation in the signal. Note that the error variance is five and ten times the size of the penultimate and last eigenvalue, respectively.
Testing for the Rank of a Covariance Operator

Table 6
Table showing the true rank (in bold) and the empirical distribution of the estimated rank under Models A1–A5 with heteroskedastic errors for \((n, L) = (150, 25)\)

| Selected rank | 1 | 2 | 3 | 4 | \(\geq 5\) | 1 | 2 | 3 | 4 | \(\geq 5\) | 1 | 2 | 3 | 4 | \(\geq 5\) |
|---------------|---|---|---|---|----------|---|---|---|---|----------|---|---|---|---|----------|
| Proposed      | 0 | 0 | 95 | 4 | \(\geq 5\) | 0 | 0 | 94 | 5 | 1 | 0 | 0 | 93 | 6 | 1 |
| AIC\(_{pa}\)   | 0 | 0 | 25 | 57 | 18 | 0 | 0 | 39 | 56 | 5 | 0 | 0 | 22 | 62 | 16 |
| AIC\(_{m}\)    | 21 | 45 | 34 | 0 | 0 | 27 | 51 | 22 | 0 | 0 | 93 | 7 | 0 | 0 | 0 |
| BIC\(_{m}\)    | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| \(PC_{p1}\)   | 60 | 38 | 2 | 0 | 0 | 63 | 37 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| \(IC_{p1}\)   | 33 | 47 | 20 | 0 | 0 | 38 | 52 | 10 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| Ladle         | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| ED            | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 100 | 0 | 0 |

Model SF2
Model (A5) is modified to have \((\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (5, 4, 0.2, 0.2, 0.1, 0.1)\) and \(\sigma^2_j = 1\). Here the top two eigenvalues explain 93.75% of the total variation in the signal, and there are three more trailing eigenvalues are of order between 1/5 and 1/10 the size of the noise variance.

Model SF3
Model (A5) is modified to have \((\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (4, 3.5, 3, 0.3, 0.2, 0.1)\) and \(\sigma^2_j = 3\). Here the top three eigenvalues explain about 95% of the total variation in the signal, and the last three are between 1/10 and 1/300 the size of the of the noise variance. This is a very challenging setup which features several trailing eigenvalues the last of which has size negligible relative to the noise variance.
Table 8
Table showing the true rank (in bold) and the empirical distribution of the estimated rank under Models S1–S5 with heteroskedastic errors for \((n, L) = (150, 25)\)

| Selected rank | 1 | 2 | 3 | 4 | 5 | 6 | ≥ 6 | 1 | 2 | 3 | 4 | 5 | 6 | ≥ 6 |
|---------------|---|---|---|---|---|---|-----|---|---|---|---|---|---|-----|
| Proposed      | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| AIC<sub>yao</sub> | 0 | 0 | 0 | 0 | 0 | 14 | 57  | 0 | 0 | 0 | 0 | 0 | 12 | 62  |
| AIC<sub>m</sub> | 8 | 20| 35| 35| 2 | 0 | 1   | 7 | 22| 37| 33| 0 | 0  | 0   |
| BIC<sub>m</sub> | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 100| 0  | 0   |
| PC<sub>p1</sub> | 38| 38| 23| 1 | 0 | 0 | 38  | 34| 27| 1 | 0 | 0 | 4 | 23  |
| IC<sub>p1</sub> | 64| 31| 5  | 0 | 0 | 66 | 31  | 3 | 0 | 0 | 0 | 0 | 8 | 30  |
| Ladle         | 0 | 0 | 11| 89| 0 | 0 | 0   | 0 | 9 | 91| 0 | 0 | 0 | 100 |
| ED            | 0 | 0 | 100| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| Model S3      | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| AIC<sub>yao</sub> | 0 | 0 | 0 | 0 | 0 | 4  | 96  | 0 | 0 | 0 | 0 | 0 | 5 | 95  |
| AIC<sub>m</sub> | 1 | 18| 81 | 0 | 0 | 0 | 13  | 87| 0 | 0 | 0 | 4 | 96| 0    |
| BIC<sub>m</sub> | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| PC<sub>p1</sub> | 2 | 18| 80 | 0 | 0 | 0 | 16  | 84| 0 | 0 | 0 | 5 | 95| 0    |
| IC<sub>p1</sub> | 4 | 34| 62 | 0 | 0 | 2  | 28  | 70| 0 | 0 | 0 | 8 | 92| 0    |
| Ladle         | 0 | 0 | 0 | 100| 0 | 0 | 0   | 0 | 0 | 100| 0 | 0 | 0 | 100 |
| Model S4      | 0 | 0 | 0 | 0 | 100| 0 | 0   | 0 | 0 | 100| 0 | 0 | 0 | 100 |

Table 9
Table showing the true rank (in bold) and the empirical distribution of the estimated rank under Models S1–S5 with heteroskedastic errors for \((n, L) = (150, 50)\)

| Selected rank | 1 | 2 | 3 | 4 | 5 | 6 | ≥ 6 | 1 | 2 | 3 | 4 | 5 | 6 | ≥ 6 |
|---------------|---|---|---|---|---|---|-----|---|---|---|---|---|---|-----|
| Proposed      | 0 | 0 | 0 | 100| 0 | 0 | 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| AIC<sub>yao</sub> | 0 | 0 | 0 | 0 | 0 | 14 | 57  | 0 | 0 | 0 | 0 | 0 | 12 | 62  |
| AIC<sub>m</sub> | 8 | 20| 35| 35| 2 | 0 | 1   | 7 | 22| 37| 33| 0 | 0  | 0   |
| BIC<sub>m</sub> | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 100| 0  | 0   |
| PC<sub>p1</sub> | 38| 38| 23| 1 | 0 | 0 | 38  | 34| 27| 1 | 0 | 0 | 4 | 23  |
| IC<sub>p1</sub> | 64| 31| 5  | 0 | 0 | 66 | 31  | 3 | 0 | 0 | 0 | 8 | 30 | 62  |
| Ladle         | 0 | 0 | 11| 89| 0 | 0 | 0   | 0 | 9 | 91| 0 | 0 | 0 | 100 |
| ED            | 0 | 0 | 100| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| Model S3      | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| AIC<sub>yao</sub> | 0 | 0 | 0 | 0 | 0 | 4  | 96  | 0 | 0 | 0 | 0 | 0 | 5 | 95  |
| AIC<sub>m</sub> | 1 | 18| 81 | 0 | 0 | 0 | 13  | 87| 0 | 0 | 0 | 4 | 96| 0    |
| BIC<sub>m</sub> | 0 | 0 | 0 | 0 | 0 | 100| 0   | 0 | 0 | 0 | 0 | 0 | 100| 0   |
| PC<sub>p1</sub> | 2 | 18| 80 | 0 | 0 | 0 | 16  | 84| 0 | 0 | 0 | 5 | 95| 0    |
| IC<sub>p1</sub> | 4 | 34| 62 | 0 | 0 | 2  | 28  | 70| 0 | 0 | 0 | 8 | 92| 0    |
| Ladle         | 0 | 0 | 0 | 100| 0 | 0 | 0   | 0 | 0 | 100| 0 | 0 | 0 | 100 |
| Model S5      | 0 | 0 | 0 | 0 | 100| 0 | 0   | 0 | 0 | 100| 0 | 0 | 0 | 100 |

Table 10 gives the empirical distribution of the selected rank for each of the above three models when \((n, L) = (150, 25)\) and \((150, 50)\) – sparse and dense grids, contrasting our proposed methods with the benchmark methods. The value of \(M = 10\) was used again in these settings. Intriguingly, it is observed that the proposed method yields near perfect estimation of the rank under all of the above models. The only exception is the most challenging model (SF3), and this only when the grid is sparse, where our method returns the true rank or the true rank minus in approximately a 50-50 split. Note that this is the scenario where the smallest eigenvalue is 1/300 the size of the error variance and the grid is sparse. The results
suggests a certain degree of robustness of the proposed procedure against extreme forms of the spectrum.

By comparison, the benchmark procedures markedly underperform relative to our procedure. The $AIC_{m}$, $PC_{p1}$ and $IC_{p1}$ procedures yield poor results as in the previous two subsections. The performance of $AIC_{yao}$ performs similarly as in earlier simulations, namely, it does well when the rank is large and the grid is sparse and the error variance is small. Its performance degrades significantly if the grid is dense or the rank is small, which results in either overestimation. When the error variance is not small, $AIC_{yao}$ underestimates the true rank. The most striking difference in performance is observed for the $BIC_{m}$ procedure, which now heavily underestimates the true rank in the spiked regime. The situation does not improve much even if we take dense grids (here $L = 50$). While this can be explained by the fact that $BIC_{m}$ is a model selection procedure targeting parsimonious models, it does also show that the consistency of $BIC_{m}$ may be slow to manifest in unbalanced spectra. The Ladle estimator is unable to pick up the true rank and chooses the "effective rank", particularly in the two high rank spiked models SF2 and SF3. The ED procedure is accurate for Models SF1 and SF2 but underestimates under Model SF3.

8.4. Infinite dimensional models. We now probe the finite sample performance of the procedure when the data are truly infinite dimensional, even prior to noise contamination; and we compare this with the output of model selection-based alternative procedures in such
situations. To this aim, we consider infinite dimensional models \( X(t) = \sum_{j=1}^{\infty} Y_j \varphi_j(t), t \in [0, 1] \) with \( E(Y_j) = 0, \text{Var}(Y_j) = \lambda_j > 0 \) for all \( j = 1, 2, \ldots \). The measurement error again satisfies \( \epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_j^2) \) for each \( i \). We consider four settings:

**Model 11** \( X \) is a standard Brownian motion, which features polynomial decay of eigenvalues and non-differentiable sample paths. Also, \( \sigma_j^2 = 1 \) for \( 1 \leq j \leq L \).

**Model 12** \( X \) is a Gaussian process with \( k_X(t, s) = \exp\{-((t-s)^2)/10\} \) – which features exponential decay of eigenvalues and infinitely smooth paths. Also, \( \sigma_j^2 = 1 \) for \( 1 \leq j \leq L \).

**Model 13** \( X \) is as in Model (I1). However, \( \sigma_j^2 = t_j \) for \( 1 \leq j \leq L \), where \( t_1 < t_2 \ldots < t_L < 1 \) is the observation grid.

**Model 14** \( X \) is as in Model (I2). However, \( \sigma_j^2 = t_j \) for \( 1 \leq j \leq L \), where \( 0 < t_1 < t_2 \ldots < t_L \) is the observation grid.

Inspection of the off-diagonal scree plot in a trial run from each scenario suggested no evident elbow below \( [(L - 1)/2] \), and so as per the recommendations of Section 7.1, we chose \( M = [(L - 1)/2] \) in each case. Tables 11 and 12 give the estimated ranks in 100 iterations under Models (I1)-(I4) for \( (n, L) = (150, 25) \) and \( (150, 50) \).

It is observed that the model selection procedures like AIC_{yoo}, AIC_m and BIC_m target some level of parsimonious representation of the data, the degree of parsimony depending on the method used. Unsurprisingly, they fail to inform us on whether the model is truly infinite dimensional or not (similar to having low power error in the testing paradigm). In the majority of cases, regardless of scenario, the chosen rank is between 1 and 3, in fact. The same is also true for the Ladle as well as the ED estimator. In contrast, the proposed method exhibits very good performance in terms of power, typically rejecting low-dimensional representations across all scenarios. In the case of dense grids \( (L = 50) \), the procedure never chose a rank below 15. In the case of a sparser grid \( (L = 25) \), the results varied somewhat between homoskedastic and heteroskedastic noise settings. In the two heteroskedastic scenarios, the procedure chose a rank of at least ten in 75% and 85% of runs. In the two homoskedastic scenarios, these percentages were modestly lower at about 56% and 62%. When we incorporated the assumption of homoskedasticity in the procedure (as per the comment in Section 6, at the top of p. 10), the performance surged in the two homoskedastic scenarios, with a rank of at least ten being chosen in 95% and 96% of runs. This suggests that, when operating with sparse grids, it can be beneficial in terms of power to make use of homoskedasticity if this can indeed be assumed.

**9. Data Analysis.** We will apply the bootstrap technique for estimating the rank to some benchmark data sets. The first of these is the well-known Tecator dataset which contains spectrometric curves for \( n = 215 \) samples of finely chopped meat (see Ferraty and Vieu (2006)). Each curve corresponds to the absorbances measured over \( L = 100 \) wavelengths. The second data set that we consider concerns the number of eggs laid by each of 1000 female Mediterranean fruit flies (medflies), Ceratitis capitata, in a fertility study described in Carey et al. (1998). The data\(^2\) contain the total number of eggs laid by each medfly as well as the daily breakup of the number of eggs laid. It is discussed in Carey et al. (1998) that there is a change in the pattern of egg production at day 51 post birth for those medflies which lived past that age. Also, the variation in the number of eggs laid from day 51 onwards is in general much larger than that before day 51. Taking these observations into account, it seems more pertinent to look at the egg-laying data till the age 50 days for those medflies that live

\(^2\)Accessible at http://anson.ucdavis.edu/~mueller/data/medfly1000.txt
past that age. This results in a sample of \( n = 145 \) medflies. Since the number of eggs laid in days 1 to 3 for these medflies equal zero, we only keep the number of eggs laid from day 4 onwards for our analysis. The Tecator data set has very smooth curves that are measured without any error while the fruit-fly data set is observed discretely and is rough/erratic.

9.1. Analysis of the Tecator data. A standard functional PCA of the Tecator data followed by a scree plot of the eigenvalues reveal a nearly finite dimensional structure since the eigenvalues decay to zero very fast (in fact super-exponentially). Indeed, the top eight eigenvalues are \( 2.613 \times 10^{-1}, 2.385 \times 10^{-3}, 7.845 \times 10^{-4}, 3.019 \times 10^{-4}, 1.523 \times 10^{-5}, \ldots \)
Table 13
Table showing the estimated rank of the Tecator data set under different error variances

| Error variance | 1 | 0.5 | 0.1 | 0.05 | 0.01 | 0.005 | 0.001 | 0.0005 | 0.0001 |
|----------------|---|-----|-----|------|------|-------|-------|--------|--------|
| Proposed method| 2 | 2   | 2   | 2    | 3    | 3     | 4     | 4      | 6      |
| AIC<sub>yao</sub> | 7 | 8   | 11  | 12   | 12   | 9     | 1     | 1      | 1      |
| AIC<sub>m</sub> | 2 | 2   | 1   | 1    | 1    | 1     | 1     | 1      | 1      |
| BIC<sub>m</sub> | 1 | 1   | 1   | 1    | 1    | 1     | 1     | 1      | 1      |
| PC<sub>p1</sub> | 2 | 2   | 2   | 1    | 1    | 1     | 1     | 1      | 1      |
| IC<sub>p1</sub> | 2 | 2   | 1   | 1    | 1    | 1     | 1     | 1      | 1      |

6.662 × 10⁻⁶, 2.052 × 10⁻⁶ and 1.084 × 10⁻⁶. The percentage of total variation explained by these principal components are 98.679%, 0.901%, 0.296%, 0.114%, 0.0058%, 0.0025%, 0.0008% and 0.0004%, respectively. Almost any model selection procedure would thus tend to choose ranks between 1 (most parsimonious choice) or perhaps 4 (which explains about 99.99% of total variation). However, there is still some numerically non-zero signal variation beyond the fourth eigenvalue and a plot of the eigenfunctions show that there is significant “pattern” till the 8th eigenfunction (past which they are numerically essentially zero). This is a very challenging scenario, where there is an obvious most parsimonious choice but the rank is still greater than that.

Since these data are recorded to high precision, and the curves are very smooth, it may be safely assumed that the measurements are essentially error-free. We will artificially add i.i.d. noise to the data and then apply our method and the alternative procedures considered in the previous section to evaluate their performance. Also, we will vary the error variance to investigate the effect of the magnitude of the signal-to-noise ratio on the rank selection algorithms. The errors are taken to be i.i.d. centered Gaussian with variances 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005 and 0.0001. These values range from “noise dominating signal completely” to “noise smaller than fourth largest eigenvalue”. It is thus expected that estimate of the rank corresponding to the error variance 0.0001 would be closest to the true numerical rank of the Tecator data. For our procedure and each value of the noise variance, we choose \( M = 10 \) as suggested by the off-diagonal scree plot.

Table 13 shows the estimated ranks obtained from the different procedures under the chosen levels of the error variance. The procedure proposed in the paper estimates the rank to be two, three or four in all cases where the error variance is large/interlaced. Only when the error variance is 0.0001 (less than the fourth eigenvalue), is the rank estimated to be six. Incidentally, the proposed procedure selected the rank to be seven for the original data without any error. This shows that our procedure is able to detect even very faint signals in the data provided the noise component is small compared to the signal (unlike most model selection procedures). We would also like to point out that deflating the error variance is akin to artificially inflating the same size. So, given the consistency of our procedure as the sample size grows, it is not surprising that the estimate of the rank when the error variance is 0.0001 and that for the data without error are very close. On the other hand, all of AIC<sub>m</sub>, PC<sub>p1</sub> and IC<sub>p1</sub> select the rank to be one or at most two. The BIC<sub>m</sub> procedure always selects the rank as one. This is expected since model selection procedures typically aim at achieving parsimony. These observations can be explained by noting that the Tecator data is an example of a nearly spiked functional dataset and the behaviour of these model selection procedures for such data was found to exhibit such behaviour in Section 8.3. The rank selected by the AIC<sub>yao</sub> procedure behaves somewhat erratically (both over and underestimation) as we decrease the error variance.
9.2. Analysis of the fruit-fly data. We implemented all the procedures for the fruit-fly data. Among the competing procedures, \( \text{AIC}_{\text{yao}} \) estimates the rank of the data to be equal to 9 while \( \text{BIC}_m \) selects the rank to be 7. All of \( \text{AIC}_m, \text{PC}_{p1} \) and \( \text{IC}_{p1} \) select the rank to be one, which appears way off based on a visual inspection of the data and the scree plot. The bootstrap procedure proposed in this paper is carried out by selecting \( M = 10 \). In fact, the off-diagonal scree plot as well as the results obtained from the competing methods indicate that the rank is likely smaller than 10. Our procedure selects the rank to be 7 at significance level \( \alpha = 5\% \) as well as 1\%. Further, our bootstrap test rejects the hypotheses \( H_{0,q} \) for \( q = 1, 2, \ldots, 6 \) with \( p \)-values that are numerically zero.

Our procedure thus yields the same result as the \( \text{BIC}_m \) approach, in this case, in addition to providing a confidence level. To understand the loss incurred in representing the data by the finite rank truncation, we compared the \( \text{AIC}_m \), the \( \text{BIC}_m \) and the \( \text{AIC}_{\text{yao}} \) approaches by computing the average relative squared error

\[
\text{ARSE} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{L} \frac{\sum_{j=1}^{L} W_{ij} - \hat{X}_{ij}(t_j))^2}{\sum_{j=1}^{L} W_{ij}^2},
\]

where \( \hat{X}_{ij}(\cdot) = \hat{\mu}(\cdot) + \sum_{j=1}^{\hat{r}} \hat{\xi}_{ij}(\cdot) \) is the prediction of \( X_{ij}(\cdot) \) using the PACE estimates of \( \mu \), \( \phi \) and \( \xi_{ij} \)’s (see Yao et al. (2005)). For computing the ARSE for each approach, we use the estimated value \( \hat{r} \) of the rank obtained from the corresponding approach. It is found that the ARSE for the \( \text{AIC}_{\text{yao}} \) approach (with \( \hat{r} = 9 \)) equals 0.200 and the ARSE for the \( \text{BIC}_m \) approach (with \( \hat{r} = 7 \)) is 0.204. Note that since our approach yields the same estimate of the rank as the \( \text{BIC}_m \) approach, the ARSE for our approach is also equal to 0.204. Thus, there is no significant improvement in the ARSE by considering 9 principal components (obtained using \( \text{AIC}_{\text{yao}} \)) instead of 7 (obtained using our approach or \( \text{BIC}_m \)). The ARSE of the \( \text{AIC}_m \) approach (as well as that of the \( \text{PC}_{p1} \) and the \( \text{IC}_{p1} \) approaches) equals 4.258. It would seem that these three approaches perform poorly in determining the true rank of the process in this example. It is easy to observe that the ARSE method considered above would automatically decrease with an increase in the value of the estimated rank. To take into account this caveat, we consider a penalised ARSE measure given by “ARSE + \( \lambda \times \) estimated rank” with a choice of \( \lambda = 0.01 \) to keep parity between the contribution from the values of the ARSE and the ranks. The following table shows a range of possible estimates of the true rank along with the corresponding penalised ARSE values. It is seen that the optimal choice of rank is 6 based on the above measure. The proposed method and \( \text{BIC}_m \) are the closest to this choice. We should mention that the difference between the values of the penalised ARSE at the ranks 6 and 7 is very marginal. Indeed, the proposed method (when performed at level \( \alpha = 0.1\% \)) selects 6 to be the true rank.

| Rank: | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Penalised ARSE: | 4.268 | 1.292 | 0.737 | 0.356 | 0.314 | 0.270 | 0.274 | 0.281 | 0.290 | 0.299 |

10. Extension to the Case of Irregular and/or Sparse Measurements. The framework defined in Section 2 assumed that we observe the measurements

\[(10.1) \quad W_{ij} = X_{t_j} + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, L,\]

for \( L \) regularly spaced nodes (i.e. \((j-1)/L \leq t_j < j/L\)). At first sight this may seem restrictive, as it does not cover the case where each curve may be measured at different and
irregularly distributed nodes. However, we show here that this case can in fact be reduced to the setting given in Section 2, and is thus covered by our methodology with only slight modification.

Given \( n \geq 2 \) i.i.d. replications \( \{X_1, \ldots, X_n\} \) of \( X \), suppose that we observe \( J \geq 2 \) irregular measurements per curve

\[
y_{ij} = X_i(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, J,
\]

where:

1. the \( n \times J \) random variables \( U_{ij} \)’s are independent uniform random variables on \([0,1]\), independent of the \( X_i \)’s.
2. conditional on \( U = (U_{ij}; 1 \leq i \leq n, 1 \leq j \leq J) \), the \( n \times J \) random variables \( \varepsilon_{ij} \)’s are uncorrelated continuous random variables which are independent of the \( X_i \)’s, with moments up to second order given by

\[
\mathbb{E}[\varepsilon_{ij} | U] = \mathbb{E}[\varepsilon_{ij} | U_{ij}] = 0, \quad & \\
\text{var}[\varepsilon_{ij} | U] = \text{var}[\varepsilon_{ij} | U_{ij}] = \sigma_{U_{ij}}^2 < \infty, \quad i = 1, \ldots, n, \quad j = 1, \ldots, J.
\]

The number of measurements per curve \( J \) can be as small as \( J = 2 \), so measurements are allowed to be sparse in addition to being irregular (we cannot admit \( J = 1 \) as in that case we have no information on covariance). Given this data we wish to test the hypothesis pair

\[
\begin{align*}
H_0 & : \text{rank}(k_X) \leq d \\
H_1 & : \text{rank}(k_X) > d
\end{align*}
\]

To this aim, let \( L = L_\perp \) as in Theorem 1 (for example, \( L = 2d + 1 \) when assuming the eigenfunctions satisfy the conditions given Remark 2). Now let \( \{I_k = (a_k, b_k)\}_{k=1}^L \) be the partition of \([0,1]\) into \( L \) subintervals of length \(|I_k| = 1/L\). Let \( \{t_k\}_{k=1}^L \) be a grid such that \( t_k \in I_k \) and

\[
k(t_k, t_{k'}) = \frac{1}{|I_k| \times |I_{k'}|} \int_{I_k} \int_{I_{k'}} k(x, y) dx dy, \quad k, k' \in \{1, \ldots, L\}.
\]

The existence of such \( \{t_k\}_{k=1}^L \) is guaranteed by the mean value theorem, \( k(\cdot, \cdot) \) being continuous on \([0,1]^2\). By definition these grid nodes are regularly spaced, i.e. \((j-1)/L \leq t_j < j/L\).

Writing \( \delta(x,y) \) for the Dirac delta at \((x,y)\) \( \in \mathbb{R}^2 \), define \( n \) independent marked point processes on the Borel subsets of \([0,1]^2\) as

\[
G_i = \frac{1}{J^2} \sum_{j=1}^J \sum_{j'=1}^J y_{ij}y_{ij'} \delta((U_{ij}, U_{ij'}), i = 1, \ldots, n.
\]

The average \( n^{-1} \sum_i G_i \) is typically referred to as the “raw covariance” in functional data analysis. Note that,

\[
G_i(I_k \times I_{k'}) = \frac{1}{J^2} \sum_{j,j'} y_{ij}y_{ij'} \mathbf{1}((U_{ij}, U_{ij'}) \in I_k \times I_{k'})
\]

Now we calculate

\[
\mathbb{E}[G_i(I_k \times I_{k'})] = \mathbb{E}[\mathbb{E}[G_i(I_k \times I_{k'}) | U_{ij}]_{j=1}^J]
\]

\[
= \frac{1}{J^2} \sum_{j=1}^J \sum_{j'=1}^J \mathbb{E}[\mathbf{1}((U_{ij}, U_{ij'}) \in I_k \times I_{k'}) \mathbb{E}[y_{ij}y_{ij'} | U_{ij}]_{j=1}^J]
\]
\[
\frac{1}{J^2} \sum_{j \neq j'} \mathbb{E} \left[ 1 \{ (U_{ij}, U_{ij'}) \in I_k \times I_{k'} \} k(U_{ij}, U_{ij'}) \right] + \frac{1}{J^2} \sum_{j=1}^J \mathbb{E} \left[ 1 \{ (U_{ij}, U_{ij}) \in I_k \times I_{k'} \} k(U_{ij}, U_{ij}) + \sigma_{U_{ij}}^2 \} \right]
\]

\[
= \frac{1}{J^2} \sum_{j \neq j'} \int_{I_k} \int_{I_{k'}} k(u, v) \, du \, dv + \frac{1}{J^2} \sum_{j=1}^J \mathbb{E} \left[ 1 \{ (U_{ij}, U_{ij}) \in I_k \times I_{k'} \} k(U_{ij}, U_{ij}) + \sigma_{U_{ij}}^2 \} \right] 1 \{ k = k' \}
\]

\[
= \frac{1}{L^2} k(t_k, t_{k'}) + 1 \{ k = k' \} \frac{1}{J} \int_{I_k} \{ k(u, u) + \sigma_u^2 \} \, du
\]

\[
= \frac{1}{L^2} k(t_k, t_{k'}) + 1 \{ k = k' \} \sigma_k^2.
\]

So, if we define the matrices

\[ A_i(k, k') = L^2 G_i(I_k \times I_{k'}), \]

we can note that the \( \{ A_i \}_{i=1}^n \) are iid random matrices with mean \( k(t_k, t_{k'}) + 1 \{ k = k' \} L^2 \sigma_k^2 = K_{X,L} + D \) for a diagonal matrix \( D \). In other words, we arrive at the precise framework of Section 2, with the regularly spaced grid \( \{ t_1, \ldots, t_L \} \) defined implicitly by the mean value-theorem. This implicit definition of the grid is immaterial, since we don’t actually need to know the grid values, we only need to be able to form the matrices \( A_i(k, k') = L^{-2} G_i(I_k \times I_{k'}) \) which play the exact same role as the matrices \( (W_{i1}, \ldots, W_{iL})(W_{i1}, \ldots, W_{iL})^\top \) of Section 2. The testing procedure is now implemented exactly as described in Section 4, simply replacing the empirical covariance \( \hat{K}_{W,L} \) by the \( L \times L \) pixelation of the “raw covariance” \( \hat{A} = n^{-1} \sum_{i=1}^n A_i \).

We do just this in a simulation study presented as part of the Supplementary Material, and it is seen that the method retains its performance in the sparse/irregular case.

11. Concluding remarks. Whether a null hypothesis can ever hold exactly true is a matter of philosophical debate. But even if a null cannot be precisely true, hypothesis testing can always make practical sense (including in the present context) when interpreted through a goodness-of-fit perspective: i.e. whether or not our data could be discerned from data coming from the null model, with a given level of confidence. Indeed significance testing is a very well established approach to order/rank determination in PCA, factor analysis, and even random processes going as far back as Fisher (1929). In light of our identifiability result (and test procedure) being non-asymptotic in the sampling resolution \( L \), our test can enjoy the same interpretation: can we distinguish the observed data from data generated from a “rank \( \leq d \) plus noise” model at the given resolution? Said differently, is the observable variation beyond dimension \( d \) discernible from that of pure noise?

Regardless of the interpretation of hypothesis testing itself, many core examples of random processes at the very foundations of harmonic analysis and (statistical) signal processing are known to be finite rank from physical first principles (i.e., the null hypothesis is known to be true). This is particularly well documented in the case of processes such as line spectra, waves, and vibrations, in concrete examples like communications (e.g. channel identification Xu et al. (1995)), geophysics (e.g. array detectors Bresler and Macovski (1986)), spectroscopy (e.g. NMR deconvolution Umesh and Tufts (1996)) and optics (e.g. interferometry Patil et al. (2005)). A particularly succinct overview is provided in Chapter 4 of the classic textbook by Stoica and Moses (2005). These examples fall within a class of processes of the form

\[
X(t) = \sum_{k=1}^r Y_k \psi_k(t - \xi_k), \quad r < \infty,
\]

(11.1)
where $Y_k$ are random coefficients, and $\{\xi_j\}$ are independent random shifts. For example, when $\psi_k(u) = \exp\{i\theta_k u\}$ we get the so-called random harmonic model. When $\psi_k(u) = \psi(u/\sigma_k)$ we get a random spectral convolution model. Because $r < \infty$, the random signal $X(t)$ has rank at most $r$. To see this, note that $E[\psi_k(t - \xi_j)] = \psi_k(t - u)f_{\xi_j}(u)du = [\psi_k * f_{\xi_j}](t) = \phi_k(t)$, and so $\text{Cov}(X(u), X(v)) = \sum_{j=1}^{r} \sum_{j'=1}^{r} \text{Cov}(Y_i, Y_{j'}) \phi_i(u) \phi_{j'}(v)$ and the rank is that of the $r \times r$ positive semidefinite matrix $\text{Cov}(Y_i, Y_{j'})$. Yet more general versions are obtained when replacing the random shifts $\psi_k(x) \rightarrow \psi_k(x - \xi_k)$ by random deformations $\psi_k(x) \rightarrow \psi_k(T_k^{-1}(x))$, and a similar calculation shows the rank to be at most $r$.

Finally, it is worth re-emphasizing out that one need not choose between hypothesis testing or model selection. As discussed in Section 7.1 our method controls the FWER even under data-snooping. Hence our method can always be used in conjunction with a selection procedure, as a confirmatory analysis.

**ACKNOWLEDGEMENT**

We are thankful to the Editors, the Associate Editor, and four referees for their many and constructive comments.

**REFERENCES**

Amini, A. A. and Wainwright, M. J. (2012). Sampled forms of functional PCA in reproducing kernel Hilbert spaces. *Ann. Statist.*, 40(5):2483–2510.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bathia, N., Yao, Q., and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Ann. Statist.*, 38(6):3352–3386.

Bresler, Y. and Macovski, A. (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust.*, 34(5):1081–1089.

Carey, J., Liedo, P., Müller, H., Wang, J., and Chiu, J. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of mediterranean fruit fly females. *J. of Gerontology - Biological Sciences*, 53:245–251.

Chen, Y. and Wainwright, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. Tech. Report arXiv:1509.03025.

Choi, Y., Taylor, J., and Tibshirani, R. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *Ann. Statist.*, 45(6):2590 – 2617.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Math. Phys. Eng. Sci.*, 125(796):54–59.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(4):689–705.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Kneip, A. (1994). Nonparametric Estimation of Common Regressors for Similar Curve Data. *nn. Statist.*, 22(3):1386 – 1427.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.*, 40(2):694 – 726.

Li, Y., Wang, N., and Carroll, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.*, 108(504):1284–1294.

Luo, W. and Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, 103(4):875–887.

Lynch, G., Guo, W., Sarkar, S. K., and Finner, H. (2017). The control of the false discovery rate in fixed sequence multiple testing. *Electron. J. Stat.*, 11(2):4649–4673.
Maurer, W., Hothorn, L., and Lehmacher, W. (1995). *Multiple comparisons in drug clinical trials and preclinical assays: A-priori ordered hypotheses*. Biometrie in der chemische-pharmazeutischen Industrie, J. Vollman ed. Fischer Verlag, Stuttgart. Vo. 6.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.*, 92(4):1004–1016.

Patil, A., Rastogi, P., and Raphael, B. (2005). Phase-shifting interferometry by a covariance-based method. *Appl. Opt.*, 44(27):5778–5785.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642.

Peres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput. Statist. Data Anal.*, 49(4):974 – 997.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Stoica, P. and Moses, R. L. (2005). *Spectral Analysis of Signals*. Pearson/Prentice Hall, Inc.

Stoica, P., Moses, R. L., et al. (2005). *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ.

Umesh, S. and Tufts, D. (1996). Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to nmr spectroscopy data. *IEEE Trans. Signal Process.*, 44(9):2245–2259.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3):321–327.

Xu, G., Liu, H., Tong, L., and Kailath, T. (1995). A least-squares approach to blind channel identification. *IEEE Trans. Signal Process.*, 43(12):2982–2993.

Yao, F., Muller, H.-G., and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.*, 33(6):2873–2903.