IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 291–302

DOI: 10.1214/12-IMSCOLL921

Some asymptotic theory for functional regression with stationary regressor

Frits Ruymgaart, Jing Wang and Shih-Hsuan Wei

Texas Tech University

Abstract: The general asymptotic distribution theory for the functional regression model in Ruymgaart et al. [Some asymptotic theory for functional regression and classification (2009) Texas Tech University] simplifies considerably if an extra assumption on the random regressor is made. In the special case where the regressor is a stochastic process on the unit interval, Johannes [Privileged communication (2008)] assumes the regressor to be stationary, in which case the eigenfunctions of their covariance operator turn out to be known, so that only the eigenvalues are to be estimated. In the present paper we will also assume the eigenvectors to be known, but within an abstract setting. The simplification mentioned above is due to the circumstance that the covariance operator of the regressor commutes with its estimator as it can be constructed under the current conditions. Moreover, it is now possible to test linear hypotheses for the regression parameter that correspond to linear subspaces spanned by a finite number of the known eigenvectors.

1. Introduction

Several functional regression models can be found in the monograph by Ramsay and Silverman [7]. One of these models, concerning functional regression for a real dependent variable and a single functional independent variable as specified in (2.1) below, is considered by Hall and Horowitz [2]. These authors establish rate optimality of the mean integrated square error of an estimator based on a Tikhonov type inverse of the sample covariance operator, when both the random regressor and the regression function are assumed to be in $L^2(0,1)$. In this paper we will also restrict ourselves to this model, that we will briefly refer to as "the functional regression model". Mas and Pumo [6] argue that in certain practical situations it might be preferable to assume they are elements of a Sobolev space such as $W^{2,1}(0,1)$. Johannes [3] obtains results similar to those in [2]: more general in the sense that general Sobolev norms are used, and more restrictive because this author assumes the regressor to be a stationary second order process; for further comments on this interesting special case see below.

The asymptotic distribution of the regression estimator is obtained in Ruymgaart et al. [8] for abstract Hilbert spaces and without any specific assumption on the regressor except a standard moment condition. Not included in that paper was the special case where the eigenvectors of the covariance operator of the regressor are

Texas Tech University, Lubbock, Texas 79409, USA, e-mail: h.ruymgaart@ttu.edu; eliza.wang@ttu.edu; scottie.wei@ttu.edu

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: Functional regression, stationary regressor, asymptotic distribution, testing linear hypotheses

assumed to be known and only the unknown eigenvalues are to be estimated. This special case, referred to above, is motivated by an important instance, considered by Johannes [3]. This author observes that any stationary second order regressor with time in the unit interval has a covariance operator with known eigenfunctions

$$\begin{cases} p_1(t) = 1, \\ p_{2k}(t) = \sqrt{2}\cos 2\pi kt, \\ p_{2k+1}(t) = \sqrt{2}\sin 2\pi kt, \end{cases} \qquad \text{for } 0 \le t \le 1, \text{ and } k \in \mathbb{N}.$$

If the eigenvectors are known a different, simpler estimator of the covariance operator is possible that leads to a different and simpler estimator of the regression function as well. Because the covariance operator and its simplified estimator commute, derivatives of analytic functions of these operators also simplify considerably, which in turn leads to simpler asymptotics. Also linear hypotheses can be formulated in terms of the known basis of eigenvectors which further reduces the complexity of the asymptotic distribution of the test statistic. Considering the general results in [8] such a reduction seems certainly welcome and it is the purpose of this paper to pursue this well motivated special case. Although direct proofs could be provided we prefer to derive the results from the general theory in [8].

In Section 2 the model and the basic assumptions are formulated and some notation is introduced. Section 3 is devoted to the modified sample covariance operator. It may be obtained by just estimating the eigenvalues, but it also turns out to be the projection of the usual sample covariance operator onto the subspace of the Hilbert-Schmidt operators spanned by a suitable collection of known tensor products (see (3.3)). Consequently its asymptotic distribution is immediate from the continuous mapping theorem and the convergence in distribution of the usual sample covariance operator although, indeed, a simple direct proof could be given as well. The modified estimator of the regression function is presented in Section 4. Some tools from [8] for functions of operators, in particular a derivative and an ensuing delta-method are briefly reviewed in Section 5. Finally, in Section 6, these results are applied to obtain tests for certain linear hypotheses.

2. The model, assumptions, and some notation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. It will be assumed that all random elements are defined on this space. Let $\mathbb{H}$ be a separable, infinite dimensional Hilbert space over the real numbers, equipped with the σ-field of Borel sets $\mathcal{B}_{\mathbb{H}}$, and $X : \Omega \to \mathbb{H}$ a random vector, i.e. an $(\mathcal{F}, \mathcal{B}_{\mathbb{H}})$-measurable mapping. It will be assumed throughout that

$$\mathbb{E}\,\|X\|^4 < \infty.$$

Suppose that $X_1, \dots, X_n$ are independent copies of X. For an unknown number $\alpha \in \mathbb{R}$ and vector $f \in \mathbb{H}$ we observe n pairs (X_i, η_i) that are i.i.d. copies of (X, η), given by

$$\eta = \alpha + \langle X,\ f\rangle + \epsilon, \qquad \alpha \in \mathbb{R},\ \ f \in \mathbb{H}, \tag{2.1}$$

where ϵ is a real valued error variable such that

$$\epsilon \perp\!\!\!\perp X, \qquad \mathbb{E}\epsilon = 0, \qquad \mathbb{E}\epsilon^2 = v^2 < \infty. \tag{2.2}$$

The random variable X is called the regressor and the unknown parameter $f \in \mathbb{H}$ the regression parameter (usually a function as most Hilbert spaces will be function spaces). Hall and Horowitz [2] study this model when $\mathbb{H} = L^2(0,1)$, and Johannes [3] considers Sobolev inner products. These authors focus on the (mean) integrated squared error. Mas and Pumo [6] consider the regression model in the Hilbert space $W^{2,1}(0,1)$.

The mean $\mu = \mathbb{E}X$ of X is defined as the representer of the bounded functional

$$\mathbb{E}\langle a,\ X\rangle = \langle a,\ \mu\rangle, \qquad \forall a \in \mathbb{H},$$

see [5]. In the same vein the covariance operator of X can be considered as a mean in the Hilbert space $\mathcal{L}_{HS}$ of all Hilbert-Schmidt operators mapping $\mathbb{H}$ into itself. A linear operator $U : \mathbb{H} \to \mathbb{H}$ is Hilbert-Schmidt if in any orthonormal basis $e_1, e_2, \ldots$ of $\mathbb{H}$ we have

$$\sum_{k=1}^{\infty} \|Ue_k\|^2 < \infty.$$

The inner product in $\mathcal{L}_{HS}$ is then defined by

$$\langle U,\ V\rangle_{HS} = \sum_{k=1}^{\infty} \langle Ue_k,\ Ve_k\rangle. \tag{2.3}$$

This inner product does not depend on the choice of basis, see [4]. We can now define the covariance operator as

$$\Sigma = \mathbb{E}(X-\mu) \otimes (X-\mu), \tag{2.4}$$

where for $a, b \in \mathbb{H}$, the operator $a \otimes b$ is defined by $(a \otimes b)x = \langle x, b\rangle a$, $x \in \mathbb{H}$. Such operators are clearly Hilbert-Schmidt. It is easily seen that Σ in (2.4) is also uniquely determined by the requirement

$$\mathbb{E}\langle a,\ X-\mu\rangle\langle X-\mu,\ b\rangle = \langle a,\ \Sigma b\rangle, \qquad \forall a, b \in \mathbb{H}.$$

The latter definition can be found in [5].

It is well known that Σ is nonnegative, Hermitian, with finite trace $tr\Sigma = \mathbb{E}\|X-\mu\|^2$, and hence Hilbert-Schmidt and therefore compact. It will be assumed that Σ is one-to-one or, equivalently, strictly positive. For simplicity of presentation let us also assume that all eigenvalues have multiplicity one. They can be written in decreasing order with limit 0 and for Σ we have the spectral representation

$$\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 \cdot p_k \otimes p_k, \qquad \sigma_1^2 > \sigma_2^2 > \cdots \downarrow 0, \tag{2.5}$$

where $p_1, p_2, \ldots$ is the corresponding orthonormal basis of eigenvectors.

Assumption 2.1. Throughout this paper it will be assumed that the eigenvectors

$$p_1, p_2, \ldots \ \text{ are known.} \tag{2.6}$$

Occasionally the stronger assumption that the

$$p_1, p_2, \ldots \ \text{ are known and } \ X \stackrel{d}{=} \text{Gaussian}(\mu, \Sigma), \tag{2.7}$$

will be made.

The eigenvalues are still unknown parameters. Under assumption (2.7), $X - \mu$ has the property that the

$$\frac{1}{\sigma_j}\langle X-\mu,\ p_j\rangle \ \text{ are i.i.d. Normal (0,1).} \tag{2.8}$$

3. The modified sample covariance operator

The traditional sample covariance operator is

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) \otimes (X_i - \bar{X}).$$

With the eigenvectors known, however, only the eigenvalues need to be estimated. Since

$$\mathbb{E}\langle X-\mu,\ p_k\rangle^2 = \langle p_k,\ \Sigma p_k\rangle = \sigma_k^2,$$

it seems natural to estimate these eigenvalues by

$$\widehat{\sigma}_k^2 = \frac{1}{n}\sum_{i=1}^{n}\langle X_i - \bar{X},\ p_k\rangle^2. \tag{3.1}$$

This yields a modified estimator of Σ, viz.

$$\widehat{\widehat{\Sigma}} = \sum_{k=1}^{\infty}\widehat{\sigma}_k^2 \cdot p_k \otimes p_k. \tag{3.2}$$

There exists a nice relation between the two, that are both in $\mathcal{L}_{HS}$. An orthonormal basis for this space (with respect to the inner product in (2.3)) is the collection of all tensor products $p_j \otimes p_k$ $(j,k \in \mathbb{N})$. Let us write

$$\mathcal{M} = [\![p_1 \otimes p_1, p_2 \otimes p_2, \ldots]\!] \subset \mathcal{L}_{HS}, \tag{3.3}$$

for the closed linear subspace spanned by the projection operators $p_j \otimes p_j$, $(j \in \mathbb{N})$, and define $\mathcal{P} : \mathcal{L}_{HS} \to \mathcal{M}$ to be the orthogonal projection of $\mathcal{L}_{HS}$ onto $\mathcal{M}$. We are now in a position to describe the relation.

Lemma 3.1. *In the above notation we have*

$$\widehat{\widehat{\Sigma}} = \mathcal{P}\widehat{\Sigma}.$$

Proof. A Fourier expansion yields

$$\mathcal{P}\widehat{\Sigma} = \sum_{j=1}^{\infty}\langle\widehat{\Sigma},\ p_j \otimes p_j\rangle_{HS} \cdot p_j \otimes p_j.$$

Using the basis $p_1, p_2, \ldots$ for evaluating the above inner products (cf (2.3)) we obtain

$$\begin{aligned}
\langle\, \widehat{\Sigma},\ p_j \otimes p_j\rangle_{HS} &= \sum_{k=1}^{\infty}\langle\, \widehat{\Sigma}p_k,\ (p_j \otimes p_j)p_k\rangle \\
&= \langle\, \widehat{\Sigma}p_j,\ p_j\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\big\langle\big((X_i - \bar{X}) \otimes (X_i - \bar{X})\big)p_j,\ p_j\big\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\langle X_i - \bar{X},\ p_j\rangle^2 \\
&= \widehat{\sigma}_j^2,
\end{aligned}$$

and we are done. □

Let $\mathcal{L}$ be the Banach space of all bounded, linear $T : \mathbb{H} \to \mathbb{H}$. Since $\mathcal{L}_{HS}$ is continuously embedded in $\mathcal{L}$, convergence in distribution in $\mathcal{L}_{HS}$ entails convergence in distribution in $\mathcal{L}$. It is well known, and it follows at once from the central limit theorem in separable Hilbert spaces [1] that there exists a zero mean Gaussian random element $\mathcal{G}_\Sigma$ in $\mathcal{L}_{HS}$ such that

$$\sqrt{n}(\widehat{\Sigma} - \Sigma) \stackrel{d}{\to} \mathcal{G}_\Sigma, \qquad \text{as} \quad n \to \infty, \quad \text{in } \mathcal{L}_{HS} \Rightarrow \text{in } \ \mathcal{L}.$$

The covariance structure of $\mathcal{G}_\Sigma$ is given by

$$\begin{aligned}&\mathbb{E}\mathcal{G}_\Sigma \otimes_{HS} \mathcal{G}_\Sigma \\ &\quad = \mathbb{E}\left\{(X-\mu)\otimes(X-\mu) - \Sigma\right\} \otimes_{HS} \left\{(X-\mu)\otimes(X-\mu)-\Sigma\right\}.\end{aligned}$$

Since $\mathcal{P}\Sigma = \Sigma$, the continuous mapping theorem entails at once the following.

Theorem 3.1. *The modified sample covariance operator satisfies*

$$\sqrt{n}(\widehat{\widehat{\Sigma}} - \Sigma) \stackrel{d}{\longrightarrow} \mathcal{P}\mathcal{G}_\Sigma = \sum_{j=1}^{\infty} \langle \mathcal{G}_\Sigma p_j, \ p_j\rangle \cdot p_j \otimes p_j, \tag{3.4}$$

as $n \to \infty$, in $\mathcal{L}_{HS} \Rightarrow$ in $\mathcal{L}$.

Theorem 3.2. *Under assumption (2.7), the above reduces to*

$$\sqrt{n}(\widehat{\widehat{\Sigma}} - \Sigma) \stackrel{d}{\longrightarrow} \sum_{j=1}^{\infty} \sqrt{2}\sigma_j^2 Z_j \cdot p_j \otimes p_j, \tag{3.5}$$

where the $Z_1, Z_2, \ldots$ are i.i.d. Normal(0,1).

Proof. Exploiting (2.8) it has been shown in [8] that $\mathcal{G}_\Sigma$ has the Karhunen-Loève expansion

$$\mathcal{G}_\Sigma = \sum_{j=1}^{\infty} \sqrt{2}\sigma_j^2 \ Z_{j,j} \cdot p_j \otimes p_j \ + \sum\sum_{j\neq k} \sigma_j \ \sigma_k \ Z_{j,k} \cdot p_j \otimes p_k, \tag{3.6}$$

where the random variables $Z_{j,k}, (j,k \in \mathbb{N})$ are i.i.d Normal(0, 1). It follows at once from (3.6) and the expression for $\mathcal{P}\mathcal{G}_\Sigma$ in (3.4) that now $\mathcal{P}\mathcal{G}_\Sigma$ has the expansion in (3.5) when we write $Z_j = Z_{j,j}$. □

4. The modified estimator of the regression function

The basic equation from which an estimator of f can be obtained is

$$\frac{1}{n}\sum_{i=1}^{n} \eta_i \ (X_i - \bar{X}) \ = \ \widehat{\Sigma} f + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \ (X_i - \bar{X}),$$

see, for instance, [2] and [8]. Since $\widehat{\Sigma}$ is not one-to-one the estimator was based on a generalized inverse of $\widehat{\Sigma}$ (of Tikhonov type). Here we want to base the estimator on an inverse of $\widehat{\widehat{\Sigma}}$. Both Σ and $\widehat{\widehat{\Sigma}}$ are compact and have therefore unbounded inverses. Consequently we will employ regularized inverses and consider the estimator

$$\hat{\hat{f}}_\delta = (\delta I + \widehat{\widehat{\Sigma}})^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \eta_i (X_i - \bar{X}) \right), \tag{4.1}$$

where $I : \mathbb{H} \to \mathbb{H}$ is the identity operator and $\delta > 0$, with its deterministic counterpart

$$f_\delta = (\delta I + \Sigma)^{-1}\Sigma f, \qquad f \in \mathbb{H}. \tag{4.2}$$

Remark 4.1. The asymptotics for the modified estimator in (4.1) is much simpler because Σ and $\widehat{\widehat{\Sigma}}$ commute (note that Σ and $\widehat{\Sigma}$ don't), see Section 5. It has been argued in [8] that for suitable asymptotic behavior the regularization parameter $\delta > 0$ should be fixed, and the same will be done here. We will exploit in Section 6 that testing linear hypotheses about f is in fact equivalent to testing the same hypotheses about f_δ, if this hypothesis is formulated in terms of some of the $p_1, p_2, \ldots$.

We have the decomposition

$$\begin{aligned}
&\sqrt{n}(\hat{\hat{f}}_\delta - f_\delta) \\
&\quad = -(\delta I + \widehat{\widehat{\Sigma}})^{-1}\sqrt{n}(\widehat{\widehat{\Sigma}} - \widehat{\Sigma})f + \sqrt{n}\big((\delta I + \widehat{\widehat{\Sigma}})^{-1}\widehat{\widehat{\Sigma}} - (\delta I + \Sigma)^{-1}\Sigma\big)f \\
&\qquad + (\delta I + \widehat{\widehat{\Sigma}})^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i(X_i - \bar{X})\right) \\
&\quad = -U_1 + U_2 + U_3.
\end{aligned}$$

In this expansion, two operator functions in the sense of functional calculus are employed, viz.

$$\varphi_1(z) = \frac{1}{\delta + z}, \qquad \varphi_2(z) = \frac{z}{\delta + z}, \qquad z \in \mathbb{C} \setminus \{-\delta\}. \tag{4.3}$$

With the help of these we may write

$$\begin{aligned}
U_1 &= \varphi_1\big(\widehat{\widehat{\Sigma}}\big)\sqrt{n}\Big(\widehat{\widehat{\Sigma}} - \widehat{\Sigma}\Big)f, \\
U_2 &= \sqrt{n}\big(\varphi_2(\widehat{\widehat{\Sigma}}) - \varphi_2(\Sigma)\big)f, \\
U_3 &= \varphi_1\big(\widehat{\widehat{\Sigma}}\big)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i(X_i - \bar{X})\right).
\end{aligned} \tag{4.4}$$

Before deriving the asymptotic distribution of the estimators we will briefly review some results for functions of (perturbed) operators.

5. Some results from functional calculus

The spectrum of Σ in (2.5) equals $\sigma(\Sigma) = \{0, \sigma_2^2, \sigma_2^2, \ldots\}$. Let Γ be a closed contour in the complex plane $\mathbb{C}$ that contains $[-\frac{1}{2}\delta, \sigma_1^2]$ in its interior, stays at a distance of at least $\frac{1}{2}\delta$ from this interval, but keeps $-\delta$ outside (δ is the same as in (4.1)). Let Ω denote the open region enclosed by Γ, $D \supset (\Omega \cup \Gamma) = \bar{\Omega}$ an open neighborhood of $\bar{\Omega}$, and suppose that

$$\varphi : \quad D \longrightarrow \mathbb{C} \quad \text{is analytic.}$$

Let $\mathcal{L}$ be the Banach space of all bounded operators mapping $\mathbb{H}$ into itself, and let $\Pi \in \mathcal{L}$ be a perturbation satisfying

$$\|\Pi\|_{\mathcal{L}} \leq \frac{1}{4}\delta, \tag{5.1}$$

where $\|\cdot\|_{\mathcal{L}}$ is the usual operator norm on $\mathcal{L}$. For such perturbations it can be shown that $\sigma(\widetilde{\Sigma}) = \sigma(\Sigma + \Pi) \subset \Omega$, so that the resolvent sets for Σ and $\widetilde{\Sigma}$ satisfy

$$\rho(\widetilde{\Sigma}) = \rho(\Sigma + \Pi) \supset \Omega^c \supset \Gamma, \qquad \rho(\Sigma) = \big(\sigma(\Sigma)\big)^c. \tag{5.2}$$

The resolvents of Σ and $\widetilde{\Sigma}$ are defined by

$$R(z) = (zI - \Sigma)^{-1}, \qquad z \in \rho(\Sigma),$$
$$\widetilde{R}(z) = (zI - \widetilde{\Sigma})^{-1}, \qquad z \in \rho(\widetilde{\Sigma}),$$

and they are analytic on the respective resolvent sets. According to the usual functional calculus we define

$$\varphi(\Sigma) = \frac{1}{2\pi i}\oint_{\Gamma} \varphi(z) R(z)\, dz, \qquad \varphi(\widetilde{\Sigma}) = \frac{1}{2\pi i}\oint_{\Gamma} \varphi(z) \widetilde{R}(z)\, dz,$$

where the latter is well defined for the same contour because of (5.2). It is well known that these definitions reduce to recognizable expressions in special cases, as in the examples below

Example 5.1. For φ_1 as in (4.3), we have

$$\varphi_1(\Sigma) = \sum_{k=1}^{\infty} \frac{1}{\delta + \sigma_k^2}\; p_k \otimes p_k = (\delta I + \Sigma)^{-1}.$$

Example 5.2. For φ_2 as in (4.3) we have

$$\varphi_2(\Sigma) = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{\delta + \sigma_k^2}\; p_k \otimes p_k = (\delta I + \Sigma)^{-1}\Sigma.$$

Let us write

$$M_\varphi = \max_{z \in \Gamma} |\varphi(z)|\,.$$

In the present general situation, where we do not yet assume that Π and Σ commute, we have the following result.

Theorem 5.1. *Provided that (5.1) is fulfilled we have, for some number* $0 < C < \infty$*,*

$$\|\varphi(\Sigma + \Pi) - \varphi(\Sigma)\| \le C\, M_\varphi \cdot \frac{1}{\delta}\, \|\Pi\|_{\mathcal{L}}\,,$$
$$\|\varphi(\Sigma + \Pi) - \varphi(\Sigma) - \dot{\varphi}_\Sigma\, \Pi\| \le C\, M_\varphi \left(\frac{1}{\delta}\, \|\Pi\|_{\mathcal{L}}\right)^2,$$

where the derivative $\dot{\varphi}_\Sigma : \mathcal{L} \to \mathcal{L}$ *is bounded and given by*

$$\dot{\varphi}_\Sigma\, \Pi = \sum_k \varphi'(\sigma_k^2)\, P_k \Pi P_k + \sum\sum_{j \ne k} \frac{\varphi(\sigma_k^2) - \varphi(\sigma_j^2)}{\sigma_k^2 - \sigma_j^2}\, P_j \Pi P_k. \tag{5.3}$$

This result almost immediately yields a delta-method for obtaining limiting distributions for functions of $\widehat{\widehat{\Sigma}}$. In view of (3.4) the random perturbation $\widehat{\widehat{\Pi}} = \widehat{\widehat{\Sigma}} - \Sigma$ is small in probability, and because $\widehat{\widehat{\Sigma}}$ is based on the same eigenprojections as Σ it is obvious that $\widehat{\widehat{\Pi}}$ commutes with all these eigenprojections. Therefore the derivative $\dot{\varphi}_\Sigma \widehat{\widehat{\Pi}}$ reduces to the single sum on the right in (5.3) and we obtain the following result.

Theorem 5.2. *In the present situation where (2.6) is assumed to hold true and* $\widehat{\widehat{\Sigma}}$ *is given by (3.2), we have*

$$\sqrt{n}\left\{\varphi(\widehat{\widehat{\Sigma}})-\varphi(\Sigma)\right\}\xrightarrow{d}\dot{\varphi}_{\Sigma}\mathcal{P}\mathcal{G}_{\Sigma},\qquad as\ \ n\to\infty, \tag{5.4}$$

in $\mathcal{L}$*, where the derivative reduces to*

$$\dot{\varphi}_{\Sigma}\mathcal{P}\mathcal{G}_{\Sigma}=\sum_{k=1}^{\infty}\varphi'(\sigma_k^2)P_k\mathcal{P}\mathcal{G}_{\Sigma}P_k. \tag{5.5}$$

Under the assumption (2.7) we have

$$\dot{\varphi}_{\Sigma}\mathcal{P}\mathcal{G}_{\Sigma}=\sum_{k=1}^{\infty}\sqrt{2}\varphi'(\sigma_k^2)\sigma_k^2 Z_k p_k\otimes p_k. \tag{5.6}$$

6. Testing hypotheses about the regression function

6.1. Asymptotics under the null hypothesis

Since the estimator of the regression function in (4.1) differs from the one in [8] which is based on $\widehat{\Sigma}$ rather than $\widehat{\widehat{\Sigma}}$, we can not directly refer to the results in that paper. The straightforward proof of its convergence in distribution, however, is very similar to the proof in [8] and will be omitted. In order to present the result, apart from (5.4), we will need (cf. (4.4))

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i(X_i-\bar{X})\xrightarrow{d}\mathcal{G}_0,\qquad \text{in }\ \mathbb{H},\qquad \text{as }\ n\to\infty,$$

where

$$\mathcal{G}_0=\sum_{k=1}^{\infty}v\ \sigma_k Z_k'\ p_k, \tag{6.1}$$

with v as in (2.2) and the

$$Z_1',Z_2',\ldots\qquad \text{i.i.d. Normal(0,1)},$$

independent of the $Z_1,Z_2,\ldots$ in (5.6).

Theorem 6.1. *For any fixed* $\delta>0$*, we have*

$$\sqrt{n}(\hat{\hat{f}}_{\delta}-f_{\delta})\xrightarrow{d}\mathcal{H}(f),\quad as\ \ n\to\infty,$$

in $\mathbb{H}$*, where* $\mathcal{H}(f)=\mathcal{H}_1(f)+\mathcal{H}_2(f)+\mathcal{H}_3$*, and the random elements on the right are zero mean, Gaussian, and given by*

$$\begin{aligned}\mathcal{H}_1(f)&=\varphi_1(\Sigma)(\mathcal{P}^{\perp}\mathcal{G}_{\Sigma})f,\\ \mathcal{H}_2(f)&=(\dot{\varphi}_{2,\Sigma}(\mathcal{P}\mathcal{G}_{\Sigma}))f,\\ \mathcal{H}_3&=\varphi_1(\Sigma)\mathcal{G}_0,\end{aligned} \tag{6.2}$$

and where, moreover,

$$\mathcal{H}_3\perp\!\!\!\perp(\mathcal{H}_1(f),\ \mathcal{H}_2(f)).$$

For each of these processes, a more explicit expression can be obtained by exploiting the basic expressions in (3.4), (5.5) and (6.1), and by realizing that

$$\begin{aligned}\mathcal{P}^{\perp}\mathcal{G}_{\Sigma} &= \sum_{j\neq k}\sum \langle \mathcal{G}_{\Sigma},\ p_j \otimes p_k\rangle_{HS}\ p_j \otimes p_k \\ &= \sum_{j\neq k}\sum \langle \mathcal{G}_{\Sigma} p_k,\ p_j\rangle\ p_j \otimes p_k.\end{aligned}$$

Theorem 6.2. *We have, more explicitly,*

$$\mathcal{H}_1(f) = \sum_{k=1}^{\infty} \frac{1}{\delta + \sigma_k^2}\big(\langle \mathcal{G}_{\Sigma} p_k,\ f\rangle - \langle \mathcal{G}_{\Sigma} p_k,\ p_k\rangle\ \langle f,\ p_k\rangle\big) p_k, \tag{6.3}$$

$$\mathcal{H}_2(f) = \sum_{k=1}^{\infty} \frac{\delta}{(\delta + \sigma_k^2)^2} \langle \mathcal{G}_{\Sigma} p_k,\ p_k\rangle\ \langle f,\ p_k\rangle\ p_k, \tag{6.4}$$

$$\mathcal{H}_3 = \sum_{k=1}^{\infty} \frac{v\sigma_k}{\delta + \sigma_k^2} Z_k'\ p_k.$$

If the assumption in (2.7) is made, the above random elements can be even made more explicit. Substitution of the r.h.s of (3.6) into (6.3) and (6.4) yields the following.

Theorem 6.3. *If (2.7) is fulfilled we have*

$$\mathcal{H}_1(f) = \sum_{k=1}^{\infty} \frac{\sigma_k}{\delta + \sigma_k^2} \left(\sum_{j\neq k} \sigma_j\ Z_{j,k} \langle f,\ \ p_j\rangle \right) p_k, \tag{6.5}$$

$$\mathcal{H}_2(f) = \sum_{k=1}^{\infty} \frac{\sqrt{2}\delta\sigma_k^2}{(\delta + \sigma_k^2)^2}\ Z_k\ \langle f,\ \ p_k\rangle p_k, \tag{6.6}$$

$$\mathcal{H}_3 = \sum_{k=1}^{\infty} \frac{v\sigma_k}{\delta + \sigma_k^2}\ Z_k' p_k, \tag{6.7}$$

where the $Z_{j,k} (j \neq k)$, the Z_k and the Z_k' are all mutually independent with Normal(0,1) distribution. This entails in particular that

$$\mathcal{H}_1(f),\ \ \mathcal{H}_2(f),\ \ \mathcal{H}_3\ \ \textit{are mutually independent.}$$

Remark 6.1. The important difference between (6.3), (6.4) on the one hand and (6.5), (6.6) on the other is that, although the $\langle \mathcal{G}_{\Sigma} p_j,\ \ p_k\rangle$ are always normal, they are independent if (2.7) is satisfied.

Let us now apply these results to testing hypotheses. In [8] the simple hypothesis $f = f_0$ was reduced to $f = 0$, which in turn was seen to be equivalent with $f_\delta = 0$. Assuming just (2.6), that the $p_1, p_2, \ldots$are known, we will here more generally test the null hypothesis

$$H_0: \qquad f \in \mathbb{M} = [\![p_1, p_2, \ldots, p_m]\!], \qquad m \in \mathbb{N}. \tag{6.8}$$

This seems rather natural in the present situation. It is clear from (4.2) that we have

$$H_0: \qquad f \in \mathbb{M} \iff f_\delta \in \mathbb{M} \iff \mathcal{P}_{\mathbb{M}}^{\perp} f_\delta = 0, \tag{6.9}$$

where $\mathcal{P}_{\mathbb{M}}$, $\mathcal{P}_{\mathbb{M}}^{\perp}$ denote the orthogonal projection onto $\mathbb{M}$ and $\mathbb{M}^{\perp}$ respectively.

Let us note that

$$f \in \mathbb{M} \iff \langle f, \ p_k \rangle = 0, \ \forall\, k \geq m+1, \tag{6.10}$$

$$\mathcal{P}_{\mathbb{M}}^{\perp}\, p_k = 0, \ \ \forall k \leq m. \tag{6.11}$$

This entails that

$$\begin{aligned}
&\mathcal{P}_{\mathbb{M}}^{\perp}\, \mathcal{H}_2(f) \\
&\quad = \mathcal{P}_{\mathbb{M}}^{\perp}\left(\sum_{k=1}^{\infty} \frac{\sqrt{2}\delta\sigma_k^2}{(\delta+\sigma_k^2)^2}\, Z_k\, \langle f, \ p_k\rangle p_k \right) \\
&\quad = \sum_{k=1}^{m} \frac{\sqrt{2}\delta\sigma_k^2}{(\delta+\sigma_k^2)^2}\, Z_k\, \langle f, \ p_k\rangle \mathcal{P}_{\mathbb{M}}^{\perp}\, p_k + \sum_{k=m+1}^{\infty} \frac{\sqrt{2}\delta\sigma_k^2}{(\delta+\sigma_k^2)^2}\, Z_k\, \langle f, \ p_k\rangle \mathcal{P}_{\mathbb{M}}^{\perp}\, p_k \\
&\quad = 0, \qquad \forall f \in \mathbb{M}.
\end{aligned}$$

It is now immediate from (6.9), Theorem 6.1 and the continuous mapping theorem that

$$\begin{aligned}
n\left\| \mathcal{P}_{\mathbb{M}}^{\perp}\, \hat{\hat{f}}_\delta \right\|^2 & \\
&= n\left\| \mathcal{P}_{\mathbb{M}}^{\perp}\, (\hat{\hat{f}}_\delta - f_\delta) \right\|^2 \xrightarrow{d} \left\| \mathcal{P}_{\mathbb{M}}^{\perp}\, \Big(\mathcal{H}_1(f) + \mathcal{H}_3 \Big) \right\|^2, \qquad \text{as } n \to \infty,
\end{aligned} \tag{6.12}$$

where the l.h.s of (6.12) can be employed as a test statistic for testing H_0 in (6.9). Again, we can be more specific about the limiting distribution if we assume (2.7).

Theorem 6.4. *If (2.7) is satisfied we have*

$$n\left\| \mathcal{P}_{\mathbb{M}}^{\perp}\, \hat{\hat{f}}_\delta \right\|^2 \xrightarrow{d} \left(v^2 + \sum_{j=1}^{m} \sigma_j^2 \langle f, p_j\rangle^2 \right) \sum_{k=m+1}^{\infty} \frac{\sigma_k^2}{(\delta+\sigma_k^2)^2} U_k^2, \tag{6.13}$$

as $n \to \infty$ where $U_1, U_2, \ldots,$ are i.i.d. Normal(0,1).

Proof. It is immediate from (6.5) and (6.7) that

$$\mathcal{P}_{\mathbb{M}}^{\perp}\, \mathcal{H}_1(f) = \sum_{k=m+1}^{\infty} \frac{\sigma_k}{\delta+\sigma_k^2}\, V_k\, p_k,$$

where the V_k are independent and

$$V_k = \sum_{j=1}^{m} \sigma_j\, Z_{j,k} \langle f, p_j\rangle \stackrel{d}{=} \text{Normal}\left(0, \sum_{j=1}^{m} \sigma_j^2\, \langle f, p_j\rangle^2 \right),$$

$$\mathcal{P}_{\mathbb{M}}^{\perp}\, \mathcal{H}_3 = \sum_{k=m+1}^{\infty} \frac{\sigma_k}{\delta+\sigma_k^2}\, W_k\, p_k,$$

where the W_k are mutually independent and independent of the V_k and

$$W_k \ = \ v\, Z_k' \stackrel{d}{=} \text{Normal}(0, v^2).$$

The result follows by observing that

$$V_k + W_k = \left(v^2 + \sum_{j=1}^{m} \sigma_j^2 \langle f, p_j\rangle^2 \right)^{1/2} U_k.$$

□

Let $F(x; f, v^2, \sigma_1^2, \sigma_2^2, \ldots)$, $x \geq 0$ denote the c.d.f. of the random variable on the right of (6.13). Consistent estimators for the unknown parameters are ($\delta_n \downarrow 0, as\ n \to \infty$)

$$\hat{\hat{f}}_{\delta_n}, \quad \hat{v}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_i - \langle X_i - \bar{X}, \hat{\hat{f}}_{\delta_n}\rangle\right)^2,$$

and the $\hat{\sigma}_k^2$ as in (3.1). Let us write $\hat{F}(x)$, $x \geq 0$ for this c.d.f. when the parameters are replaced with the above estimators.

Theorem 6.5. *An approximate level $\alpha \in (0,1)$ test for testing H_0 in (6.8) rejects the null hypothesis for $n\|\mathcal{P}_{\mathbb{M}}^{\perp}\hat{\hat{f}}_\delta\|^2 > \hat{F}^{-1}(1-\alpha)$.*

6.2. Asymptotics under local alternatives

Suppose that

$$f_n = f + \frac{1}{\sqrt{n}}g, \qquad f, g \in \mathbb{H}. \tag{6.14}$$

For such f_n only minor changes in the asymptotics are required, because the conditions on the X_i and ϵ_i still are the same and don't depend on n. In the notation of (4.2) we have

$$f_{n,\delta} = f_\delta + \frac{1}{\sqrt{n}}g_\delta, \tag{6.15}$$

where $f_\delta = (\delta I + \Sigma)^{-1}\Sigma f$, $g_\delta = (\delta I + \Sigma)^{-1}\Sigma g$. The following is immediate from a minor modification of Theorem 6.1.

Theorem 6.6. *Under the local alternative in (6.14) we have*

$$\sqrt{n}(\hat{\hat{f}}_\delta - f_\delta) \xrightarrow{d} g_\delta + \mathcal{H}_1(f) + \mathcal{H}_2(f) + \mathcal{H}_3, \qquad as\ \ n \to \infty\ \ in\ \ \mathbb{H},$$

where $\mathcal{H}_1(f) = \varphi_1(\Sigma)(\mathcal{P}^{\perp}\mathcal{G}_\Sigma)f$, $\mathcal{H}_2(f) = (\dot{\varphi}_{2,\Sigma}(\mathcal{P}\mathcal{G}_\Sigma))f$, $\mathcal{H}_3$ is the same as in (6.2), and $\mathcal{H}_3 \perp\!\!\!\perp (\mathcal{H}_1(f), \mathcal{H}_2(f))$.

Let us apply this result to derive the power of the test described in Theorem 6.5 against local alternatives. For this purpose suppose that

$$f \in \mathbb{M}, \qquad g \perp \mathbb{M}, \tag{6.16}$$

so that $f_\delta \in \mathbb{M}$ and $P_{\mathbb{M}}^{\perp}f_\delta = 0$, and $g_\delta \perp \mathbb{M}$, and hence $P_{\mathbb{M}}^{\perp}g_\delta = g_\delta$.

Theorem 6.7. *Under the local alternatives specified in (6.15) and (6.16) we have*

$$n\left\|P_{\mathbb{M}}^{\perp}\hat{\hat{f}}_\delta\right\|^2 \xrightarrow{d} \left\|g_\delta + P_{\mathbb{M}}^{\perp}(\mathcal{H}_1(f) + \mathcal{H}_3)\right\|^2, \qquad as\ \ n \to \infty.$$

If (2.7) is satisfied the distribution on the right can be further specified similar to (6.13). Parameters can be estimated as before.

Acknowledgments

The first author is particularly grateful for the opportunity to present this paper at the conference “From Probability to Statistics and Back: High-Dimensional Models and Processes”, in honor of Jon Wellner's 65th birthday. The authors wish to thank Jan Johannes who drew their attention to the special case of stationary regressors which yields examples of covariance operators with known eigenfunctions. They are also thankful to a referee for some useful comments.

References

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

[2] Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91.

[3] Johannes, J. (2008). Privileged communication.

[4] Lax, P. D. (2002). *Functional Analysis.* Wiley-Interscience, New York.

[5] Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory.* Wiley, New York.

[6] Mas, A. and Pumo, B. (2006). Functional linear regression with derivatives. *J. Nonpar. Statist.* **21** 19–40.

[7] Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd edn, Springer, New York.

[8] Ruymgaart, F., Wang, J., Wei, S.-H. and Yu, L. (2009). Some asymptotic theory for functional regression and classification. Tech. Report, Dep. of Math. and Stat., Texas Tech University. Available at http://www.math.ttu.edu/~fruymgaa/gat_for_functional_regression_and_classification.pdf.