The Econometrics Journal

Econometrics Journal (2017), volume **20**, pp. 86–117.
doi: 10.1111/ectj.12075

Change point tests in functional factor models with application to yield curves

PATRICK BARDSLEY[†], LAJOS HORVÁTH[‡], PIOTR KOKOSZKA[§] AND GABRIEL YOUNG[‖]

[†]*Institute for Computational Engineering and Sciences, University of Texas, Austin TX, USA*.
E-mail: bardsley@ices.utexas.edu

[‡]*Department of Mathematics, University of Utah, Salt Lake City UT, USA*.
E-mail: horvath@math.utah.edu

[§]*Department of Statistics, Colorado State University, Fort Collins CO, USA*.
E-mail: Piotr.Kokoszka@colostate.edu

[‖]*Department of Statistics, Columbia University, New York NY, USA*.
E-mail: gjy2107@columbia.edu

First version received: December 2015; final version accepted: October 2016

Summary Motivated by the problem of the detection of a change point in the mean structure of yield curves, we introduce several methods to test the null hypothesis that the mean structure of a time series of curves does not change. The mean structure does not refer merely to the level of the curves, but also to their range and other aspects of their shape, most prominently concavity. The performance of the tests depends on whether possible break points in the error structure, which refers to the random variability in the aspects of the curves listed above, are taken into account or not. If they are not taken into account, then an existing change point in the mean structure may fail to be detected with a large probability. The paper contains a complete asymptotic theory, a simulation study and illustrative data examples, as well as details of the numerical implementation of the testing procedures.

Keywords: *Change point, Functional time series, Yield curve.*

1. INTRODUCTION

Recent advances in yield curve modelling have brought to the fore a class of functional data models, which we call functional factor models. In this paper, we propose several methods for testing the hypothesis of the existence of change points in such models. While there has been extensive research on yield curve prediction, modelling via regime switching processes and even change point estimation, we are not aware of any previous research concerned with a change point testing problem.

Yield curve modelling has been an important direction of economic research over many decades. A solid account of the classical theory related to the so-called affine models is presented in Chapter 8 of Campbell et al. (1997); Filipović (2009) presents a continuous finance theory perspective. An approach that has gained wide acceptance in recent years is the Nelson–Siegel

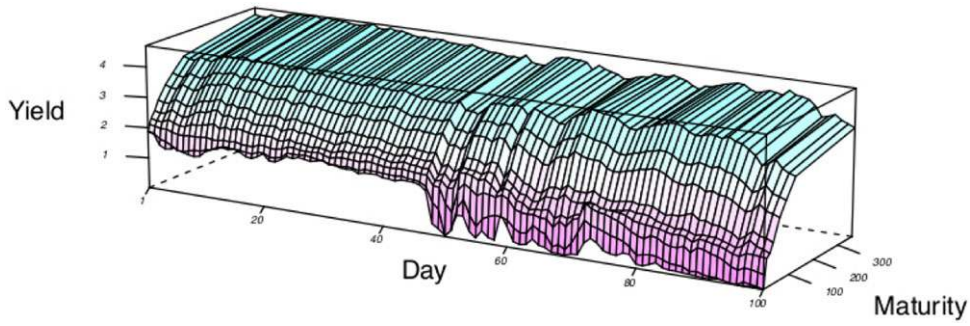

Figure 1. US yield curves: 8 July 2008 to 28 November 2008.

model and its dynamic modification; see Diebold and Rudebusch (2013). This paper is concerned with the detection of change points in models that generalize the dynamic Nelson–Siegel model. The most direct motivation for our research comes from the recent work of Chen and Niu (2014) who show that accounting for possible change points in the term structure improves yield curve predictions. The purpose of this work is to develop formal tests of significance for the existence of change points in a general class of functional models for yield curves. We now elaborate on our contribution.

Denote by t_j, $1 \leq j \leq J$, the maturities ordered from the shortest (one month) to the longest (10 years). The general form of the dynamic Nelson–Siegel model can be written as

$$X_i(t_j) = \sum_{k=1}^{K} \beta_{i,k} f_k(t_j) + \varepsilon_i(t_j). \tag{1.1}$$

The index i refers to time periods at which the curves are available; it typically indexes days or months. The functions f_k are postulated to have a specific parametric form; in the standard model, $K = 3$ functions are used. The function f_1 is equal to one, and its weight $\beta_{i,1}$ represents the general level of yields in period i. The function f_2 is decreasing, and for negative $\beta_{i,2}$ models the increase of yields with maturity. The function f_3 has a hump at maturities of two to three years, and models the curvature of the yield curve over such maturities; detailed formulae are presented in Section 5. The attribute 'dynamic' stems from the fact that the weights $\beta_{i,k}$ are time series; in a static model, $\beta_{i,k} = \beta_k$ does not depend on period i. The objective of this work is to develop significance tests whose null hypothesis is that the mean structure of the K series $\{\beta_{i,k}\}$ is constant over a time period under consideration against the alternative that it changes at unknown change points. Our approach allows the error structure to change at pre-specified points. Precise definitions of the mean and error structures are given in Section 2. Before proceeding further, we give an illustrative example. Figure 1 shows 100 yield curves centred at the height of the financial crisis of 2008; the central time point corresponds to the collapse of Lehman Brothers. Visual inspection shows the typical shape of the curves, which we quantify as the mean structure, changed in mid-September. If we apply our test assuming that the error structure is the same over the sample period, then the test does not reject the null hypothesis that the mean structure has not changed. If, however, the error structure is allowed to change on 16 September, then the test rejects the null hypothesis with a very high significance.

This example illustrates the need for flexibility in modelling the error structure when testing for change points in the mean structure. The pre-specified break points in the error structure reflect exogenous information. In Figure 1, the mean structure is reflected by the general level and range of the curves, and the error structure by the 'wiggliness' in the top, middle and bottom parts of the curves. The errors, volatility, of the curves increased after the middle of the sample. The change in mean structure is fairly obvious in Figure 1; this period is used to emphasize a main message of this paper that without taking instability in the error structure into account, the decision on the existence of change points in the mean structure can be incorrect.

Similar observations have been made and suitably quantified in the context of scalar time series. Zhou (2013) investigates properties of the CUSUM test when there are changes in the structure of the innovations. Our Assumptions 2.1 and 2.2 can be considered functional versions of the main conditions in Zhou (2013). A similar approach is taken by Dette et al. (2015) to detect a change in the correlation structure. Dalla et al. (2015) study the change point in the mean problem under heteroscedastic errors, but use the statistic of Giraitis et al. (2003). As in Zhou (2013), in our approach, the nonstationarity (or errors) is modelled by assuming stationarity on some segments. In Busetti and Taylor (2004), Cavaliere and Taylor (2008) and Cavaliere et al. (2011), the source of nonstationarity is a unit root in the error sequence. These papers are not concerned with testing the change in the mean structure; different types of changes are studied. We are not aware of research on change point inference in regression or factor models with heteroscedastic errors.

The fact that the stochastic structure of the yield curve changes has been recognized for some two decades. Using eigenvalue analysis, akin to the factor models we study, Rogers and Stummer (2000) provide evidence that there are periods of over which parameters remain constant. A currently established approach is to use hidden Markov chains to estimate the structural changes together with the parameters of the affine structure; see Nieh et al. (2010). In addition to the aforementioned paper of Chen and Niu (2014), the only other paper concerned with change point estimation is Chib and Kang (2013). The main difference between hidden Markov models and change point models is that in the former only a few (typically two) states are assumed and the system moves between them. In a change point model, no such assumption is made; the parameters can take any values between the change points. This is the paradigm advocated by Chib and Kang (2013) who use a Bayesian approach to estimate the change points. This paper is concerned with testing for the presence of change points. We propose frequentist procedures based on asymptotic distributions of test statistics.

The methodology and theory we develop does not depend on the specific form of the curves f_k, which we call factor curves, extending the usual terminology of multivariate statistics. Lengwiler and Lenz (2010), Hays et al. (2012) and Sen and Klüppelberg (2015), among others, argue that the standard Nelson–Siegel factor curves are not optimal in some respects. In our methodological and theoretical exposition, we merely assume that the f_k are square integrable and linearly independent. We then go a step further, and assume that the yield curves follow the model $X_i(t_j) = \tau_i(t_j) + \eta_i(t_j)$, where the curves τ_i describe the mean structure and η_i are error curves. We will derive tests that allow to test for the presence of change points in the form of the functions τ_i without any parametric assumptions.

The remainder of the paper is organized as follows. In Section 2, we specify the general functional factor model. Testing for the presence of change point in this model is addressed in Section 3. In Section 4, we extend our approach to the nonparametric setting. Finite sample performance of the proposed methods is studied by a simulation study in Section 5. The main conclusions are summarized in Section 6. The Appendix contains proofs of the asymptotic results

on which the tests are based. An algorithmic description of the procedures and details of their numerical implementation are presented in the online Appendix.

2. FUNCTIONAL FACTOR MODEL

In this section, we introduce the general functional factor model. In Section 4, we introduce the fully functional model, which does not assume a factor structure.

We consider the functional term structure model

$$X_i(t) = \sum_{k=1}^{K} \beta_{i,k} f_k(t) + \varepsilon_i(t), \qquad 1 \le i \le N. \tag{2.1}$$

This is the same model as (1.1), but the maturity t is modelled as a continuous variable, following the Nelson–Siegel paradigm. An empirical justification is that fractions of bonds with standard maturities can be traded at any time. However, reconciling the continuous time formulation (2.1) with the actual data model (1.1) requires some attention. The factor curves f_k are fundamentally continuous time functions. For example, the most commonly used Nelson–Siegel factors are built of exponential and rational functions. However, the data $X_i(t_j)$ are observed only at discrete times t_j, typically not more than 10 for yield curves, and up to 30 for FX forward rates. Lebesgue's L^2 Hilbert space provides a convenient mathematical framework in which theory can be developed, which covers both discretely observed data and continuous factor curves. Denoting $\Delta_j = t_j - t_{j-1}$, the inner product in this space is

$$\langle f, g \rangle = \sum_{j=1}^{J} f(t_j) g(t_j) \Delta_j = \int f(t) g(t) \nu_J(dt),$$

where ν_J is a counting measure. The norm generated by this inner product will be denoted by $\|\cdot\|_2$. The norm in the Euclidean space will be denoted by $\|\cdot\|$. We emphasize that our results remain valid if the counting measure ν_J is replaced by any other positive measure with respect to which all functions are square integrable and norms of random functions have sufficiently high moments, as specified in the statements that follow. In particular, the countable measure ν_J could be replaced by the Lebesgue measure, if a purely continuous time framework is desired.

We assume that the functions f_k are known. This corresponds to the currently accepted practice of using fixed Nelson–Siegel factors. In Section A.3 in the Appendix, we show that our results remain valid if only a parametric form is assumed, and the parameters are consistently estimated. We assume that the factors f_k are linearly independent, i.e. $\sum_{k=1}^{K} a_k f_k = 0$ in L^2 implies $a_1 = a_2 = \ldots = a_K = 0$. The assumption of linear independence is satisfied by any practically used system of factors, but is actually not needed to establish the asymptotic results of Section 3. It is used to justify our testing approach and to claim the consistency of the tests under arbitrary changes.

The random coefficients $\beta_{i,k}$ can be decomposed as

$$\beta_{i,k} = \mu_{i,k} + b_{i,k}, \qquad E[b_{i,k}] = 0.$$

We want to test the change point hypothesis

$$H_0 : \quad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_N, \tag{2.2}$$

where

$$\boldsymbol{\mu}_i = [\mu_{i,1}, \mu_{i,2}, \ldots, \mu_{i,K}]^\top.$$

Thus, we want to test if the first-order structure of the process (2.1) changes at some unknown points. Under the alternative, there are at most R possible changes in the mean structure at times

$$1 = r_0 < r_1 < r_2 < \ldots < r_R < r_{R+1} = N.$$

When testing for a change in the mean, it is generally assumed in change point analysis – see, e.g. Csörgő and Horváth (1997) and Horváth and Rice (2014) – that the second-order properties do not change. Even in the simplest case of independent normal observations, allowing for a change in both mean and variance leads to quite complex asymptotic theory; see Horváth (1993). In the setting of model (2.1), the error structure is captured by the terms $\sum_{k=1}^{K} b_{i,k} f_k(t) + \varepsilon_i(t)$. We allow their stochastic structure to change at specific points

$$1 = i_0 < i_1 < i_2 < \ldots < i_M < i_{M+1} = N.$$

In application to yield curves, i_m can be determined as dates of central bank intervention or times of events of economic impact. The dates i_m reflect available exogenous information and are treated as known. We refer to them as *break points*. By contrast, the change points r_ℓ are unknown, and their existence is to be tested. In Figure 1, the break point visually practically coincides with the change point. The break point is reflected by the higher variability of the yield curves in the second half of the sample. The top and bottom parts of the graph are more variable, and have larger errors. The change point is reflected by the wider range of the curves. In the Nelson–Siegel model, the decreasing curve f_2 describes the spread of yields between short and long maturities. For $\mu_2^* < \mu_2 < 0$, the yield curves containing $\mu_2^* f_2$ will have a larger spread that the curves containing $\mu_2 f_2$. This paper focuses on the formal testing framework in which the break points play a crucial role. It is hoped that it will stimulate empirical research on the selection of break points among the events that may potentially affect interest rate volatility.

We now formulate model assumptions. Introduce the error vectors

$$\mathbf{a}_i(t) = [b_{i,1}, \ldots, b_{i,K}, \varepsilon_i(t)]^\top, \qquad 1 \le i \le N. \tag{2.3}$$

The vectors $\mathbf{a}_i$ are stationary on each interval $(i_m, i_{m+1}]$, and have mean zero. Their dependence structure is described by the following assumption.

ASSUMPTION 2.1. *Assume that the vectors* $\mathbf{a}_i$ *defined by* (2.3) *admit the representation*

$$\mathbf{a}_i = g_m(\delta_i, \delta_{i-1}, \ldots), \qquad i_m < i \le i_{m+1}, \qquad m = 0, 1, \ldots, M,$$

where $g_0, g_1, \ldots, g_M : S^\infty \mapsto L^2$ *are unknown deterministic measurable functions. The random elements* $\{\delta_i, -\infty < i < \infty\}$ *are i.i.d. with values in a measurable space* S.

Broadly speaking, Assumption 2.1 requires that on the segments of stationarity are some abstract (nonlinear) moving averages of abstract errors. Representations of this type impose a very flexible dependence structure and have become popular over the last decade; see, e.g. Wu (2005), Shao and Wu (2007), Aue et al. (2009), Hörmann and Kokoszka (2010), Hörmann et al. (2013) and Kokoszka and Reimherr (2013). Observe that the same sequence $\{\delta_i\}$ is used in all functions g_m. This reflects the intuition that even though the stochastic structure can change from segment to segment, there is a dependence between the segments; Kokoszka and Leipus (2000), among others, used this paradigm in a change point problem for scalar ARCH models.

Next, we formalize the assumption that on each subinterval $(i_m, i_{m+1}]$ the sequence $\{\mathbf{a}_i, 1 \leq i \leq N\}$ is weakly dependent. We use the notion of approximability, which has been recently used in the analysis of time series of functions. Chapter 16 of Horváth and Kokoszka (2012) puts this notion in a historical context and provides a number of its applications. To formulate it in our context of segment-stationarity, we must extend the part-sequences $\{\mathbf{a}_i, i_m < i \leq i_{m+1}\}$ to the full domain of all integers. We thus denote by $\{\mathbf{a}_i^{(m)}\}$ the above sequence extended to all integers, i.e.

$$\mathbf{a}_i^{(m)} = g_m(\delta_i, \delta_{i-1}, \ldots), \qquad -\infty < i < \infty. \tag{2.4}$$

ASSUMPTION 2.2. *For some $\delta > 0$ and $\kappa > 2 + \delta$,*

$$\max_{1 \leq m \leq M+1} \sum_{\ell=1}^{\infty} \left(E[\|\mathbf{a}_{i,\ell}^{(m)} - \mathbf{a}_i^{(m)}\|^{2+\delta}]\right)^{1/\kappa} < \infty, \tag{2.5}$$

where $\mathbf{a}_{i,\ell}^{(m)}$ is defined by $\mathbf{a}_{i,\ell}^{(m)} = g_m(\delta_i, \delta_{i-1}, \ldots, \delta_{i-\ell+1}, \delta_{i-\ell}^*, \delta_{i-\ell-1}^*, \ldots)$ and δ_k^* are independent copies of δ_i, independent of $\{\delta_i, -\infty < i < \infty\}$.

The essence of Assumption 2.2 is that the impact of innovations δ_i far back in the past becomes negligible; replacing them by independent copies does not affect the distribution of $\mathbf{a}_i^{(m)}$ much. Condition (2.5) quantifies the magnitude of the effect of such a replacement. It allows us to control the remainder terms arising by replacing the sequence $\mathbf{a}_i^{(m)}$ by sequences consisting of variables that are independent for sufficiently large lags (m-dependent sequences). The arbitrary constants δ and κ in (2.5) are needed to guarantee that a weak approximation theorem of Berkes et al. (2013), which we use in our proofs, holds for every segment.

Our last assumption states that the segments of stationarity have asymptotically comparable lengths.

ASSUMPTION 2.3. *We assume that $i_m = i_m(N)$ and*

$$\lim_{N \to \infty} N^{-1} i_m(N) = \theta_m, \qquad 1 \leq m \leq M, \tag{2.6}$$

with $0 = \theta_0 < \theta_1 < \theta_2 < \ldots \theta_M < \theta_{M+1} = 1$.

3. DETECTION THROUGH PROJECTIONS ONTO FACTORS

The method presented in this section directly exploits representation (2.1). A method that does not use the factor structure is presented in Section 4.

Test statistics can be derived either using the vector of projections

$$\mathbf{z}_i = [\langle X_i, f_1 \rangle, \ldots, \langle X_i, f_K \rangle]^\top \tag{3.1}$$

or cumulative estimates $\hat{\boldsymbol{\mu}}_k$ computed using the first k observations. These two approaches are equivalent, as we now explain. We begin with the CUSUM process of the vectors (3.1). Introduce the matrix

$$\mathbf{C} = [\langle f_k, f_j \rangle, \ 1 \leq k, j \leq K].$$

The matrix $\mathbf{C}$ is deterministic and known. As f_k are linearly independent, the columns of $\mathbf{C}$ are linearly independent, so $\mathbf{C}^{-1}$ exists. By (2.1),

$$\mathbf{z}_i = \mathbf{C}\boldsymbol{\mu}_i + \boldsymbol{\gamma}_i, \tag{3.2}$$

where

$$\boldsymbol{\gamma}_i = \mathbf{C}\mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

and where

$$\mathbf{b}_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,K}]^\top, \qquad \boldsymbol{\varepsilon}_i = [\langle \varepsilon_i, f_1\rangle, \langle \varepsilon_i, f_2\rangle, \ldots, \langle \varepsilon_i, f_K\rangle]^\top.$$

Because the matrix $\mathbf{C}$ is deterministic and invertible, the relation (3.2) implies that a change in the vectors $\boldsymbol{\mu}_i$ is equivalent to a change in $\mathbf{z}_i$ at the same change points. Test statistics will thus be based on the CUSUM process:

$$\boldsymbol{\alpha}_N(x) = N^{-1/2}\Big(\sum_{i=1}^{[Nx]} \mathbf{z}_i - \frac{[Nx]}{N}\sum_{i=1}^{N} \mathbf{z}_i\Big), \qquad 0 \le x \le 1. \tag{3.3}$$

Another route that leads to the process (3.3) is through cumulative least-squares estimators of the vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_K]^\top$, which does not depend on i under H_0. The estimator based on the whole sample minimizes the least-squares criterion

$$U_N(\boldsymbol{\mu}) = U_N(\mu_1, \mu_2, \ldots, \mu_K) = \sum_{i=1}^{N} \|X_i - \sum_{k=1}^{K} \mu_k f_k\|_2^2.$$

It is given by $\hat{\boldsymbol{\mu}}_N = N^{-1}\mathbf{C}^{-1}\sum_{i=1}^{N} \mathbf{z}_i$. Denote by $\hat{\boldsymbol{\mu}}_k$ the estimator based on the first k functions, i.e. $\hat{\boldsymbol{\mu}}_k = k^{-1}\mathbf{C}^{-1}\sum_{i=1}^{k} \mathbf{z}_i$. Then

$$N^{-1/2}k(\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}}_N) = \mathbf{C}^{-1}\boldsymbol{\alpha}_N(kN^{-1}).$$

Thus, functionals of the process $N^{1/2}x(\hat{\boldsymbol{\mu}}_{[Nx]} - \hat{\boldsymbol{\mu}}_N)$, $0 \le x \le 1$, are the same as those of the process (3.3), up to the multiplication by a known deterministic matrix.

Our next goal is to specify the limit distribution of the process $\boldsymbol{\alpha}_N$. Notice that $\boldsymbol{\gamma}_i = \mathbf{w}(\mathbf{a}_i)$, where $\mathbf{w}$ is a known function and $\mathbf{a}_i$ is defined in (2.3). This allows us to introduce the $M + 1$ infinite domain stationary sequences

$$\boldsymbol{\gamma}_i^{(m)} = \mathbf{w}(\mathbf{a}_i^{(m)}), \qquad -\infty < i < \infty,$$

and to define their long-run covariance matrices

$$\mathbf{V}_m = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\boldsymbol{\gamma}_i^{(m)}, \boldsymbol{\gamma}_{i+\ell}^{(m)}).$$

THEOREM 3.1. *If H_0 (2.2) and Assumptions 2.1, 2.2 and 2.3 hold, then*

$$\boldsymbol{\alpha}_N \xrightarrow{d} \mathbf{G}^0, \qquad \text{in } D^K([0, 1]),$$

where the process $\mathbf{G}^0$ is defined by

$$\mathbf{G}^0(x) = \mathbf{G}(x) - x\mathbf{G}(1), \tag{3.4}$$

and $\mathbf{G}(x)$, $x \in [0, 1]$, *is a mean zero* $\mathbb{R}^K$*-valued Gaussian process with covariances*

$$E[\mathbf{G}(x)\mathbf{G}^\top(y)] = \sum_{j=1}^{m}(\theta_j - \theta_{j-1})\mathbf{V}_j + (x - \theta_m)\mathbf{V}_{m+1}, \qquad \theta_m \le x \le \theta_{m+1},\ y \ge x.$$

The covariances of the process $\mathbf{G}^0(x)$ can be computed explicitly. They are given by

$$\begin{aligned}\mathbf{Q}(x, y) &:= E[\mathbf{G}^0(x)\mathbf{G}^0(y)^\top] \\ &= (1 - y)\Big(\sum_{j=1}^{m}(\theta_j - \theta_{j-1})\mathbf{V}_j + (x - \theta_m)\mathbf{V}_{m+1}\Big) \\ &\quad -x\Big(\sum_{j=1}^{m'}(\theta_j - \theta_{j-1})\mathbf{V}_j + (y - \theta_{m'})\mathbf{V}_{m'+1}(t, s)\Big) \\ &\quad +xy\sum_{j=1}^{M+1}(\theta_j - \theta_{j-1})\mathbf{V}_j, \end{aligned} \tag{3.5}$$

where $0 \le x \le y \le 1$, $\theta_m \le x \le \theta_{m+1}$ and $\theta_{m'} \le y \le \theta_{m'+1}$.

Tests can be based on the Cramér–von-Mises functional

$$\mathcal{C}_N = \int_0^1 \|\boldsymbol{\alpha}_N(x)\|^2 dx, \tag{3.6}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^K$, or the Kolmogorov–Smirnov functional

$$\mathcal{K}_N = \sup_{0\le x\le 1} \|\boldsymbol{\alpha}_N(x)\|, \tag{3.7}$$

or their weighted versions. In finite samples, Cramér–von-Mises tests generally perform better, and we therefore focus on the statistic $\mathcal{C}_N$. By Theorem 3.1,

$$\mathcal{C}_N \xrightarrow{d} \int_0^1 \|\mathbf{G}^0(x)\|^2 dx. \tag{3.8}$$

To perform the test, we must simulate the distribution of the right-hand side of (3.8). We propose two methods. The first is based on the following general result.

PROPOSITION 3.1. *Let* $\boldsymbol{\Gamma}(x)$, $x \in [0, 1]$, *be a mean zero* $\mathbb{R}^K$*-valued Gaussian processes with covariances* $\mathbf{R}(x, y) = E[\boldsymbol{\Gamma}(x)\boldsymbol{\Gamma}(y)^\top]$. *Then,*

$$\int_0^1 \|\boldsymbol{\Gamma}(x)\|^2 dx \overset{d}{=} \sum_{j=1}^{\infty}\lambda_j Z_j^2, \tag{3.9}$$

where Z_j *are independent standard normal random variables and* λ_j *are the eigenvalues of covariance kernel* $\mathbf{R}(\cdot, \cdot)$, *i.e.*

$$\int_0^1 \mathbf{R}(x, y)\boldsymbol{\phi}_j(y)dy = \lambda_j\boldsymbol{\phi}_j(x), \tag{3.10}$$

where $\phi_j(x)$ *are orthonormal eigenfunctions defined on the unit interval and taking values in* $\mathbb{R}^K$.

Proposition 3.1 follows from the Karhunen–Loéve decomposition of a Gaussian element in a separable Hilbert space. To enhance the understanding of this result, we present a proof in the Appendix. It shows that to approximate the distribution of $\int_0^1 \|\mathbf{G}^0(x)\|^2 dx$, it is enough to compute the eigenvalues λ_j in (3.10), with $\mathbf{Q}$ given by (3.5) in place of $\mathbf{R}$. To this end, we must estimate the long-run covariance matrices $\mathbf{V}_m$. These estimates are also needed in the second method of approximating the limit in (3.8), which we now discuss.

The second method is based on generating replications of the process $\mathbf{G}^0$. By (3.4), this reduces to generating replications of the process $\mathbf{G}$. In the course of the proof of Theorem 3.1, it is shown that

$$\mathbf{G}(x) = \sum_{j=1}^{m} (\theta_j - \theta_{j-1})^{1/2} \mathbf{G}_j(1) + (\theta_{m+1} - \theta_m)^{1/2} \mathbf{G}_{m+1}\Big(\frac{x - \theta_m}{\theta_{m+1} - \theta_m}\Big), \qquad x \in (\theta_m, \theta_{m+1}]. \tag{3.11}$$

Each process $\mathbf{G}_j$ is a mean zero Gaussian process with $E[\mathbf{G}_j(x)\mathbf{G}_j(y)^\top] = \min(x, y)\mathbf{V}_j$. It can be simulated as $\mathbf{G}_j(x) = \mathbf{L}_j \mathbf{W}(x)$, where $\mathbf{L}_j \mathbf{L}_j^\top = \mathbf{V}_j$ and $\mathbf{W} = [W_1, W_2, \ldots, W_K]^\top$ consists of K independent standard Wiener processes. The decomposition of the long-run variance matrix uses either the upper or lower Cholesky decomposition. This representation always exists because $\mathbf{V}_j$ is non-negative definite. In order to simulate the Gaussian processes $\mathbf{G}_j$, we must compute the estimated long-run covariance matrices $\widehat{\mathbf{V}}_j$, for which computationally efficient `R` implementations exist; details are described in Section 2 of the online Appendix.

We summarize the above discussion in the following corollary.

COROLLARY 3.1. *Denote by $\hat{c}_N(\alpha)$ the level α critical value obtained by any of the two methods proposed above. Under the assumptions of Theorem 3.1,*

$$\lim_{N\to\infty} P\{\mathcal{C}_N \geq \hat{c}_N(\alpha)\} = \alpha.$$

We conclude this section by stating the consistency of the test based on convergence (3.8). To keep the statement simple, we consider only one change point, but it can be shown by the same technique, merely with a more complex notation, that the statistic $\mathcal{C}_N$ diverges if there are more than one change points. Thus, we assume that for some $r \in (0, 1)$,

$$\boldsymbol{\mu} = \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_{[Nr]}, \qquad \boldsymbol{\mu}^* = \boldsymbol{\mu}_{[Nr]+1} = \ldots = \boldsymbol{\mu}_N. \tag{3.12}$$

We allow the size of the change to depend on the sample size N, and we assume that for some sequence $\{a_N\}$ and a vector $\mathbf{u} \neq \mathbf{0}$,

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + a_N \mathbf{u}. \tag{3.13}$$

THEOREM 3.2. *Suppose Assumptions 2.1, 2.2 and 2.3 and relations* (3.12) *and* (3.13) *hold: (a) if*

$$\lim_{N\to\infty} N^{1/2} a_N = \infty, \tag{3.14}$$

then

$$\Big(\frac{1}{N^{1/2} a_N}\Big)^2 \mathcal{C}_N \xrightarrow{p} \|\mathbf{C}\mathbf{u}\|^2 \int_0^1 \{g^*(x, r)\}^2 dx, \tag{3.15}$$

with the function g^* defined in (A.6); (b) if $a_N = N^{-1/2}$, then with $\bar{\mathbf{G}}^0(x) = \mathbf{G}^0(x) - g^*(x, r)\mathbf{Cu}$, $0 \le x \le 1$, we have $\boldsymbol{\alpha}_N \xrightarrow{d} \bar{\mathbf{G}}^0$ in $D^K([0, 1])$. Hence

$$\mathcal{C}_N \xrightarrow{d} \int_0^1 \|\bar{\mathbf{G}}^0(x)\|^2 dx.$$

Relations (3.14) and (3.15) imply that $\mathcal{C}_N \xrightarrow{p} \infty$, provided $\mathbf{Cu}$ is not a zero vector. This is the case if $\mathbf{C}^{-1}$ exists, i.e. if f_k are linearly independent. Because the limit distribution in (3.8) does not depend on the vectors $\boldsymbol{\mu}_i$, it follows that at any fixed significance level the probability of rejection approaches one, as $N \to \infty$. It follows from Anderson (1955) that

$$P\Big\{\int_0^1 \|\bar{\mathbf{G}}^0(x)\|^2 dx \ge t\Big\} \ge P\Big\{\int_0^1 \|\bar{\mathbf{G}}(x)\|^2 dx \ge t\Big\} \qquad \text{for all } t,$$

and therefore the level of rejection is at least α when $a_N = N^{-1/2}$.

We summarize this discussion in the following corollary.

COROLLARY 3.2. *Denote by $\hat{c}_N(\alpha)$ the level α critical value obtained by any of the two methods proposed in this section. Under the assumptions of part (a) of Theorem 3.2,*

$$\lim_{N\to\infty} P\{\mathcal{C}_N \ge \hat{c}_N(\alpha)\} = 1.$$

Under the assumptions of part (b) of Theorem 3.2

$$\lim_{N\to\infty} P\{\mathcal{C}_N \ge \hat{c}_N(\alpha)\} \ge \alpha.$$

4. A NONPARAMETRIC FUNCTIONAL APPROACH

In this section, we consider a testing procedure that does not assume model (2.1). To motivate it, we rewrite model (2.1) as

$$X_i(t) = \tau_i(t) + \eta_i(t), \qquad 1 \le i \le N, \tag{4.1}$$

where

$$\tau_i(t) = \sum_{k=1}^{K} \mu_{i,k} f_k(t), \qquad \eta_i(t) = \sum_{k=1}^{K} b_{ik} f_k(t) + \varepsilon_i(t). \tag{4.2}$$

In (4.2), the mean functions τ_i and the error functions η_i are expressed in terms of the components of model (2.1). However, such a specific form is not assumed in (4.1). Functional factor models, and earlier affine models, postulate some form of parametric dependence of the yield curve on a small number of parameters. Formulation (4.1) can be viewed as a nonparametric model emphasizing the main first-order structure described by the functions τ_i, which are all equal to a function τ under H_0. The form of the function τ can be arbitrary. It can be estimated by nonparametric methods, but our focus here is on testing if it does not change. In this setting, the null hypothesis becomes

$$H_0: \quad \tau_1 = \tau_2 = \ldots = \tau_N, \tag{4.3}$$

with an alternative formulated analogously as in Section 2. Observe that under model (2.1), in light of (4.2), the null hypothesis (4.3) is equivalent to the null hypothesis (2.2).

The problem of testing for a change point in the functional means τ_i, without any reference to yield curves, was addressed by Berkes et al. (2009) who assumed that the error curves η_i are i.i.d. Hörmann and Kokoszka (2010) extended that test by allowing to η_i to be stationary and weakly dependent. In both cases, the tests are based on projections of the data on the estimated functional principal components of the errors η_i.[1] In our setting, η_i are not stationary; the functional principal components and, hence, the corresponding projections cannot be defined. It is however possible to derive tests based directly on the functional CUSUM process:

$$\alpha_N(x,t) = \frac{1}{\sqrt{N}}\Big(\sum_{1=1}^{[Nx]} X_i(t) - \frac{[Nx]}{N}\sum_{i=1}^{N} X_i(t)\Big), \qquad 0 \le x \le 1, \quad (t \in [0,1]). \tag{4.4}$$

In the remainder of this section, we develop the required theory and derive the tests.

We begin by stating assumptions analogous to Assumptions 2.1 and 2.2. The error functions η_i are mean zero and form stationary sequences on the intervals $(i_m, i_{m+1}]$. By $\eta_i^{(m)}$ we denote their extensions to the infinite domain consisting of all integers.

ASSUMPTION 4.1. *The functions η_i in* (4.1) *have mean zero and admit the representation*

$$\eta_i = g_m(\delta_i, \delta_{i-1}, \ldots), \qquad i_m < i \le i_{m+1}, \quad m = 0, 1, \ldots, M,$$

where the functions g_m and the errors δ_i satisfy conditions of Assumption 2.1.

ASSUMPTION 4.2. *For some $\delta > 0$ and $\kappa > 2 + \delta$,*

$$\max_{1\le m\le M+1} \sum_{\ell=1}^{\infty} \big(E[\|\eta_{i,\ell}^{(m)} - \eta_i^{(m)}\|_2^{2+\delta}]\big)^{1/\kappa} < \infty, \tag{4.5}$$

where $\eta_{i,\ell}^{(m)}$ is defined by $\eta_{i,\ell}^{(m)} = g_m(\delta_i, \delta_{i-1}, \ldots, \delta_{i-\ell+1}, \delta_{i-\ell}^*, \delta_{i-\ell-1}^*, \ldots)$ and δ_k^* are independent copies of δ_i, independent of $\{\delta_i, -\infty < i < \infty\}$.

Consider the long-run covariance kernels defined by

$$D_m(t,s) = \sum_{\ell=-\infty}^{\infty} E[\eta_0^{(m)}(t)\eta_\ell^{(m)}(s)], \qquad 0 \le t, s \le 1. \tag{4.6}$$

The existence of the L^2-limit $D_m(\cdot,\cdot)$ was established by Horváth et al. (2013). The following theorem is an analogue of Theorem 3.1.

THEOREM 4.1. *Under H_0* (4.3) *and Assumptions 4.1, 4.2 and 2.3, we can define Gaussian processes $\Gamma_N^0(x,t)$, $0 \le x, t \le 1$, such that*

$$\sup_{0\le x\le 1} \|\alpha_N(x,\cdot) - \Gamma_N^0(x,\cdot)\|_2 \xrightarrow{p} 0.$$

Each process Γ_N^0 is defined by

$$\Gamma_N^0(x,t) = \Gamma_N(x,t) - x\Gamma_N(1,t),$$

[1] Both tests are described, respectively, in Chapters 6 and 16 of Horváth and Kokoszka (2012).

where Γ_N *is mean zero Gaussian with covariances*

$$E[\Gamma_N(x,t)\Gamma_N(y,s)] = \sum_{j=1}^{m}(\theta_j - \theta_{j-1})D_j(t,s) + (x-\theta_m)D_{m+1}(t,s),\ \theta_m \le x \le \theta_{m+1}, x \le y.$$

The covariances $E[\Gamma_N^0(x,t)\Gamma_N^0(y,s)]$ can be computed explicitly. Assuming

$$0 \le x \le y \le 1, \qquad \theta_m \le x \le \theta_{m+1}, \qquad \theta_{m'} \le y \le \theta_{m'+1},$$

$$\begin{aligned}
E[\Gamma_N(x,t)\Gamma_N(1,s)] &= \sum_{j=1}^{m}(\theta_j - \theta_{j-1})D_j(t,s) + (x-\theta_m)D_{m+1}(t,s),\\
E[\Gamma_N(y,s)\Gamma_N(1,t)] &= \sum_{j=1}^{m'}(\theta_j - \theta_{j-1})D_j(t,s) + (y-\theta_{m'})D_{m'+1}(t,s),\\
E[\Gamma_N(1,s)\Gamma_N(1,t)] &= \sum_{j=1}^{M+1}(\theta_j - \theta_{j-1})D_j(t,s). \qquad (4.7)
\end{aligned}$$

Assuming (4.7), we thus obtain

$$\begin{aligned}
U^0(x,y;t,s) &:= E[\Gamma_N^0(x,t)\Gamma_N^0(y,s)]\\
&= E[(\Gamma_N(x,t) - x\Gamma_N(1,t))(\Gamma_N(y,s) - y\Gamma_N(1,s))]\\
&= (1-y)\Big(\sum_{j=1}^{m}(\theta_j - \theta_{j-1})D_j(t,s) + (x-\theta_m)D_{m+1}(t,s)\Big)\\
&\quad -x\Big(\sum_{j=1}^{m'}(\theta_j - \theta_{j-1})D_j(t,s) + (y-\theta_{m'})D_{m'+1}(t,s)\Big)\\
&\quad +xy\sum_{j=1}^{M+1}(\theta_j - \theta_{j-1})D_j(t,s). \qquad (4.8)
\end{aligned}$$

The complex structure of U^0 is what distinguishes the present research from the methods developed by Berkes et al. (2009) and Hörmann and Kokoszka (2010). The later work can be considered as a special case, in which $D_1 = D_2 = \ldots = D_{M+1}(= D)$. In that case, $U^0(x,y;t,s) = (\min(x,y) - xy)D(t,s)$. Consequently, new approaches are needed to implement tests based on Theorem 4.1, even though they use the usual functionals. Denote by $\{\Gamma^0(x,t), 0 \le x,t \le 1\}$ a process with the same distribution as each Γ_N^0. As in Section 3, we focus on the Cramér–von-Mises functional

$$\mathcal{V}_N = \int_0^1 \int \alpha_N^2(x,t)\nu_J(dt)\,dx, \qquad (4.9)$$

and the convergence

$$\mathcal{V}_N \xrightarrow{d} \int_0^1 \int (\Gamma^0(x,t))^2 \nu_J(dt)\,dx. \tag{4.10}$$

There is no explicit formula for the distribution of the limit in (4.10). It depends on the unknown long-run covariance kernels D_j, $1 \le j \le M+1$, whose estimation is discussed below. Once these estimates are computed, the approximation of the right-hand side of (4.10) is based on the relation

$$\int_0^1 \int (\Gamma^0(x,t))^2 \nu_J(dt)\,dx = \sum_{j=1}^{\infty} \lambda_j Z_j^2, \tag{4.11}$$

where Z_j are i.i.d. standard normal and λ_j are the eigenvalues of U^0 given by (4.8). If $\widehat{U}^0$ is an estimator of U^0, then λ_j are approximated by the eigenvalues $\hat{\lambda}_j$ defined by

$$\int_0^1 \int \widehat{U}^0(x,y;t,s)\hat{\varphi}_j(y,s)\nu_J(ds)\,dy = \hat{\lambda}_j \hat{\varphi}_j(x,t),$$

and $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ by $\sum_{j=1}^{N} \hat{\lambda}_j Z_j^2$. The numerical computation of $\hat{\lambda}_j$ is not trivial; details are described in Section 2 of the online Appendix.

We now turn to the estimation of the covariance kernels D_j. We describe the method proposed by Horváth et al. (2013). Set $N_j = i_j - i_{j-1}$. Before defining the sample covariance kernel $\widehat{D}_{j,N_j}$, we need to introduce more notation. Let $X_{i_{j-1}+1}, X_{i_{j-1}+2}, \ldots, X_{i_j}$ denote the jth subset of observations. Define the nth residual of the jth subset by

$$e_{j,n}(t) = X_{i_{j-1}+n}(t) - \bar{X}_{N_j}(t), \quad 1 \le n \le N_j, \quad 1 \le j \le M,$$

where $\bar{X}_{N_j}(t)$ is the subset's sample mean defined by

$$\bar{X}_{N_j}(t) = \frac{1}{N_j}(X_{i_{j-1}+1} + X_{i_{j-1}+2} + \ldots + X_{i_j}).$$

The jth subset's autocovariances are defined by

$$\hat{\gamma}_{j,\ell,N_j}(t,s) = \frac{1}{N_j} \sum_{i=\ell+1}^{N_j} e_{j,i}(t) e_{j,i-\ell}(s), \qquad 1 \le j \le M.$$

We then define the jth subset's long-run kernel estimator by

$$\widehat{D}_{j,N_j}(t,s) = \hat{\gamma}_{j,0,N_j}(t,s) + \sum_{\ell=1}^{N_j-1} K\Big(\frac{\ell}{h}\Big)(\hat{\gamma}_{j,\ell,N_j}(t,s) + \hat{\gamma}_{i,\ell,N_j}(s,t)). \tag{4.12}$$

Using Theorem 2 of Horváth et al. (2013), it is easy to verify that $\|\widehat{U}_N^0 - U^0\| \xrightarrow{p} 0$, where $\widehat{U}_N^0$ is defined analogously to U^0 (4.8) with D_j replaced by $\widehat{D}_{j,N_j}$ and θ_j by $N^{-1}i_j$. The kernel K must satisfy the following assumption.

ASSUMPTION 4.3. *The function K is continuous, bounded, $K(0) = 1$ and $K(u) = 0$ if $|u| > c$, for some $c > 0$. The smoothing bandwidth $h = h(N)$ satisfies $h(N) \to \infty$, $N^{-1}h(N) \to 0$, as $N \to \infty$.*

Table 1. Testing procedures.

Method	Description
ProjSim	Projections onto factors approach of Section 3; limit approximated by simulating process (3.11)
ProjEigen	Projections onto factors approach of Section 3; limit approximated by simulating the right-hand side of (3.9)
NFEigen	Nonparametric functional approach of Section 4; limit approximated by simulating the right-hand side of (4.11)

We conclude this section with a consistency result proven in Section A.2. We assume that there is a change point $r \in (0, 1)$ such that $\tau \neq \tau^*$, where $\tau = \tau_1 = \ldots = \tau_{[Nr]}$, $\tau^* = \tau_{[Nr]+1} = \ldots = \tau_N$ and

$$\tau^*(t) = \tau(t) + a_N u(t), \tag{4.13}$$

with some nonzero function u.

THEOREM 4.2. *Suppose Assumptions 2.3, 4.1 and 4.2, and relation* (4.13), *hold: (a) if* (3.14) *holds, then*

$$\Big(\frac{1}{N^{1/2}a_N}\Big)^2 \mathcal{V}_N \xrightarrow{p} \|u\|_2^2 \int_0^1 (g^*(x, r))^2 dx, \tag{4.14}$$

with the function g^* *defined in* (A.6)*; (b) if* $a_N = N^{-1/2}$, *then with* $\bar{G}_N^0(x, t) = G_N^0(x) - g^*(x, r)u(t)$,

$$\sup_{0\leq x\leq 1} \|\alpha_N(x, \cdot) - \bar{G}_N^0(x, \cdot)\|_2 \xrightarrow{p} 0,$$

and hence

$$\mathcal{V}_N \xrightarrow{d} \int_0^1 \|\bar{G}_N^0(x)\|_2^2 dx.$$

It is interesting to compare Theorems 3.2 and 4.2. Writing the jump vector $\mathbf{u}$ in (3.13) as $\mathbf{u} = [u_1, u_2, \ldots, u_K]^\top$, assume that $u(t) = \sum_{k=1}^K u_k f_k(t)$. Under this assumption, $\|u\|_2^2 = \mathbf{u}^\top \mathbf{C}\mathbf{u}$. Because $\|\mathbf{C}\mathbf{u}\|^2 = \mathbf{u}^\top \mathbf{C}^2 \mathbf{u}$, the limits in Theorems 3.2 and 4.2 are the same, if $\mathbf{C}$ is the identity matrix (i.e. if $f_1, f_2, \ldots, f_K$ are orthonormal).

Analogues of Corollaries 3.1 and 3.2 can be stated for the test of this section.

For ease of reference, we list in Table 1 the procedures we implemented numerically.

5. FINITE SAMPLE PERFORMANCE

In this section, we assess the empirical size and power of the procedures introduced in Sections 3 and 4. We emphasize the importance of incorporating a break point in the error structure. This section is not meant to be an extensive empirical study of yield curves, but rather a study of the proposed statistical methods. However, we take care to use simulated data that closely resemble actual yield curves. We work with zero coupon US yield curves defined at

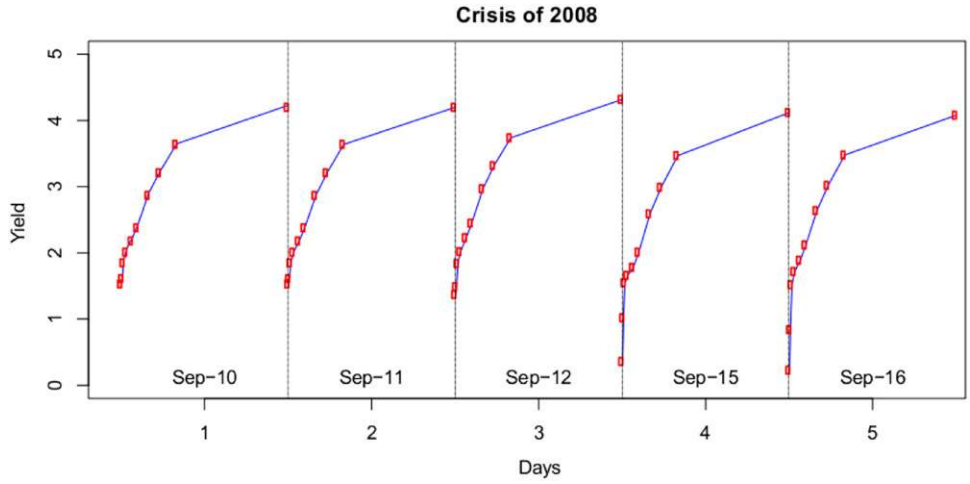

Figure 2. US yield curves around the bankruptcy of Lehman Brothers.

maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months. The data set has been obtained from a Federal Reserve web site.[2] Figure 2 shows five consecutive yield curves on five business days around the bankruptcy of Lehman Brothers on 15 September 2008.

We begin by illustrating the behaviour of the methods using three sampling periods listed in the first column of Table 2. In the first two periods, whose central 100 days are shown in Figures 1 and 3, respectively, we expect a change point in the mean structure. In the third period (see Figure 4), we do not expect such a change point. The periods were chosen in such a way that the potential break point in the error structure is in the middle of the sampling period. To perform the tests, we assume either no break point (designation 'no' in the 'Break point' column of Table 2) or one break point in the middle of the sample ($\theta_1 = 1/2$, designation 'yes'). We emphasize that a break point in the error structure should be viewed as an option in the application of the tests; we test for change points in the 'main' mean structure.

Table 2 shows that a likely change point in the mean structure may not be detected if a break point in the error structure is not taken into account. We cannot be sure which dates correspond to actual change points, but this finding is confirmed by the simulation study that we now describe.

As simulated data we use realizations of the dynamic Nelson–Siegel model

$$X_n(t) = \beta_{i,1} f_1(t, \lambda) + \beta_{i,2} f_2(t, \lambda) + \beta_{i,3} f_3(t, \lambda) + \varepsilon_i(t), \tag{5.1}$$

where

$$f_1(t, \lambda) = 1, \qquad f_2(t, \lambda) = \frac{1 - e^{-\lambda t}}{\lambda t}, \qquad f_3(t, \lambda) = \frac{1 - e^{-\lambda t}}{\lambda t} - e^{-\lambda t}. \tag{5.2}$$

Curves (5.2) are shown in Figure 5. The left panel shows curves with $\lambda = 0.0609$ corresponding to the domain of real yield curves, and the right panel shows curves with $\lambda = 21.52$ corresponding to the unit interval. The value of the parameter λ is chosen to maximize $f_3(t, \lambda)$

[2] See https://www.federalreserve.gov/releases/h15/data.htm.

Table 2. Yield curve P-values.

Sampling period	Method	Break point	P-value
(1) 20/03/2008–19/03/2009	ProjSim	Yes	1.5%
	ProjEigen	Yes	1.7%
	NFEigen	Yes	0.1%
	ProjSim	No	87.9%
	ProjEigen	No	85.2%
	NFEigen	No	26.2%
(2) 30/06/2005–29/06/2006	ProjSim	Yes	0.1%
	ProjEigen	Yes	0.0%
	NFEigen	Yes	0.2%
	ProjSim	No	56.7%
	ProjEigen	No	50.5%
	NFEigen	No	57.2%
(3) 16/02/2012–14/02/2013	ProjSim	Yes	68.1%
	ProjEigen	Yes	66.9%
	NFEigen	Yes	55.8%
	ProjSim	No	80.4%
	ProjEigen	No	77.7%
	NFEigen	No	76.3%

Note: Application of the test procedures to yield curves over three sampling periods. We expect small P-values in periods (1) and (2), and large P-values in period (3).

at the maturity of 30 months; see Diebold and Li (2003). Because we simulate data on the rescaled interval [0, 1], we use $\lambda = 21.5194$, which maximizes $f_3(t, \lambda)$ at $t = 30/360 = 0.0833$. To assess the sensitivity of the results to the specific form of the factors, we also performed simulations using factor curves $f_1(t) = 1$, $f_2(t) = t$ and $f_3(t) = t(1 - t)$. The properties of the methods and the empirical rejection rates are very similar, so we do not include the additional tables.

The coefficients $\beta_{i,k}$ are generated as AR(1) processes

$$\beta_{i,k} = \mu_k(1 - \varphi_k) + \varphi_k \beta_{i-1,k} + u_{i,k}, \qquad u_{i,k} \sim N(0, \sigma_k^2), \quad i = 1, \ldots, N, \quad k = 1, 2, 3.$$

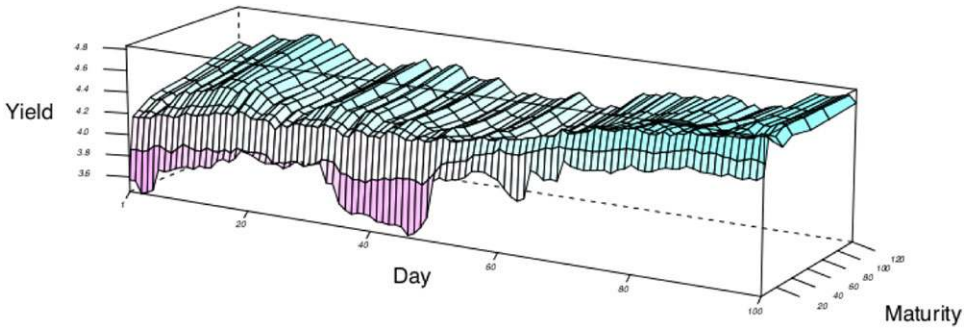

Figure 3. Yield curves: 18 October 2005 to 14 March 2006.

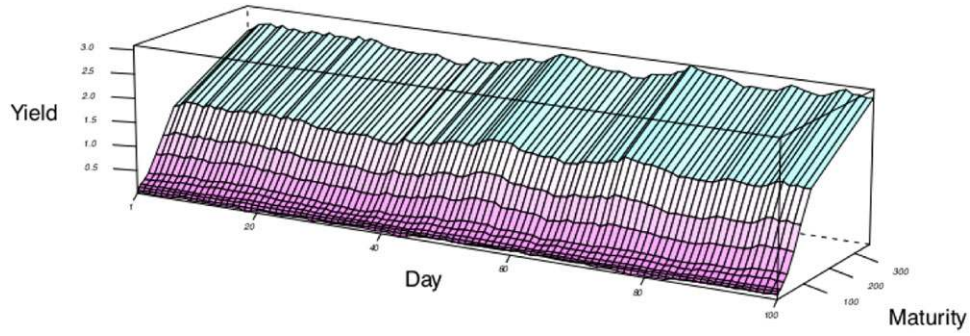

Figure 4. Yield curves: 4 June 2012 to 24 October 2012.

To resemble real data, the values of model parameters used in simulations are obtained as follows. For a specified sampling period, we compute least-squares estimates of $\beta_{i,1}$, $\beta_{i,2}$ and $\beta_{i,3}$ for each day i. This is the approach recommended by Diebold and Rudebusch (2013). Treating these estimates as a realization of an AR(1) processes, we estimate μ_k, φ_k and σ_k^2 by maximum likelihood. Following Bech and Lengwiler (2012), the presence of a break point in the error structure is simulated by using different AR(1) error variances σ_k^2 before and after the break point. Finally, the error curves ε_i are simulated as random curves

$$\varepsilon_i(t) = \frac{2}{25}\zeta_{i1} + \frac{1}{25}\zeta_{i2}\sin(2\pi t), \qquad t \in [0, 1], \tag{5.3}$$

chosen so that they are of the same size as the actual residuals; see Figure 6, where the left panel shows the residuals of the dynamic Nelson–Siegel model estimated over $N = 250$ business days from 20 March 2008 to 19 March 2008 and the right panel shows $N = 100$ error curves simulated using (5.3). The series $\zeta_{i,j}$ are autoregressions defined by $\zeta_{i,j} = 0.9\zeta_{i-1,j} + Z_{i,j}$,

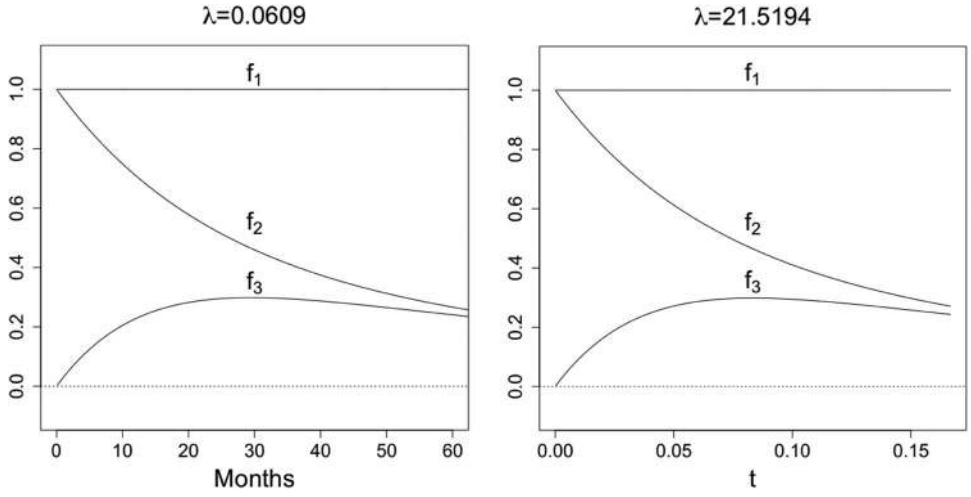

Figure 5. Nelson–Siegel factors $f_1(t, \lambda)$, $f_2(t, \lambda)$ and $f_3(t, \lambda)$.

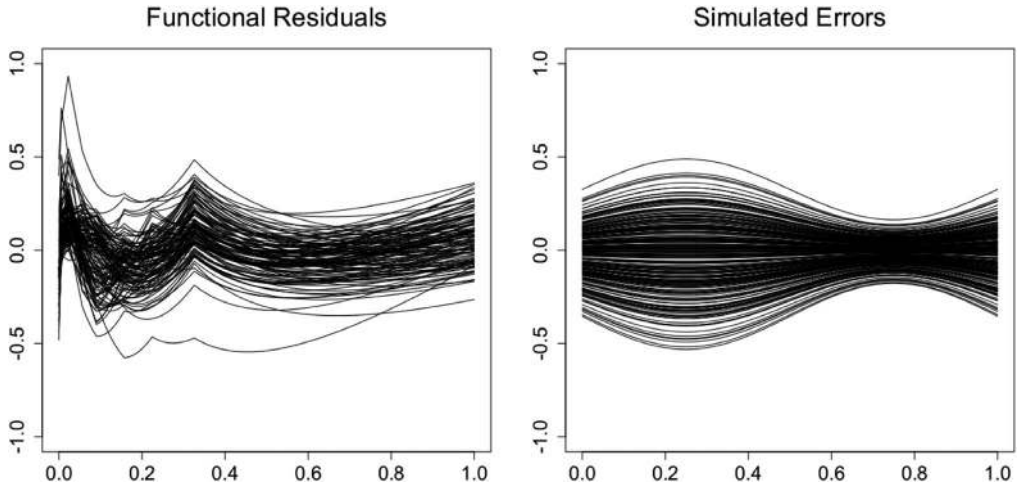

Figure 6. Dynamic Nelson–Siegel model residuals and simulated error curves.

$Z_{i,j} \sim N(0, 1)$, $j = 1, 2$, $i = 1, \ldots, N$. The parameters chosen above represent well the error curves over periods where no break point in the error structure is visible.

Using the approach described above, it is possible to specify a large number of realistic data-generating processes. We estimated several segments consisting of 250 and 125 consecutive yield curves. The following values are fairly representative of values obtained for many time periods, and we use them to simulate data.

AR(1) coefficients: $\varphi_1 = 0.90$, $\varphi_2 = 0.90$, $\varphi_3 = 0.90$.

AR(1) error variances:

$$\text{Var}[u_{i,1}] = 0.003,\ i \le i_1, \qquad \text{Var}[u_{i,1}] = 0.012,\ i > i_1,$$

$$\text{Var}[u_{i,2}] = 0.006,\ i \le i_1, \qquad \text{Var}[u_{i,2}] = 0.026,\ i > i_1,$$

$$\text{Var}[u_{i,3}] = 0.063,\ i \le i_1, \qquad \text{Var}[u_{i,3}] = 0.095,\ i > i_1.$$

We use two different locations of the break point θ_1, 1/2 and 2/3, and the following mean structures.
Means under H_0:

$$\boldsymbol{\mu} = \begin{pmatrix} 4.54 & -2.82 & -3.03 \end{pmatrix}^{\top}. \tag{5.4}$$

Means under $H_{A(1)}$:

$$\boldsymbol{\mu}_i = \begin{pmatrix} 4.54 \\ -2.82 \\ -3.03 \end{pmatrix},\ i \le N/2; \qquad \boldsymbol{\mu}_i = \begin{pmatrix} 4.20 \\ -3.00 \\ -3.20 \end{pmatrix},\ i > N/2. \tag{5.5}$$

Table 3. Empirical size and power for ProjSim.

Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$	$H_{A(3)}$
		$CP = BP$				
Yes	5%	$N = 250$	7.8%	61.1%	97.3%	79.6%
Yes	5%	$N = 500$	5.2%	84.6%	100.0%	97.4%
Yes	10%	$N = 250$	11.6%	70.7%	98.1%	87.3%
Yes	10%	$N = 500$	8.5%	90.3%	100.0%	99.0%
No	5%	$N = 250$	1.6%	9.8 %	2.8%	8.1%
No	5%	$N = 500$	2.8%	50.8%	39.9%	64.8%
No	10%	$N = 250$	5.7%	25.6%	12.9%	27.8%
No	10%	$N = 500$	6.3%	73.7%	77.1%	88.2%
		$CP \neq BP$				
Yes	5%	$N = 250$	6.8%	65.8%	98.0%	85.2%
Yes	5%	$N = 500$	5.1%	87.5%	100.0%	98.6%
Yes	10%	$N = 250$	12.4%	75.9%	98.9%	90.7%
Yes	10%	$N = 500$	8.7%	91.5%	100%	99.6%
No	5%	$N = 250$	0.9%	7.8%	0.8%	6.5%
No	5%	$N = 500$	2.6%	48.7%	25.2%	59.7%
No	10%	$N = 250$	4.3%	23.3%	7.3%	24.5%
No	10%	$N = 500$	5.3%	72.5%	61.2%	87.3%

Means under $H_{A(2)}$:

$$\boldsymbol{\mu}_i = \begin{pmatrix} 4.54 \\ -2.82 \\ -3.03 \end{pmatrix},\ i \leq N/2; \qquad \boldsymbol{\mu}_i = \begin{pmatrix} 3.89 \\ -3.32 \\ -3.32 \end{pmatrix},\ i > N/2. \tag{5.6}$$

Means under $H_{A(3)}$:

$$\begin{aligned} \boldsymbol{\mu}_i &= \begin{pmatrix} 4.54 \\ -2.82 \\ -3.03 \end{pmatrix},\ i \leq N/2; \qquad \boldsymbol{\mu}_i = \begin{pmatrix} 4.2 \\ -3.0 \\ -3.2 \end{pmatrix},\ N/2 < i \leq 3N/4; \\ \boldsymbol{\mu}_i &= \begin{pmatrix} 4.2 \\ -3.1 \\ -3.0 \end{pmatrix},\ 3N/4 < i \leq N. \end{aligned} \tag{5.7}$$

The alternative $H_{A(2)}$ is a larger departure from H_0 than $H_{A(1)}$. Based on exploratory analysis for several time periods, $H_{A(2)}$ somewhat exaggerates the change points that can be expected in real data; $H_{A(1)}$ is very realistic. For the simulation study, we investigate how the power behaves when the change point coincides with the break point ($CP = BP$), and when the change point is different from the break point ($CP \neq BP$). In the first scenario, we use $CP = BP = N/2$, and in the second $CP = N/2$ and $BP = 2N/3$. Under $H_{A(3)}$, there are two change points, so $CP = BP$ means that the break point is at $\theta_1 = 1/2$ and $CP \neq BP$ means that $\theta_1 = 2/3$.

The empirical rejection rates, based on 1,000 replications are displayed in Tables 3, 4 and 5. For $N = 500$, if a break point is taken into account, all methods have the correct size, within the chance error. For $N = 250$, the size is overinflated. As preliminary examples indicated,

Table 4. Empirical size and power for ProjEigen.

Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$	$H_{A(3)}$
		$CP = BP$				
Yes	5%	$N = 250$	8.5%	68.7%	99.2%	85.2%
Yes	5%	$N = 500$	5.1%	88.6%	100%	98.3%
Yes	10%	$N = 250$	13.7%	76.7%	99.5%	90.4%
Yes	10%	$N = 500$	9.4%	92.3%	100%	99.5%
No	5%	$N = 250$	2.6%	19.3%	12.6%	25.7%
No	5%	$N = 500$	3.4%	68.5%	81.5%	86.7%
No	10%	$N = 250$	8.0%	43.3%	41.7%	53.0%
No	10%	$N = 500$	7.3%	84.9%	97.6%	96.4%
		$CP \neq BP$				
Yes	5%	$N = 250$	8.7%	71.9%	97.2%	88.1%
Yes	5%	$N = 500$	5.8%	89.2%	100%	98.3%
Yes	10%	$N = 250$	14.9%	79.2%	98.9%	93.1%
Yes	10%	$N = 500$	11.9%	94.7%	100%	99.1%
No	5%	$N = 250$	3.2%	19.4%	10.5%	26.0%
No	5%	$N = 500$	4.8%	70.2%	76.0%	87.6%
No	10%	$N = 250$	8.9%	43.3%	36.6%	54.4%
No	10%	$N = 500$	9.6%	86.6%	97.9%	97.4%

Table 5. Empirical size and power for NFEigen.

Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$	$H_{A(3)}$
		$CP = BP$				
Yes	5%	$N = 250$	7.9%	67.7%	98.3%	84.6%
Yes	5%	$N = 500$	5.7%	90.7%	100%	98.2%
Yes	10%	$N = 250$	12.3%	76.5%	99.7%	90.6%
Yes	10%	$N = 500$	10.9%	94.2%	100%	99.2%
No	5%	$N = 250$	2.2%	13.8%	3.8%	12.3%
No	5%	$N = 500$	4.2%	64.8%	54.6%	75.9%
No	10%	$N = 250$	7.4%	37.2%	22.4%	38.1%
No	10%	$N = 500$	7.8%	83.6%	88.5%	93.0%
		$CP \neq BP$				
Yes	5%	$N = 250$	9.7%	71.8%	99.1%	86.7%
Yes	5%	$N = 500$	6.1%	92.5%	100%	99.4%
Yes	10%	$N = 250$	14.6%	79.4%	99.6%	92.5%
Yes	10%	$N = 500$	12.6%	96.0%	100%	99.9%
No	5%	$N = 250$	3.0%	13.1%	2.4%	10.0%
No	5%	$N = 500$	5.1%	63.5%	34.1%	71.3%
No	10%	$N = 250$	7.7%	38.3%	15.6%	33.5%
No	10%	$N = 500$	10.1%	83.8%	80.1%	92.5%

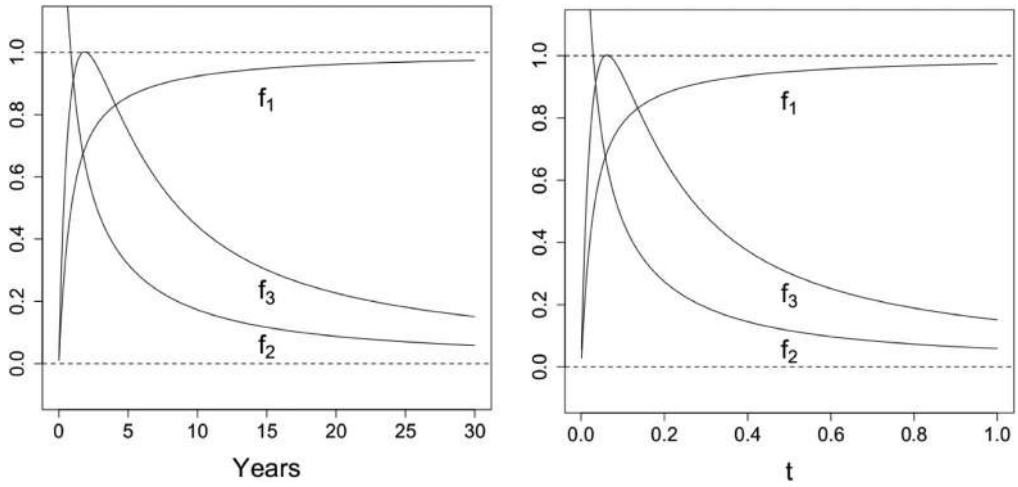

Figure 7. Intelligible factor curves.

if the break point is not taken into account, all methods can fail to detect an existing change point with a large probability. In that case, all procedures also suffer from nonmonotonic power (i.e. power is smaller for the larger departure from H_0). Taking the break point into account preserves monotonicity, in addition to leading to tests of practically usable power and correct size for sufficiently large sample size.

In the setting described so far, the projection methods ProjSim and ProjEigen have an automatic advantage because they use projections on the same factor curves that are used to generate the data. To further investigate the performance of these two approaches, we impose a different factor structure on the data-generating process than the structure used to compute test statistic (3.6). We use the same data-generating process as before, which imposes the factor structure in the dynamic Nelson–Siegel model (5.1). However, the tests are applied using the intelligible factors introduced by Lengwiler and Lenz (2010). Their factor curves have the form

$$f_1(t,\alpha_1,\alpha_2,b_1,b_2,b_3) = 1 + \frac{(1/t)(b_3 - \log(\alpha_3)b_1)(1-\alpha_2^t) - (1/m)(b_2 - \log(\alpha_2)b_1)(1-\alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4},$$

$$f_2(t,\alpha_1,\alpha_2,b_1,b_2,b_3) = \frac{-(1/t)b_3(1-\alpha_2^t) - (1/m)b_2(1-\alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4},$$

$$f_3(t,\alpha_1,\alpha_2,b_1,b_2,b_3) = \frac{(1/t)\log(\alpha_3)(1-\alpha_2^t) - (1/t)\log(\alpha_2)(1-\alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4}.$$

Estimating the parameters α_2, α_3, b_1, b_2 and b_3 requires a nested optimization. In our simulation, we do not estimate them, but use the values obtained by Lengwiler and Lenz (2010), i.e. $\alpha_2 = 0.1133$, $\alpha_3 = 0.6798$, $b_1 = 0.2674$, $b_2 = -0.4343$ and $b_3 = -0.2584$. The resulting factor curves are displayed in Figure 7: the left panel shows the curves corresponding to real yield curves and the right panel shows those transformed to the unit interval. To make the intelligible factors conform to the simulated yield curves generated on the unit interval, we transformed them in a similar manner as the Nelson–Siegel factors. The parameters of the transformed

Table 6. Empirical size and power for ProjSim with misspecified factors.

Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
		$CP = BP$			
Yes	5%	$N = 250$	6.3%	62.3%	97.7%
Yes	5%	$N = 500$	6.2%	85.7%	100.0%
Yes	10%	$N = 250$	11.4%	71.6%	99.2%
Yes	10%	$N = 500$	10.8%	91.2%	100.0%
No	5%	$N = 250$	0.7%	10.3%	0.3%
No	5%	$N = 500$	3.1%	53.6%	41.1%
No	10%	$N = 250$	3.1%	26.6%	13.6%
No	10%	$N = 500$	7.2%	73.6%	77.5%
		$CP \neq BP$			
Yes	5%	$N = 250$	6.3%	62.1%	97.45%
Yes	5%	$N = 500$	4.5%	87.8%	100.0%
Yes	10%	$N = 250$	11.3%	72.4%	98.9%
Yes	10%	$N = 500$	8.2%	92.4%	100.0%
No	5%	$N = 250$	1.2%	8.7%	0.9%
No	5%	$N = 500$	2.2%	51.9%	23.0%
No	10%	$N = 250$	4.4%	24.7%	6.5%
No	10%	$N = 500$	5.4%	73.9%	65.7%

intelligible factors are $\alpha_2 = 0.1133^{30}$, $\alpha_3 = 0.6798^{30}$, $b_1 = 0.2674$, $b_2 = -0.4343 * 30$ and $b_3 = -0.2584 * 30$. These transformed intelligible factors are displayed in the right panel of Figure 7.

Tables 6 and 7 show the rejection rates in the case of a misspecified factor structure. Compared to the correctly specified structure, the sizes become somewhat overinflated, especially in the case of the ProjEigen method. However, employing the intelligible factors in the ProjSim test improves the empirical size for $N = 250$.

Following queries raised by the referees, we studied several additional issues. We discuss the findings. Suppose a researcher implements a test assuming a break point when in truth there is no break point. Generally, in finite samples, we see a slight overrejection under H_0. However, if no break point is correctly assumed, the methods may fail to detect an existing change point with a large probability. This may be due to the *spurious* change in variance which a break point may capture. Related to this, the break point may be misplaced without affecting the size and power of the tests much. If $N = 500$, placing a break point even 200 days away from a true break point, leads to acceptable performance. Finally, how many breaks points can one assume before the method breaks down? Since the methods are based on asymptotic approximations, one can expect that each segment between the break points should be sufficiently long. If $N = 500$, the methods have acceptable size and power if not more than three break points are assumed. A general conclusion is that methods work best if about one break point per year is assumed, and it does not matter much where it is placed. However, the size is controlled more precisely if it is placed closer to the actual break point.

We conclude this section by displaying examples of densities used to obtain the critical values or to compute the P-values. The simulated data follow the alternative $H_{A(1)}$. Figure 8 shows the impact of not assuming a break point. The distributions in the figure are computed under the

Table 7. Empirical size and power for ProjEigen with misspecified factors.

Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
		$CP = BP$			
Yes	5%	$N = 250$	8.4%	66.7%	98.1%
Yes	5%	$N = 500$	6.9%	89.1%	100%
Yes	10%	$N = 250$	13.5%	75.1%	99.1%
Yes	10%	$N = 500$	10.5%	93.1%	100%
No	5%	$N = 250$	1.9%	21.2%	13.2%
No	5%	$N = 500$	3.3%	65.7%	82%
No	10%	$N = 250$	7.1%	43.4%	39.4%
No	10%	$N = 500$	9.5%	84.9%	97.3%
		$CP \neq BP$			
Yes	5%	$N = 250$	8.6%	69.9%	98.9%
Yes	5%	$N = 500$	7.4%	91.9%	100%
Yes	10%	$N = 250$	14.5%	77.6%	99.5%
Yes	10%	$N = 500$	11.7%	95.5%	100%
No	5%	$N = 250$	3.5%	20%	8.8%
No	5%	$N = 500$	6.2%	70.9%	77.6%
No	10%	$N = 250$	9.6%	42.9%	34.6%
No	10%	$N = 500$	9.6%	86.2%	97.2%

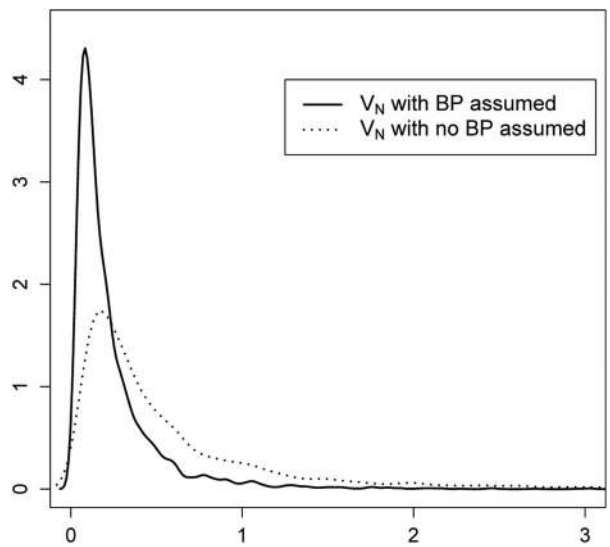

Figure 8. Approximate asymptotic densities to test the significance of $\mathcal{V}_N$.

alternative (change point in mean exists), but they play the role of distributions from which critical values are obtained. (These are not sampling distributions of $\mathcal{V}_N$ under the alternative.) The distributions were computed in two cases: with and without assuming a break point. For example, we use the 95th percentile of the asymptotic distribution of $\mathcal{V}_N$ as the critical value. If we assume a break point (solid line), then this percentile is near 1, while if we do not assume a

break point (dotted line), then this percentile is closer to 1.5. Thus, the critical value assuming a break point will be smaller than the critical value assuming no break point, i.e. we reject more often by assuming a break point. This heuristically explains the higher power if the break point is assumed.

6. SUMMARY

We derived several asymptotic methods to test the null hypothesis that the mean structure of a sequence of curves does not change. The tests are motivated by application to yield curves. In this context, the mean structure does not refer merely to the level of the curves, but also to their range and other aspects of their shape, most prominently concavity. We have observed the importance of the error structure, which refers to the random variability in the aspects of the curves listed above.

Two tests, called ProjSim and ProjEigen, are based on projections of the factor curves, for example on the Nelson–Siegel curves or the intelligible factors of Lengwiler and Lenz (2010). The difference between them is that ProjSim is based on simulating data that approximately satisfy H_0 (no change point), while ProjEigen is based on approximating suitable eigenvalues (it also requires generating a Monte Carlo distribution). These two tests require a specification of a factor structure. The third approach, NFEigen, does not require any factor structure; it is a nonparametric version of the method ProjEigen.

Based on our data analysis and simulation study, the following conclusions can be drawn.

(a) If a possible break point in the error structure is not taken into account in any of the testing procedures, an existing change point in the mean structure can fail to be detected with a large probability.
(b) The tests are generally well calibrated if $N = 500$.
(c) The location and the number of break points can be misspecified within certain limits. Using one or two break points per year, even about half a year away from a true break point, leads to better performance than using no break points at all.
(d) If $N = 250$, all tests have a tendency to overreject at the 5% level, i.e. the empirical type I error tends to be larger than 5%.
(e) If the intelligible factors are used, the empirical size of the ProjSim test improves at the 5% nominal level, but the size of the ProjEigen test deteriorates.

ACKNOWLEDGEMENTS

We thank the editor, Professor Dennis Kristensen, and three referees for reading our work carefully and for providing detailed advice. We thank Professor Davar Khoshnevisan for directing us to the paper of Anderson (1955). This research has been partially supported by NSF Grants DMS–1462067 and DMS–1305858.

REFERENCES

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society 6*, 170–76.

Aue, A., S. Hörmann, L. Horváth and M. Reimherr (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics 37*, 4046–87.

Bech, M. and Y. Lengwiler (2012). The financial crisis and the changing dynamics of the yield curve. BIS Paper No. 65o, Bank of International Settlements.

Berkes, I., R. Gabrys, L. Horváth and P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society*, Series B *71*, 927–46.

Berkes, I., L. Horváth and G. Rice (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications 123*, 385–403.

Busetti, F. and A. M. R. Taylor (2004). Test of stationarity agains a change in persistence. *Journal of Econometrics 123*, 33–66.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Cavaliere, G. and A. M. R. Taylor (2008). Testing for a change in persistence of nonstationary volatility. *Journal of Econometrics 147*, 259–72.

Cavaliere, G., D. I. Harvey, S. Leybourne and A. M. R. Taylor (2011). Testing for unit roots in the presence of a possible break in trend and nonstationary volatility. *Econometric Theory 27*, 957–91.

Chen, Y. and L. Niu (2014). Adaptive dynamic Nelson–Siegel term structure model with applications. *Journal of Econometrics 180*, 98–115.

Chib, S. and K. H. Kang (2013). Change points in affine arbitrage-free term structure models. *Journal of Financial Econometrics 14*, 302–34.

Csörgő, M. and L. Horváth (1997). *Limit Theorems in Change-Point Analysis*. New York, NY: Wiley.

Dalla, V., L. Giraitis and P. C. B. Phillips (2015). Testing mean stability of heteroskedastic time series. Cowles Foundation Discussion Paper No. 2006, Yale University.

Dette, H., W. Wu and Z. Zhou (2015). Change point analysis of second-order characteristics in non-stationary time series. Working paper, Ruhr Universität Bochum.

Diebold, F. X. and C. Li (2003). Forecasting the term structure of government bond yields. Working Paper 10048, National Bureau of Economic Research.

Diebold, F. and G. Rudebusch (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson–Siegel Approach*. Princeton, NJ: Princeton University Press.

Filipović, D. (2009). *Term–Structure Models*. Berlin: Springer.

Giraitis, L., P. S. Kokoszka, R. Leipus and G. Teyssière (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics 112*, 265–94.

Hays, S., H. Shen and J. Z. Huang (2012). Functional dynamic factor models with application to yield curve forecasting. *Annals of Applied Statistics 6*, 870–94.

Hörmann, S., L. Horváth and R. Reeder (2013). A functional version of the ARCH model. *Econometric Theory 29*, 267–88.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *Annals of Statistics 38*, 1845–84.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics 21*, 671–80.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. Berlin: Springer.

Horváth, L. and G. Rice (2014). Extensions and some classical methods in change point analysis (with discussions). *TEST 23*, 219–90.

Horváth, L., P. Kokoszka and R. Reeder (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society*, Series B *75*, 103–22.

Kokoszka, P. and R. Leipus (2000). Change-point estimation in ARCH models. *Bernoulli 6*, 513–39.

Kokoszka, P. and M. Reimherr (2013). Predictability of shapes of intraday price curves. *Econometrics Journal 16*, 285–308.

Lengwiler, Y. and C. Lenz (2010). Intelligible factors for the yield curve. *Journal of Econometrics 157*, 481–91.

Nieh, C., S. Wu and Y. Zeng (2010). Regime shifts and the term structure of interest rates. In A. C. Lee and C. Lee (Eds.), *Handbook of Quantitative Finance and Risk Management*, 1121–33. Berlin: Springer.

Rogers, C. and W. Stummer (2000). Consistent fitting of one-factor models to interest rate data. *Insurance: Mathematics and Economics 27*, 45–63.

Sen, R. and C. Klüppelberg (2015). Time series of functional data. ISIC preprint, ASU-2015-1, Indian Statistical Institute. Available at http://www.isichennai.res.in/tr/asu/2015/1/.

Shao, X. and W. B. Wu (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics 35*, 1773–801.

Wu, W. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States 102*, 14150–4.

Zhou, Z. (2013). Heteroskedastic and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*, 726–40.

APPENDIX: PROOFS OF THE ASYMPTOTIC RESULTS

A.1. Proofs of the results of Section 3

The main result of Section 3 is Theorem 3.1, which follows from Lemmata A.1 and A.2. We use the following notation:

$$N_m = i_m - i_{m-1} \qquad \text{and} \qquad J_m(x) = \{j : i_{m-1} < j \leq \lfloor i_{m-1} + N_m x \rfloor\}.$$

LEMMA A.1. *Suppose Assumptions 2.1, 2.2 and 2.3 are satisfied. Then, for each N, we can define* $M + 1$ *independent Gaussian* $\mathbb{R}^K$*-valued processes* $\mathbf{G}_{N,m}$, $1 \leq m \leq M + 1$, *such that*

$$E[\mathbf{G}_{N,m}(x)] = 0 \qquad \text{and} \qquad E[\mathbf{G}_{N,m}(x)\mathbf{G}_{N,m}(y)^\top] = \min(x, y)\mathbf{V}_m,$$

and for all $1 \leq m \leq M + 1$,

$$\max_{0 \leq x \leq 1} \|N_m^{-1/2} \sum_{j \in J_m(x)} \boldsymbol{\gamma}_j - \mathbf{G}_{N,m}\| \xrightarrow{p} 0. \tag{A.1}$$

Proof: Lemma A.1 is a consequence of Theorem 1.1 of Berkes et al. (2013). Their approximation principle is applied to each segment of stationarity of $\boldsymbol{\gamma}_i$. The only difference is that Berkes et al. (2013) consider L^2-valued processes, whereas (A.1) involves $\mathbb{R}^K$-valued processes. All arguments used by Berkes et al. (2013) remain valid; the inner product must be interpreted as the inner product in $\mathbb{R}^K$ rather than in L^2.

The processes $N_m^{-1/2} \sum_{j \in J_m(x)} \boldsymbol{\gamma}_j$, $1 \leq M \leq M + 1$, are not independent under our assumptions. The proof of Berkes et al. (2013) shows that it is enough to consider ℓ-dependent sequences (cf. Assumption 2.2). These ℓ-dependent sequences are asymptotically independent for any $\ell \geq 1$. Therefore, we obtain the independence of the approximating sequences $\mathbf{G}_{N,m}$. □

LEMMA A.2. *Suppose Assumptions 2.1, 2.2 and 2.3 are satisfied. Then*

$$\max_{0 \leq x \leq 1} \|N^{-1/2} \sum_{1 \leq j \leq Nx} \boldsymbol{\gamma}_j - \mathbf{G}_N(x)\| \xrightarrow{p} 0,$$

where $\mathbf{G}_N$ *is a mean zero Gaussian process with covariances*

$$E[\mathbf{G}_N(x)\mathbf{G}_N^\top(y)] = \sum_{j=1}^{m}(\theta_j - \theta_{j-1})\mathbf{V}_j + (x - \theta_m)\mathbf{V}_{m+1}, \qquad \theta_m < x \le \theta_{m+1}, \quad x \le y \le 1. \quad \text{(A.2)}$$

Proof: For $\theta_m < x \le \theta_{m+1}$, we can write the partial sum process as

$$\sum_{1\le j\le Nx} \boldsymbol{\gamma}_j = \sum_{j=1}^{m} \sum_{i_{j-1}<i\le i_j} \boldsymbol{\gamma}_i + \sum_{i_m<i\le Nx} \boldsymbol{\gamma}_i. \quad \text{(A.3)}$$

Therefore, by Lemma A.1,

$$N^{-1/2} \sum_{1\le j\le Nx} \boldsymbol{\gamma}_j = \sum_{j=1}^{m} \left(\frac{N_j}{N}\right)^{1/2} N_j^{-1/2} \sum_{i_{j-1}<i\le i_j} \boldsymbol{\gamma}_i + \left(\frac{N_{m+1}}{N}\right)^{1/2} N_{m+1}^{-1/2} \sum_{i_m<i\le Nx} \boldsymbol{\gamma}_i$$

can be approximated by

$$\mathbf{G}_N(x) = \sum_{j=1}^{m}(\theta_j - \theta_{j-1})^{1/2}\mathbf{G}_{N,j}(1) + (\theta_{m+1} - \theta_m)^{1/2}\mathbf{G}_{N,m+1}\left(\frac{x - \theta_m}{\theta_{m+1} - \theta_m}\right).$$

The process $\mathbf{G}_N(x)$ is a linear combination of Gaussian processes and hence is Gaussian. The covariance structure (A.2) follows from the independence of $\mathbf{G}_{N,1}, \mathbf{G}_{N,2}, \ldots, \mathbf{G}_{N,M+1}$ and their covariances established in Lemma A.1. □

Proof of Theorem 3.1: By the triangle inequality,

$$\begin{aligned}\|\boldsymbol{\alpha}_N(x) - \mathbf{G}_N^0(x)\| &= \left\| \frac{1}{\sqrt{N}} \left(\sum_{1\le i\le Nx} \boldsymbol{\gamma}_j - \frac{[Nx]}{N} \sum_{i=1}^{N} \boldsymbol{\gamma}_j \right) - (\mathbf{G}_N(x) - x\mathbf{G}_N(1)) \right\| \\ &\le \left\| \frac{1}{\sqrt{N}} \sum_{1\le i\le Nx} \boldsymbol{\gamma}_j - \mathbf{G}_N(x) \right\| + \left\| \frac{[Nx]}{N} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\gamma}_j - x\mathbf{G}_N(1) \right\|.\end{aligned}$$

Thus, by Lemma A.2, $\max_{0\le x\le 1} \|\boldsymbol{\alpha}_N(x) - \mathbf{G}_N^0(x)\| \xrightarrow{p} 0$, and the claim follows. □

Proof of Proposition 3.1: Suppose Γ is a zero mean random element in a separable Hilbert space, which satisfies $E[\|\Gamma\|_H^2] < \infty$. Then, Γ admits the Karhunen–Loéve decomposition, $\Gamma = \sum_{j=1}^{\infty} \xi_j \varphi_j$, where φ_j are (deterministic) orthonormal eigenvectors of the covariance operator of Γ, and ξ_j are random variables, $\xi_j = \langle X, \varphi_j \rangle$. The covariance operator of Γ is defined by $f \mapsto E[\langle \Gamma, f\rangle \Gamma]$, so φ_j satisfy $E[\langle \Gamma, \varphi_j\rangle \Gamma] = \lambda_j \varphi_j$. If Γ is Gaussian, then ξ_j are independent and normal with means zero and variances λ_j.

In the setting of Proposition 3.1, we consider the Hilbert space of R^K-valued functions

$$\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_K(x))^\top, \qquad x \in [0, 1],$$

with the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{k=1}^{K} \int_0^1 f_k(x) g_k(x) dx.$$

Direct verification shows that

$$E[\langle \boldsymbol{\Gamma}, \mathbf{f}\rangle \boldsymbol{\Gamma}](x) = \int_0^1 \mathbf{R}(x, y)\mathbf{f}(y) dy.$$

Therefore, the Karhunen–Loéve decomposition with the eigenelements in (3.10) is

$$\mathbf{\Gamma}(x) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j \phi_j(x).$$

By the orthonormality of ϕ_j, we obtain

$$\int_0^1 \|\mathbf{\Gamma}(x)\|^2 dx = \langle \mathbf{\Gamma}, \mathbf{\Gamma} \rangle = \sum_{j=1}^{\infty} \lambda_j Z_j^2.$$

□

Proof of Theorem 3.2: Under the change point alternative,

$$\mathbf{z}_i = \begin{cases} \mathbf{C}\boldsymbol{\mu} + \boldsymbol{\gamma}_i, & 1 \le i \le \ell^*, \\ \mathbf{C}\boldsymbol{\mu}^* + \boldsymbol{\gamma}_i, & \ell^* < i \le N, \end{cases}$$

with $\ell^* = [Nr]$. Therefore, the CUSUM process can be expressed as

$$\boldsymbol{\alpha}_N(x) = \boldsymbol{\beta}_N(x) + N^{-1/2} g_N(x, r) \mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*), \tag{A.4}$$

where

$$\boldsymbol{\beta}_N(x) = N^{-1/2} \Bigg(\sum_{1 \le i \le Nx} \boldsymbol{\gamma}_i - \frac{[Nx]}{N} \sum_{i=1}^{N} \boldsymbol{\gamma}_i \Bigg),$$

$$g_N(x, r) = \frac{[Nx]}{N}(N - [Nr]) I(\{x \le r\}) + \frac{[Nr]}{N}(N - [Nx]) I(\{x > r\}), \tag{A.5}$$

and $I(A)$ is the indicator function of set A. The Cramér–von-Mises test statistic can therefore be expressed as

$$\begin{aligned} \mathcal{C}_N &= \int_0^1 \|\boldsymbol{\alpha}_N(x)\|^2 dx \\ &= \int_0^1 \|\boldsymbol{\beta}_N(x)\|^2 dx + \int_0^1 \|N^{-1/2} g_N(x, r) \mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|^2 dx \\ &\quad + 2 \int_0^1 N^{-1/2} g_N(x, r) \boldsymbol{\beta}_N^{\top}(x) \mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*) dx. \end{aligned}$$

We first establish (3.15). Recall the Gaussian limit process defined in Theorem 3.1. It follows that

$$\int_0^1 \|\boldsymbol{\beta}_N(x)\|^2 dx \xrightarrow{d} \int_0^1 \|\mathbf{G}^0(x)\|^2 dx = O_P(1).$$

So, by (3.14), the first term in the expansion of $\mathcal{C}_N$ does not contribute to the limit in (3.15).

Next we examine the third term. Observe that (3.13) can be written as $(\boldsymbol{\mu} - \boldsymbol{\mu}^*)/a_N = -\mathbf{u}$. We also define the function

$$g^*(x, r) = x(1 - r) I(\{x \le r\}) + r(1 - x) I(\{x > r\}), \tag{A.6}$$

and notice that for fixed $x, r \in [0, 1]$,

$$N^{-1} g_N(x, r) \to g^*(x, r), \quad \text{as } N \to \infty. \tag{A.7}$$

Observe that

$$\left(\frac{1}{N^{1/2}a_N}\right)^2 \int_0^1 N^{-1/2} g_N(x,r)\boldsymbol{\beta}_N^\top(x)\mathbf{C}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)dx$$
$$= -\frac{1}{N^{1/2}a_N}\int_0^1 N^{-1} g_N(x,r)\boldsymbol{\beta}_N^\top(x)\mathbf{C}\mathbf{u}\,dx.$$

By (A.7) and the continuous mapping theorem,

$$\int_0^1 N^{-1} g_N(x,r)\boldsymbol{\beta}_N^\top(x)dx \xrightarrow{d} \int_0^1 g^*(x,r)[\mathbf{G}^0(x)]^\top dx = O_P(1).$$

So, by (3.14), the third term does not contribute to the limit in (3.15) either.

The second term is deterministic and dominates the other two terms. Relation (3.15) follows by observing that

$$\left(\frac{1}{N^{1/2}a_N}\right)^2 \int_0^1 \|N^{-1/2} g_N(x,r)\mathbf{C}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)\|^2 dx = \|\mathbf{C}\mathbf{u}\|^2 \int_0^1 \|N^{-1} g_N(x,r)\|^2 dx$$

and applying (A.7).

The argument that establishes part (b) of the theorem is similar. If $a_N = N^{-1/2}$, then (3.13) implies that

$$N^{-1/2} g_N(x,r)\mathbf{C}(\boldsymbol{\mu}-\boldsymbol{\mu}^*) \to -g^*(x,r)\mathbf{C}\mathbf{u}.$$

Part (b) thus follows from Theorem 3.1 and (A.4). □

A.2. Proofs of the results of Section 4

The proof of Theorem 4.1 is analogous to the proof of Theorem 3.1. Recall the notation N_j and $J_m(x)$ introduced at the beginning of Section A.1. Throughout this section, we assume that the Assumptions of Theorem 4.1 hold.

Lemma A.3 follows from Theorem 1.1 of Berkes et al. (2013); the argument for the independence of the $M+1$ processes $\Gamma_{N,m}$ is presented in the proof of Lemma A.1.

LEMMA A.3. *For each N, we can define we can define* $M+1$ *independent Gaussian processes* $\Gamma_{N,1}, \Gamma_{N,2}, \ldots, \Gamma_{N,M+1}$ *such that for all* $1 \le m \le M+1$,

$$E[\Gamma_{N,m}(x,t)] = 0, \qquad E[\Gamma_{N,m}(x,t)\Gamma_{N,m}(y,s)] = \min(x,y)D_m(t,s)$$

and

$$\max_{0\le x\le 1} \|N_m^{-1/2} \sum_{j\in J_m(x)} \eta_j(\cdot) - \Gamma_{N,m}(x,\cdot)\|_2 \xrightarrow{p} 0.$$

The proof of Lemma A.4 is analogous to the proof of Lemma A.2, so it is omitted.

LEMMA A.4. *Define the process*

$$\Gamma_N(x,t) = \sum_{l=1}^{m}(\theta_j - \theta_{j-1})^{1/2}\Gamma_{N,j}(1,t) + (\theta_{m+1}-\theta_m)^{1/2}\Gamma_{N,m+1}\left(\frac{x-\theta_m}{\theta_{m+1}-\theta_m}, t\right),$$

$\theta_m \le x \le \theta_{m+1}$, $0 \le m \le M+1$. *Then*

$$\max_{0\le x\le 1} \|N^{-1/2} \sum_{j=1}^{\lfloor Nx\rfloor} \eta_j(\cdot) - \Gamma_N(x,\cdot)\|_2 \xrightarrow{p} 0.$$

Proof of Theorem 4.1: Under $H_0 : \tau_1 = \tau_2 = \cdots \tau_N = \tau$, so

$$\sum_{1\le i\le Nx} X_i(t) - \frac{[Nx]}{N}\sum_{i=1}^{N} X_i(t) = \sum_{1\le i\le Nx} \eta_i(t) - \frac{[Nx]}{N}\sum_{i=1}^{N} \eta_i(t).$$

It remains to apply Lemma A.4 and the triangle inequality, analogously as in the proof of Theorem 3.1. □

Proof of Theorem 4.2: The following proof is similar to the proof of Theorem 3.2. Under the change point alternative,

$$X_i(t) = \begin{cases} \tau(t) + \eta_i(t), & 1 \le i \le \ell^*, \\ \tau^*(t) + \eta_i(t), & \ell^* < i \le N, \end{cases}$$

with $\ell^* = [Nr]$. The CUSUM process can then be expressed as

$$\alpha_N(x,t) = \beta_N(x,t) + N^{-1/2} g_N(x,r)(\tau(t) - \tau^*(t)),$$

where

$$\beta_N(x,t) = N^{-1/2}\Big(\sum_{1\le i\le Nx} \eta_i(t) - \frac{[Nx]}{N}\sum_{i=1}^{N} \eta_i(t)\Big),$$

and $g_N(x,r)$ is the function defined in (A.5). Under H_A, the Cramér–von-Mises test statistic can be expressed as

$$\begin{aligned} \mathcal{V}_N &= \int_0^1 \int \alpha_N^2(x,t) \nu_J(dt)\, dx \\ &= \int_0^1 \int \beta_N^2(x,t)\nu_J(dt)\, dx + \|\tau - \tau^*\|_2^2 \int_0^1 N^{-1} g_N^2(x,r)\, dx \\ &\quad + 2\int_0^1 \int N^{-1/2} g_N(x,r)\beta_N(x,t)(\tau(t) - \tau^*(t))\nu_J(dt)\, dx. \end{aligned}$$

By the Gaussian limit process defined in Theorem 4.1,

$$\int_0^1 \int \beta_N^2(x,t)\nu_J(dt)\, dx \xrightarrow{d} \int_0^1 \int (\Gamma^0(x,t))^2 \nu_J(dt)\, dx = O_P(1).$$

By (A.7) and the continuous mapping theorem,

$$\begin{aligned} &\frac{1}{N^{1/2} a_N} \int_0^1 \int N^{-1/2} g_N(x,r)\beta_N(x,t)(\tau(t) - \tau^*(t))\nu_J(dt)\, dx \\ &\quad \xrightarrow{d} -\int_0^1 \int g^*(x,r)\Gamma^0(x,t)u(t)\nu_J(dt)\, dx = O_P(1). \end{aligned}$$

The second term is deterministic and dominates the other two terms. By (A.7),

$$\left(\frac{1}{N^{1/2} a_N}\right)^2 \|\tau - \tau^*\|_2^2 \int_0^1 N^{-1} g_N^2(x,r) dx \to \|u\|_2^2 \int_0^1 \{g^*(x,r)\}^2 dx.$$

Combining the above limits, we obtain relation (4.14).

Part (b) follows analogously from the above decomposition, Theorem 4.1 and (A.7). □

A.3. Effect of parameter estimation

In Section 3, following the established current practice, we assumed that the factors $f_k(t)$, $1 \le k \le K$, are fully specified functions. To expand the scope of the applicability of our methodology, we show in this section that it remains valid if only a parametric form of the factors is known, and the parameters are estimated. We thus assume that $f_k(t) = f_k(t, \boldsymbol{\lambda})$ and the true value of the parameter, $\boldsymbol{\lambda}_0$, is estimated by a consistent estimator $\hat{\boldsymbol{\lambda}}_N$. We also assume that each $f_k(t, \boldsymbol{\lambda})$ is a smooth function of the parameter $\boldsymbol{\lambda}$. To state the latter assumption, we denote by $\nabla_\lambda f_k(\cdot, \boldsymbol{\lambda})$ the functional vector of the partial derivatives of $f_k(\cdot, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$.

ASSUMPTION A.1. *The estimator $\hat{\boldsymbol{\lambda}}_N$ satisfies $\|\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}_0\| = o_P(1)$ and there is $\mathbf{U}$, a neighbourhood of $\boldsymbol{\lambda}_0$, such that*

$$\max_{1\le k\le K} \left\| \sup_{\boldsymbol{\lambda}\in\mathbf{U}} \|\nabla_\lambda f_k(\cdot, \boldsymbol{\lambda})\| \right\|_2 < \infty.$$

In this setting, the vector of projections is defined by

$$\begin{aligned}\hat{\mathbf{z}}_i &= (\hat{z}_{i,1}, \hat{z}_{i,2}, \ldots, \hat{z}_{i,K})^\top \\ &= (\langle X_i, f_1(\cdot, \hat{\boldsymbol{\lambda}}_N)\rangle, \langle X_i, f_2(\cdot, \hat{\boldsymbol{\lambda}}_N)\rangle, \ldots, \langle X_i, f_K(\cdot, \hat{\boldsymbol{\lambda}}_N)\rangle)^\top\end{aligned}$$

and the CUSUM process as

$$\hat{\boldsymbol{\alpha}}_N(x) = N^{-1/2}\left(\sum_{i=1}^{\lfloor Nx\rfloor} \hat{\mathbf{z}}_i - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} \hat{\mathbf{z}}_i \right), \quad 0 \le x \le 1.$$

The following proposition shows that the estimation of $\boldsymbol{\lambda}$ does not change the limit distribution of the test statistics introduced in Section 3.

PROPOSITION A.1. *If H_0 and Assumptions 2.1, 2.2, 2.3 and A.1 hold, then*

$$\sup_{0\le x\le 1} \|\hat{\boldsymbol{\alpha}}_N(x) - \boldsymbol{\alpha}_N(x)\| = o_P(1).$$

Before presenting a proof, we point out that in Proposition A.1 the model is correctly specified for the true parameter, i.e. for some $\mu_1, \mu_2, \ldots, \mu_K$, $E[X_i(t)] = \sum_{k=1}^K \mu_k f_k(t, \boldsymbol{\lambda}_0)$.

Proof: Observe that

$$\begin{aligned}&\left| \frac{1}{N^{1/2}}\left(\sum_{i=1}^{\lfloor Nx\rfloor} \hat{z}_{i,k} - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} \hat{z}_{i,k} \right) - \frac{1}{N^{1/2}}\left(\sum_{i=1}^{\lfloor Nx\rfloor} z_{i,k} - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} z_{i,k} \right) \right| \\ &\quad\le \left| \frac{1}{N^{1/2}} \left\langle \sum_{i=1}^{\lfloor Nx\rfloor} X_i(\cdot) - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} X_i(\cdot),\, f_k(\cdot, \hat{\boldsymbol{\lambda}}_N) - f_k(\cdot, \boldsymbol{\lambda}_0) \right\rangle \right| \\ &\quad\le \left| \frac{1}{N^{1/2}} \left\langle \sum_{i=1}^{\lfloor Nx\rfloor} X_i(\cdot) - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} X_i(\cdot),\, \nabla_\lambda f_k(\cdot, \boldsymbol{\xi})^\top (\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}_0) \right\rangle \right| \\ &\quad\le \left\| \frac{1}{N^{1/2}}\left(\sum_{i=1}^{\lfloor Nx\rfloor} X_i - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} X_i \right) \right\|_2 \left\| \nabla_\lambda f_k(\cdot, \boldsymbol{\xi})^\top (\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}_0) \right\|_2 \\ &\quad\le \left\| \frac{1}{N^{1/2}}\left(\sum_{i=1}^{\lfloor Nx\rfloor} X_i - \frac{\lfloor Nx\rfloor}{N} \sum_{i=1}^{N} X_i \right) \right\|_2 \left\| \sup_{\boldsymbol{\lambda}\in\mathbf{U}} \|\nabla_\lambda f_k(\cdot, \boldsymbol{\lambda})\| \right\|_2 I(\{\hat{\boldsymbol{\lambda}}_N \in \mathbf{U}\}) \|\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}_0\|\end{aligned}$$

$$+\left\| \frac{1}{N^{1/2}} \left(\sum_{i=1}^{\lfloor Nx \rfloor} X_i - \frac{\lfloor Nx \rfloor}{N} \sum_{i=1}^{N} X_i \right) \right\|_2 \left\| \sup_{\boldsymbol{\lambda} \in \mathbf{U}} \|\nabla_{\boldsymbol{\lambda}} f_k(\cdot, \boldsymbol{\lambda})\| \right\|_2 \|\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}_0\| I(\{\hat{\boldsymbol{\lambda}}_N \notin \mathbf{U}\})$$

$$= u_{N,1} + u_{N,2},$$

where $\boldsymbol{\xi}$ is a point satisfying $\|\boldsymbol{\xi} - \boldsymbol{\lambda}_0\| \leq \|\hat{\boldsymbol{\lambda}}_N - \boldsymbol{\lambda}\|$. Using Theorem 3.1 and Assumption A.1, we find that $u_{N,1} = o_P(1)$. According to Assumption A.1, $\lim_{N\to\infty} P\{\hat{\boldsymbol{\lambda}}_N \notin \mathbf{U}\} = 0$, and therefore $u_{N,2} = o_P(1)$. Hence, the proof of Proposition A.1 is complete. □

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix
Replication files