Dependent functional linear models with applications to monitoring structural change*

Alexander Aue [1], Siegfried Hörmann [2], Lajos Horváth [3] and Marie Hušková [4]

[1] Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[2] Départment de Mathématique, Université Libre de Bruxelles, CP 215, Boulevard du Triomphe, B-1050 Bruxelles, Belgium

[3] Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City,UT 84112-0090, USA

[4] Charles University of Prague, Department of Statistics, Sokolovská 83, CZ – 186 00 Prague, Czech Republic

Abstract

We study sequential monitoring procedures that detect instabilities of the regression operator in an underlying (fully) functional regression model allowing for dependence. These open-end and closed-end procedures are built on a functional principal components analysis of both the predictor and response functions, thus giving rise to multivariate detector functions, whose fluctuations are compared against a curved threshold function. The main theoretical result of the paper quantifies the large-sample behavior of the procedures under the null hypothesis of a stable regression operator. To establish these limit results, classical results on functional principal components analysis are generalized to a dependent setting, which may be of interest in its own sake. In an accompanying empirical study we illustrate the finite sample properties, while an application to environmental data highlights practical usefulness. To the best of our knowledge this is the first paper that combines sequential with functional data methodology.

1 Introduction

Let $\mathbb{N}$ denote the set of positive integers. We are concerned with the functional linear model

$$Y_k(t) = \int_0^1 X_k(s)\Psi_k(s,t)ds + \varepsilon_k(t), \qquad t \in [0,1],\, k \in \mathbb{N}, \tag{1.1}$$

where Y_k, X_k, ε_k are random functions defined on a compact interval, which we assume to be $[0,1]$, without loss of generality, and Ψ_k is a possibly time dependent (in k) sequence of operators. More specific assumptions are imposed in the next section. These are general enough to include the important case of functional autoregressive processes.

*This research was partially supported by NSF grants DMS 0905400, DMS 1209226 and DMS 1305858, grant GACR 201/09/J006, grant GACR P201/12/1277, Banque nationale de Belgique and Communauté française de Belgique—Actions de Recherche Concertées (2010–2015) and IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). An earlier draft of this paper has been circulated under the title "Sequential stability test for functional linear models".

Scientific and economic theories are often based directly on linear models or on approximations of a more complex truth by means of linear models. In order to apply regression techniques, the operators Ψ_k are typically required to be identical for all k. In this paper, we propose a sequential monitoring procedure for functional linear models that aims at checking the constancy of the Ψ_k. The sequential setting may be more appropriate whenever a stochastic process is to be monitored in real time and a decision made promptly in order to adjust procedures to potential instabilities in the observed system. Examples include financial markets, where stock prices, volatilities of speculative assets, and exchange rates are monitored with high frequency so that daily curves or functional observations naturally arise in this context. Related empirical work includes Cyree, Griffith, and Winters (2004) and Elazović (2009). Another area of potential applications is stochastic process control, where one monitors outputs of various industrial processes. A summary of recent developments may be found in Stoumbos et al. (2000). In this paper, we focus on an environmental data set with half-hourly observations measuring the concentration of particulate matter with an aerodynamic diameter of less than $10\mu m$ (PM10) in ambient air at Klagenfurt, Austria. Since epidemiological and toxicological studies have pointed to negative health effects, European Union (EU) regulation sets pollution standards for the level of the concentration. Policy makers have to ensure compliance with these EU rules and need reliable statistical tools to determine, and justify to the public, appropriate measures such as partial traffic regulation (see Stadlober, Hörmann, and Pfeiler (2008)). We propose a fully functional linear model to monitor PM10 concentration in Section 4.4 below.

To see whether or not the functional linear relationship suggested by (1.1) holds with the same operator for all observations, we test the null hypothesis

$$H_0: \quad \Psi \equiv \Psi_1 = \Psi_2 = \dots \tag{1.2}$$

against the general alternative

$$\begin{aligned} H_A: \quad &\text{There is } k^* \geq 1 \text{ such that } \Psi \equiv \Psi_1 = \dots = \Psi_{m+k^*-1} \\ &\qquad\qquad \text{but } \Psi^* \equiv \Psi_{m+k^*} = \Psi_{m+k^*+1} = \dots \end{aligned}$$

with $\Psi \neq \Psi^*$. The formulation of the hypotheses implicitly ensures the existence of a calibration sample of size m for which the linear relationship (1.1) is stable ($\Psi \equiv \Psi_1 = \dots = \Psi_m$). The calibration sample can be used to adjust the monitoring procedures to the null model. It also allows us to express limit results in the form $m \to \infty$.

The functional observations Y_k and X_k are infinite dimensional objects and can therefore not be entirely recovered based on a finite number of observations. For statistical estimation, we settled on using functional principal components analysis to compress the data. Moreover, since we aim to apply our methodology to data sampled sequentially in time (such as daily stock price curves), we allow the bivariate sequence (X_k, ε_k) to be a functional time series. How to establish in this context the functional principal components analysis under a set of nonrestrictive conditions is discussed in Section 2. The functional principal components analysis is performed with a fixed number of projections for both predictor and response function. To modify the analysis to allow for an increasing

number of projections is a delicate question, as one would need to re-establish the results of the present paper in a triangular array setting. While both interesting and challenging, pursuing this task is beyond the scope of this article. We refer here only to the recent paper Fremdt et al. (2013).

The rest of this paper is organized as follows. Section 3 contains the definition of the monitoring procedure which consists of a stopping rule that rejects the null if a suitably constructed detector crosses a (curved) threshold function. The large sample properties of the monitoring procedure are derived. The main difficulty in establishing limit theory is that the detector is based on asymptotically biased estimators, thereby introducing additional components in the limit that need careful balancing. Finite sample properties and an application to environmental data are presented in Section 4 , while proofs and additional technical results are given in Sections 5 and 6.

We list at the end of this section some general conventions and fix notation. All random elements are defined on a common probability space $(\Omega, \mathcal{A}, P)$. We henceforth often simply write X for a functional variable $(X(t)\colon t \in [0,1])$. All random functions are assumed to have realizations in $L^2 = L^2([0,1], \mathcal{B}[0,1], \lambda)$, the (real) Hilbert space of square integrable functions with respect to Lebesgue measure and with inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$. The latter determines the L^2-norm $\|f\| = (\langle f, f\rangle)^{1/2}$. If X is a random element with values in L^2 and if $\|X\|$ is in $L^p(\Omega, \mathcal{A}, P)$, then in case $p \geq 1$, we write $\nu_p(X) = (E[\|X\|^p])^{1/p}$ for the L^p-norm. We use $E[X]$ to abbreviate $(E[X(t)]\colon t \in [0,1])$ and, if $E[X] = 0$, the right-hand side is meant to be the zero element in L^2. All random functionals are assumed to be measurable, which implies for example that terms such as $\int_0^1 E[|X(t)|]dt$ make sense and can be computed with the help of Fubini's theorem. Another convention and slight abuse of notation is that we identify X with Y, denoted $X(t) = Y(t)$, if $\|X - Y\| = 0$. We refer the interested reader to Bosq (2000) and Gohberg, Golberg, and Kaashoek (1990) for relevant background on random elements in functions spaces and functional analysis, respectively.

2 Estimation in dependent functional linear models

In this section, we introduce functional principal component analysis in the dependent setting and introduce the biased estimators that are used in the definition of the monitoring procedure to be introduced in Section 3. The literature for linear models with functional covariates and functional responses under independence assumptions has been discussed by a number of authors. Practical and computational aspects are, for example, discussed in Ramsay and Silverman (2005), while asymptotic consistency results have been obtained by Cuevas, Febrero, and Fraiman (2002). Generalizing the existing contributions, the first assumption quantifies the notion of dependence that is allowed in our setting. Let $\mathbb{Z}$ denote the set of integers.

Assumption 2.1. *$((X_k, \varepsilon_k)\colon k \in \mathbb{Z})$ is a stationary and ergodic sequence satisfying $E[X_1] = 0$, $E[\varepsilon_1] = 0$, X_k and ε_k are pairwise independent, and $\nu_{2\mu}(X_1) + \nu_{2\mu}(\varepsilon_1) < \infty$ for some $\mu > 2$.*

Starting with the predictor functions X_k, it follows readily from Assumption 2.1 that

$$C(s,t) = E[X_k(s)X_k(t)]$$

exists and is independent of $k \in \mathbb{Z}$. The eigenvalues, say, $\lambda_1 \geq \lambda_2 \geq \ldots$ of C are necessarily non-negative, and we denote the corresponding orthonormal eigenfunctions by $v_1, v_2, \ldots$. Projecting the X_k onto the subspace spanned by $v_1, \ldots, v_p$, we can expect to explain the major contributions to the variation in the X_k provided that p is selected accordingly. In applications, C and its p largest eigenvalues $\lambda_1, \ldots, \lambda_p$ along with their eigenfunctions $v_1, \ldots, v_p$ are unknown and need to be estimated from the available data. Utilizing the calibration sample, we choose to work with

$$\hat{C}_m(s,t) = \frac{1}{m} \sum_{k=1}^{m} X_k(s) X_k(t)$$

for the statistical inference. It is clear that, for any $s, t \in [0,1]$, $\hat{C}_m(s,t)$ is an unbiased estimator of $C(s,t)$. Let $\hat{\lambda}_{1,m} \geq \hat{\lambda}_{2,m} \geq \ldots \geq \hat{\lambda}_{p,m}$ denote the p largest eigenvalues of $\hat{C}_m$ and $\hat{v}_{1,m}, \hat{v}_{2,m} \ldots, \hat{v}_{p,m}$ the corresponding orthonormal eigenfunctions. While the projections of X_k into the subspaces spanned by $\hat{v}_{1,m}, \ldots, \hat{v}_{p,m}$ or $v_1, \ldots, v_p$ explain a large part of the randomness in the predictors, they may reveal little about the variation in the response functions Y_k. Let

$$D(s,t) = E[Y_k(s) Y_k(t)],$$

and denote its eigenvalues and orthonormal eigenfunctions by $\tau_1 \geq \tau_2 \geq \ldots$ and $w_1, w_2, \ldots$, respectively. Estimation is then performed with the empirical counterpart

$$\hat{D}_m(s,t) = \frac{1}{m} \sum_{k=1}^{m} Y_k(s) Y_k(t)$$

on the calibration sample. The q largest eigenvalues of $\hat{D}_m$ and their orthonormal eigenfunctions are denoted by $\hat{\tau}_{1,m} \geq \hat{\tau}_{2,m} \geq \ldots \geq \hat{\tau}_{q,m}$ and $\hat{w}_{1,m}, \hat{w}_{2,m}, \ldots, \hat{w}_{q,m}$. We make the following standard assumption (see Bosq (2000)).

Assumption 2.2. *The p largest eigenvalues of the covariance operator C satisfy $\lambda_1 > \ldots > \lambda_p > \lambda_{p+1}$, and the q largest eigenvalues of the covariance operator D satisfy $\tau_1 > \ldots > \tau_q > \tau_{q+1}$.*

Thus, up to signs, $v_1, \ldots, v_p$ and $w_1, \ldots, w_q$ are uniquely defined.

Assumption 2.3. *The Hilbert-Schmidt norm $\|\cdot\|_{\mathcal{S}}$ of the operator Ψ satisfies*

$$\|\Psi\|_{\mathcal{S}}^2 = \int_0^1 \int_0^1 \psi^2(s,t) ds dt < \infty,$$

where ψ denotes some kernel function in $L^2([0,1]^2, \mathcal{B}[0,1]^2, \lambda^2)$.

Since the product functions $(v_i w_j : i, j \in \mathbb{N})$ constitute an orthonormal basis of $L^2([0,1]^2, \mathcal{B}[0,1]^2, \lambda^2)$ we get the L^2-expansion

$$\Psi(s,t) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \psi_{i,j} v_i(s) w_j(t)$$

with square-summable coefficients $(\psi_{i,j}\colon i,j\in\mathbb{N})$. Using inner product notation, we get, under the null hypothesis H_0, that

$$\begin{aligned} Y_k(t) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\psi_{i,j}\langle X_k, v_i\rangle w_j(t) + \varepsilon_k(t) \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q}\psi_{i,j}\langle X_k, v_i\rangle w_j(t) + e_k(t), \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} e_k(t) &= \varepsilon_k(t) + \sum_{i=p+1}^{\infty}\sum_{j=1}^{q}\psi_{i,j}\langle X_k, v_i\rangle w_j(t) + \sum_{i=1}^{\infty}\sum_{j=q+1}^{\infty}\psi_{i,j}\langle X_k, v_i\rangle w_j(t) \\ &= \varepsilon_k(t) + \rho_{k,1}(t) + \rho_{k,2}(t). \end{aligned}$$

Let $\zeta_{k,j} = \langle Y_k, w_j\rangle$ and $\xi_{k,i} = \langle X_k, v_i\rangle$ be the principal component scores, and take $\eta_{k,j} = \langle e_k, w_j\rangle$. Projecting both sides of (2.1) onto the subspace spanned by $w_1,\dots,w_q$, we obtain the multivariate regression model

$$\boldsymbol{\zeta}_k = W\boldsymbol{\xi}_k + \boldsymbol{\eta}_k, \tag{2.2}$$

where $\boldsymbol{\zeta}_k = (\zeta_{k,1},\dots,\zeta_{k,q})'$, $\boldsymbol{\xi}_k = (\xi_{k,1},\dots,\xi_{k,p})'$, $\boldsymbol{\eta}_k = (\eta_{k,1},\dots,\eta_{k,q})'$ are principal component scores vectors, and $W = (\psi_{i,j}\colon i = 1,\dots,p,\ j = 1,\dots,q)$. Here and in the following $'$ signifies transposition. Through (2.2) we have transformed the testing problem to a problem that checks whether or not the $q\times p$ matrix W remains constant. Thus, we monitor only the most important (in terms of variation in both the predictor and response functions) coefficients $\psi_{i,j}$ in the expansion of Ψ. Note that the regression in (2.2) is nonstandard, since the regressors $\boldsymbol{\xi}_k$ are random and possibly correlated with the errors $\boldsymbol{\eta}_k$.

The regression (2.2) cannot be used directly for statistical inference on W. Since the eigenfunctions $v_1,\dots,v_p$ and $w_1,\dots,w_q$ are unknown, one has to work with the sample eigenfunctions $\hat{v}_{1,m},\dots,\hat{v}_{p,m}$ and $\hat{w}_{1,m},\dots,\hat{w}_{q,m}$ instead. The latter are unique only up to random signs and one can merely expect that $\hat{v}_{i,m}$ $(\hat{w}_{j,m})$ is close $\hat{c}_{i,m}v_i$ $(\hat{d}_{j,m}w_j)$, where $\hat{c}_{i,m} = \operatorname{sign}\langle\hat{v}_{i,m}, v_i\rangle$ $(\hat{d}_{j,m} = \operatorname{sign}\langle\hat{w}_{j,m}, w_j\rangle)$. Hence (2.1) can be replaced with

$$Y_k(t) = \sum_{i=1}^{p}\sum_{j=1}^{q}\hat{c}_{i,m}\psi_{i,j}\hat{d}_{j,m}\langle X_k, \hat{v}_{i,m}\rangle\hat{w}_{j,m}(t) + \hat{e}_k(t), \tag{2.3}$$

where the new error functions can be decomposed as

$$\begin{aligned} \hat{e}_k(t) = e_k(t) &+ \sum_{i=1}^{p}\sum_{j=1}^{q}\psi_{i,j}\langle X_k, v_i - \hat{c}_{i,m}\hat{v}_{i,m}\rangle w_j(t) \\ &+ \sum_{i=1}^{p}\sum_{j=1}^{q}\psi_{i,j}\langle X_k, \hat{c}_{i,m}\hat{v}_{i,m}\rangle\big[w_j(t) - \hat{d}_{j,m}\hat{w}_{j,m}(t)\big] \end{aligned}$$

$$= \varepsilon_k(t) + \rho_{k,1}(t) + \rho_{k,2}(t) + \phi_{k,1}^{(m)}(t) + \phi_{k,2}^{(m)}(t). \tag{2.4}$$

The residual term $\hat{e}_k$ exhibits the different sources of errors: the original noise in the system given by the innovations ε_k; the error occurring through replacement of predictor and response functions with their finite dimensional projections (captured in $\rho_{k,1}$ and $\rho_{k,2}$, which do not have an additional m superscript); the error coming from the deviations of the sample eigenfunctions from their population counterparts (captured in the remaining terms $\phi_{k,1}^{(m)}$ and $\phi_{k,2}^{(m)}$).

Dropping the dependence on m, we write $\hat{v}_i = \hat{v}_{i,m}$, $\hat{w}_j = \hat{w}_{j,m}$, $\hat{c}_i = \hat{c}_{i,m}$, and $\hat{d}_j = \hat{d}_{j,m}$ in the following. Let $\hat{\zeta}_{k,i} = \langle Y_k, \hat{w}_j \rangle$ and $\hat{\xi}_{k,i} = \langle X_k, \hat{v}_i \rangle$ and $\hat{\eta}_{k,j} = \langle \hat{e}_k, \hat{w}_j \rangle$. Using the projection arguments as in (2.2), we are led to the new multivariate regression

$$\hat{\boldsymbol{\zeta}}_k = \hat{W}\hat{\boldsymbol{\xi}}_k + \hat{\boldsymbol{\eta}}_k, \tag{2.5}$$

where, in analogy to (2.2), $\hat{\boldsymbol{\zeta}}_k = (\hat{\zeta}_{k,1}, \ldots, \hat{\zeta}_{k,q})'$, $\hat{\boldsymbol{\xi}}_k = (\hat{\xi}_{k,1}, \ldots, \hat{\xi}_{k,p})'$, $\hat{\boldsymbol{\eta}}_k = (\hat{\eta}_{k,1}, \ldots, \hat{\eta}_{k,q})'$ and $\hat{W} = (\hat{c}_i \psi_{i,j} \hat{d}_j \colon i = 1, \ldots, p,\ j = 1, \ldots, q)$. The signs $\hat{c}_1, \ldots, \hat{c}_p$ and $\hat{d}_1, \ldots, \hat{d}_q$ are fixed, so that our testing procedure can be reduced to checking the stability of $\hat{W}$. To this end, let

$$\boldsymbol{\beta} = \operatorname{vec}\bigl(\hat{W}'\bigr), \tag{2.6}$$

where the vec-operator creates a column vector by stacking the columns of a matrix below one another. Our monitoring procedure is based on estimating $\boldsymbol{\beta}$ via least squares on the calibration sample $(Y_1, X_1), \ldots, (Y_m, X_m)$, and to compare the resulting calibration to least squares estimators $\tilde{\boldsymbol{\beta}}_\ell$ computed from the samples $(Y_{m+1}, X_{m+1}), \ldots, (Y_{m+\ell}, X_{m+\ell})$ as ℓ increases. A violation of H_0 should be visible in statistically significant differences of these least squares estimators.

We introduce further notation. For $n < N$, let $\hat{\boldsymbol{Z}}_{n,N} = (\hat{\boldsymbol{\zeta}}_{n+1}', \ldots, \hat{\boldsymbol{\zeta}}_N')'$, $\hat{\boldsymbol{E}}_{n,N} = (\hat{\boldsymbol{\eta}}_{n+1}', \ldots, \hat{\boldsymbol{\eta}}_N')'$, and $\hat{\Xi}_{n,N} = (\hat{\Xi}_{n+1}', \ldots, \hat{\Xi}_N')'$ with $\hat{\Xi}_k = I_q \otimes \hat{\boldsymbol{\xi}}_k'$, where $\otimes$ is the Kronecker product and I_q the $q \times q$ identity matrix. Then the regression in (2.5) for the variables $(Y_{n+1}, X_{n+1}), \ldots, (Y_N, X_N)$ is

$$\hat{\boldsymbol{Z}}_{n,N} = \hat{\Xi}_{n,N}\boldsymbol{\beta} + \hat{\boldsymbol{E}}_{n,N}, \tag{2.7}$$

for which the least squares estimator is

$$\boldsymbol{\beta}_{n,N} = (\hat{\Xi}_{n,N}'\hat{\Xi}_{n,N})^{-1}\hat{\Xi}_{n,N}'\hat{\boldsymbol{Z}}_{n,N}. \tag{2.8}$$

We use the notation $\hat{\boldsymbol{\beta}}_m = \boldsymbol{\beta}_{0,m}$ and $\tilde{\boldsymbol{\beta}}_\ell = \boldsymbol{\beta}_{m,m+\ell}$ to denote the least squares estimator of the calibration sample and the monitoring period, respectively. In addition to the dependence of the regressors $\hat{\Xi}_{n,N}$ and the errors $\hat{\boldsymbol{E}}_{n,N}$, $\boldsymbol{\beta}_{n,N}$ is no longer an unbiased estimator. This introduces additional technical challenges.

We allow for dependent stationary and ergodic sequences $(X_k \colon k \in \mathbb{Z})$ and $(\varepsilon_k \colon k \in \mathbb{Z})$, and further specify their structure.

Assumption 2.4. *There are functionals $a \colon S_1^\infty \to L^2$ and $b \colon S_2^\infty \to L^2$ such that*

$$X_k = a(\omega_k, \omega_{k-1}, \ldots) \qquad \text{and} \qquad \varepsilon_k = b(\delta_k, \delta_{k-1}, \ldots),$$

where $(\omega_k \colon k \in \mathbb{Z})$ and $(\delta_k \colon k \in \mathbb{Z})$ are independent, identically distributed (iid) sequences of random elements with values in some measurable spaces S_1 and S_2, respectively.

Assumption 2.4 states that X_k and ε_k are Hilbert space valued Bernoulli shifts. We can assume that the innovations ω_k and δ_k take values in some general space, usually $S_1 = S_2 = L^2$. We admit only weakly dependent random processes here.

Assumption 2.5. *There are $c_0 \geq 0$ and $\kappa > 2, \mu > 2$ such that, for all $\ell \in \mathbb{N}$,*

$$\nu_{2\mu}(X_k - X_k^{(\ell)}) \leq c_0 \ell^{-\kappa}, \tag{2.9}$$

$$\nu_{2\mu}(\varepsilon_k - \varepsilon_k^{(\ell)}) \leq c_0 \ell^{-\kappa}, \tag{2.10}$$

with

$$X_k^{(\ell)} = a(\omega_k, \omega_{k-1}, \ldots, \omega_{k-\ell+1}, \omega_{k-\ell}^{(\ell)}, \omega_{k-\ell-1}^{(\ell)}, \ldots),$$

where $(\omega_k^{(\ell)} \colon k, \ell \in \mathbb{Z})$ are iid copies of ω_0. The $(\varepsilon_k^{(\ell)} \colon k, \ell \in \mathbb{Z})$ are defined analogously.

From Assumption 2.4, the norms in (2.9) and (2.10) do not depend on k . Assumption 2.5 states that (X_k, ε_k) can be approximated with the ℓ-dependent sequences $(X_k^{(\ell)}, \varepsilon_k^{(\ell)})$, $k \in \mathbb{Z}$, and this approximation improves at rate $\ell^{-\kappa}$ as ℓ increases. This method to describe dependence is in the spirit of Ibragimov (1962) and Billingsley (1968). The crucial idea is to approximate the original process with auxiliary processes whose asymptotics are known. This approach was recently used in Aue et al. (2009a) and Hörmann and Kokoszka (2010) to examine, respectively, multivariate time series and functional time series. The latter paper and Aue et al. (2012) also contain a number of examples of processes that satisfy Assumptions 2.4 and 2.5, these notably including functional autoregressive processes.

We have a result on the long-run behavior of $\hat{\boldsymbol{\beta}}_m$ and $\tilde{\boldsymbol{\beta}}_\ell$.

Theorem 2.1. *If Assumptions 2.1–2.5 are satisfied, then it holds under H_0 that*

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_m &= \boldsymbol{\beta} + \boldsymbol{\Delta}_{m,1} + \boldsymbol{R}_m + O_p\left(m^{-1}\right) \\
\tilde{\boldsymbol{\beta}}_\ell &= \boldsymbol{\beta} + \boldsymbol{\Delta}_{m,\ell,1} + \boldsymbol{R}_m + O_p\left(m^{-1} + (\log \ell)^4 \ell^{-1}\right),
\end{aligned}$$

where $\sqrt{m}\hat{B}\boldsymbol{\Delta}_{m,1}$ and $\sqrt{\ell}\hat{B}\boldsymbol{\Delta}_{m,\ell,1}$ are asymptotically normal ($m, \ell \to \infty$), with $\hat{B}$ a matrix depending only on the random signs $\hat{c}_i$ and $\hat{d}_j$, $|\boldsymbol{R}_m| = O_P(m^{-1/2})$.

Thus the bias of our estimators, caused by the estimation of eigenfunctions, is not negligible. A common part of the bias for $\hat{\boldsymbol{\beta}}_m$ and $\tilde{\boldsymbol{\beta}}_\ell$ is captured in the vector $\boldsymbol{R}_m$ and depends on the differences $\hat{v}_i - \hat{c}_i v_i$ and $\hat{w}_j - \hat{d}_j w_j$. Even if this bias is taken into account, the differences $\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}$ are not asymptotically normal because the estimation of the eigenfunctions is unique only up to the random signs. To accommodate these findings, our functional monitoring procedure is based on a CUSUM-type detector, for which the bias of Theorem 2.1 and the random signs cancel.

3 Functional monitoring

In this section, we give the monitoring procedure and state its large sample properties. We define a detector function that measures the variation in suitably normalized differences $\tilde{\boldsymbol{\beta}}_\ell - \hat{\boldsymbol{\beta}}_m$ for increasing

ℓ. Under H_0 these fluctuations are said to be in control. Significant deviations from the in-control level that lead to a rejection of H_0 are quantified in terms of a threshold function. The approach we take is in the spirit of Chu et al. (1996), a paper which has been the foundation for many subsequent developments in the sequential monitoring of stochastic processes. Contributions to the area include, but are not limited to, Leisch, Hornik, and Kuan (2000), Zeileis et al. (2005), Aue et al. (2006), and Hušková, Prašková, and Steinebach (2007). A recent survey of research in the area is given in Aue and Horváth (2013).

To begin, by (2.7) and (2.8) we have

$$\boldsymbol{\beta}_{n,N} = \left(\hat{\Xi}'_{n,N}\hat{\Xi}_{n,N}\right)^{-1}\hat{\Xi}'_{n,N}\hat{\boldsymbol{Z}}_{n,N} = \boldsymbol{\beta} + \left(\hat{\Xi}'_{n,N}\hat{\Xi}_{n,N}\right)^{-1}\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N}, \tag{3.1}$$

and thus we need to investigate the large sample behavior of $\hat{\Xi}'_{n,N}\hat{\Xi}_{n,N}$ and $\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N}$. We show in Lemma 6.1 that $\frac{1}{N-n}\hat{\Xi}'_{n,N}\hat{\Xi}_{n,N}$ is close to the matrix $Q = I_q \otimes \Lambda$, with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. So roughly speaking, if $\hat{Q}_m$ is an estimator for Q obtained from the calibration sample, then $\hat{Q}_m(\hat{\boldsymbol{\beta}}_m - \tilde{\boldsymbol{\beta}}_\ell)$ is close to

$$\frac{1}{m}\hat{\Xi}'_{0,m}\hat{\boldsymbol{E}}_{0,m} - \frac{1}{\ell}\hat{\Xi}'_{m,m+\ell}\hat{\boldsymbol{E}}_{m,m+\ell}. \tag{3.2}$$

Both terms in (3.2) are vector-valued partial sums processes. The summands are of the form

$$\hat{\boldsymbol{g}}_k = (\hat{\xi}_{k,1}\hat{\eta}_{k,1}, \ldots, \hat{\xi}_{k,p}\hat{\eta}_{k,1}, \ldots, \hat{\xi}_{k,1}\hat{\eta}_{k,q}, \ldots, \hat{\xi}_{k,p}\hat{\eta}_{k,q})'.$$

To derive asymptotic normality we have to take their dependence into account in the standardization. Let $\boldsymbol{\gamma}_k = \mathrm{vec}(\Gamma_k)$, where the matrices $\Gamma_k = (\Gamma_k(i,j)\colon i = 1, \ldots, p, j = 1, \ldots, q)$ are

$$\Gamma_k(i,j) = \xi_{k,i}\eta_{k,j} = \xi_{k,i}(\eta^*_{k,j} + \theta_{k,j}), \tag{3.3}$$

where $\eta^*_{k,j} = \langle \varepsilon_k, w_j \rangle$, $\theta_{k,j} = \langle X_k, u_j \rangle$, and

$$u_j = \sum_{i=p+1}^{\infty} \psi_{i,j} v_i. \tag{3.4}$$

Thus the $\boldsymbol{\gamma}_k$ are the theoretical counterparts of $\hat{\boldsymbol{g}}_k$, with $\hat{v}_i$ and $\hat{w}_j$ replaced by the true eigenfunctions v_i and w_j. Since $E[\xi_{k,i}\theta_{k,j}] = \lambda_i(\psi_{p+1,j}\langle v_i, v_{p+1}\rangle + \psi_{p+2,j}\langle v_i, v_{p+2}\rangle + \ldots) = 0$ and $E[\xi_{k,i}\eta^*_{k,j}] = 0$, each $\Gamma_k(i,j)$ is centered. Let

$$\Sigma = E[\boldsymbol{\gamma}_0\boldsymbol{\gamma}'_0] + \sum_{k=1}^{\infty} E[\boldsymbol{\gamma}_0\boldsymbol{\gamma}'_k + \boldsymbol{\gamma}_k\boldsymbol{\gamma}'_0].$$

To establish an estimator (up to random signs) for Σ, also based on the calibration sample, let

$$\hat{\varepsilon}_k = Y_k - \sum_{i=1}^{p}\sum_{j=1}^{q} \hat{\psi}_{i,j}\hat{\xi}_{k,i}\hat{w}_j$$

denote the residuals, where $\mathrm{vec}((\hat{\psi}_{i,j}\colon i = 1, \ldots, p, j = 1, \ldots, q)') = \hat{\boldsymbol{\beta}}_m$. Then we can define the sample version of Γ_k by $\hat{\Gamma}_k = (\hat{\Gamma}_k(i,j)\colon i = 1, \ldots, p, j = 1, \ldots, q)$ via

$$\hat{\Gamma}_k(i,j) = \hat{\xi}_{k,i}\hat{\eta}^*_{k,j} - \frac{1}{m}\sum_{k'=1}^{m} \hat{\xi}_{k',i}\hat{\eta}^*_{k',j},$$

where $\hat{\eta}^*_{k,j} = \langle \hat{\varepsilon}_k, \hat{w}_j \rangle$. It follows from Lemma 6.2 that $\hat{\Gamma}_k(i,j)$ is close to $\hat{c}_i \hat{d}_j \Gamma_k(i,j)$. Let $\hat{\boldsymbol{\gamma}}_k = \mathrm{vec}(\hat{\Gamma}_k)$. We define the lag-window estimator

$$\hat{\Sigma}_m = \hat{G}_{0,m} + \sum_{k=1}^{b_m} w_{k,m}(\hat{G}_{k,m} + \hat{G}'_{k,m}),$$

where, for $k = 0, 1, \ldots, b_m$,

$$\hat{G}_{k,m} = \frac{1}{m} \sum_{k'=1}^{m-k} \hat{\boldsymbol{\gamma}}_{k'+k} \hat{\boldsymbol{\gamma}}'_{k'}.$$

The $w_{k,m}$ are weights and b_m is a bandwidth that tends to ∞ slowly compared to m. For further information on lag-window estimators and specific examples, such as the Bartlett estimator and the flat-top kernel, we refer to Section 10.4 of Brockwell and Davis (1991). If the weights $w_{k,m}$ and the smoothing parameter b_m are chosen appropriately, it can be shown that $\hat{\Sigma}_m$ is close to $(\hat{d} \otimes \hat{c})\Sigma$, where $\hat{c} = \mathrm{diag}(\hat{c}_1, \ldots \hat{c}_p)$ and $\hat{d} = \mathrm{diag}(\hat{d}_1, \ldots \hat{d}_q)$. Instead of providing explicit conditions for this, we make the following assumption.

Assumption 3.1. $|\hat{\Sigma}_m - (\hat{d} \otimes \hat{c})\Sigma| = o_P(1)$ *as* $m \to \infty$.

Utilizing this, it makes sense to define the detector

$$\hat{V}_\ell = \hat{V}_{\ell,m} = (\tilde{\boldsymbol{\beta}}_\ell - \hat{\boldsymbol{\beta}}_m)' \hat{Q}_m \hat{\Sigma}_m^{-1} \hat{Q}_m (\tilde{\boldsymbol{\beta}}_\ell - \hat{\boldsymbol{\beta}}_m).$$

Due to the quadratic nature of $\hat{V}_\ell$, it is independent of the random signs $\hat{c}_i$ and $\hat{d}_j$. The task is now to compare the detector against a boundary function, so if $h_\ell = h_{\ell,m}$ is the value of the threshold function at lag ℓ, then H_0 is rejected if $\hat{V}_\ell > h_\ell$. Thus one is interested in finding the first (random) time lag, say τ, for which this is the case. Typically τ is referred to as a stopping time and one generally distinguishes between open-end procedures, τ^o, for which monitoring continues as long as H_0 is not rejected, and closed-end procedures, τ^c, for which monitoring is terminated after a pre-specified maximal number of observations have not led to a violation of H_0. In the present setting this amounts to

$$\begin{aligned} \tau^o_m &= \inf\{\ell \in \mathbb{N} \colon \hat{V}_{\ell,m} > h_{\ell,m}\}, \\ \tau^c_m &= \min\{\ell = 1, \ldots, \lfloor mT \rfloor \colon \hat{V}_{\ell,m} > h_{\ell,m}\}, \end{aligned}$$

emphasizing the dependence on the calibration sample size m. Using a parameter $T > 0$, the maximal number $\lfloor mT \rfloor$ of observations obtained before termination is parametrized in terms of m as well. We use the general specification

$$h_\ell = h_{\ell,m} = \frac{cm}{\ell^2} \left(1 + \frac{\ell}{m}\right)^2 g^2\left(\frac{\ell}{m+\ell}\right) \tag{3.5}$$

for the threshold, where $c = c(\alpha)$ is a critical constant ensuring that the monitoring procedure has level $\alpha \in (0,1)$ and $g \colon [0,1] \to \mathbb{R}$ is a function. We assume the following conditions on g to be satisfied.

Assumption 3.2. *(i) g is continuous on $[0,1]$;*
(ii) there are $C>0$ and $\gamma<\frac{1}{2}$ such that $Cx^{\gamma}\le g(x)$ for all $x\in(0,1]$.

Examples of threshold functions can be found in Chu et al. (1996), Horváth et al. (2004) and Aue et al. (2009b), among others. We have the limit behavior of both the open-end and the closed-end procedures under the null hypothesis.

Theorem 3.1. *If Assumptions 2.1–2.5, 3.1, and 3.2 are satisfied, then, under H_0,*

$$\lim_{m\to\infty} P\left(\sup_{\ell\ge 1}\frac{\hat{V}_{\ell,m}}{h_{\ell,m}}>1\right)=P\left(\sup_{0\le x\le 1}\frac{\mathcal{V}(x)}{cg^2(x)}>1\right)$$

and, for any $T>0$,

$$\lim_{m\to\infty} P\left(\max_{1\le\ell\le\lfloor mT\rfloor}\frac{\hat{V}_{\ell,m}}{h_{\ell,m}}>1\right)=P\left(\sup_{0\le x\le T/(1+T)}\frac{\mathcal{V}(x)}{cg^2(x)}>1\right),$$

where

$$(\mathcal{V}(x)\colon x\in[0,1])\overset{\mathcal{D}}{=}\left(\sum_{j=1}^{pq}W_j^2(x)\colon x\in[0,1]\right),$$

with $(W_1(x)\colon x\in[0,1]),\dots,(W_{pq}(x)\colon x\in[0,1])$ denoting independent standard Brownian motions.

Theorem 3.2. *Let Assumptions 2.1–2.5, 3.1, and 3.2 be satisfied. If W^* is analogous to W in (2.2) and H_A holds with $W\neq W^*$, then $P(\tau_m^o<\infty)\to 1$ and $P(\tau_m^c<mT)\to 1$ as $m\to\infty$.*

The proofs of Theorems 3.1 and 3.2 are presented in Section 6. Theorem 3.2 could be formulated so that k^* is allowed to depend on m in a certain way. It would also be of interest to derive a theoretical result on the delay time of the detector as was done in Aue and Horváth (2004) for a simple change in location model, and in Aue et al. (2009b) for more complex time series regressions. However, this is beyond the scope of the current account. A heuristic argument can be provided as follows. First, assume for simplicity that the change-point is at $k^*=1$, the case of arbitrary k^* following from similar reasoning. The detector $\hat{V}_{\ell,m}$ is based on a quadratic form with argument $\hat{\boldsymbol{\beta}}_m-\tilde{\boldsymbol{\beta}}_\ell$. Referring to Theorem 2.1,

$$\hat{\boldsymbol{\beta}}_m-\tilde{\boldsymbol{\beta}}_\ell=\boldsymbol{\delta}+O_P\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{\ell}}\right),$$

where $\boldsymbol{\delta}$ denotes the difference between the pre-change and post-change values of $\boldsymbol{\beta}$. When

$$\|\boldsymbol{\delta}\|\succ\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{\ell}},\tag{3.6}$$

we have that

$$\hat{V}_{\ell,m}=(\hat{\boldsymbol{\beta}}_m-\tilde{\boldsymbol{\beta}}_\ell)'\hat{Q}_m\hat{\Sigma}_m^{-1}\hat{Q}_m(\hat{\boldsymbol{\beta}}_m-\tilde{\boldsymbol{\beta}}_\ell)\approx\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}.$$

Note that $\|\boldsymbol{\delta}\|^2\ll\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}\ll\|\boldsymbol{\delta}\|^2$, since $\|x\|_\Sigma:=\sqrt{x'\Sigma^{-1}x}$ defines a norm in $\mathbb{R}^{pq}$. Now suppose $g(x)=x^\gamma$, with $\gamma<1/2$. Then a change-point is indicated once, for some ℓ,

$$\frac{m\hat{V}_{\ell,m}}{c_\gamma\,(1+m/\ell)^{2(1-\gamma)}}>1.\tag{3.7}$$

The constant c_γ is determined via the limiting result in Theorem 3.1. Assume that $\|\boldsymbol{\delta}\| \gg m^{-\alpha}$, for some $0 \leq \alpha < 1/2$. Further let $1/2 > \alpha' > \alpha$. There are two possibilities: either (3.7) happens with $\ell \leq m^{2\alpha'}$, in which case the delay time is $\leq m^{2\alpha'}$. If the change has not yet been detected for $\ell \leq m^{2\alpha'}$, then we are in situation (3.6), meaning roughly that $\hat{V}_{\ell,m}$ in (3.7) can be replaced by $\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}$. Then, basic manipulations imply that ℓ has to satisfy

$$\ell > m\left[\left(\frac{m\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}}{c_\gamma}\right)^{1/[2(1-\gamma)]} - 1\right]^{-1} \approx \left(c_\gamma \times \frac{m^{1-2\gamma}}{\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta}}\right)^{1/[2(1-\gamma)]}.$$

For example, if $\|\boldsymbol{\delta}\| \gg 1$, one can choose $\alpha' = [1-2\gamma]/[4(1-\gamma)]$. For $\gamma < 1/2$ we see that $0 < \alpha' < 1/2$ as required. It follows that the standardized delay time is proportional to $m^{-1/[2(1-\gamma)]}$. Hence, asymptotically, the closer γ to $1/2$, the faster the change-point is detected. In turn, the risk of a false alarm also increases with γ. The other interesting situation is when α approaches $1/2$. Then, by similar arguments, one can see that the standardized delay time converges to a positive constant.

4 Empirical performance

In this section, we evaluate the finite sample performance of the monitoring procedure with a particular focus on the impact of functional principal components. We do not aim for completeness in this. The section is organized as follows. In Section 4.1, we provide a table with critical values necessary in order to apply the procedure. In Section 4.2, we describe the setting used for the simulations, while Section 4.3 presents and discusses the results of the Monte Carlo study. An application to environmental data is given in Section 4.4, and Section 4.5 offers a brief conclusion.

4.1 Critical values

The limit distribution in Theorem 3.1 depends on the choice of threshold function h, variously the choice of the function $g\colon [0,1] \to \mathbb{R}$, for which only general requirements are formulated in Assumption 3.2. We focus here on the class of functions

$$g(x) = x^\gamma, \qquad x \in [0,1],\ \gamma < 1/2. \tag{4.1}$$

These functions are parameterized in terms of a sensitivity parameter γ which can be adjusted with respect to the specific goal of monitoring. For more details on this see, for example, Aue et al. (2009b). More precisely, critical values are provided for the values $\gamma = -0.5, 0.1, 0.25$, and 0.45 in the following.

The process

$$\mathcal{V}(x) = \sum_{j=1}^{pq} W_j^2(x), \qquad x \in [0,1],$$

was simulated by an approximation of the Brownian motions $(W_j(x)\colon x \in [0,1])$ on a grid of 5,000 equidistant points on $[0,1]$. We chose a number of reasonable values for the upper limit pq which

is related to the number of functional principal components used to evaluate the detector. The last parameter to choose is the length of monitoring quantified in terms of T. We report critical values for $T = 1, 2, 3$ (closed-end procedures) and $T = \infty$ (open-end procedure). The results for the three nominal levels $\alpha = 0.10, 0.05$, and 0.01 are summarized in Table 1, based on 50,000 repetitions.

4.2 Simulation setting

In the simulation study, we used the following specifications for the functional linear model (1.1). The predictor functions were

$$X_k(t) = \sum_{j=1}^{5} \vartheta_{kj} j^{-1/2} \sin(j\pi t), \qquad t \in [0,1], \tag{4.2}$$

with the ϑ_{kj} standard normal. This specification has been used, albeit for the errors, in Gabrys et al. (2010). Here, we generated the errors from the equations

$$\varepsilon_k(t) = \int_0^1 B_k(s)\Psi_E(s,t)ds, \qquad t \in [0,1], \tag{4.3}$$

where $\Psi_E(s,t) = d\cos(2\pi(s+t) - \frac{1}{2})$, with d chosen such that $\|\Psi_E\|_{\mathcal{S}} = 1$, and $(B_k(t) \colon t \in [0,1])$ a standard Brownian bridge. All functions were converted from discrete observations to functional objects represented in a cubic B-spline basis with 20 basis functions (using the R package `fda`). We used as regression operator the banded cosine-taper integral kernel

$$\Psi(s,t) = \begin{cases} d\{1 + \cos(b^{-1}\pi[s-t])\}, & s,t \in [0,1] \colon |s-t| \le b, \\ 0, & \text{else}, \end{cases} \tag{4.4}$$

where the bandwidth parameter b controls which "off-diagonal" values contribute to the regression and d is again used to normalize Ψ. We obtain the response functions from

$$Y_k(t) = \int_0^1 X_k(s)\Psi(s,t)ds + \varepsilon_k(t),$$

as prescribed by (1.1). Different realizations of these functions are displayed in Figure 1 of the supplement Aue et al. (2013). It can be seen there that the bandwidth parameter b controls how close the response is to the predictor. If $b = 0.2$, as in the left panel of the figure, then response functions can be viewed as a somewhat smoother version of the predictors. Response functions are more different from (and smoother than) the predictors if $b = 0.5$.

4.3 Empirical level and power

To assess the empirical level under the null hypothesis (2.1), we conducted simulations with the functional data as described in the previous section. We took the calibration sample sizes $m = 120$ and $m = 240$. In both cases, monitoring was stopped according to the stopping rule τ_m^c with monitoring horizon specified by $T = 1$. We used the threshold function from (4.1) with sensitivity

parameter $\gamma = -0.50$. To hold to nominal levels, we recommend an adjustment to the threshold function h_ℓ in (3.5). We let $\tilde{h}_{\ell,m} = a_m h_\ell$ where $a_m \to 1$ does not change the large-sample behavior, and worked with $a_{120} = 1.4$ and $a_{240} = 1.2$. The resulting empirical levels are in Tables 2 (for the bandwidth parameter $b = 0.2$) and 3 (for $b = 0.5$) and are based on 1,000 repetitions of the experiment. The nominal levels are held quite nicely as long as the amount of variability explained by the chosen principal components is in reasonable balance, and in particular for the cases $p = q$.

To assess the power, we investigated how sensitive our procedure is to changes in the bandwidth parameter b in (4.4) in relation to the time of change k^*. Throughout, we used $p = 3$ and $q = 3$. We chose $b = 0.2$ as the null bandwidth prior to k^* and $b = 0.3, 0.4$, and 0.5 as post-break values. We evaluated the procedures for early changes, $k^* = 1$, intermediate changes, $k^* = m/2 + 1$, and late changes, $k^* = 2m/3 + 1$. With calibration sample sizes of $m = 120$ or $m = 240$, and based on 1,000 repetitions, results are reported in Tables 4 and 5. Table 4 summarizes the power of the procedures and shows the percentage of correctly identified and not identified changes. It can be seen that, even for $m = 120$, the power is always high, with the exception of the cases for which the bandwidth change is small and the change does not occur early. Most of the low power cases are alleviated in the $m = 240$ scenario. In Table 5 we give estimates for the standardized mean and median delay times

$$\bar{\tau}_m^c = \frac{E[\tau_m^c] - k^*}{m}$$

conditional on the correct identification of a break, thereby omitting the false alarm and no detection cases in the computations. Overall, there is a marked decrease in the (relative) average detection time when the test period is increased from $m = 120$ to $m = 240$. Moreover, mean and median average delay times are always close, hinting at a somewhat symmetric distribution. This conjecture is underlined in Figures 1 and 2 that display histograms for a few selected situations.

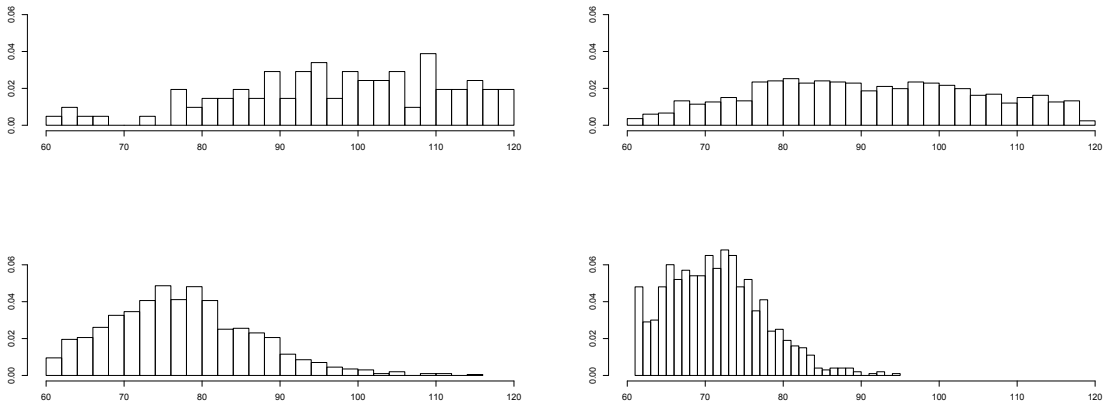

Figure 1: Histograms of false alarms under H_0 (top left) and of delay times under H_A with $b = 0.3$ (top right), $b = 0.4$ (bottom left), and $b = 0.5$ (bottom right) for $m = 120$, $p = 3$, $q = 3$, $\gamma = -0.50$. Shown are only the parts after the intermediate change occurs at $k^* = 61$.

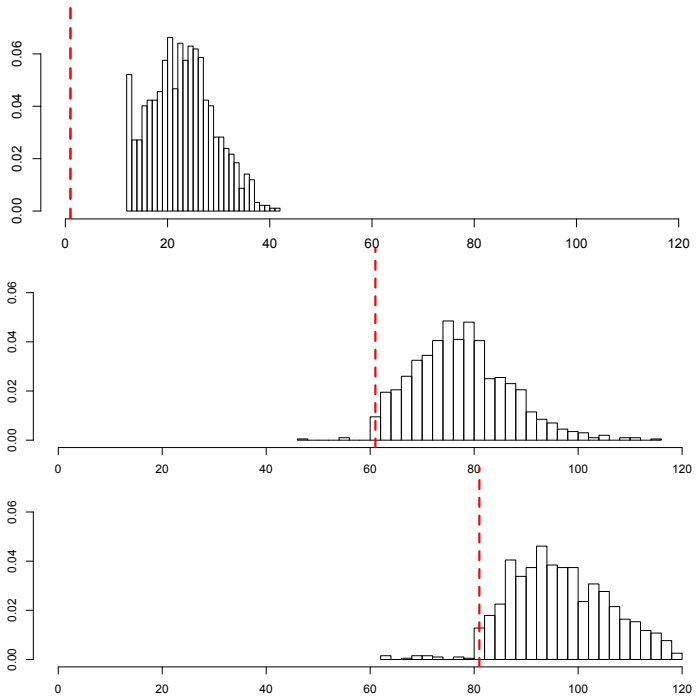

Figure 2: Histograms of delay times under H_A with $b = 0.4$ for the early change $k^* = 1$ (top), the intermediate change $k^* = 61$ (middle) and late change $k^* = 81$ (bottom), and for $m = 120$, $p = 3$, $q = 3$, and $\gamma = -0.50$.

In the final part of the simulation study, we looked at the influence of the numbers of principal components p and q on the performance of the procedure. We chose situations for which essentially all stopping rules involved have power one, namely a change from $b = 0.2$ to $b = 0.5$ at either $k^* = 1$ or $k^* = m/2 + 1$. The results can be found in Tables 6 and 7. They reveal this message: in order to detect changes quickly, one should use $p = 3$ and $q = 3$; this combination provides the quickest reaction times in all cases. We only considered this case in the power study above.

Additional simulations with the processes used in the empirical study of Gabrys et al. (2010), namely various combinations of Brownian motions and Brownian bridges as predictors and errors in conjunction with Wiener, Gaussian, and parabolic regression kernels, have led to similar conclusions and the results are not reported here.

4.4 Monitoring particle matter concentrations

European Union (EU) regulation aims at limiting the concentration (measured in $\mu g m^{-3}$) of particulate matter of aerodynamic diameter less than $10\mu m$ (PM10) in ambient air. At locations suffering from severe temperature inversions, such as the basin areas of the Alps, the limits set by the EU are often exceeded and local policy makers are then forced to impose additional measures (for example traffic regulation) to counteract high PM10 concentrations and the ensuing negative health effects,

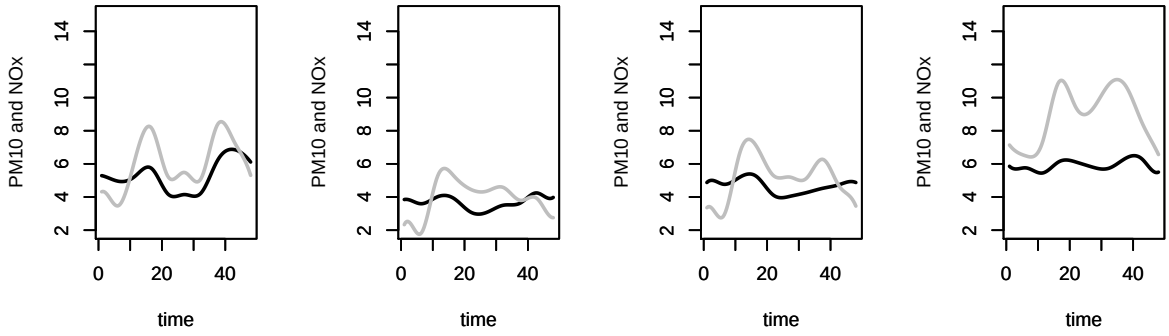

Figure 3: Five-day averages (Monday-Friday) of diurnal square root PM10 (black) and square root NOx (gray) curves at Völkermarkterstrasse for different seasonal periods: (from left to right) March 14-18, June 13-17, September 12-16 and December 12-16, 2011.

especially in the cold season (see Stadlober, Hörmann, and Pfeiler (2008)). PM10 can, similar to NO and NO2 (NOx), be attributed to a large amount to traffic. While NOx is caused by combustion processes, PM10 concentrations are additionally amplified by abration (of breaks, tarmac, road salt, etc.) and road dust resuspension. Preventive methods are therefore aimed at reducing the pollution levels of the sources (e.g. using Diesel particle filters), but also at avoiding resuspension through an efficient road cleaning process or by lowering speed limits during periods of high contamination. In the context of the EU-Life project CMA+, the city of Klagenfurt, Austria has started to use Calcium Magnesium Acetate, CMA, as a deicer in place of sodium chloride road salt. While the latter is itself a source of PM10, one hopes that CMA may help lower pollution levels by virtually 'glueing' dust particles to the ground and weaken the distortion effect. Since January 11, 2012 CMA is used area-wide in the central city. In practice it may be difficult to evaluate the effect of such measures. The main reason is that meteorological processes play a dominant role in determining concentration levels. Unfavorable meteorological conditions, for example long temperature inversion periods or substantial periods without precipitation, may completely obliterate the efficacy of any preventive method.

What we suggest, then, is to set up a functional monitoring scheme for two pollutant concentrations, in our case obtained at a measuring station located in Völkermarkterstraße of Klagenfurt, Austria. The data consists of 12,960 half-hourly PM10 concentrations (response) and NOx concentrations (predictor) from 05/01/2011 to 01/25/2012. The pollutants have a similar diurnal pattern (see Figure 3) but NOx, being an aerially pollutant, is not affected by 'road treatment' (such as application of CMA). In view of this, we are interested in checking whether the assumed functional relationship between PM10 and NOx is constant over time. If a structural change is detected, it is hoped that it can be related to application of a particular measure such as CMA, street sweeping, or partial speed limits. The raw data is skewed to the right and a square root transformation is applied to symmetrize the observations. The observations have also been detrended (concentrations are higher

in the cold season) and deseasonalized (we have a clear 'weekend effect') by application of a moving average filter. Particulate matter concentrations follow a diurnal pattern and we therefore view the collection of 48 intra-daily data points as one of a combined 270 functional observations. As in the simulation study, functional objects are represented in a cubic B-spline basis using 25 basis functions. The resulting centered functions are displayed in Figure 2 of the supplement Aue et al. (2013).

After a preliminary retrospective analysis, we chose the calibration set size $m = 90$ (05/01/2011–07/29/2011) to estimate the functional regression parameter $\hat{\boldsymbol{\beta}}_m$, and then, using h_ℓ in (3.5) with (4.1) and $\gamma = -0.50, 0.10, 0.25$, and 0.45, ran the closed-end monitoring procedures

$$\tau_m^c = \min\{\ell = 1, \ldots, L \colon \hat{V}_{\ell,m} > h_\ell\}, \qquad \hat{V}_\ell = (\tilde{\boldsymbol{\beta}}_\ell - \hat{\boldsymbol{\beta}}_m)' \hat{Q}_m \hat{\Sigma}_m^{-1} \hat{Q}_m (\tilde{\boldsymbol{\beta}}_\ell - \hat{\boldsymbol{\beta}}_m),$$

on the remaining $L = 180$ observations. We used $p = 3$ and $q = 3$ functional principal components to explain roughly 80% of the variation in both predictor and response functions. Critical values can be obtained from Table 1 with $T = 2$. We followed the recommendation of Section 4.3 and multiplied these by 1.4. An in-depth correlation analysis, that may be requested from the authors, revealed that the estimator $\hat{\Sigma}_m$ of Assumption 3.1 can be computed reasonably well with a bandwidth set to $b_m = 3$. Our findings corroborated the outcomes of the simulation study. First, with the exception of the procedure based on $\gamma = 0.45$, all procedures rejected at the 10% significance level, thereby providing evidence for a structural change in the relationship between PM10 and NOx concentrations in Völkermarkterstraße. Only the procedure based on $\gamma = -.50$ rejected at the 5% significance level. All other procedures were close to but did not reject. The results are displayed in Figure 4. There is a visible jump in the sample path of the test statistic shortly after CMA was introduced as a deicer (vertical line in the graph). Our analysis therefore seems to weakly support the hypothesis of a change in functional relationship between PM10 and NOx shortly after this measure was realized. It is worthwhile noting that this conclusion was obtained from only 15 available observations with CMA. A deeper analysis on the potential reasons for, and implications of, this behavior should be performed once more recent data are available. Results obtained using NO2 as predictor in place of NOx lead to the same conclusions and are thus not reported here.

4.5 Conclusions

To the best of our knowledge this paper is the first to combine sequential methodologies with functional data techniques. We have proposed a monitoring scheme for data that can be described by a possibly dependent functional linear model, and have given a comprehensive theoretical treatment. In this empirical section, we have demonstrated that the procedure can be applied in practice, for example to environmental data, and that its performance can be satisfactory. More empirical work is yet to be done.

5 Functional principal components in the dependent setting

We summarize some basic properties of random processes in Hilbert spaces. Throughout this section we assume that H_0 holds. In the first lemma, we provide convergence rates for the eigenvalue and

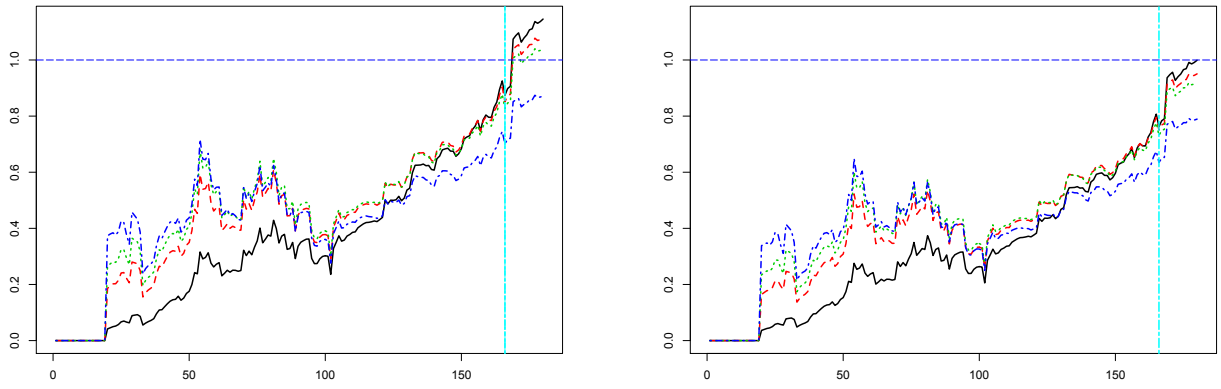

Figure 4: Behavior of stopping rules τ_m^c for $m = 90$ at the 10% level (left) and the 5% level (right). Plotted are the ratios $\hat{V}_{\ell,m}/h_\ell$, where h_ℓ is determined by (4.1), for $\gamma = -0.50$ (solid), 0.10 (dashed), 0.25 (dotted), and 0.45 (dash-dots). Crossing of the horizontal line indicates rejection of the null. The vertical line indicates the start of CMA use for deicing.

eigenfunction estimators. Under the assumptions made here, it turns out that they are the same as in the independent case.

Lemma 5.1. *If Assumptions 2.1, 2.2, 2.4, and 2.5 are satisfied, then*

$$\max_{1\le i\le p} |\hat{\lambda}_{i,m} - \lambda_i| = O_P\left(\frac{1}{\sqrt{m}}\right), \qquad \max_{1\le i\le p} \|\hat{v}_{i,m} - \hat{c}_{i,m} v_i\| = O_P\left(\frac{1}{\sqrt{m}}\right),$$

$$\max_{1\le j\le q} |\hat{\tau}_{j,m} - \tau_j| = O_P\left(\frac{1}{\sqrt{m}}\right), \qquad \max_{1\le j\le q} \|\hat{w}_{j,m} - \hat{d}_{j,m} w_j\| = O_P\left(\frac{1}{\sqrt{m}}\right),$$

where $\hat{c}_{j,m} = \operatorname{sign}\langle \hat{v}_{i,m}, v_i\rangle$ *and* $\hat{d}_{j,m} = \operatorname{sign}\langle \hat{w}_{j,m}, w_j\rangle$ *are random signs.*

Proof. All assertions follow from Theorem 3.1 in Hörmann and Kokoszka (2010), and Corollary 1.6 of Gohberg, Golberg, and Kaashoek (1990, p. 99). □

We now give uniform approximations of C and D by $\hat{C}_\ell$ and $\hat{D}_\ell$, respectively.

Lemma 5.2. *If Assumptions 2.1, 2.4, and 2.5 hold, then*

$$\sup_{\ell>1} \frac{\ell}{(\log \ell)^\beta} \|\hat{C}_\ell - C\|_{\mathcal{S}}^2 = O_P(1) \qquad \textit{and} \qquad \sup_{\ell>1} \frac{\ell}{(\log \ell)^\beta} \|\hat{D}_\ell - D\|_{\mathcal{S}}^2 = O_P(1).$$

for any $\beta > 3$.

Proof. We prove only the first statement as the second follows along the same lines. It suffices to show that

$$P\left(\|\hat{C}_\ell - C\|_{\mathcal{S}}^2 > \frac{(\log \ell)^\beta}{\ell} \text{ i.o.}\right) = 0, \tag{5.1}$$

where i.o. stands for infinitely often. A moment's reflection yields that (5.1) follows from

$$P\left(\max_{2^{n-1}\le \ell\le 2^n} \|\ell(\hat{C}_\ell - C)\|_{\mathcal{S}}^2 > n^\beta 2^{n-1} \text{ i.o.}\right) = 0.$$

We verify the stronger statement

$$P\left(\max_{1\le\ell\le 2^n}\|\ell(\hat{C}_\ell - C)\|_{\mathcal{S}}^2 > n^\beta 2^{n-1} \text{ i.o.}\right) = 0. \tag{5.2}$$

Let $U_k(s,t) = X_k(s)X_k(t) - E[X_k(s)X_k(t)]$ so $U_1 + \ldots + U_\ell = \ell(\hat{C}_\ell - C)$. Following the proof of Theorem 3.1 in Hörmann and Kokoszka (2010), define a non-negative function $\rho(s,t)$ such that, for all $0 \le L < M$,

$$E\left[\big(U_{L+1}(s,t) + \ldots + U_M(s,t)\big)^2\right] \le (M-L)\rho(s,t)$$

and $\|\rho\|_{\mathcal{S}}^2 < \infty$. Hence by Menshov's inequality, see Billingsley (1968, p. 102),

$$E\left[\max_{1\le\ell\le N}\big(U_1(s,t) + \ldots + U_\ell(s,t)\big)^2\right] \le 4N(\log 4N)^2\rho(s,t).$$

In combination with an application of Markov's inequality,

$$P\left(\max_{1\le\ell\le 2^n}\|U_1 + \ldots + U_\ell\|_{\mathcal{S}}^2 > n^\beta 2^{n-1}\right) \le \frac{2^{n+2}(\log 2^{n+2})^2}{n^\beta 2^{n-1}}\|\rho\|_{\mathcal{S}}^2.$$

Since the right-hand side is summable in n on account of $\beta > 3$, (5.2) is established with the Borel-Cantelli lemma. □

We consider approximations for the sample correlations of the projections.

Lemma 5.3. *If Assumptions 2.1, 2.4, and 2.5 hold, and if x_1 and x_2 are elements in L^2, then there is a constant C_0 which does not depend on ℓ such that*

$$\left(E\left[\left|\langle X_1, x_1\rangle\langle\varepsilon_1, x_2\rangle - \langle X_1^{(\ell)}, x_1\rangle\langle\varepsilon_1^{(\ell)}, x_2\rangle\right|^\mu\right]\right)^{1/\mu} \le C_0\left(\nu_{2\mu}(X_1 - X_1^{(\ell)}) + \nu_{2\mu}(\varepsilon_1 - \varepsilon_1^{(\ell)})\right), \tag{5.3}$$

$$\left(E\left[\left|\langle X_1, x_1\rangle\langle X_1, x_2\rangle - \langle X_1^{(\ell)}, x_1\rangle\langle X_1^{(\ell)}, x_2\rangle\right|^\mu\right]\right)^{1/\mu} \le C_0\,\nu_{2\mu}(X_1 - X_1^{(\ell)}). \tag{5.4}$$

Proof. We prove only (5.3), the proof of (5.4) is the same. By the Minkowski and Cauchy-Schwarz inequalities we have

$$\begin{aligned}
&\left(E\left[\left|\langle X_1, x_1\rangle\langle\varepsilon_1, x_2\rangle - \langle X_1^{(\ell)}, x_1\rangle\langle\varepsilon_1^{(\ell)}, x_2\rangle\right|^\mu\right]\right)^{1/\mu} \\
&\quad\le \left(E\left[\left|\langle X_1 - X_1^{(\ell)}, x_1\rangle\langle\varepsilon_1, x_2\rangle\right|^\mu\right]\right)^{1/\mu} + \left(E\left[\left|\langle\varepsilon_1 - \varepsilon_1^{(\ell)}, x_2\rangle\langle X_1^{(\ell)}, x_1\rangle\right|^\mu\right]\right)^{1/\mu} \\
&\quad\le \left(E\left[\left|\|X_1 - X_1^{(\ell)}\|\|x_1\|\|\varepsilon_1\|\|x_2\|\right|^\mu\right]\right)^{1/\mu} + \left(E\left[\left|\|\varepsilon_1 - \varepsilon_1^{(\ell)}\|\|x_2\|\|X_1^{(\ell)}\|\|x_1\|\right|^\mu\right]\right)^{1/\mu}.
\end{aligned}$$

The result follows from another application of the Cauchy-Schwarz inequality. □

Lemma 5.4. *If Assumptions 2.1, 2.4, and 2.5 hold, then, for each m, there are two independent, identically distributed Gaussian processes $(\mathbf{\Gamma}_m^{(1)}(t)\colon t \ge 0)$ and $(\mathbf{\Gamma}_m^{(2)}(t)\colon t \ge 0)$, with $E[\mathbf{\Gamma}_m^{(1)}(t)] = E[\mathbf{\Gamma}_m^{(2)}(t)] = \mathbf{0}$ and $E[\mathbf{\Gamma}_m^{(1)}(s)(\mathbf{\Gamma}_m^{(1)}(t))'] = E[\mathbf{\Gamma}_m^{(2)}(s)(\mathbf{\Gamma}_m^{(2)}(t))'] = \min\{s,t\}\Sigma$, such that*

$$\left|\sum_{k=1}^m \boldsymbol{\gamma}_k - \mathbf{\Gamma}_m^{(1)}(m)\right| = O_P(m^{1/2-\kappa}),$$

$$\sup_{\ell\geq 1}\frac{1}{\ell^{1/2-\kappa}}\left|\sum_{k=m+1}^{m+\ell}\boldsymbol{\gamma}_k-\boldsymbol{\Gamma}_m^{(2)}(\ell)\right|=O_P(1),$$

with some $\kappa>0$, where $\boldsymbol{\gamma}_k=\mathrm{vec}(\Gamma_k)$ with Γ_k as in (3.3).

Proof. An application of Lemma 5.3 and Assumption 2.5 shows that the vectors $(\boldsymbol{\gamma}_k\colon k\in\mathbb{Z})$ satisfy the conditions of Theorem S2.1 in the supplement Aue et al. (2013). The proof of that theorem also yields that the partial sums $\sum_{k=1}^{m}\boldsymbol{\gamma}_k$ and $\sum_{k=m+1}^{m+\ell}\boldsymbol{\gamma}_k$ can be approximated with independent processes. Since the strong approximation of Theorem S2.1 in the supplement Aue et al. (2013) implies the (uniform) weak approximations of this lemma, the proof is complete. □

6 Proofs

We provide the proofs for Theorems 2.1, 3.1 and 3.2. It is assumed throughout that the null hypothesis H_0 holds. We use $|A|=\sup|a_{i,j}|$ for a generic matrix A whose entries are $a_{i,j}$.

Lemma 6.1. *If Assumptions 2.1–2.5 hold, then*

$$\left|\frac{1}{m}\hat{\Xi}_{0,m}'\hat{\Xi}_{0,m}-Q\right|=O_P\left(\frac{1}{\sqrt{m}}\right),\tag{6.1}$$

$$\sup_{\ell>1}\left(\frac{(\log\ell)^{\beta/2}}{\sqrt{\ell}}+\frac{1}{\sqrt{m}}\right)^{-1}\left|\frac{1}{\ell}\hat{\Xi}_{m,m+\ell}'\hat{\Xi}_{m,m+\ell}-Q\right|=O_P(1),\tag{6.2}$$

for all $\beta>3$, where $Q=I_q\otimes\Lambda$ and $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_p)$.

Proof. Recalling the definition of $\hat{\Xi}_{m,m+\ell}$ above (2.7), it can be seen that the only non-zero elements of $\hat{\Xi}_{m,m+\ell}'\hat{\Xi}_{m,m+\ell}$ are sums of estimated principal component scores, $\sum_{k=m+1}^{m+\ell}\hat{\xi}_{k,i}\hat{\xi}_{k,i'}$. Elementary arguments give the decomposition

$$\sum_{k=m+1}^{m+\ell}\hat{\xi}_{k,i}\hat{\xi}_{k,i'}=\hat{c}_i\hat{c}_{i'}\sum_{k=m+1}^{m+\ell}\xi_{k,i}\xi_{k,i'}+\sum_{k=m+1}^{m+\ell}\big(\hat{\xi}_{k,i}-\hat{c}_i\xi_{k,i}\big)\hat{\xi}_{k,i'}+\sum_{k=m+1}^{m+\ell}\hat{c}_i\xi_{k,i}\big(\hat{\xi}_{k,i'}-\hat{c}_j\xi_{k,i'}\big).$$

Since the $(v_i\colon i\in\mathbb{N})$ are the orthonormal eigenfunctions of the covariance operator C, we conclude that $E[\xi_{n,i}^2]=E[\langle X_n,v_i\rangle^2]=\lambda_i$ and $E[\xi_{n,i}\xi_{n,i'}]=E[\langle X_n,v_i\rangle\langle X_n,v_{i'}\rangle]=0$ for $i\neq i'$. The stationarity of $(X_k\colon k\in\mathbb{Z})$ and an application of Lemma 5.2 imply

$$\begin{aligned}\sup_{\ell>1}&\frac{1}{\sqrt{\ell(\log\ell)^{\beta}}}\left|\sum_{k=m+1}^{m+\ell}\big(\xi_{k,i}\xi_{k,i'}-E[\xi_{k,i}\xi_{k,i'}]\big)\right|\\&\overset{\mathcal{D}}{=}\sup_{\ell>1}\frac{1}{\sqrt{\ell(\log\ell)^{\beta}}}\left|\sum_{k=1}^{\ell}\big(\xi_{k,i}\xi_{k,i'}-E[\xi_{k,i}\xi_{k,i'}]\big)\right|\\&=O_P(1).\end{aligned}$$

Similarly, by Lemmas 5.1 and 5.2, we have via the Cauchy-Schwarz inequality

$$\sup_{\ell>1}\frac{1}{\ell}\left|\sum_{k=m+1}^{m+\ell}\big(\hat{\xi}_{k,i}-\hat{c}_i\xi_{k,i}\big)\hat{\xi}_{k,i'}\right|=\|\hat{v}_i-\hat{c}_iv_i\|\sup_{\ell>1}\frac{1}{\ell}\sum_{k=m+1}^{m+\ell}\|X_k\|^2=O_P\left(\frac{1}{\sqrt{m}}\right),$$

since $\frac{1}{\ell}\sum_{k=m+1}^{m+\ell}\|X_k\|^2 \stackrel{\mathcal{D}}{=} \frac{1}{\ell}\sum_{k=1}^{\ell}\|X_k\|^2 = O_P(1)$ by stationarity and the Ergodic Theorem. Using the same arguments one can also verify that

$$\sup_{\ell>1}\frac{1}{\ell}\left|\sum_{k=m+1}^{m+\ell}\hat{c}_i\xi_{k,i}\big(\hat{\xi}_{k,i'}-\hat{c}_{i'}\xi_{k,i'}\big)\right| = O_P(1),$$

completing the proof of (6.2). The proof of (6.1) is similar to that of (6.2) if we use Theorem 3.1 of Hörmann and Kokoszka (2010) in place of Lemma 5.2. □

We state a lemma that establishes the asymptotics for the terms $\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N}$. The proof requires a number of auxiliary results that are given in the supplement Aue et al. (2013). We notice the explicit form of $\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N} = \mathrm{vec}(\hat{G})$, where the matrix $\hat{G}$ consists of the entries $\hat{G}_{i,j} = \sum_{k=n+1}^{N}\hat{\xi}_{k,i}\hat{\eta}_{k,j}$. Let

$$T_{n,N}(i,j) = \hat{c}_i\hat{d}_j\sum_{k=n+1}^{N}\xi_{k,i}\eta_{k,j}. \tag{6.3}$$

Lemma 6.2. *If Assumptions 2.1–2.5 hold, then*

$$\sum_{k=1}^{m}\hat{\xi}_{k,i}\hat{\eta}_{k,j} = T_{0,m}(i,j) + mR_m(i,j) + O_P(1),$$

$$\sum_{k=m+1}^{m+\ell}\hat{\xi}_{k,i}\hat{\eta}_{k,j} = T_{m,m+\ell}(i,j) + \ell R_m(i,j) + U^{(1)}_{\ell,m}(i,j) + U^{(2)}_{\ell,m}(i,j),$$

where $R_m(i,j) = O_P(m^{-1/2})$,

$$\sup_{\ell>1}\frac{|U^{(1)}_{\ell,m}(i,j)|}{\sqrt{\ell(\log\ell)^\beta}} = O_P\left(\frac{1}{\sqrt{m}}\right) \qquad \textit{and} \qquad \sup_{\ell>1}\frac{|U^{(2)}_{\ell,m}(i,j)|}{\ell} = O_P\left(\frac{1}{m}\right).$$

Proof. Recognizing that $T_{n,N} = T^{(1)}_{n,N} + T^{(2)}_{n,N}$ with $T^{(1)}_{n,N}$ and $T^{(2)}_{n,N}$ defined in Lemmas S3.1 and S3.2 of the supplement Aue et al. (2013), that S_m is defined in Lemma S3.1, and that one can set $R_m = R^{(1)}_m + \ldots + R^{(4)}_m$, $U^{(1)}_{\ell,m} = U_{m,1}(\ell) + U_{m,3}(\ell) + U_{m,5}(\ell) + U_{m,7}(\ell)$ and $U^{(2)}_{\ell,m} = U_{m,2}(\ell) + U_{m,4}(\ell) + U_{m,6}(\ell) + U_{m,8}(\ell)$, with the terms on the respective right-hand sides being defined in Lemmas S3.1–S3.5, the statements of this lemma are implied by the results of Appendix S3 in the supplement Aue et al. (2013). □

We start to establish Theorems 2.1 and 3.1. To simplify notation, let

$$f(x) = (1+x)g\left(\frac{x}{1+x}\right).$$

Lemma 6.3. *If Assumptions 2.1–2.5 and 3.2 hold, then*

$$\sup_{\ell\geq 1}\frac{\ell}{\sqrt{m}f(\ell/m)}\big|(\hat{\Xi}'_{0,m}\hat{\Xi}_{0,m})^{-1} - (mQ)^{-1}\big|\big|\hat{\Xi}'_{0,m}\hat{\boldsymbol{E}}_{0,m}\big| = O_P\left(\frac{1}{\sqrt{m}}\right), \tag{6.4}$$

$$\sup_{\ell>1}\frac{\ell}{\sqrt{m}f(\ell/m)}\big|(\hat{\Xi}'_{m,m+\ell}\hat{\Xi}_{m,m+\ell})^{-1} - (\ell Q)^{-1}\big|\big|\hat{\Xi}'_{m,m+\ell}\hat{\boldsymbol{E}}_{m,m+\ell}\big| = o_P(1), \tag{6.5}$$

where Q is defined in Lemma 6.1.

Proof. We prove only (6.5), since (6.4) follows from similar arguments. Lemma 6.1 implies that $|(\hat{\Xi}'_{m,m+\ell}\hat{\Xi}_{m,m+\ell})^{-1} - (\ell Q)^{-1}| \le r_{m,1}(\ell) + r_{m,2}(\ell)$, where

$$\sup_{\ell>1} \frac{\ell^{3/2}}{\sqrt{(\log \ell)^{\beta}}} |r_{m,1}(\ell)| = O_P(1) \qquad \text{and} \qquad \sup_{\ell \ge 1} \ell |r_{m,2}(\ell)| = O_P\left(\frac{1}{\sqrt{m}}\right).$$

Invoking Lemmas 5.4 and 6.2 yields $|\hat{\Xi}'_{m,m+k}\hat{\boldsymbol{E}}_{m,m+k}| \le r_{m,3}(k) + r_{m,4}(k)$, where

$$\sup_{\ell>1} \frac{1}{\sqrt{\ell(\log \ell)^{\beta}}} |r_{m,3}(\ell)| = O_P(1) \qquad \text{and} \qquad \sup_{\ell>1} \frac{1}{\ell} |r_{m,4}(\ell)| = O_P\left(\frac{1}{\sqrt{m}}\right).$$

Combining the previous statements, we get

$$\begin{aligned}
&\sup_{\ell>1} \frac{\ell}{\sqrt{m} f(\ell/m)} \left|(\hat{\Xi}'_{m,m+\ell}\hat{\Xi}_{m,m+\ell})^{-1} - (\ell Q)^{-1}\right| \left|\hat{\Xi}'_{m,m+\ell}\hat{\boldsymbol{E}}_{m,m+\ell}\right| \\
&\quad = O_P(1) \sup_{\ell \ge 1} \frac{\ell}{\sqrt{m} f(\ell/m)} \left(\frac{\sqrt{(\log \ell)^{\beta}}}{\ell^{3/2}} + \frac{1}{\sqrt{m}\ell}\right) \left(\sqrt{\ell(\log \ell)^{\beta}} + \frac{\ell}{\sqrt{m}}\right) \\
&\quad = O_P(1)(r_{m,5} + r_{m,6} + r_{m,7}),
\end{aligned}$$

where

$$r_{m,5} = \sup_{\ell>1} \frac{(\log \ell)^{\beta}}{\sqrt{m} f(\ell/m)}, \quad r_{m,6} = \sup_{\ell>1} \frac{\sqrt{\ell(\log \ell)^{\beta}}}{m f(\ell/m)} \quad \text{and} \quad r_{m,7} = \sup_{\ell>1} \frac{\ell}{m^{3/2} f(\ell/m)}.$$

It remains to show that $r_{m,5}$, $r_{m,6}$, and $r_{m,7}$ converge to zero as m tends to ∞. Using Assumption 3.2 in various places, in addition to a number of basic estimation steps, one can show that

$$\begin{aligned}
r_{m,5} &\le \sup_{\ell \le m} \frac{(\log \ell)^{\beta}}{\sqrt{\ell}} \frac{\sqrt{\ell/m}}{f(\ell/m)} + \sup_{\ell>m} \frac{\sqrt{m}(\log \ell)^{\beta}}{\ell} \sup_{1/2 \le x \le 1} \frac{1}{g(x)} \\
&= O(1) \left(\sup_{\ell \le m} \frac{(\log \ell)^{\beta}}{\sqrt{\ell}} \left(\frac{\ell}{m}\right)^{1/2-\gamma} + \frac{(\log m)^{\beta}}{\sqrt{m}} \right) \\
&= O(1)(\log m)^{\beta} m^{\gamma - 1/2} \\
&= o(1).
\end{aligned}$$

Using similar arguments, it follows that

$$r_{m,6} \le \sqrt{\frac{(\log m)^{\beta}}{m}} \sup_{1/m \le x \le 1} \frac{\sqrt{x}}{g(1+x)} + \sup_{\ell>m} \sqrt{\frac{(\log \ell)^{\beta}}{\ell}} \sup_{1/2 \le x \le 1} \frac{1}{g(x)} = o(1).$$

Since the function $x \mapsto x/f(x)$ is bounded on $[0, \infty)$ we conclude that

$$r_{m,7} \le \frac{1}{\sqrt{m}} \sup_{x>0} \frac{x}{f(x)} = o(1),$$

thus completing the proof of (6.5). □

Lemma 6.4. *If Assumptions 2.1–2.5 and 3.2 hold, then*

$$\sup_{\ell \ge 1} \frac{1}{\sqrt{m} f(\ell/m)} \left| \left(\hat{\Xi}'_{m,m+\ell}\hat{\boldsymbol{E}}_{m,m+\ell} - \frac{\ell}{m}\hat{\Xi}'_{0,m}\hat{\boldsymbol{E}}_{0,m} \right) - \left(\sum_{k=m+1}^{m+\ell} \boldsymbol{\gamma}_k - \frac{\ell}{m} \sum_{k=1}^{m} \boldsymbol{\gamma}_k \right) \right| = o_P(1). \tag{6.6}$$

Proof. With $\boldsymbol{\gamma}_k = \mathrm{vec}(\Gamma_k)$, where $\Gamma_k(i,j) = \xi_{k,i}(\eta^*_{k,j} + \theta_{k,j})$, utilizing the result of Lemma 6.2 it suffices to verify that

$$\sup_{\ell \geq 1} \frac{1}{\sqrt{m} f(\ell/m)} \left(\sqrt{\frac{\ell (\log \ell)^\beta}{m}} + \frac{\ell}{m} \right) = o(1).$$

But the rate on the left-hand side is bounded from above by $r_{m,6} + r_{m,7}$, as defined in Lemma 6.3, and the claim follows. □

Lemma 6.5. *If Assumptions 2.1–2.5 and 3.2 hold, then*

$$\sup_{\ell \geq 1} \frac{1}{\sqrt{m} f(\ell/m)} \left| \left(\sum_{k=m+1}^{m+\ell} \boldsymbol{\gamma}_k - \frac{\ell}{m} \sum_{k=1}^{m} \boldsymbol{\gamma}_k \right) - \left(\boldsymbol{\Gamma}^{(2)}_m(\ell) - \frac{\ell}{m} \boldsymbol{\Gamma}^{(1)}_m(m) \right) \right| = o_P(1),$$

where $(\boldsymbol{\Gamma}^{(1)}_m(x) \colon x \in [0,1])$ *and* $(\boldsymbol{\Gamma}^{(2)}_m(x) \colon x \in [0,1])$ *are the Gaussian processes defined in Lemma 5.4.*

Proof. The rate of convergence of the respective partial sums of the $\boldsymbol{\gamma}_k$ to the Gaussian processes $(\boldsymbol{\Gamma}^{(1)}_m(x) \colon x \in [0,1])$ and $(\boldsymbol{\Gamma}^{(2)}_m(x) \colon x \in [0,1])$ established in Lemma 5.4 dictates that we only need to show that

$$\sup_{\ell \geq 1} \frac{(k/m) m^{1/2-\kappa}}{\sqrt{m} f(\ell/m)} = o(1) \qquad \text{and} \qquad \sup_{\ell \geq 1} \frac{\ell^{1/2-\kappa}}{\sqrt{m} f(\ell/m)} = o(1). \tag{6.7}$$

To prove the second statement, by Assumption 3.2 we have

$$\sup_{\ell \leq m} \frac{\ell^{1/2-\kappa}}{\sqrt{m} f(\ell/m)} = O(1) \sup_{\ell \leq m} \frac{\ell^{1/2-\kappa}}{\sqrt{m} (\ell/m)^\gamma} = O(1) \sup_{\ell \leq m} \frac{\ell^{1/2-\kappa-\gamma}}{m^{1/2-\gamma}} = O(1) \sup_{\ell \leq m} \ell^{-\kappa} = o(1)$$

as well as

$$\sup_{\ell > m} \frac{\ell^{1/2-\kappa}}{\sqrt{m} f(\ell/m)} = O(1) \sqrt{m} \sup_{\ell > m} \ell^{-1/2-\kappa} = o(1).$$

Since similar arguments also yield the first statement in (6.7), the lemma is established. □

Proof of Theorem 2.1. We only derive the asymptotics of $\hat{\boldsymbol{\beta}}_m$. Take $\Xi_{0,m}$ as the theoretical version of $\hat{\Xi}_{0,m}$. The matrix $\boldsymbol{E}_{0,m}$ is defined analoguously. Then $\Xi'_{0,m} \boldsymbol{E}_{0,m} = \sum_{j=k}^{m} \boldsymbol{\gamma}_k$. By (3.1) we have $\hat{\boldsymbol{\beta}}_m = \boldsymbol{\beta} + \left(\hat{\Xi}'_{0,m} \hat{\Xi}_{0,m} \right)^{-1} \hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m}$ and

$$\begin{aligned}
&\left(\hat{\Xi}'_{0,m} \hat{\Xi}_{0,m} \right)^{-1} \hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m} \\
&\quad = (mQ)^{-1} \Xi'_{0,m} \boldsymbol{E}_{0,m} (\hat{d} \otimes \hat{c}) + \left[\left(\hat{\Xi}'_{0,m} \hat{\Xi}_{0,m} \right)^{-1} - (mQ)^{-1} \right] \hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m} \\
&\qquad + (mQ)^{-1} \left[\hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m} - \Xi'_{0,m} \boldsymbol{E}_{0,m} (\hat{d} \otimes \hat{c}) \right].
\end{aligned}$$

Using (6.4) with $\ell \to \infty$ we see that

$$\left| \left[\left(\hat{\Xi}'_{0,m} \hat{\Xi}_{0,m} \right)^{-1} - (mQ)^{-1} \right] \hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m} \right| = O_P\left(\frac{1}{m} \right).$$

By Lemma 6.2

$$(mQ)^{-1} \left[\hat{\Xi}'_{0,m} \hat{\boldsymbol{E}}_{0,m} - \Xi'_{0,m} \boldsymbol{E}_{0,m} (\hat{d} \otimes \hat{c}) \right] = Q^{-1} \mathrm{vec}(R_m) + O_P\left(\frac{1}{m} \right),$$

where $R_m = (R_m(i,j)\colon i = 1,\dots,p, j = 1,\dots,q)$. Take $\boldsymbol{R}_m = Q^{-1}\mathrm{vec}(R_m)$. The asymptotic normality of $(mQ)^{-1}\Xi'_{0,m}\boldsymbol{E}_{0,m}$ can be established from Lemma 5.4. Combining these results completes the proof. □

Proof of Theorem 3.1. We prove only the first part of Theorem 3.1, the proof of the second part is somewhat simpler. Combining the results of Appendix S3 in the supplement Aue et al. (2013) with Lemmas 6.2–6.5 and Assumption 3.1 shows that, in order to establish the limit distribution for $\sup_\ell \hat{V}_{\ell,m}/h_{\ell,m}$, it suffices to show that

$$\sup_{\ell\geq 1}\frac{V_{\ell,m}}{mf^2(\ell/m)} \xrightarrow{\mathcal{D}} \sup_{x>0}\frac{\mathcal{V}(x)}{f^2(x)}, \tag{6.8}$$

where

$$V_{\ell,m} = \left(\boldsymbol{\Gamma}_m^{(2)}(\ell) - \frac{\ell}{m}\boldsymbol{\Gamma}_m^{(1)}(m)\right)' \Sigma^{-1}\left(\boldsymbol{\Gamma}_m^{(2)}(\ell) - \frac{\ell}{m}\boldsymbol{\Gamma}_m^{(1)}(m)\right),$$
$$\mathcal{V}(x) = \big(\boldsymbol{\Gamma}^{(2)}(x) - x\boldsymbol{\Gamma}^{(1)}(1)\big)'\Sigma^{-1}\big(\boldsymbol{\Gamma}^{(2)}(x) - x\boldsymbol{\Gamma}^{(1)}(1)\big),$$

with $((\boldsymbol{\Gamma}^{(1)}(x),\boldsymbol{\Gamma}^{(2)}(x))\colon x\geq 0) \stackrel{\mathcal{D}}{=} ((\boldsymbol{\Gamma}_m^{(1)}(x),\boldsymbol{\Gamma}_m^{(2)}(x))\colon x\geq 0)$ for all m. Since the scale transformation implies that

$$\left(\frac{1}{\sqrt{m}}\big(\boldsymbol{\Gamma}_m^{(1)}(\ell),\boldsymbol{\Gamma}_m^{(2)}(\ell)\big)\colon \ell\geq 1\right) \stackrel{\mathcal{D}}{=} \left(\big(\boldsymbol{\Gamma}^{(1)}(\ell/m),\boldsymbol{\Gamma}^{(2)}(\ell/m)\big)\colon \ell\geq 1\right), \tag{6.9}$$

we can use the continuity of $(\boldsymbol{\Gamma}^{(1)}(x)\colon x\geq 0)$ and $(\boldsymbol{\Gamma}^{(2)}(x)\colon x\geq 0)$ to establish, for all fixed $0<T_1<T_2<\infty$, the convergence

$$\sup_{mT_1\leq\ell\leq mT_2}\frac{V_{\ell,m}}{mf^2(\ell/m)} \xrightarrow{\mathcal{D}} \sup_{T_1\leq x\leq T_2}\frac{\mathcal{V}(x)}{f^2(x)}. \tag{6.10}$$

On the other hand, the Law of Iterated Logarithm and Assumption 3.2 yield that

$$\limsup_{T_1\to\infty}\sup_{0<x\leq T_1}\frac{\mathcal{V}(x)}{f^2(x)} = 0 \quad \text{a.s.} \qquad \text{and} \qquad \limsup_{T_2\to\infty}\sup_{T_2\leq x<\infty}\frac{\mathcal{V}(x)}{f^2(x)} = 0 \quad \text{a.s.},$$

so that (6.8) follows from (6.9) and (6.10). Computing the covariance function, one can easily verify that the coordinates of $(\Sigma^{-1/2}(\boldsymbol{\Gamma}^{(2)}(x) - x\boldsymbol{\Gamma}^{(1)}(1))\colon x\geq 0)$ are independent and identically distributed processes having the same distribution as $(W(x) - xZ\colon x\geq 0)$, where $(W(x)\colon x\geq 1)$ is a standard Brownian motion and Z a standard normal random variable, independent of $(W(x) - xZ\colon x\geq 0)$. It is well known that

$$\left((1+x)W\left(\frac{x}{1+x}\right)\colon x\geq 0\right) \stackrel{\mathcal{D}}{=} \big(W(x) - xZ\colon x\geq 0\big).$$

Hence

$$\left(\sup_{x>0}\frac{\mathcal{V}(x)}{f^2(x)}\colon x\geq 0\right) \stackrel{\mathcal{D}}{=} \left(\sum_{j=1}^{pq} W_j^2\left(\frac{x}{1+x}\right) g^{-2}\left(\frac{x}{1+x}\right)\colon x\geq 0\right),$$

where $(W_1(x)\colon x\geq 0),\dots,(W_{pq}(x)\colon x\geq 0)$ are independent standard Brownian motions. A change of variables from $x/(1+x)$ to x completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Under the specified alternative we have that $\tilde{\beta}_\ell$ converges to $\beta^* = \mathrm{vec}\,((W^*)')$. This is because of Theorem 3.1 and the fact that $\tilde{\beta}_\ell$ is an average (see (3.1)). The first k^* terms in this average that yield convergence to β have no impact on the limit. Hence

$$(\hat{\beta}_m - \tilde{\beta}_{Tm})/V_{Tm,m}^{1/2} = \frac{\sqrt{m}\,T}{T+1}\frac{\hat{\beta}_m - \tilde{\beta}_{mT}}{g\,(T/(T+1))} \sim \frac{\sqrt{m}\,T}{T+1}\frac{\beta - \beta^*}{g\,(T/(T+1))}.$$

The latter term is divergent, so the proof follows. □

References

Aue, A., Hörmann, S., Horváth, L., and Hušková, M. (2013). Online supplementary material to "Dependent functional linear models with applications to monitoring structural change". Available at `http://www.stat.sinica.edu.tw/statistica`.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J.G. (2012). Sequential testing for the stability of high frequency portfolio betas. *Econometric Theory* **28**, 804–837.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009a). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**, 4046–4087.

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters* **67**, 221–231.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis* **34**, 1–16.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *Econometrics Journal* **9**, 373–403.

Aue, A., Horváth, L., and Reimherr, M.L. (2009b). Delay times of sequential monitoring procedures for multiple time series regression models. *Journal of Econometrics* **149**, 174–190.

Berkes, I., Hörmann, S., and Schauer, J. (2011). Split invariance principles for stationary sequences. *The Annals of Probability* **39**, 2441–2473.

Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *The Annals of Probability* **7**, 29–54.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods (2nd ed.).* Springer, New York.

Chu, C.-S.J., Stinchcombe, J., and White, H. (1996). Monitoring structural change. *Econometrica* **64**, 1045–1065.

Cuevas, A., Febrero, M., and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *Canadian Journal of Statistics* **30**, 285–300.

Cyree, K.K., Griffiths, M.D., and Winters, D.B. (2004). An empirical examination of intraday volatility in euro-dollar rates. *The Quaterly Review of Economics and Finance* **44**, 44–57.

Elazović, S. (2009). Functional modeling of volatility in the Swedish limit order book. *Computational Statistics & Data Analysis* **53**, 2107–2118.

Fremdt, S., Horváth, L., Kokoszka, P., and Steinebach, J.G. (2013). Functional data analysis with increasing number of projections. Preprint, available on `arXiv:1302.6102`.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association* **106**, 1113–1125.

Gohberg, I., Golberg, S., and Kaashoek, M.A. (1990). *Classes of Linear Operators.* Operator Theory: Advances and Applications **49**. Birkhäuser, Boston.

Hörmann, S., and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845–1884.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126**, 225–251.

Hušková, M., Prašková, Z., and Steinebach, J. (2007). On the detection of changes in autoregressive time series, I. Asymptotics. *Journal of Statistical Planning and Inference* **137**, 1243–1259.

Ibragimov, A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7**, 349–382.

Leisch, F., Hornik, K., and Kuan, C.M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* **16**, 835–854.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis (2nd ed.)*. Springer, New York.

Stadlober, E., Hörmann, S., and Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment* **42**, 1098–1109.

Stoumbos, Z.G., Reynolds Jr., M.R., Ryan,T.P., and Woodall, W.H. (2000). The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association* **95**, 992–998.

Von Bahr, B. (1967). Multi-dimensional integral limit theorems. *Arkiv för Matematik* **7**, 71–88.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* **20**, 99–121.

		$\gamma = -0.50$			$\gamma = 0.10$			$\gamma = 0.25$			$\gamma = 0.45$		
r	$T\backslash\alpha$	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
1	1	0.832	1.121	1.816	2.288	2.962	4.611	3.121	3.984	5.998	5.881	7.068	9.810
	2	1.493	2.008	3.281	2.885	3.738	5.737	3.597	4.597	6.919	6.118	7.343	10.346
	3	1.892	2.541	4.177	3.152	4.092	6.391	3.812	4.806	7.413	6.155	7.409	10.409
	∞	3.358	4.496	7.303	3.975	5.168	8.077	4.402	5.613	8.543	6.369	7.704	10.920
2	1	1.324	1.681	2.463	3.425	4.241	6.126	4.584	5.592	7.867	8.091	9.431	12.550
	2	2.366	2.986	4.377	4.332	5.368	7.627	5.283	6.436	9.134	8.383	9.804	13.092
	3	2.983	3.795	5.623	4.772	5.919	8.524	5.657	6.892	9.722	8.508	9.959	13.171
	∞	5.257	6.643	9.838	6.001	7.376	10.578	6.532	7.936	11.109	8.722	10.217	13.692
3	1	1.720	2.114	2.968	4.454	5.369	7.363	5.826	6.938	9.332	9.944	11.412	14.767
	2	3.094	3.780	5.378	5.618	6.759	9.253	6.774	8.067	10.985	10.356	11.944	15.243
	3	3.914	4.795	6.734	6.144	7.437	10.218	7.184	8.551	11.596	10.369	11.903	15.552
	∞	6.964	8.545	12.154	7.741	9.322	12.843	8.291	9.918	13.553	10.814	12.296	16.015
4	1	2.114	2.534	3.457	5.317	6.334	8.559	6.955	8.190	10.819	11.680	13.240	16.758
	2	3.750	4.520	6.306	6.713	7.948	10.705	8.071	9.472	12.494	12.046	13.660	17.333
	3	4.798	5.779	7.913	7.387	8.726	11.704	8.529	9.998	13.261	12.205	13.940	17.549
	∞	8.577	10.280	14.091	9.290	10.998	14.557	9.870	11.576	15.306	12.635	14.283	17.969
6	1	2.834	3.310	4.389	7.033	8.166	10.643	9.076	10.419	13.342	14.736	16.504	20.267
	2	5.079	5.923	7.834	8.874	10.246	13.268	10.514	12.029	15.597	15.169	17.033	20.996
	3	6.387	7.469	9.852	9.704	11.312	14.709	11.178	12.916	16.463	15.318	17.120	21.147
	∞	11.415	13.442	17.610	12.226	14.149	18.553	12.885	14.839	19.007	15.877	17.833	21.856
9	1	3.864	4.433	5.636	9.369	10.662	13.267	12.030	13.551	16.762	18.845	20.852	25.232
	2	6.866	7.874	9.978	11.863	13.464	16.737	13.906	15.729	19.591	19.500	21.536	25.941
	3	8.641	9.862	12.489	12.985	14.754	18.443	14.719	16.591	20.885	19.710	21.805	26.326
	∞	15.411	17.677	22.333	16.374	18.629	23.176	17.063	19.246	24.007	20.334	22.459	26.889

Table 1: Asymptotic critical values for various configurations with $r = pq$.

q/p	1	2	3
1	0.115	0.091	0.078
	0.061	0.051	0.043
	0.016	0.009	0.004
2	0.111	0.095	0.084
	0.045	0.057	0.059
	0.013	0.013	0.014
3	0.069	0.092	0.107
	0.038	0.041	0.070
	0.010	0.009	0.022

q/p	1	2	3
1	0.129	0.091	0.089
	0.069	0.049	0.045
	0.019	0.005	0.008
2	0.095	0.107	0.084
	0.052	0.044	0.044
	0.007	0.013	0.014
3	0.068	0.093	0.099
	0.033	0.046	0.055
	0.011	0.010	0.008

Table 2: Empirical levels for $m = 120$ (left) and $m = 240$ (right), $b = 0.2$, $\gamma = -0.50$, and various values of p and q. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.

q/p	1	2	3
1	0.132	0.079	0.069
	0.069	0.088	0.042
	0.018	0.008	0.013
2	0.085	0.088	0.070
	0.051	0.036	0.052
	0.007	0.010	0.010
3	0.076	0.094	0.096
	0.034	0.043	0.053
	0.007	0.012	0.007

q/p	1	2	3
1	0.110	0.098	0.086
	0.064	0.041	0.048
	0.015	0.004	0.022
2	0.074	0.101	0.088
	0.043	0.035	0.038
	0.007	0.017	0.014
3	0.094	0.087	0.082
	0.043	0.040	0.040
	0.009	0.010	0.007

Table 3: Empirical levels for $m = 120$ (left) and $m = 240$ (right), $b = 0.5$, $\gamma = -0.50$, and various values of p and q. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.

k^*	1		61		81	
b	cd	nd	cd	nd	cd	nd
0.3	1.000	0.000	0.831	0.169	0.583	0.405
	1.000	0.000	0.805	0.195	0.552	0.447
	0.997	0.003	0.670	0.330	0.398	0.602
0.4	1.000	0.000	0.997	0.000	0.961	0.024
	1.000	0.000	0.999	0.001	0.948	0.047
	1.000	0.000	0.997	0.003	0.900	0.099
0.5	1.000	0.000	0.999	0.000	0.987	0.000
	1.000	0.000	0.999	0.000	0.996	0.000
	1.000	0.000	1.000	0.000	0.993	0.004

k^*	1		121		161	
b	cd	nd	cd	nd	cd	nd
0.3	1.000	0.000	0.991	0.009	0.861	0.132
	1.000	0.000	0.987	0.013	0.834	0.165
	1.000	0.000	0.968	0.032	0.770	0.229
0.4	1.000	0.000	1.000	0.000	0.989	0.000
	1.000	0.000	1.000	0.000	0.996	0.000
	1.000	0.000	1.000	0.000	0.998	0.002
0.5	1.000	0.000	0.998	0.000	0.996	0.000
	1.000	0.000	1.000	0.000	0.999	0.000
	1.000	0.000	1.000	0.000	1.000	0.000

Table 4: Empirical percentages for correctly detected (cd) changes and not detected (nd) changes for $m = 120$ (top) and $m = 240$ (bottom), $p = 3$, $q = 3$, $\gamma = -0.50$, and various values of k^* and b. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.

k^*	1		61		81	
b	mean	med	mean	med	mean	med
0.3	0.342	0.333	0.250	0.250	0.175	0.175
	0.350	0.342	0.283	0.292	0.183	0.183
	0.400	0.392	0.283	0.292	0.200	0.208
0.4	0.183	0.183	0.141	0.133	0.133	0.125
	0.192	0.192	0.158	0.150	0.142	0.142
	0.208	0.208	0.183	0.175	0.167	0.167
0.5	0.133	0.133	0.092	0.092	0.092	0.083
	0.141	0.141	0.100	0.100	0.100	0.092
	0.150	0.150	0.117	0.108	0.117	0.117

k^*	1		121		161	
b	mean	med	mean	med	mean	med
0.3	0.213	0.213	0.183	0.167	0.154	0.150
	0.229	0.225	0.196	0.183	0.167	0.167
	0.254	0.250	0.225	0.217	0.183	0.183
0.4	0.125	0.125	0.088	0.083	0.083	0.079
	0.133	0.133	0.100	0.096	0.096	0.092
	0.142	0.142	0.108	0.104	0.113	0.108
0.5	0.992	0.096	0.058	0.054	0.054	0.050
	0.096	0.100	0.067	0.063	0.063	0.058
	0.104	0.108	0.071	0.071	0.075	0.075

Table 5: Standardized conditional empirical mean and median delay times $(E[\tau]-k^*)/m$ for $m = 120$ (top) and $m = 240$ (bottom), $p = 3$, $q = 3$, $\gamma = -0.50$, and various values of k^* and b. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.

p	1		2		3	
q	mean	med	mean	med	mean	med
1	0.192	0.183	0.200	0.192	0.200	0.192
	0.217	0.200	0.225	0.208	0.217	0.208
	0.267	0.250	0.258	0.250	0.250	0.242
2	0.208	0.200	0.175	0.175	0.175	0.175
	0.225	0.217	0.192	0.183	0.192	0.183
	0.275	0.258	0.217	0.217	0.208	0.208
3	0.217	0.208	0.175	0.175	0.133	0.133
	0.283	0.225	0.183	0.183	0.142	0.142
	0.275	0.258	0.217	0.208	0.150	0.150

p	1		2		3	
q	mean	med	mean	med	mean	med
1	0.125	0.121	0.138	0.138	0.142	0.142
	0.142	0.138	0.154	0.150	0.154	0.154
	0.167	0.163	0.179	0.175	0.179	0.175
2	0.146	0.142	0.125	0.121	0.129	0.129
	0.154	0.154	0.133	0.133	0.138	0.138
	0.183	0.179	0.150	0.150	0.154	0.154
3	0.150	0.150	0.129	0.129	0.092	0.096
	0.163	0.163	0.142	0.138	0.096	0.100
	0.188	0.183	0.154	0.154	0.104	0.108

Table 6: Standardized conditional empirical mean and median delay times $(E[\tau]-k^*)/m$ for $m = 120$ (top) and $m = 240$ (bottom), $k^* = 1$, $\gamma = -0.50$, and various values of p and q. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.

p	1		2		3	
q	mean	med	mean	med	mean	med
1	0.192	0.175	0.183	0.175	0.167	0.158
	0.217	0.200	0.208	0.200	0.192	0.183
	0.258	0.250	0.242	0.233	0.225	0.217
2	0.200	0.192	0.150	0.142	0.125	0.125
	0.217	0.217	0.200	0.158	0.150	0.142
	0.258	0.250	0.192	0.183	0.175	0.167
3	0.200	0.183	0.133	0.133	0.083	0.083
	0.217	0.200	0.150	0.142	0.092	0.092
	0.258	0.250	0.175	0.167	0.100	0.108

p	1		2		3	
q	mean	med	mean	med	mean	med
1	0.113	0.108	0.113	0.108	0.108	0.104
	0.129	0.121	0.133	0.129	0.125	0.121
	0.158	0.146	0.158	0.154	0.150	0.146
2	0.117	0.113	0.092	0.086	0.086	0.083
	0.138	0.133	0.104	0.100	0.100	0.096
	0.167	0.163	0.121	0.117	0.117	0.113
3	0.125	0.121	0.092	0.086	0.054	0.050
	0.142	0.133	0.100	0.100	0.063	0.058
	0.167	0.158	0.121	0.117	0.071	0.067

Table 7: Standardized conditional empirical mean and median delay times $(E[\tau]-k^*)/m$ for $m = 120$ (top) and $m = 240$ (bottom), $k^* = m/2 + 1$, $\gamma = -0.50$, and various values of p and q. In each cell, the top, middle, and bottom values refer to the nominal levels $\alpha = 0.10, 0.05$, and 0.01, respectively.