KPSS test for functional time series

Piotr Kokoszka*
Colorado State University

Gabriel Young
Colorado State University

Abstract

We develop extensions of the KPSS test to functional time series. The null hypothesis of the KPSS test for scalar data is that the series follows the model $x_t = \alpha+\beta t+\eta_t$, where $\{\eta_t\}$ is a stationary time series. The alternative is the model that includes a random walk: $x_t = \alpha + \beta t + \sum_{i\le t} u_i + \eta_t$. A functional time series (FTS) is a collection of curves observed consecutively over time. Examples include intraday price curves, term structure curves, and intraday volatility curves. We define the relevant testing problem for FTS, formulate the required assumptions, derive test statistics and their asymptotic distributions. These distributions are used to construct effective tests, both Monte Carlo and pivotal, which are applied to series of yield curves and examined in a simulation study.

Keywords: Functional data, Integrated time series, Random walk, Trend stationarity, Yield curve.

1 Introduction

The KPSS test of Kwiatkowski *et al.* (1992) has become one of the standard tools in the analysis of econometric time series. Its null hypothesis is that the series follows the model $x_t = \alpha + \beta t + \eta_t$, where $\{\eta_t\}$ is a stationary time series. The alternative is the model that includes a random walk: $x_t = \alpha + \beta t + \sum_{i\le t} u_i + \eta_t$, which then dominates the long term behavior of the series. The alternative is thus a series known in econometrics as a unit root or an integrated series. The work of Kwiatkowski *et al.* (1992) was in fact motivated by the unit root tests of Dickey and Fuller (1979, 1981) and Said and Dickey (1984). In these tests, the null hypothesis is that the series has a unit root. Since such tests have low power

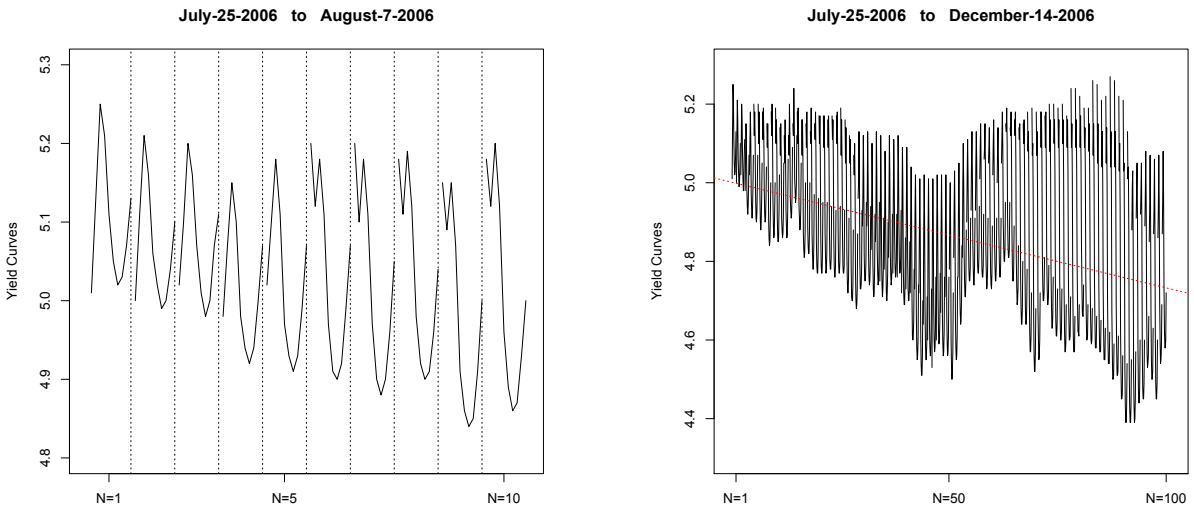
*Address for correspondence: Piotr Kokoszka, Department of Statistics, Colorado State University, Fort Collins, CO 80522, USA. E-mail: Piotr.Kokoszka@colostate.edu

in samples of sizes occurring in many applications, Kwiatkowski *et al.* (1992) proposed that trend stationarity should be considered as the null hypothesis, and the unit root should be the alternative. Rejection of the null hypothesis could then be viewed as a convincing evidence in favor of a unit root. It was soon realized that the KPSS test has a much broader utility. For example, Lee and Schmidt (1996) and Giraitis *et al.* (2003) used it to detect long memory, with short memory as the null hypothesis; de Jong *et al.* (1997) developed a robust version of the KPSS test. The work of Lo (1991) is crucial because he observed that under temporal dependence, to obtain parameter free limit null distributions, statistics similar to the KPSS statistic must be normalized by the long run variance rather than by the sample variance.

We develop extensions of the KPSS test to time series of curves, which we call functional time series (FTS). Many financial data sets form FTS. The best known and most extensively studied data of this form are yield curves. Even though they are reported at discrete maturities, in financial theory and practice they are viewed as continuous functions, one function per month or per day, see Figure 1. This is because fractions of bonds can be traded which can have an arbitrary maturity up to 30 years. Other well known examples include intraday price, volatility or volume curves. Intraday price data are smooth, volatility and volume data are noisy, and must be smoothed before they can be effectively treated as curves. Figure 2 shows series of price curves. Over a specific period of time, FTS of this type often exhibit a visual trend, and obviously cannot be treated as stationary. The question is if trend plus stationarity is enough or if a nonstationary component must be included. If the time period is sufficiently long, trend stationarity will not be a realistic assumption due to periods of growth and recession and changes in monetary policy of central banks. As in the context of scalar time series, the question is if a specific finite realization can be assumed to be generated by a trend stationary model.

We develop the required theory in the framework of functional data analysis (FDA). Application of FDA to time series analysis and econometrics is not new. Among recent contributions, we note Antoniadis *et al.* (2006), Kargin and Onatski (2008), Horváth *et al.* (2010), Müller *et al.* (2011), Panaretos and Tavakoli (2013), Kokoszka and Reimherr (2013), Hörmann *et al.* (2015), Aue *et al.* (2015), with a strong caveat that this list is far from exhaustive. A contribution most closely related to the present work is Horváth *et al.* (2014) who developed a test of level stationarity. The FTS that motivate this work are visually not level stationary, but can be potentially trend stationary. Incorporating a possible trend, changes the structure of functional residuals and leads to different limit distributions. It also requires the asymptotic analysis of the long run variance function of these residuals, which was not required in the level stationary case. A spectral approach to testing stationarity of multivariate time series has recently been developed by Jentsch and Subba Rao (2015). It is possible that it could be extended to a test of *trend* stationarity of FTS, but in this paper

Figure 1: Right Panel: Ten consecutive yield curves of bonds issued by the United States Federal Reserve; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period.



we focus on the time domain approach in the spirit of the original work of Kwiatkowski *et al.* (1992).

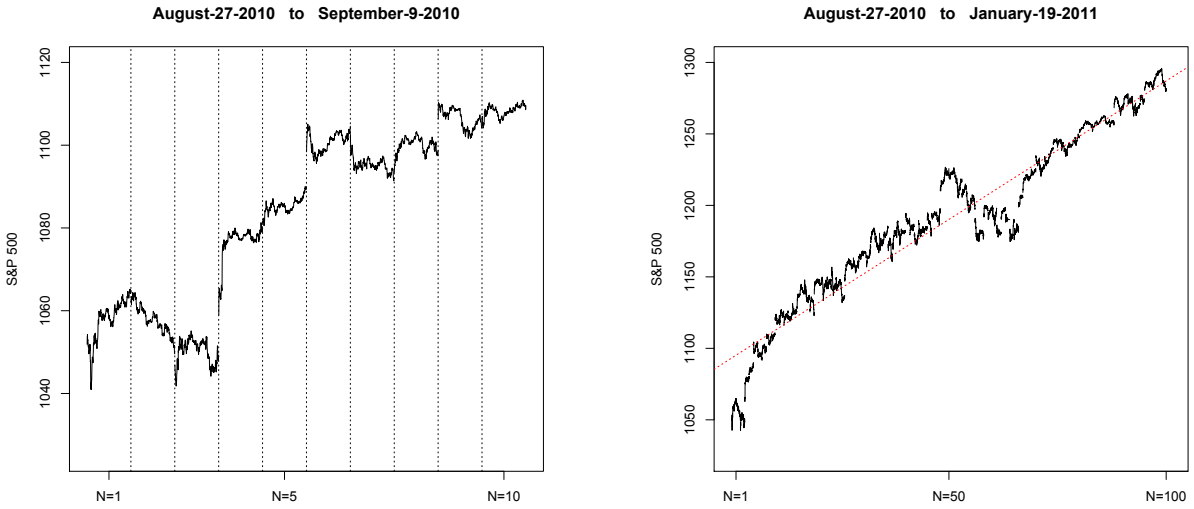
The remainder of the paper is organized as follows. Section 2 introduces the problem and assumptions. Test statistics and their asymptotic distributions are presented in Section 3. Section 4 contains an application to yield curves and a small simulation study. Proofs of the theorems stated in Section 3 are developed in Section 5.

2 Problem statement, definitions and assumptions

In FDA, the index t is used to denote "time" within a function. For example, for price curves, t is the time (e.g. in minutes) within a trading day; for yield curves, t is time to maturity. Functional observations are indexed by n; it is convenient to think of n as a trading day. Using this convention, the null hypothesis of trend stationarity is stated as follows:

$$H_0 : \qquad X_n(t) = \mu(t) + n\xi(t) + \eta_n(t). \tag{2.1}$$

Figure 2: Left Panel: Ten consecutive price curves of the SP500 index; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period



The functions μ and ξ correspond, respectively, to the intercept and slope. The errors η_n are also functions. Under the alternative, the model contains a random walk component:

$$H_A: \qquad X_n(t) = \mu(t) + n\xi(t) + \sum_{i=1}^{n} u_i(t) + \eta_n(t). \tag{2.2}$$

Our theory requires only that the sequences $\{\eta_n\}$ and $\{u_i\}$ be stationary in a function space, they do not have to be iid. Our tests have power against other alternatives, for example change–points or heteroskedasticity. We focus on the alternative (2.2) to preserve the context of the scalar KPSS test.

All random functions and deterministic functional parameters μ and ξ are assumed to be elements of the Hilbert space $L^2 = L^2([0,1])$ with the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$. This means that the domain of all functional observations, e.g. of the daily price or yield curves, has been normalized to be the unit interval. If the limits of integration are omitted, integration is over the interval $[0,1]$. All random functions are assumed to be square integrable, i.e. $E\|\eta_n\|^2 < \infty$, $E\|u_n\|^2 < \infty$. Further background on random elements of L^2 is given in Chapter 2 of Horváth and Kokoszka (2012); a more extensive theoretical treatment is presented in Hsing and Eubank (2015).

We quantify the weak dependence of the errors via the following assumption:

ASSUMPTION 2.1 The errors η_j are Bernoulli shifts, i.e. $\eta_j = g(\epsilon_j, \epsilon_{j-1}, ...)$ for some measurable function $g: S^\infty \to L^2$ and iid functions ϵ_j, $-\infty \le j \le \infty$, with values in a measurable

space S. The functions $(t,\omega) \mapsto \eta_j(t,\omega)$ are product measurable.
$E\eta_0 = 0$ in L^2 and $E\|\eta_0\|^{2+\delta} < \infty$ for some $0 < \delta < 1$.
The sequence $\{\eta_n\}_{n=-\infty}^{\infty}$ can be approximated by ℓ-dependent sequences $\{\eta_{n,\ell}\}_{n=-\infty}^{\infty}$ in the sense that

$$\sum_{\ell=1}^{\infty} (E\|\eta_n - \eta_{n,\ell}\|^{2+\delta})^{1/\kappa} < \infty \text{ for some } \kappa > 2+\delta, \tag{2.3}$$

where $\eta_{n,\ell}$ is defined by

$$\eta_{n,\ell} = g(\epsilon_n, \epsilon_{n-1}, ..., \epsilon_{n-\ell+1}, \epsilon^*_{n-\ell}, \epsilon^*_{n-\ell-1}, \ldots)$$

where the ϵ^*_k are independent copies of ϵ_0, independent of $\{\epsilon_i, -\infty < i < \infty\}$.

Assumption 2.1 has been shown to hold for all known models for temporally dependent functions, assuming the parameters of these models satisfy nonrestrictive conditions, see Hörmann and Kokoszka (2010, 2012), or Chapter 16 of Horváth and Kokoszka (2012). Its gist is that the dependence of the function g on the iid innovations ϵ_j far in the past decays so fast that these innovations can be replaced by their independent copies. Such a replacement is asymptotically negligible in the sense quantified by (2.3). For scalar time series, conditions similar in spirit were used by Pötscher and Prucha (1997), Wu (2005), Shao and Wu (2007) and Berkes *et al.* (2011), to name just a few references. In this paper, Assumption 2.1 is needed to ensure that the partial sums (3.2) can be approximated by a two–parameter Gaussian process. In particular, (2.3) is not used directly; it is a condition used by Berkes *et al.* (2013) to prove Theorem 5.2. To establish the results of Section 3, one can, in fact, replace Assumption 2.1 by the conclusions of Theorem 5.2 and the existence of an estimator $\hat{c}(t,s)$ such that

$$\iint [\hat{c}(t,s) - c(t,s)]^2 dtds \xrightarrow{P} 0, \quad \text{as } N \to \infty, \tag{2.4}$$

with the kernel c defined by (2.5). Assumption 2.1 is a general weak dependence condition under which these conclusions hold. While we expect that our limit results can be proven under different weak dependence conditions, the general theorems we use have been so far proven only under Assumption 2.1.

We now define the bivariate functions appearing in (2.4). The long–run covariance function of the errors η_n is defined as

$$c(t,s) = E\eta_0(t)\eta_0(s) + \sum_{i=1}^{\infty} \left(E\eta_0(t)\eta_i(s) + E\eta_0(s)\eta_i(t)\right). \tag{2.5}$$

The series defining the function $c(t,s)$ converges in $L^2([0,1]\times[0,1])$, see Horváth *et al.* (2013). The function $c(t,s)$ is positive definite. Therefore there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq 0$, and orthonormal eigenfunctions $\phi_i(t)$, $0 \leq t \leq 1$, satisfying

$$\lambda_i \phi_i(t) = \int c(t,s)\phi_i(s)ds, \quad 0 \leq i \leq \infty. \tag{2.6}$$

To ensure that the ϕ_i corresponding to the d largest eigenvalues are uniquely defined (up to a sign), we assume that

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1} > 0. \tag{2.7}$$

The eigenvalues λ_i play a crucial role in our tests. They are estimated by the sample, or empirical, eigenvalues defined by

$$\hat{\lambda}_i \hat{\phi}_i(t) = \int \hat{c}(t,s)\hat{\phi}_i(s)ds, \quad 0 \leq i \leq N, \tag{2.8}$$

where $\hat{c}(\cdot,\cdot)$ is an estimator of (2.5). We use a kernel estimator similar to that introduced by Horváth *et al.* (2013), but with suitably defined residuals in place of the centered observations X_n. To define model residuals, consider the least squares estimators of the functional parameters $\xi(t)$ and $\mu(t)$ in model (2.1):

$$\hat{\xi}(t) = \frac{1}{s_N}\sum_{n=1}^{N}\left(n - \frac{N+1}{2}\right)X_n(t) \tag{2.9}$$

with

$$s_N = \sum_{n=1}^{N}\left(n - \frac{N+1}{2}\right)^2 \tag{2.10}$$

and

$$\hat{\mu}(t) = \bar{X}(t) - \hat{\xi}(t)\Big(\frac{N+1}{2}\Big). \tag{2.11}$$

The functional residuals are therefore

$$e_n(t) = (X_n(t) - \bar{X}(t)) - \hat{\xi}(t)\Big(n - \frac{N+1}{2}\Big), \quad 1 \leq n \leq N. \tag{2.12}$$

Defining their empirical autocovariances by

$$\hat{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N} e_j(t)e_{j-i}(s), \quad 0 \leq i \leq N-1, \tag{2.13}$$

leads to the kernel estimator

$$\hat{c}(t,s) = \hat{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)(\hat{\gamma}_i(t,s) + \hat{\gamma}_i(s,t)). \tag{2.14}$$

The following assumption is imposed on kernel function K and the bandwidth h.

ASSUMPTION 2.2 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

$$h(N) \to \infty, \ \frac{h(N)}{N} \to 0, \quad \text{as} \ \ N \to \infty. \tag{2.15}$$

The assumption that K vanishes outside a compact interval is not crucial to establish (2.4). It is a simplifying condition which could be replaced by a sufficiently fast decay condition, at the cost of technical complications in the proof of (2.4).

Recall that if $\{W(x), 0 \le x \le 1\}$ is a standard Brownian motion (Wiener process), then the Brownian bridge is defined by $B(x) = W(x) - xW(x), \ 0 \le x \le 1$. The second–level Brownian bridge is defined by

$$V(x) = W(x) + \Big(2x - 3x^2\Big)W(1) + \Big(-6x + 6x^2\Big)\int_0^1 W(y)dy, \quad 0 \le x \le 1. \tag{2.16}$$

Both the Brownian bridge and the second–level Brownian bridge are special cases of the generalized Brownian bridge introduced by MacNeill (1978) who studied the asymptotic behavior of partial sums of polynomial regression residuals. Process (2.16) appears as the null limit of the KPSS statistic of Kwiatkowski *et al.* (1992). We will see in Section 3 that for functional data the limit involves an infinite sequence of independent and identically distributed second-level Brownian bridges $V_1(x), V_2(x), \ldots$.

3 Test statistics and their limit distributions

We will work with the partial sum process of the curves $X_1(t), X_2(t), \ldots, X_N(t)$ defined by

$$S_N(x,t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t), \quad 0 \le t, x \le 1, \tag{3.1}$$

and the partial sum process of the unobservable errors defined by

$$V_N(x,t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t), \quad 0 \le t, x \le 1. \tag{3.2}$$

Test statistic are based on the partial sum process of residuals (2.12). Observe that

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor} e_n(t) &= \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor}\Big((X_n(t)-\bar{X}(t))-\hat{\xi}(t)(n-(N+1)/2)\Big)\\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor} X_n(t) - \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor}\bar{X}(t) - \frac{\hat{\xi}(t)}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor}(n-(N+1)/2)\\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nx\rfloor} X_n(t) - \frac{\lfloor Nx\rfloor}{N}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} X_n(t) - \frac{\hat{\xi}(t)}{2\sqrt{N}}\Big(\lfloor xN\rfloor(\lfloor xN\rfloor - N)\Big)\\ &= S_N(x,t) - \frac{\lfloor Nx\rfloor}{N}S_N(1,t) - \frac{1}{2}N^{3/2}\hat{\xi}(t)\left(\frac{\lfloor xN\rfloor}{N}\left(\frac{\lfloor xN\rfloor}{N}-1\right)\right).\end{aligned}$$

A suitable test statistic is therefore given by

$$R_N = \iint Z_N^2(x,t)dtdx = \int \|Z_N(x,\cdot)\|^2 dx, \quad 0\le t,x\le 1, \tag{3.3}$$

where

$$Z_N(x,t) = S_N(x,t) - \frac{\lfloor Nx\rfloor}{N}S_N(1,t) - \frac{1}{2}N^{3/2}\hat{\xi}(t)\Big(\frac{\lfloor Nx\rfloor}{N}\Big(\frac{\lfloor Nx\rfloor}{N}-1\Big)\Big) \tag{3.4}$$

and $S_N(x,t)$ and $\hat{\xi}(t)$ are, respectively, defined in equations (3.1) and (2.9). The null limit distribution of test statistic (3.3) is given in the following theorem.

THEOREM 3.1 *If Assumption 2.1 holds, then under null model* (2.1),

$$R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty}\lambda_i\int V_i^2(x)dx,$$

where λ_1, λ_2,..., *are eigenvalues of the long–run covariance function* (2.5), *and* $V_1, V_2, \ldots$ *are iid second–level Brownian bridges.*

The proof of Theorem 3.1 is given in Section 5. We now explain the issues arising in the functional case by comparing our result to that obtained by Kwiatkowski *et al.* (1992). If all curves are constant functions ($X_i(t) = X_i$ for $t\in[0,1]$), the statistic R_N given by (3.3) is the numerator of the KPSS test statistic of Kwiatkowski *et al.* (1992), which is given by

$$\text{KPSS}_N = \frac{1}{N^2\hat{\sigma}_N^2}\sum_{n=1}^{N} S_n^2 = \frac{R_N}{\hat{\sigma}_N^2},$$

where $\hat{\sigma}_N^2$ is a consistent estimator of the long-run variance σ^2 of the residuals. In the scalar case, Theorem 3.1 reduces to $R_N \xrightarrow{D} \sigma^2\int_0^1 V^2(x)dx$, where $V(x)$ is a second–level

Brownian bridge. If $\hat{\sigma}_N^2$ is a consistent estimator of σ^2, the result of Kwiatkowski *et al.* (1992) is recovered, i.e. $\text{KPSS}_N \xrightarrow{D} \int_0^1 V^2(x)dx$. In the functional case, the eigenvalues λ_i can be viewed as long–run variances of the residual curves along the principal directions determined by the eigenfunctions of the kernel $c(\cdot,\cdot)$ defined by (2.5). To obtain a test analogous to the scalar KPSS test, with a parameter free limit null distribution, we must construct a statistic which involves a division by consistent estimators of the λ_i. We use only d largest eigenvalues in order not to increase the variability of the statistic caused by division by small empirical eigenvalues. A suitable statistic is

$$R_N^0 = \sum_{i=1}^{d} \frac{1}{\hat{\lambda}_i} \int_0^1 \langle Z_N(x,\cdot), \hat{\phi}_i \rangle^2 dx, \tag{3.5}$$

where the sample eigenvalues $\hat{\lambda}_i$ and eigenfunctions $\hat{\phi}_i$ are defined by (2.8). Statistic (3.5) extends the statistic KPSS_N. Its limit distribution is given in the next theorem.

THEOREM 3.2 *If Assumptions 2.1, 2.2 and* (2.7) *hold, then under null model* (2.1),

$$R_N^0 \xrightarrow{\mathcal{D}} \sum_{i=1}^{d} \int_0^1 V_i^2(x)dx,$$

with the $V_i, 1 \le i \le d$, *the same as in Theorem 3.1.*

Theorem 3.2 is proven in Section 5. Here we only note that that the additional Assumption 2.2 is needed to ensure that (2.4) holds which is known to imply $\hat{\lambda}_i \xrightarrow{P} \lambda_i,\ 1 \le i \le d$.

We conclude this section by discussing the consistency of the tests based on the above theorems. Theorem 3.3 implies that under H_A statistic R_N of Theorem 3.1 increases like N^2. The critical values increase at the rate not greater than N. The test based on Theorem 3.1 is thus consistent. The exact asymptotic behavior under H_A of the normalized statistic R_N^0 appearing in Theorem 3.2 is more difficult to study due to almost intractable asymptotics (under H_A) of the empirical eigenvalues and eigenfunctions of the kernel $\hat{c}(\cdot,\cdot)$. The precise asymptotic behavior under H_A is not known even in the scalar case, i.e. for the statistic KPSS_N. We therefore focus on the asymptotic limit under H_A of the statistic R_N whose derivation is already quite complex. This limit involves iid copies of the process

$$\Delta(x) = \int_0^x W(y)dy + (3x^2 - 4x)\int_0^1 W(y)dy + (-6x^2 + 6x)\int_0^1 yW(y)dy, \quad 0 \le x \le 1, \tag{3.6}$$

where $W(\cdot)$ is a standard Brownian motion.

THEOREM 3.3 *If the errors u_i satisfy Assumption 2.1, then under the alternative* (2.2),

$$\frac{1}{N^2} R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x) dx,$$

where R_N is the test statistic defined in (3.3) *and $\Delta_1, \Delta_2(x), \ldots$ are iid processes defined by* (3.6). *The weights τ_i are the eigenvalues of the long-run covariance kernel of the errors u_i defined analogously to* (2.5) *by*

$$c_u(t,s) = E[u_0(t)u_0(s)] + \sum_{l=1}^{\infty} Eu_0(t)u_l(s) + \sum_{l=1}^{\infty} Eu_0(s)u_l(t). \tag{3.7}$$

The proof of Theorem 3.3 is given in Section 5.

4 Application to yield curves and a simulation study

In this section, we illustrate the theory developed in this paper with an application to yield curves followed by a simulation study. Applications to other asset classes, including currency exchange rates, commodities and equities are presented in Kokoszka and Young (2015) and Young (2016), which also contain details of numerical implementation.

We consider a time series of daily United States Federal Reserve yield curves constructed from discrete rates at maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months. Yield curves are discussed in many finance textbooks, see e.g. Chapter 10 of Campbell *et al.* (1997) or Diebold and Rudebusch (2013). The left panel of Figure 1 shows ten consecutive yield curves. Following the usual practice, each yield curve is treated as a single functional observation, and so the yield curves observed over a period of many days form a functional time series. The right panel of Figure 1 shows the sample period we study, which covers 100 consecutive trading days. It shows a downward trend in interest rates, and we want to test if these curves also contain a random walk component. The tests were performed using $d = 2$. The first two principal components of $\hat{c}$ explain over 95% of variance and provide excellent visual fit. Our selection thus uses three principal shapes to describe the yield curves, the mean function and the first two principal components. It is in agreement with with recent approaches to modeling the yield curve, cf. Hays *et al.* (2012) and Diebold and Rudebusch (2013), which are based on the three component Nelson–Siegel model.

We first apply both tests to the time series of $N = 100$ yield curves shown in the right panel of Figure 1. The test based on statistic R_N, yields the P–value of 0.0282 and the test based on R_N^0, 0.0483, indicating the presence of random walk in addition to a downward trend. Extending the sample by adding the next 150 business days, so that $N = 250$, yields the respective P–values 0.0005 and 0.0013. In all computation the bandwidth $h = N^{2/5}$

was used. Examination of different periods shows that trend stationarity does not hold if the period is sufficiently long. This agrees with the empirical finding of Chen and Niu (2014) whose method of yield curve prediction, based on utilizing periods of approximate stationarity, performs better than predictions based on the whole sample; random walk is not predictable. Even though our tests are motivated by the alternative of a random walk component, they reject any serious violation of trend stationarity. Broadly speaking, our analysis shows that daily yield curves can be treated as a trend stationary functional time series only over certain short periods of time, generally not longer than a calendar quarter.

We complement our data example with a small simulation study. There is a multitude of data generating process that could be used. The following quantities could vary: shapes of the mean and the principal components functions, the magnitude of the eigenvalues, the distribution of the scores and their dependence structure. In this paper, concerned chiefly with theory, we merely want to present a very limited simulation study that validates the conclusions of the data example. We therefore attempt to simulate curves whose shapes resemble those of the real data, and for which either the null or the alternative holds. The artificial data is therefore generated according to the following algorithm.

ALGORITHM 4.1 [Yield curves simulation under H_0]

1. Using real yield curves, calculate the estimates $\hat{\xi}(t)$ and $\hat{\mu}(t)$ defined, respectively, by (2.9) and (2.11). Then compute the residuals $e_n(t)$ defined in (2.12).
2. Calculate the first two empirical principal components $\hat{\phi}_1(t)$ and $\hat{\phi}_2(t)$ using the empirical covariance function

$$\hat{\gamma}_0(s,t) = \frac{1}{N}\sum_{n=1}^{N}(e_n(t) - \bar{e}(t))(e_n(s) - \bar{e}(s)). \tag{4.1}$$

   This step leads to the approximation

$$e_n(t) \approx a_{1,n}\hat{\phi}_1(t) + a_{2,n}\hat{\phi}_2(t), \qquad n = 1, 2, \ldots, N,$$

   where $a_{1,n}$ and $a_{2,n}$ are the first two functional scores. The functions $\hat{\phi}_1(t)$ and $\hat{\phi}_2(t)$ are treated as deterministic, while the scores $a_{1,n}$ and $a_{2,n}$ form random sequences indexed by n.
3. To simulate temporally independent residuals e_n, generate independent in n scores $a'_{1,n} \sim N(0, \sigma^2_{a_1})$ and $a'_{2,n} \sim N(0, \sigma^2_{a_2})$, where $\sigma^2_{a_1}$ and $\sigma^2_{a_2}$ are the sample variances of the real scores, and set

$$e'_n(t) = a'_{1,n}\hat{\phi}_1(t) + a'_{2,n}\hat{\phi}_2(t), \qquad n = 1, 2, \ldots, N.$$

   To simulate dependent residual curves, generate scores $a'_{1,n}, a'_{2,n} \sim \mathrm{AR}(1)$, where each autoregressive process has parameter 0.5.

4. Using the estimated functional parameters $\hat{\mu}(t)$, $\hat{\xi}(t)$ and the simulated residuals $e'_n(t)$, construct the simulated data set

$$X'_n(t) = \hat{\mu}(t) + \hat{\xi}(t)n + e'_n(t), \quad n = 1, 2, \ldots, N. \tag{4.2}$$

Table 1 shows empirical sizes based on 1000 replication of the data generating process described in Algorithm 4.1. We use two ways of estimating the eigenvalues and eigenfunctions. The first one uses the function $\hat{\gamma}_0$ defined by (4.1) (in the scalar case this corresponds to using the usual sample variance rather than estimating the long–run variance). The second uses the estimated long-run covariance function (2.14) with the bandwidth h specified in Table 1. The covariance kernel $\hat{\gamma}_0(t, s)$ is appropriate for independent error curves. The bandwidth $h = N^{1/3}$ is too small, not enough temporal dependence is absorbed. The bandwidth $h = N^{2/5}$ gives fairly consistent empirical size, typically within one percent of the empirical size. The bandwidth h is not relevant when the kernel $\hat{\gamma}_0$ is used. The different empirical sizes reflect random variability due to three different sets of 1000 replications being used. This indicates that with 1000 replications, a difference of one percent in empirical sizes is not significant.

Table 1: Empirical sizes for functional time series generated using Algorithm 4.1.

	Test statistic	R_N			R_N^0		
	DGP	iid normal	iid normal	AR(1)	iid normal	iid normal	AR(1)
N	Cov-kernel	$\hat{\gamma}_0(t,s)$	$\hat{c}(t,s)$	$\hat{c}(t,s)$	$\hat{\gamma}_0(t,s)$	$\hat{c}(t,s)$	$\hat{c}(t,s)$
100	$h = N^{1/3}$	6.3	5.6	9.4	5.9	5.2	9.1
	$h = N^{2/5}$	5.6	4.4	6.6	5.8	3.6	6.5
	$h = N^{1/2}$	5.1	4.8	3.5	4.5	5.1	2.9
250	$h = N^{1/3}$	5.0	4.3	10.2	5.8	5.2	9.4
	$h = N^{2/5}$	5.5	4.9	7.2	4.5	4.1	5.6
	$h = N^{1/2}$	5.5	5.9	4.3	4.8	3.4	3.5
1000	$h = N^{1/3}$	4.8	4.2	7.0	5.9	5.6	7.1
	$h = N^{2/5}$	6.1	6.3	6.3	6.0	5.1	5.7
	$h = N^{1/2}$	5.8	4.9	4.6	5.6	4.7	3.9

To evaluate power, instead of (4.2), the data generating process is

$$X'_n(t) = \hat{\mu}(t) + \hat{\xi}(t)n + \sum_{i=1}^{n} u_i(t) + e'_n(t), \quad n = 1, 2, \ldots, N, \tag{4.3}$$

where the increments u_i are defined by

$$u_i(t) = aN_{i1} \sin\Big(\pi t\Big) + aN_{i2} \sin\Big(2\pi t\Big),$$

Table 2: Empirical power based on the DGP (4.3) and $h = N^{2/5}$.

	Test statistic	R_N			R_N^0		
	DGP	iid normal	iid normal	AR(1)	iid normal	iid normal	AR(1)
	Cov-kernel	$\hat{\gamma}_0(t,s)$	$\hat{c}(t,s)$	$\hat{c}(t,s)$	$\hat{\gamma}_0(t,s)$	$\hat{c}(t,s)$	$\hat{c}(t,s)$
$a = 0.1$	$N = 125$	100.0	89.9	10.1	100.0	87.9	10.4
	$N = 250$	100.0	97.0	27.7	100.0	96.0	21.9
$a = 0.5$	$N = 125$	100.0	91.5	83.1	100.0	89.7	71.2
	$N = 250$	100.0	97.3	96.4	100.0	97.4	92.4

with standard normal $N_{ij},\ j = 1, 2,\ 1 \leq i \leq N$, totally independent of each other. The scalar a quantifies the distance from H_0; i$a = 0$, corresponds to H_0. For all empirical power simulations, we use a 5% size critical value and $h = N^{2/5}$. The empirical power reported in Table 2 increases as the sample size and the distance from H_0 increase. It is visibly higher for iid curves as compared to dependent curves.

5 Proofs of the results of Section 3

5.1 Preliminary results

For ease of reference, we state in this section two theorems which are used in the proofs of the results of Section 3. Theorem 5.1 is well-known, see Theorem 4.1 in Billingsley (1968). Theorem 5.2 was recently established in Berkes *et al.* (2013).

THEOREM 5.1 *Suppose Z_N, Y_N, Y are random variables taking values in a separable metric space with the distance function ρ. If $Y_N \xrightarrow{\mathcal{D}} Y$ and $\rho(Z_N, Y_N) \xrightarrow{P} 0$, then $Z_N \xrightarrow{\mathcal{D}} Y$.*

In our setting, we work in the metric space $D([0,1], L^2)$ which is the space of right-continuous functions with left limits taking values in $L^2([0,1])$. A generic element of $D([0,1], L^2)$ is

$$z = \{z(x,t), 0 \leq x \leq 1, 0 \leq t \leq 1\}.$$

For each fixed x, $z(x,\cdot) \in L^2$, so $\|z(x,\cdot)\|^2 = \int z^2(x,t)dt < \infty$. The uniform distance between $z_1, z_2 \in D([0,1], L^2)$ is

$$\rho(z_1, z_2) = \sup_{0 \leq x \leq 1} |z_1(x,\cdot) - z_2(x,\cdot)\| = \sup_{0 \leq x \leq 1} \Big\{ \int (z_1(x,t) - z_2(x,t))^2 dt \Big\}^{1/2}.$$

In the following, we work with the space $D([0,1], L^2)$ equipped with the uniform distance.

THEOREM 5.2 *If Assumption 2.1 holds, then* $\sum_{i=1}^{\infty} \lambda_i < \infty$, *and we can construct a sequence of Gaussian processes* $\Gamma_N(x,t)$ *such that for every* N

$$\{\Gamma_N(x,t), 0 \le x,t \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma(x,t), 0 \le x,t \le 1\},$$

where

$$\Gamma(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x)\phi_i(t), \tag{5.1}$$

and

$$\kappa_N = \sup_{0\le x\le 1} \|V_N(x,\cdot) - \Gamma_N(x,\cdot)\| = o_p(1). \tag{5.2}$$

Recall that the W_i are independent standard Wiener processes, λ_i and ϕ_i are defined in (2.6) and $V_N(x,t)$ is defined in (3.2).

5.2 Proof of Theorem 3.1

The proof of Theorem 3.1 is constructed from several lemmas decomposing the statistic R_N into a form suitable for the application of the results of Section 5.1, i.e. to leading and asymptotically negligible terms. Their proofs are presented in the on–line supplement to this paper. Throughout this section, we assume that the null model (2.1) and Assumption 2.1 hold.

LEMMA 5.1 *For the* s_N *is defined in* (2.10),

$$\frac{s_N}{N^3} \to \frac{1}{12}, \quad \text{as } N \to \infty.$$

LEMMA 5.2 *For the functional slope estimate* $\hat{\xi}$ *defined by* (2.9),

$$\hat{\xi}(t) - \xi(t) = \frac{1}{s_N} \sum_{n=1}^{N} \Big(n - \frac{N+1}{2}\Big)\eta_n(t).$$

LEMMA 5.3 *The following identity holds*

$$\frac{N^{3/2}}{s_N} \sum_{n=1}^{N} \Big(n - \frac{N+1}{2}\Big)\eta_n(t) = \frac{1}{N^{-3}s_N} \left\{ \Big(\frac{N-1}{2N}\Big) V_N(1,t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\Big(\frac{k}{N}, t\Big) \right\},$$

where $V_N(x,t)$ *is the partial sum process of the errors defined in* (3.2).

LEMMA 5.4 *The process $Z_N(x,t)$ defined by* (3.4) *admits the decomposition*

$$Z_N(x,t) = V_N(x,t) - \frac{\lfloor Nx \rfloor}{N} V_N(1,t) \\ - \frac{1}{2}\frac{1}{N^{-3}s_N}\left\{ \Big(\frac{N-1}{2N}\Big) V_N(1,t) - \frac{1}{N}\sum_{k=1}^{N-1} V_N\Big(\frac{k}{N},t\Big)\right\} \frac{\lfloor Nx \rfloor}{N}\left(\frac{\lfloor Nx \rfloor}{N} - 1 \right).$$

LEMMA 5.5 *The following convergence holds*

$$\int \Big\{ \frac{1}{N}\sum_{k=1}^{N} V_N\Big(\frac{k}{N},t\Big) - \int_0^1 \Gamma_N(y,t)dy \Big\}^2 dt \xrightarrow{P} 0,$$

where the Γ_N are the Gaussian processes in Theorem 5.2.

LEMMA 5.6 *Consider the process $\Gamma(\cdot,\cdot)$ defined by* (5.1) *and set*

$$\Gamma^0(x,t) = \Gamma(x,t) + \Big(2x - 3x^2\Big)\Gamma(1,t) + \Big(-6x + 6x^2\Big) \int_0^1 \Gamma(y,t)dy. \tag{5.3}$$

Then

$$\int_0^1 \|\Gamma^0(x,\cdot)\|^2 dx = \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x)dx.$$

LEMMA 5.7 *For the processes $Z_N(\cdot,\cdot)$ and $\Gamma_N^0(\cdot,\cdot)$ defined, respectively, in* (3.4) *and* (5.3),

$$\sup_{0\le x\le 1} \|Z_N(x,\cdot) - \Gamma_N^0(x,\cdot)\| \xrightarrow{P} 0. \tag{5.4}$$

Using the above lemmas, we can now present a compact proof of Theorem 3.1.

PROOF OF THEOREM 3.1: Recall that the test statistic R_N is defined by $R_N = \iint Z_N^2(x,t)dxdt$, where

$$Z_N(x,t) = S_N(x,t) - \frac{\lfloor Nx \rfloor}{N} S_N(1,t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \Big(\frac{\lfloor Nx \rfloor}{N} \Big(\frac{\lfloor Nx \rfloor}{N} - 1 \Big) \Big)$$

with $S_N(x,t)$ and $\hat{\xi}(t)$ are respectively defined in equations (3.1) and (2.9). Recall that

$$\Gamma_N^0(x,t) = \Gamma_N(x,t) + \Big(2x - 3x^2\Big)\Gamma_N(1,t) + \Big(-6x + 6x^2\Big) \int_0^1 \Gamma_N(y,t)dy,$$

and

$$\Gamma^0(x,t) = \Gamma(x,t) + \Big(2x - 3x^2\Big)\Gamma(1,t) + \Big(-6x + 6x^2\Big) \int_0^1 \Gamma(y,t)dy.$$

From Lemma 5.7, we know that

$$\rho(Z_N(x,\cdot),\Gamma_N^0(x,\cdot)) = \sup_{0\le x\le 1}\|Z_N(x,\cdot)-\Gamma_N^0(x,\cdot)\| \xrightarrow{P} 0.$$

By Theorem 5.2, $\Gamma_N^0(x,t) \stackrel{\mathcal{D}}{=} \Gamma^0(x,t)$. Thus, Theorem 5.1 implies that

$$Z_N(x,t) \xrightarrow{\mathcal{D}} \Gamma^0(x,t).$$

By Lemma 5.6,

$$\iint (\Gamma^0(x,t))^2 dxdt \stackrel{D}{=} \sum_{i=1}^{\infty}\lambda_i \int_0^1 V_i^2(x)dx.$$

Thus, by the continuous mapping theorem,

$$R_N = \iint (Z_N(x,t))^2 dxdt \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty}\lambda_i \int V_i^2(x)dx,$$

which proves the desired result. □

5.3 Proof of Theorem 3.2

The key fact needed in the proof is the consistency of the sample eigenvalues $\hat{\lambda}_i$ and eigenfunctions $\hat{\phi}_i$. The required result, stated in (5.5), follows fairly directly from (2.4). However, the verification that (2.4) holds for the kernel estimator (2.14) is not trivial. The required result can be stated as follows.

THEOREM 5.3 *Suppose Assumption 2.1 holds with $\delta = 0$ and $\kappa = 2$. If H_0 and Assumption 2.2 hold, then relation* (2.4) *holds.*

Observe that assuming that relation (2.3) in Assumption 2.1 holds with $\delta = 0$ and $\kappa = 2$ weakens the universal assumption that it holds with some $\delta > 0$ and $\kappa > 2 + \delta$.

We first present the proof of Theorem 3.2, which uses Theorem 5.3, and then turn to a rather technical proof of Theorem 5.3.

PROOF OF THEOREM 3.2: If Assumptions 2.1, 2.2, condition (2.7) and H_0 hold, then

$$\max_{1\le i\le d}|\hat{\lambda}_i - \lambda_i| = o_p(1) \quad \text{and} \quad \max_{1\le i\le d}\|\hat{\phi}_i - \hat{c}_i\phi_i\| = o_p(1), \tag{5.5}$$

where $\hat{c}_1$, $\hat{c}_2$, ..., $\hat{c}_d$ are unobservable random signs defined as $\hat{c}_i = \text{sign}(\langle\hat{\phi}_i, \phi_i\rangle)$. Indeed, Theorem 5.3 states that relation (2.4) holds under H_0 and Assumptions 2.1 and 2.2. Relations (5.5) follow from (2.4) and Lemmas 2.2. and 2.3 of Horváth and Kokoszka (2012)

which state that the differences of the eigenvalues and eigenfunctions are bounded by the Hilbert–Schmidt norm of the difference of the corresponding operators.

Using (5.1), it is easy to see that for all N

$$\{\langle \Gamma_N^0(x,\cdot),\phi_i\rangle, 0\le x\le 1, 1\le i\le d\} \stackrel{\mathcal{D}}{=} \{\sqrt{\lambda_i}V_i(x), 0\le x\le 1, 1\le i\le d\}. \tag{5.6}$$

We first show that

$$\sup_{0\le x\le 1} |\langle Z_N(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N^0(x,\cdot),\hat{c}_i\phi_i\rangle| \xrightarrow{P} 0. \tag{5.7}$$

By the Cauchy-Schwarz inequality and Lemma 5.7, we know

$$\sup_{0\le x\le 1} |\langle Z_N(x,\cdot) - \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle| \le \sup_{0\le x\le 1} \|Z_N(x,\cdot) - \Gamma_N^0(x,\cdot)\| = o_p(1).$$

Again by the Cauchy-Schwarz inequality and (5.5), we have

$$\sup_{0\le x\le 1} |\langle \Gamma_N^0(x,\cdot),\hat{\phi}_i - \hat{c}_i\phi_i\rangle| \le \sup_{0\le x\le 1} \|\Gamma_N^0(x,\cdot)\|\|\hat{\phi}_i - \hat{c}_i\phi_i\| = o_p(1).$$

Then using the triangle inequality and inner product properties,

$$\begin{aligned}
&\sup_{0\le x\le 1} |\langle Z_N(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N(x,\cdot),\hat{c}_i\phi_i\rangle| \\
&\quad= \sup_{0\le x\le 1} |\langle Z_N(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle + \langle \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N^0(x,\cdot),\hat{c}_i\phi_i\rangle| \\
&\quad\le \sup_{0\le x\le 1} |\langle Z_N(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle| + \sup_{0\le x\le 1} |\langle \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle - \langle \Gamma_N^0(x,\cdot),\hat{c}_i\phi_i\rangle| \\
&\quad= \sup_{0\le x\le 1} |\langle Z_N(x,\cdot) - \Gamma_N^0(x,\cdot),\hat{\phi}_i\rangle| + \sup_{0\le x\le 1} |\langle \Gamma_N^0(x,\cdot),\hat{\phi}_i - \hat{c}_i\phi_i\rangle| \\
&\quad= o_p(1),
\end{aligned}$$

which proves relation (5.7). Thus by Theorem 5.1, (5.5), (5.7), (5.6) and the continuous mapping theorem,

$$R_N^0 = \sum_{i=1}^d \frac{1}{\hat{\lambda}_i}\int \langle Z_N(x,\cdot),\hat{\phi}_i\rangle^2 dx \xrightarrow{D} \sum_{i=1}^d \int V_i^2(x)dx.$$

□

Proof of Theorem 5.3: Recall definitions of the kernels c and $\hat{c}$ given, respectively, in (2.5) and (2.14). The claim will follow if we can show that

$$\iint \{\hat{\gamma}_0(t,s) - E[\eta_0(t)\eta_0(s)]\}^2 dtds = o_P(1) \tag{5.8}$$

and

$$\iint \left\{ \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \hat{\gamma}_i(t,s) - \sum_{i\geq 1} E[\eta_0(s)\eta_i(t)] \right\}^2 dtds = o_P(1). \tag{5.9}$$

These relations are established in a sequence of Lemmas which split the argument by isolating the terms related to the estimation of trend from those related to the autocovariances of the η_i. The latter terms were treated in Horváth *et al.* (2013), so the present proof focuses on the extra terms appearing in our context. The proofs of Lemmas 5.8 and 5.9 are presented in the on–line supplement.

LEMMA 5.8 *Relation* (5.8) *holds under the assumptions of Theorem 5.3.*

LEMMA 5.9 *Relation* (5.9) *holds under the assumptions of Theorem 5.3.*

5.4 Proof of Theorem 3.3

The proof of Theorem 3.3 is constructed from several lemmas which are proven in the on–line supplement.

LEMMA 5.10 *Under the alternative* (2.2)*, for the functional slope estimate* $\hat{\xi}$ *defined by* (2.9)*,*

$$N^{3/2}(\hat{\xi}(t) - \xi(t)) = \frac{1}{N^{-3}s_N}\left\{ \Big(\frac{N-1}{2N}\Big) V_N(1,t) - \frac{1}{N}\sum_{k=1}^{N-1} V_N\Big(\frac{k}{N},t\Big) \right\}$$
$$+ \frac{1}{N^{-3}s_N}\left\{ \sum_{n=1}^{N} \frac{n}{N} Y_N\Big(\frac{n}{N},t\Big) - \Big(\frac{N+1}{2N}\Big) \sum_{n=1}^{N} Y_N\Big(\frac{n}{N},t\Big) \right\},$$

where $V_N(x,t)$ *is the partial sum process of the errors* η_n *defined in* (3.2) *and* $Y_N(x,t)$ *is the partial sum process of the random walk errors* u_n *defined by*

$$Y_N(x,t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} u_n(t). \tag{5.10}$$

LEMMA 5.11 *Under the alternative, $Z_N(x,t)$ defined in* (3.4) *can be expressed as*

$$
\begin{aligned}
Z_N(x,t) = V_N(x,t) - \frac{\lfloor Nx \rfloor}{N} V_N(1,t) \\
- \frac{1}{2}\frac{1}{N^{-3}s_N}\left\{ \Big(\frac{N-1}{2N}\Big) V_N(1,t) - \frac{1}{N}\sum_{k=1}^{N-1} V_N\Big(\frac{k}{N},t\Big) \right\} \left(\frac{\lfloor Nx \rfloor}{N}\left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
+ \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\Big(\frac{n}{N},t\Big) - \frac{\lfloor Nx \rfloor}{N}\sum_{n=1}^{N} Y_N\Big(\frac{n}{N},t\Big) \\
- \frac{1}{2}\frac{1}{N^{-3}s_N}\left\{ \sum_{n=1}^{N} \frac{n}{N} Y_N\Big(\frac{n}{N},t\Big) - \Big(\frac{N+1}{2N}\Big)\sum_{n=1}^{N} Y_N\Big(\frac{n}{N},t\Big) \right\} \left(\frac{\lfloor Nx \rfloor}{N}\left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}
$$

Since the u_i satisfy Assumption 2.1, an analog of Theorem 5.2 holds, i.e. there exist Gaussian processes Λ_N equal in distribution to

$$
\Lambda(x,t) = \sum_{i=1}^{\infty} \tau_i^{1/2} W_i(x)\psi_i(t), \tag{5.11}
$$

where τ_i, ψ_i are, respectively, the eigenvalues and the eigenfunctions of the kernel (3.7). Moreover, for the partial sum process Y_N defined by (5.10),

$$
l_n = \sup_{0\le x\le 1} ||Y_N(x,\cdot) - \Lambda_N(x,\cdot)|| = o_p(1). \tag{5.12}
$$

LEMMA 5.12 *Under the alternative, the following convergence holds*

$$
\sup_{0\le x\le 1} \|N^{-1} Z_N^A(x,\cdot) - \Delta_N^0(x,\cdot)\| \xrightarrow{P} 0,
$$

where the processes $Z_N^A(\cdot,\cdot)$ and $\Delta_N^0(\cdot,\cdot)$ are respectively defined by

$$
\begin{aligned}
Z_N^A(x,t) = \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\Big(\frac{n}{N},t\Big) - \frac{\lfloor Nx \rfloor}{N}\sum_{n=1}^{N} Y_N\Big(\frac{n}{N},t\Big) \\
-\frac{1}{2}\frac{1}{N^{-3}s_N}\left\{ \sum_{n=1}^{N} \frac{n}{N} Y_N\Big(\frac{n}{N},t\Big) - \Big(\frac{N+1}{2N}\Big)\sum_{n=1}^{N} Y_N\Big(\frac{n}{N},t\Big) \right\} \left(\frac{\lfloor Nx \rfloor}{N}\left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right)
\end{aligned} \tag{5.13}
$$

and

$$
\Delta_N^0(x,t) = \int_0^x \Lambda_N(y,t)dy + (3x^2 - 4x)\int_0^1 \Lambda_N(y,t)dy + (-6x^2 + 6x)\int_0^1 y\Lambda_N(y,t)dy. \tag{5.14}
$$

LEMMA 5.13 *Under the alternative, the following convergence holds*

$$\sup_{0\le x\le 1} \|N^{-1}Z_N(x,\cdot) - \Delta_N^0(x,\cdot)\| \xrightarrow{P} 0,$$

where the process $Z_N(\cdot,\cdot)$ *is defined in* (3.4) *and the process* $\Delta_N^0(\cdot,\cdot)$ *in* (5.14).

LEMMA 5.14 *Consider the process* $\Lambda(\cdot,\cdot)$ *defined by* (5.11) *and set*

$$\Delta^0(x,t) = \int_0^x \Lambda(y,t)dy + (3x^2-4x)\int_0^1 \Lambda(y,t)dy + (-6x^2+6x)\int_0^1 y\Lambda(y,t)dy. \tag{5.15}$$

Then

$$\int_0^1 \|\Delta^0(x,\cdot)\|^2 dx = \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x)dx,$$

where $\tau_1, \tau_2, \ldots$ *are eigenvalues of the long–run covariance function of the* u_i*, i.e.* (3.7)*, and* $\Delta_1, \Delta_2, \ldots$ *are independent copies of the process* Δ *defined in* (3.6).

PROOF OF THEOREM 3.3: Recall that the test statistic R_N is defined by $R_N = \iint Z_N^2(x,t)dxdt$, with Z_N defined by (3.4). We want to show that under the alternative model (2.2),

$$\frac{1}{N^2}R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x)dx,$$

where $\Delta_1, \Delta_2, \ldots$ are independent copies of the process defined by (3.6) and $\tau_1, \tau_2, \ldots$ are the eigenvalues of the long-run covariance kernel (3.7). By Lemma 5.13,

$$\rho(N^{-1}Z_N(x,\cdot), \Delta_N^0(x,\cdot)) = \sup_{0\le x\le 1} \|N^{-1}Z_N(x,\cdot) - \Delta_N^0(x,\cdot)\| \xrightarrow{P} 0.$$

By construction, $\Delta_N^0 \stackrel{D}{=} \Delta^0$, so Theorem 5.1 implies that $N^{-1}Z_N \xrightarrow{D} \Delta^0$. By Lemma 5.14,

$$\iint (\Delta^0(x,t))^2 dxdt \stackrel{D}{=} \sum_{i=1}^{\infty} \lambda_i \int_0^1 \Delta_i^2(x)dx.$$

Thus by the continuous mapping theorem,

$$\frac{1}{N^2}R_N = \iint (N^{-1}Z_N(x,t))^2 dxdt \xrightarrow{D} \sum_{i=1}^{\infty} \lambda_i \int \Delta_i^2(x)dx.$$

□

References

Antoniadis, A., Paparoditis, E. and Sapatinas, T. (2006). A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series* B, **68,** 837–857.

Aue, A., Dubard-Norinho, D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110,** 378–392.

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

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

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

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

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

Dickey, D. A. and Fuller, W. A. (1979). Distributions of the estimattors for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74,** 427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with unit root. *Econometrica*, **49,** 1057–1074.

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

Giraitis, L., Kokoszka, P. S., Leipus, R. and Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, **112,** 265–294.

Hays, S., Shen, H. and Huang, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, **6,** 870–894.

Hörmann, S., Kidzinski, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society (B)*, **77,** 319–348.

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

Hörmann, S. and Kokoszka, P. (2012). Functional time series. In *Time Series* (eds C. R. Rao and T. Subba Rao), Handbook of Statistics, volume 30. Elsevier.

Horváth, L., Hušková, M. and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, **101,** 352–367.

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

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, **75,** 103–122.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, **179,** 66–82.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley.

Jentsch, C. and Subba Rao, S. (2015). A test for second order stationarity of a multivariate time series. *Journal of Econometrics*, **185,** 124–161.

de Jong, R. M., Amsler, C. and Schmidt, P. (1997). A robust version of the KPSS test based on indicators. *Journal of Econometrics*, **137,** 311–333.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, **99,** 2508–2526.

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, **34,** 116–129.

Kokoszka, P. and Young, G. (2015). Testing trend stationarity of functional time series with application to yield and daily price curves. Technical Report. Colorado State University.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, **54,** 159–178.

Lee, D. and Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally integrated alternatives. *Journal of Econometrics*, **73,** 285–302.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, **59,** 1279–1313.

MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics*, **6,** 422–433.

Müller, H-G., Sen, R. and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, **165,** 233–245.

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, **41,** 568–603.

Pötscher, B. and Prucha, I. (1997). *Dynamic Non–linear Econonometric Models. Asymptotic Theory*. Springer.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive–moving average models of unknown order. *Biometrika*, **71,** 599–608.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, **35,** 1773–1801.

Wu, W. (2005). *Nonlinear System Theory: Another Look at Dependence*. Proceedings of The National Academy of Sciences of the United States, volume 102. National Academy of Sciences.

Young, G. (2016). Inference for functional time series with applications to yield curves and intraday cumulative returns. Ph.D. Thesis. Colorado State University, Fort Collins, CO, USA.