Testing for Parallelism Among Trends in Multiple Time Series

David Degras, Zhiwei Xu, Ting Zhang, and Wei Biao Wu

***Abstract*—This paper considers the inference of trends in multiple, nonstationary time series. To test whether trends are parallel to each other, we use a parallelism index based on the L^2-distances between nonparametric trend estimators and their average. A central limit theorem is obtained for the test statistic and the test's consistency is established. We propose a simulation-based approximation to the distribution of the test statistic, which significantly improves upon the normal approximation. The test is also applied to devise a clustering algorithm. Finally, the finite-sample properties of the test are assessed through simulations and the test methodology is illustrated by a cell phone download data collected in the United States.**

***Index Terms*—Central limit theorem, clustering, multiple time series, nonstationary time series, testing for parallelism.**

I. Introduction

COMPARISON of trends or regression curves is a common problem in applied sciences. For example in longitudinal clinical studies, evaluators are interested in comparing response curves for treatment and control groups. In agriculture, it may be relevant to compare at different spatial locations the relationship between yield per plant and plant density [1]. In biology, assessing parallelism between sets of dose-response data allows to determine if the biological response to two substances is similar or if two different biological environments give similar dose-response curves to the same substance [2]. Also in economics, a standard problem is to compare the yield over time of US Treasury bills at different maturities, or the evolution of long-term rates in several countries [3].

The statistical methodology developed in this paper is motivated by a collection of time series of cell phone download activity (applications, audio, images, ringtones, and wall papers) collected in the United States between September 2005 and June 2006. The measurements were collected hourly and aggregated at the area code level (129 area codes were observed in total). A question of considerable interest is to determine whether the download trends in the area codes are identical up to scale differences. (Scale differences can be expected because of the differences in numbers of phone users for each area code.) If this hypothesis was true, it could be asserted that for those area codes whose time series temporarily display slower growth rates than average, the growth deficit is nonstructural (i.e. it is caused by local random variations and not by the deterministic trend). By placing more advertising efforts and commercial incentives in these areas, the phone company and its commercial partners could thus expect cell phone downloads to increase. Another interesting application of comparing trends in cell phone activity pertains to the allocation of bandwidth in phone networks.

After a pilot study revealed a multiplicative structure in the trend, seasonality, and irregularities of the time series, a logarithmic transform was applied to the data so as to stabilize the variance and obtain an additive (signal+noise) model. In this context, an efficient way to test for proportionality between the trends in the initial data is to consider the alternative problem of testing for parallelism between the trends in the log-transformed time series. From here on, we consider an additive nonparametric time-series model that, in our analysis of the cell phone download data, pertains to the log-transformed data. Suppose that we observe N time series $\{X_{it}\}_{t=1}^{T}$, $i = 1, \ldots, N$, according to the model

$$X_{it} = \mu_i(t/T) + e_{it}, \quad t = 1, \ldots, T \tag{1}$$

where the μ_i are unknown smooth (deterministic) functions defined over [0, 1] and the $\{e_{it}\}_{t=1}^{T}$ are mean-zero error processes. The regression functions μ_i are deterministic and represent the trends of the time series $\{X_{it}\}_{t=1}^{T}$. Throughout the paper, we shall deal with deterministic trends. The scaling device $\frac{t}{T}$ in (1) indicates that the means $\mathbb{E}X_{it}$ change smoothly in time, due to the smoothness of $\mu_i(\cdot)$. It is widely used in statistics and econometrics; see, for example, [4] and [5]. We are interested in testing whether the μ_i, $i = 1, \ldots, N$, are parallel, namely, whether there exists a function μ and numbers c_i such that

$$H_0: \quad \mu_i(u) = c_i + \mu(u), \quad i = 1, \ldots, N, \quad u \in [0, 1]. \tag{2}$$

Under H_0, the c_i represent vertical shifts between the curves μ_i and the reference curve μ. They can be viewed as nuisance parameters for testing purposes. Note that testing for parallelism is closely related to testing for equality as, on the one hand, H_0 is formally equivalent to equality of the N centered functions $(\mu_i - \int_0^1 \mu_i(u)du)$ and, on the other hand, the scalars $\int_0^1 \mu_i(u)du$ can be considered as known since they are estimated at parametric rates while the functions μ_i and μ are estimated at slower nonparametric rates.

Manuscript received May 12, 2011; revised September 18, 2011 and November 08, 2011; accepted November 10, 2011. Date of publication December 02, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kainam Thomas Wong. This work is supported in part by NSF Grants DMS-0906073 and DMS-1106970.

D. Degras is with the Department of Mathematical Sciences, DePaul University, Chicago, IL 60614 USA (e-mail: ddegrasv@depaul.edu).

Z. Xu is with the Department of Computer and Information Science, University of Michigan-Dearborn, Dearborn, MI 48128 USA (e-mail: zwxu@umich.edu).

T. Zhang and W. B. Wu are with the Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: tzhang@galton.uchicago.edu; wbwu@galton.uchicago.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2177831

Various tests for comparing mean functions can be found in the regression literature. Reference [6] compares two nonparametric regression curves by testing whether one of them is a parametric transformation of the other. To test the equality of $N = 2$ regression curves in the setup of independent errors, [7] proposes a bootstrap test, [8] devises a procedure based on the L^2-distance between kernel regression estimators, and [9] applies a wavelet-based method. For $N \geq 2$, assuming independent errors, [10] uses a test based on weighted L^2-distances that requires no smoothing parameter selection. To test whether a nonparametric mean curve has a certain parametric shape, [11] and [12] appeal to signal processing theory and the Whittaker–Shannon sampling theorem under independent errors while [13] utilizes approximate simultaneous confidence bands in the functional data setup. Under model and design conditions, their tests can be adapted to assess parallelism for two mean curves. Reference [1] builds ANOVA-type tests for equality and parallelism in $k \geq 2$ regression curves under independent and identically distributed (i.i.d.) errors. In the time-series setup, to infer an equality of two trends, [3] applies a graphical device assuming stationary, weakly correlated errors; [14] builds a test based on the cumulative regression functions, assuming long-memory moving average errors; [15] uses an adaptive Neyman test with stationary Gaussian linear error processes. For random designs of observations, contributions to the comparison of regression curves include [16]–[18].

The present work brings several contributions to the statistical problem of testing parallelism between trends in multiple time series. First, studies to date are based on one or both of the following assumptions: i) the error processes $\{e_{it}\}_{t=1}^{T}$ in (1) are independent in time, or more generally stationary; and ii) the number N of time series is fixed and usually small. In this paper, we relax both assumptions: the $\{e_{it}\}_{t=1}^{T}$ can be nonstationary and N can be arbitrarily large. We describe the dependence of the $\{e_{it}\}_{t=1}^{T}$ in terms of the physical dependence model of [19], which represents errors as being generated by series of i.i.d. innovations. The data-generating mechanism may be nonlinear with respect to the innovation process and may vary smoothly over time. This nonstationary dependence is realistic in practice and it generalizes the parametric or stationarity assumptions of the literature. Second, we devise a method of independent interest to estimate consistently the long-run variance function of a locally stationary time series. Third, we exploit a strong invariance principle to build a simulation-based method that approximates the finite sample distribution of the test statistic. The resulting approximation is more accurate than the limiting normal distribution and its implementation is faster than bootstrap alternatives. Fourth, we apply the test to an iterative clustering algorithm that groups time series according to the parallelism of their trends. A nice feature of the algorithm is that it does not require the specification of the number of clusters: time series that are very different from all others may form groups of their own. For this reason, the algorithm provides valuable insights in the data that complement standard approaches like k-means clustering. Another attractive by-product of the clustering algorithm is that it readily provides significance levels for all clusters.

The rest of the paper is organized as follows. Section II presents a test statistic based on the L^2-distances between the estimators of the individual trends $(\mu_i - c_i)$ and the estimator of the global trend μ in (2). The test statistic estimates a parallelism index. Its asymptotic properties are discussed in Section III for both fixed N and $N \to \infty$. A central limit theorem is derived under (2) and the test is shown to be consistent against local alternatives. Section IV deals with the test implementation and provides methods for bandwidth selection and long-run variance estimation, as well as a simulation-based method to approximate the finite-sample distribution of the test statistic. Simulations are carried out in Section V to assess the empirical significance level and statistical power of the test procedure. The clustering algorithm is described in Section VI and illustrated in Section VII with the cell phone download data. Proofs of the main results are deferred to the Appendix.

II. Test Statistic

To ensure model identifiability under the null hypothesis H_0 in (2), we assume that

$$\sum_{i=1}^{N} c_i = 0. \tag{3}$$

A natural way to test H_0 is to compare the curves $\hat{\mu}_i$ estimated under the general model (1) to the curves $\hat{c}_i + \hat{\mu}$ estimated under H_0. To estimate the common trend μ under H_0, we can use the averaged process $\bar{X}_{\cdot t} = \frac{\sum_{i=1}^{N} X_{it}}{N}$ for $t = 1, \ldots, T$:

$$\bar{X}_{\cdot t} = \mu(t/T) + \bar{e}_{\cdot t}. \tag{4}$$

Similarly, define $\bar{X}_{i\cdot} = \frac{\sum_{t=1}^{T} X_{it}}{T}$, $\bar{X}_{\cdot\cdot}$, $\bar{e}_{\cdot t}$ and $\bar{e}_{i\cdot}$. In this paper, we adopt the popular local linear smoothing procedure [20] to estimate the trends. Let K be a Lipschitz continuous, bounded, symmetric kernel function with support $[-1, 1]$ and satisfies $\int_{-1}^{1} K(u)du = 1$; let $b > 0$ be the bandwidth. Then the local linear estimator of μ is

$$\hat{\mu}(u) = \sum_{t=1}^{T} w_b(t, u)\bar{X}_{\cdot t}, \quad 0 \leq u \leq 1, \tag{5}$$

with the weights w_b defined by

$$w_b(t, u) = K\left((u - t/T)/b\right) \frac{S_{b,2}(u) - (u - t/T)S_{b,1}(u)}{S_{b,2}(u)S_{b,0}(u) - S_{b,1}^2(u)} \tag{6}$$

where

$$S_{b,j}(u) = \sum_{t=1}^{T} (u - t/T)^j K\left((u - t/T)/b\right), \quad u \in [0, 1]. \tag{7}$$

For simplicity of the procedure, it is advantageous to estimate μ_i with the same bandwidth used for μ. This also simplifies mathematical derivations (see Section IV-A). In this case, the local linear estimate for μ_i is

$$\hat{\mu}_i(u) = \sum_{t=1}^{T} w_b(t, u)X_{it}. \tag{8}$$

The intercepts c_i are estimated by

$$\hat{c}_i = \frac{1}{T} \sum_{t=1}^{T} \left[\hat{\mu}_i(t/T) - \hat{\mu}(t/T)\right]. \tag{9}$$

Since the same bandwidth b is used in (5) and (8), we have the interesting observation that $\hat{\mu}(u) = N^{-1}\sum_{i=1}^{N}\hat{\mu}_i(u)$. Therefore, the $\hat{c}_i$ naturally satisfy the constraint (3).

There are many ways to measure the distance between the curves $\hat{c}_i + \hat{\mu}(\cdot)$ and $\hat{\mu}_i(\cdot)$. In this paper, we adopt the L^2-distance

$$\widehat{\Delta}_{N,T} = \sum_{i=1}^{N}\int_0^1 \left[\hat{\mu}_i(u) - \hat{c}_i - \hat{\mu}(u)\right]^2 du. \tag{10}$$

Clearly $\widehat{\Delta}_{N,T}$ is a natural estimate for the parallelism index

$$\Delta_N = \min_{\substack{\mu, c_1,\ldots,c_N \\ \sum_i c_i = 0}} \sum_{i=1}^{N}\int_0^1 \left[\mu_i(u) - c_i - \mu(u)\right]^2 du, \tag{11}$$

where the explicit solutions are $\mu(u) = \frac{\sum_{i=1}^{N}\mu_i(u)}{N}$ and $c_i = \int_0^1[\mu_i(u) - \mu(u)]du$.

III. Asymptotic Theory

Here, we shall discuss the limiting distribution and consistency of the test. In our framework we allow both N and T to go to infinity, and the error processes $\{e_{it}\}_{t=1}^{T}$ can be nonstationary. To establish the asymptotic normality of $\widehat{\Delta}_{N,T}$, we impose structural conditions on the error processes $\{e_{it}\}_{t=1}^{T}$, following the ideas of [19]. Specifically, we assume that the $\{e_{it}\}_{t=1}^{T}$, $i = 1,\ldots,N$, are i.i.d. as a process $\{e_t\}_{t=1}^{T}$ of the form

$$e_t = G(t/T; \mathcal{F}_t), \tag{12}$$

where $\mathcal{F}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$, $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ is an innovation process with i.i.d. elements, and $G(\cdot;\cdot)$ is a measurable function. Equation (12) can be interpreted as an input/output physical system where the $\{\varepsilon_j\}_{j=-\infty}^{t}$ are the inputs and e_t is the output. Assuming that $G(u;\mathcal{F}_t)$ has a finite pth moment for some $p > 0$, define the physical dependence measure

$$\delta_p(t) = \sup_{0\le u\le 1}\left\|G(u;\mathcal{F}_t) - G\left(u;\mathcal{F}_t'\right)\right\|_p \tag{13}$$

where $\mathcal{F}_t' = (\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$ and ε_0' is a random variable such that $\varepsilon_0', \varepsilon_t, t \in \mathbb{Z}$, are i.i.d. The index $\delta_p(t)$ quantifies the dependence of the output e_t on the inputs $\mathcal{F}_t$ by measuring the distance between $G(\cdot;\mathcal{F}_t)$ and its coupled version $G(\cdot;\mathcal{F}_t')$. Furthermore, assume that $G(u;\mathcal{F}_t)$ is stochastically Lipschitz continuous (SLC), that is, there exists a constant C such that

$$\left\|G(u_1;\mathcal{F}_t) - G(u_2;\mathcal{F}_t)\right\|_p \le C|u_1 - u_2| \tag{14}$$

for all $u_1, u_2 \in [0,1]$, which we denote by $G \in SLC$. The previous equation expresses the fact that the data generating mechanism can change smoothly over time (local stationarity). It includes stationarity as a special case. Our asymptotic theory may fail if the data-generating mechanism drastically changes over time. Note that the $\{e_{it}\}_{t=1}^{T}$ can be represented in the following manner: let $\varepsilon_{ik}, i = 1,\ldots,N, k \in \mathbb{Z}$, be i.i.d. random variables; let $\mathcal{F}_{it} = (\ldots, \varepsilon_{i,t-1}, \varepsilon_{it})$, then

$$e_{it} = G(t/T; \mathcal{F}_{it}). \tag{15}$$

Assuming that $\mathbb{E}e_k = 0$ for all $k \in \mathbb{Z}$, let

$$\gamma_k(u) = \mathbb{E}\left[G(u;\mathcal{F}_k)G(u;\mathcal{F}_0)\right], \quad 0 \le u \le 1. \tag{16}$$

Define the long-run variance function

$$g(u) = \sum_{k\in\mathbb{Z}}\gamma_k(u) \tag{17}$$

and its squared integral

$$\sigma^2 = \int_0^1 g^2(u)du. \tag{18}$$

Recall that the kernel function K is symmetric and Lipschitz continuous on its support $[-1,1]$. Let

$$K^*(x) = \int_{-1}^{1-2|x|} K(v)K\left(v + 2|x|\right)dv$$

and $K_2^* = \int_{-1}^{1}\left[K^*(v)\right]^2 dv$. We have the following result.

Theorem 1: Let $N = N(T)$ be such that either i) $N \to \infty$ as $T \to \infty$ or ii) N is fixed. Let $b = b(T)$ be a bandwidth sequence such that $Tb^{\frac{3}{2}} \to \infty$ and $b \to 0$. Assume that $G \in SLC$ and that, for some $p > 4$, the following short-range dependence condition holds

$$\sum_{t=0}^{\infty}\delta_p(t) < \infty. \tag{19}$$

Then, under the null hypothesis H_0, we have

$$Tb^{1/2}(N-1)^{-1/2}(\widehat{\Delta}_{N,T} - \mathbb{E}\widehat{\Delta}_{N,T}) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2 K_2^*\right). \tag{20}$$

It is worth observing that the limit distribution in (20) is the same whether i) $N \to \infty$ or ii) N is fixed. However, the proofs for these two cases are different; see Appendix A. Here, we provide intuitions of the proofs. If $N \to \infty$, the estimates $\hat{c}_i$ and $\hat{\mu}$ will both be close to their true values. Hence, the $\int_0^1[\hat{\mu}_i(u) - \hat{c}_i - \hat{\mu}(u)]^2du$, $i = 1,\ldots,N$, in (10) can be approximated by the $\int_0^1[\hat{\mu}_i(u) - c_i - \mu(u)]^2du$, which are i.i.d., and the classical Lindeberg–Feller central limit theorem (CLT) applies.

In case ii), the Lindeberg–Feller CLT is no longer applicable since $N = N(T)$ is bounded; however, we can apply the m-dependent and martingale approximations as in [21] and still obtain (20). Note that the factor $(N-1)$ in (20) is due to the fact that we average the N independent streams to get the function estimate $\hat{\mu}$, thus losing one degree of freedom.

We now look into test consistency. Recall that $\widehat{\Delta}_{N,T}$ serves as an estimate of the parallelism index Δ_N defined in (11) under both H_0 in (2) and alternatives. Our test rejects H_0 at level α if $\widehat{\Delta}_{N,T}$ exceeds the $(1-\alpha)$th quantile of its distribution. (The precise implementation of the test is provided in Section IV.) The next theorem asserts that this test is consistent against local alternatives approaching (2) such that $N^{-1}(Tb + b^{-2})\Delta_N \to \infty$, namely under the latter condition the power goes to 1.

Theorem 2: Assume conditions of Theorem 1. Also assume that the $\mu_i, i = 1,\ldots,N$, in (1) have uniformly bounded second derivatives on [0, 1]. Then, the parallelism test based on $\widehat{\Delta}_{N,T}$ has unit asymptotic power if $N^{-1}(Tb + b^{-2})\Delta_N \to \infty$.

IV. Test Implementation

We address here the implementation of Theorem 1 for hypothesis testing. In particular, we discuss the issues of bandwidth selection and variance estimation, and we propose a simulation-based procedure that improves upon the normal approximation for the test statistic $\widehat{\Delta}_{N,T}$.

A. Bandwidth Selection

As seen in Section II, the same bandwidth b is used in the test procedure to estimate both μ and the μ_i, $i = 1, \ldots, N$. In addition to simplifying the test implementation and theoretical study, this choice automatically corrects biases under H_0 as noted by [22]

$$\mathbb{E}\left[\hat{\mu}_i(v) - \hat{\mu}(v)\right] = \sum_{t=1}^{T} w_b(t, v)\left[\mu_i(t/T) - \mu(t/T)\right] = c_i.$$

To select the bandwidth b, we propose a generalized cross-validation (GCV) procedure that can adjust for the dependence of the time series. The simulation study of Section V suggests that our test procedure is reasonably robust to the choice of b and the GCV method (21) performs reasonably well. Since our test procedure aims at reconstructing the mean function differences $\mu_i - \mu$, $i = 1, \ldots, N$, and assess whether they are constant over time, it is natural to base the GCV score on the $\mathbf{Y}_i = \{X_{it} - \bar{X}_{\cdot t}\}_{t=1}^{T}$ rather than on the original time series $\mathbf{X}_i = \{X_{it}\}_{t=1}^{T}$. Let $\mathbf{\Gamma} = (\gamma_{t,t'})_{1 \le t,t' \le T}$ be the covariance matrix of the error process, where $\gamma_{t,t'} = \mathbb{E}(e_t e_{t'})$, and let $\mathbf{H}(b)$ be the $T \times T$ "hat" matrix associated to the local linear smoother with bandwidth b, namely $\mathbf{H}(b)_{st} = w_b(t, \frac{s}{T})$, $1 \le s, t \le T$. Denoting by $\widehat{\mathbf{Y}}_i = \mathbf{H}(b)\mathbf{Y}_i$ the estimator of $\mu_i - \mu$ at the design points, we propose to choose b by minimizing the GCV score

$$\mathrm{GCV}(b) = \sum_{i=1}^{N} \frac{(\widehat{\mathbf{Y}}_i - \mathbf{Y}_i)^{\top}\mathbf{\Gamma}^{-1}(\widehat{\mathbf{Y}}_i - \mathbf{Y}_i)}{\left[1 - \mathrm{tr}\left(\mathbf{H}(b)\right)/T\right]^2}. \tag{21}$$

We now consider the estimation of the covariance matrix $\mathbf{\Gamma} = (\gamma_{t,t'})_{1 \le t,t' \le T}$. In view of the local stationarity of e_{it}, we use the local linear smoothing [20] technique and naturally estimate $\gamma_{t,t+k}$, $0 \le k < T$, by

$$\hat{\gamma}_{t,t+k} = \frac{1}{N}\sum_{i=1}^{N}\sum_{v=1}^{T-k} \hat{e}_{iv}\hat{e}_{i,v+k} w_b\left(v, t/(T-k)\right) \tag{22}$$

where $w_b(t, u)$ are the local linear weights defined by (6) with T therein replaced by $T - k$, and $\hat{e}_{iv} = X_{iv} - \hat{\mu}_i(\frac{v}{T})$, $i = 1, \ldots, N$, $v = 1, \ldots, T$, are the estimated residuals. Since $\gamma_{t,t'}$ is small if $|t - t'|$ is large, using the regularization method of banding [23], we estimate $\mathbf{\Gamma}$ by $\left(\hat{\gamma}_{t,t'}\mathbb{I}_{\{|t-t'| \le T^{\frac{4}{15}}\}}\right)_{1 \le t,t' \le T}$.

B. Estimation of the Long-Run Variance Function

In order to apply Theorem 1, we need to estimate the critical quantity σ^2 in (18) which serves as the asymptotic variance (up to a known scalar) of the test statistic (10), or more essentially we need to estimate the long-run variance function g. For each $u \in [0, 1]$, let

$$\mathcal{N}_\tau(u) = \{t : |t/T - u| \le \tau\} \tag{23}$$

where $\tau = \tau(T)$ is a window size satisfying $\tau \to 0$ and $T\tau \to \infty$ as $T \to \infty$. The points of $\mathcal{N}_\tau(u)$, suitably rescaled by $\frac{1}{T}$, become increasingly dense in $[u - \tau, u + \tau]$ as $T \to \infty$. By the local stationarity (14), the process $\{e_{it}\}_{t \in \mathcal{N}_\tau(u)}$ can be approximated by the stationary process $(G(u, \mathcal{F}_{it}))_{t \in \mathcal{N}_\tau(u)}$ in the sense that

$$\sup_{0 \le u \le 1} \max_{t \in \mathcal{N}_\tau(u)} \left\|e_{it} - G(u, \mathcal{F}_{it})\right\|_p = O(\tau). \tag{24}$$

Denote by $\hat{\gamma}_{ik}(u)$ the sample auto-covariance of $\{e_{it}\}_{t \in \mathcal{N}_\tau(u)}$ at lag k and average these quantities over i to estimate the auto-covariance (16) by

$$\hat{\gamma}_k(u) = \frac{1}{N}\sum_{i=1}^{N} \hat{\gamma}_{ik}(u). \tag{25}$$

Then, $g(u)$ can be simply estimated by

$$\hat{g}(u) = \sum_{k=-K_T}^{K_T} \hat{\gamma}_k(u) \tag{26}$$

for some truncation parameter $K_T = \lfloor T\tau\varrho \rfloor$ with bandwidth $\varrho \to 0$ and $T\tau\varrho \to \infty$. Indeed, $\gamma_k(u)$ will be close to zero for large k and for all $u \in [0, 1]$ under the local stationarity condition (14) and the short-range dependence assumption (19). More precisely, we need to specify the decay rate of the physical dependence measure (13) to characterize the bias caused by truncation. Also, the error processes $\{e_{it}\}$, $i = 1, \ldots, N$, are not observable in practice and we recommend plugging the residuals $\hat{e}_{it} = X_{it} - \hat{\mu}_i(\frac{t}{T})$ from (8) into (26) to get an estimate $\tilde{g}$ of the long-run variance function. The following theorem provides error bounds for both $\hat{g}$ and $\tilde{g}$.

Theorem 3: Assume that $G \in SLC$, $g \in \mathcal{C}^2[0, 1]$, $\sum_{t=0}^{\infty} \delta_4(t) < \infty$, and $\sum_{t=T}^{\infty} \delta_2(t) = \mathcal{O}(T^{-\varsigma})$ for some $\varsigma > 0$. Let $\varpi = \sqrt{\frac{\varrho}{N}} + (T\tau\varrho)^{-\varsigma} + (\tau\rho)^{\frac{\varsigma}{(1+\varsigma)}} + \tau^2 + \varrho$. Then

$$\sup_{u \in [0,1]} \left\|\hat{g}(u) - g(u)\right\|_2 = \mathcal{O}(\varpi). \tag{27}$$

If in addition $\iota = (T\tau\varrho)^{\frac{1}{2}}\left(b^2 + T^{\frac{-1}{2}}b^{\frac{-1}{2}}\right) \to 0$, then

$$\sup_{u \in [0,1]} \left\|\tilde{g}(u) - g(u)\right\|_2 = \mathcal{O}(\iota + \varpi). \tag{28}$$

The choice of banding parameters τ and ϱ that minimize the bound on the right hand side of (27) can depend on N, T and ς in a highly complicated fashion. Nevertheless, when $\varsigma \ge 2$, we have the following dichotomy:

- if $T^{\frac{2\varsigma}{(3\varsigma+2)}} = \mathcal{O}(N)$, the optimal bound in (27) is $\mathcal{O}\left(T^{\frac{-2\varsigma}{(3\varsigma+2)}}\right)$ for $\tau \asymp T^{\frac{-\varsigma}{(3\varsigma+2)}}$ and $\varrho \asymp T^{\frac{-2\varsigma}{(3\varsigma+2)}}$;
- if $N = \mathcal{O}\left(T^{\frac{2\varsigma}{(3\varsigma+2)}}\right)$ in which case N is not required to blow up, the optimal bound in (27) is $\mathcal{O}\left((TN)^{\frac{-2\varsigma}{(5\varsigma+2)}}\right)$ for $\tau \asymp (TN)^{\frac{-\varsigma}{(5\varsigma+2)}}$ and $\varrho \asymp T^{\frac{-4\varsigma}{(5\varsigma+2)}}N^{\frac{(\varsigma+2)}{(5\varsigma+2)}}$.

In particular, when the errors satisfy the geometric moment contraction condition, that is, $\delta_2(k)$ decays geometrically quickly as in the case of an autoregressive process, the optimal bound for (27) is $\mathcal{O}(T^{\frac{-2}{3}}\log T)$ if $\frac{N}{T^{\frac{2}{3}}} \to \infty$ and $\mathcal{O}(T^{\frac{-2}{5}}\log T)$ otherwise.

TABLE I
EMPIRICAL ACCEPTANCE PROBABILITIES (IN PERCENTAGE) AT (90%, 95%, 99%) NOMINAL LEVELS WITH DIFFERENT COMBINATIONS OF N, T, AND b

		b				
N	T	0.1	0.2	0.3	0.4	0.5
50	100	(94.0, 97.7, 99.8)	(92.7, 97.2, 99.6)	(92.2, 96.3, 99.6)	(91.4, 95.9, 99.4)	(91.4, 95.9, 99.2)
	300	(91.5, 96.1, 99.4)	(90.6, 95.5, 99.2)	(90.7, 95.8, 99.1)	(90.2, 95.3, 99.0)	(90.2, 94.9, 99.2)
	500	(91.2, 95.7, 99.3)	(90.3, 95.3, 99.3)	(90.7, 95.5, 99.2)	(90.0, 95.1, 99.0)	(90.3, 95.5, 99.2)
100	100	(94.5, 98.0, 99.8)	(93.3, 97.1, 99.6)	(93.0, 96.5, 99.5)	(92.5, 96.5, 99.5)	(92.4, 96.7, 99.4)
	300	(92.1, 96.4, 99.4)	(91.2, 95.5, 99.2)	(91.0, 95.9, 99.1)	(91.6, 95.8, 99.1)	(91.5, 95.6, 99.1)
	500	(91.0, 95.6, 98.9)	(91.2, 96.0, 99.3)	(89.6, 94.5, 99.0)	(90.6, 95.6, 99.2)	(89.8, 95.1, 99.1)
150	100	(95.6, 98.3, 99.8)	(94.0, 97.4, 99.6)	(93.5, 97.1, 99.5)	(92.4, 96.3, 99.5)	(92.0, 96.2, 99.3)
	300	(91.8, 96.5, 99.5)	(91.1, 95.8, 99.4)	(91.1, 95.7, 99.2)	(90.5, 95.2, 98.9)	(91.4, 95.9, 99.3)
	500	(90.8, 95.3, 98.9)	(90.3, 95.3, 99.0)	(90.8, 95.3, 99.2)	(90.2, 95.3, 99.1)	(90.6, 95.3, 99.0)

Note that the bound in (28) goes to zero at a slower rate than the one in (27) and reaches $\mathcal{O}\left(T^{\frac{-2}{5}}\log T\right)$ when the geometric moment contraction condition is satisfied.

C. Simulation-Based Approximation to the Distribution of the Test Statistic

The normal convergence in (20) can be quite slow. A popular way to improve the convergence speed is via bootstrap; see, for example, [7] and [24]. Here, we propose an alternative simulation-based method, which is easily implementable and has a better finite-sample performance.

Let $Z_{it}, i = 1, \ldots, N, t = 1, \ldots, T$, be i.i.d. standard normal random variables. If the long-run variance function g is known, let $X_{it}^{\diamond} = g(\frac{t}{T})Z_{it}$ and otherwise, use the estimate $\tilde{g}$ to define $X_{it}^{\diamond} = \tilde{g}(\frac{t}{T})Z_{it}$. Let $\widehat{\Delta}_{N,T}^{\diamond}$ be the test statistic associated to the $X_{it}^{\diamond}$, assuming that $c_i \equiv 0$. By Theorem 1, $\widehat{\Delta}_{N,T}$ and $\widehat{\Delta}_{N,T}^{\diamond}$ have the same asymptotic distribution under the assumptions of Theorems 1 and 2. Hence, the distribution of $\widehat{\Delta}_{N,T}$ can be assessed by simulating $\widehat{\Delta}_{N,T}^{\diamond}$. Specifically, one can generate many realizations of $(X_{it}^{\diamond})_{t=1}^{T}$, $i = 1, \ldots N$, and compute the corresponding $\widehat{\Delta}_{N,T}^{\diamond}$ from which one can obtain the estimated $(1-\alpha)$th quantile $\hat{q}_{1-\alpha}$. Based on this, one can reject at level α the null hypothesis if $\widehat{\Delta}_{N,T} > \hat{q}_{1-\alpha}$, and accept otherwise. The validity of this method is guaranteed by the invariance principle (see [25]) which asserts that partial sums of dependent random vectors can be approximated by Gaussian processes.

V. SIMULATION STUDY

A. Acceptance Probabilities

In this section, we present a simulation study to assess the performance of our test procedure. Consider the model

$$X_{it} = c_i + \mu(t/T) + e_{it} \tag{29}$$

with $c_i = 0$ and $\mu(u) = 2\sin(2\pi u)$. Note that under (29), the test procedure is independent of the c_i. The error process $\{e_{it}\}$ is generated by $e_{it} = \zeta_{i,t}(\frac{t}{T})$, where for all $i \in \mathbb{Z}$ and $u \in [0,1]$, the process $(\zeta_{i,t}(u))_{t\in\mathbb{Z}}$ follows the recursion

$$\zeta_{i,t}(u) = \rho(u)\zeta_{i,t-1}(u) + \sigma\varepsilon_{i,t}, \tag{30}$$

with the $\varepsilon_{i,t}$ being i.i.d. random variables satisfying $\mathbb{P}(\varepsilon_{i,t} = -1) = \mathbb{P}(\varepsilon_{i,t} = 1) = \frac{1}{2}$. Thus, $\{e_{it}\}_{t\in\mathbb{Z}}$ for $i = 1, \ldots, N$ are i.i.d. AR(1) processes with time-varying coefficients. Let $\rho(u) = 0.2 - 0.3u$ and $\sigma = 1$. Easy calculations show that $\mathbb{E}(\zeta_{i,t}(u)) = 0$, $\mathrm{Var}(\zeta_{i,t}(u)) = \frac{\sigma^2}{(1-\rho(u)^2)}$ and the long-run variance function $g(u) = \frac{\sigma^2}{(1-\rho(u))^2}$.

In our simulation the Epanechnikov kernel $K(v) = \frac{3\max(0,1-v^2)}{4}$ is used. We simulate 10 000 realizations of (30) and, for each realization, 10 000 simulations of $\widehat{\Delta}_{N,T}^{\diamond}$ are performed as in Section IV-C. We are interested in the proportion of realizations for which the null hypothesis is correctly accepted. Acceptance probabilities are presented in Table I for different choices of T, N, and b. This suggests that the acceptance probabilities are reasonably close to their (90%, 95%, 99%) nominal levels and become more robust to the size of bandwidth as the sample size T gets bigger.

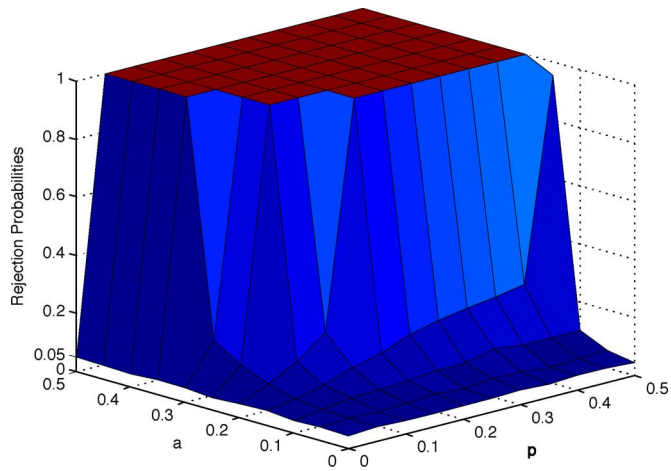


Fig. 1. Rejection probabilities at 5% nominal levels with different choices of deviation proportion p and distortion magnitude a.

B. Statistical Power

In the setting of Section V-A with $T = 300$ and $N = 100$, we study the statistical power of our testing procedure. For a certain proportion (say p) of the N time series $\{X_{it}\}_{t=1}^{T}$ for $i \in \{1, \ldots, N\}$, we add a distortion $a\mu_d(\frac{t}{T})$ in addition to (29), where $\mu_d(u) = 2\cos(2\pi u)$ and a denotes the corresponding magnitude. We investigate on the rejection probabilities at 5% nominal levels with different choices of p and a and the results are summarized in Fig. 1. It can be easily seen that the power goes to one very quickly as the magnitude of distortion a gets large and the proportion of different trends p approaches 0.5.

VI. Application of the Test to Clustering

The test procedure developed in the previous sections can be applied to cluster collections of time series based on their similarity in terms of parallelism. In the sequel we identify the time series in (1) with their indexes. To build our iterative clustering algorithm, we start by finding the largest cluster G_1 in $U^{(0)} = \{1, \ldots, N\}$ for which the parallelism assumption H_0 is retained at level α. The cluster G_1 is obtained by progressively removing from the analysis the sample units that contribute most to the test statistic (10). Specifically, H_0 is first tested on $U^{(0)}$, then on a subset $U^{(1)} \subset U^{(0)}$ if rejected on $U^{(0)}$, and so on so forth until H_0 is accepted or $U^{(k)}$ is reduced to a single element for some k, in which case the algorithm ends without any cluster being found. At the second iteration, the procedure is repeated with the remaining time series (set $U^{(0)} := \{1, \ldots, N\} \setminus G_1$) and so on so forth until either all time series are clustered (i.e. $\{1, \ldots, N\} = G_1 \cup \cdots \cup G_L$ for some L) or no more clusters of size >1 can be formed.

We now give a precise description of the algorithm. For the test implementation, the user must provide a significance level α, bandwidth b (cf. Sections II and IV-A), and parameters (τ, ϱ) (cf. Section IV-B). The user must also specify the number n of sample units to remove at each step of the cluster search. As long as n is small (say $n = 1$ for small N and, say $n \leq \frac{5N}{100} = \frac{N}{20}$ for moderate to large N), this tuning parameter does not affect the outcome of the clustering algorithm; it however influences the computational time. Note that when moving from working index set $U^{(k)}$ to $U^{(k+1)}$ during the cluster search, the algorithm removes at least one and at most $(N_k - 2)$ sample units from $U^{(k)}$, where N_k is the size of $U^{(k)}$, so that there remains at least two units to compare at the next step. As a result the effective number of removed units is $n^* = \max(1, \min(n, N_k - 2))$. Also, if H_0 is rejected on $U^{(k)}$ and accepted on $U^{(k+1)}$, this may mean that too many units (n^* of them) have been removed from $U^{(k)}$ and that H_0 can be retained on an intermediate set $U^{(k+1)} \subseteq U' \subset U^{(k)}$. In this case a flag F is activated and the algorithm starts a dichotomic search, returning to the previous working index set $U^{(k)}$ and attempting to remove less units at each subsequent step (i.e. roughly $\frac{n}{2}$, then $\frac{n}{4}$, etc.). The following notations are needed for the formal statement of the algorithm:

- $* \; k$: step counter; l: group counter, F: flag.
- $* \; \bar{X}_{\cdot t}^{(k)} = N_k^{-1} \sum_{i \in U^{(k)}} X_{it}$, $\bar{X}_{\cdots}^{(k)} = T^{-1} \sum_{t=1}^{T} \bar{X}_{\cdot t}^{(k)}$, $\hat{\mu}^{(k)}(u) = \sum_{t=1}^{T} w_b(t,u) \bar{X}_{\cdot t}^{(k)}$, and $\hat{c}_i^{(k)} = T^{-1} \sum_{i \in U^{(k)}} [\hat{\mu}_i(\frac{t}{T}) - \hat{\mu}^{(k)}(\frac{t}{T})]$. Recall that N_k is the size of $U^{(k)}$.

The algorithm works as follows:

Initialization.

1) Set $U^{(0)} := \{1, \ldots, N\}$, $k := 0$, $l := 1$, and $F := 0$.
2) Initialize the parameters α, b, τ, ϱ, and n, with $n < N_0$.
3) Perform the parallelism test on $\{X_{it}\}_{t=1}^{T}$ for $i \in U^{(0)}$ and compute the p-value.
   a) Case $p > \alpha$.
      - Compute the $\Delta_i := \int_0^1 [\hat{\mu}_i(u) - \hat{c}_i^{(0)} - \hat{\mu}^{(0)}(u)]^2 du$ for $i \in U^{(0)}$ and sort them as $\Delta_{\sigma(1)} \leq \cdots \leq \Delta_{\sigma(N_0)}$.
      - Write $n^* := \max(1, \min(n, N_0 - 2))$ and set $U^{(1)} := U^{(0)} \setminus \{\sigma(1), \ldots, \sigma(n^*)\}$ and $k := 1$.
   b) Case $p \leq \alpha$.
      - Set $G_1 := U^{(0)}$ and stop the algorithm.

Determination of the cluster G_l.

If $N_k = 1$, then stop the algorithm.
Perform the parallelism test on the $\{X_{it}\}_{t=1}^{T}$ for $i \in U^{(k)}$ and compute the p-value.

a) Case $p > \alpha$.
   - Compute the $\Delta_i := \int_0^1 [\hat{\mu}_i(u) - \hat{c}_i^{(k)} - \hat{\mu}^{(k)}(u)]^2 du$ for $i \in U^{(k)}$ and sort them as $\Delta_{\sigma(1)} \leq \cdots \leq \Delta_{\sigma(N_k)}$.
   - If $F = 1$, set $n := \max(1, \lfloor \frac{n}{2} \rfloor)$.
   - Write $n^* := \max(1, \min(n, N_k - 2))$ and set $U^{(k+1)} := U^{(k)} \setminus \{\sigma(1), \ldots, \sigma(n^*)\}$ and $k := k + 1$.
   - Return to step 4.

b) Case $p \leq \alpha$.
   - Set $F := 1$.

   Case $n = 1$.
   - Set $G_l := U^{(k)}$.
   - If $\{1, \ldots, N\} = (G_1 \cup \cdots \cup G_l)$, then stop the algorithm.
     Else return to step 1 and set $U^{(0)} := \{1, \ldots, N\} \setminus (G_1 \cup \cdots \cup G_l)$, $k := 0$ and $l := l + 1$.

   Case $n > 1$.
   - Set $n := \max(1, \lfloor \frac{n}{2} \rfloor)$ and $U^{(k)} := U^{(k-1)} \setminus \{\sigma(1), \ldots, \sigma(n^*)\}$, where $n^* := \max(1, \min(n, N_{k-1} - 2))$.
   - Return to step 4.

The R implementation of the algorithm can be obtained from the authors upon request.

VII. Data Analysis

To illustrate our parallelism test and clustering procedure, we consider a data set of hourly volumes of downloads from cell phones (in byte) in $N = 129$ U.S. area codes (24 area codes are in Center America, 87 in Eastern America, 1 in Hawaii, and 17 in Pacific America). Rather than studying the original data, we take their daily sums ($\mathrm{T} = 327$ days) so as to remove daily periodicity (which produces long-range dependence). Since the area codes have different numbers of phone users, we also apply a logarithmic transform (base 10) to the data to adjust for the scale effect. Thus, multiplicative differences in the time series become additive, which makes it relevant to test for parallelism in the trends of area codes. Examples of time series as well as the estimated global trend function $\mu = N^{-1} \sum_{i=1}^{N} \mu_i$ and long-run variance function g are displayed in Fig. 2.

Prior to statistical analysis, the validity conditions of our theoretical results have been verified on the data set. In particular, the rapid decrease observed in the autocorrelation functions of the detrended time series (see Fig. 3) indicates that the short-range dependence assumption (19) is very plausible. Also global similarity in the autocovariance functions across the time series make the assumption of identically distributed error processes look reasonable. Finally, the 129×129 cross-correlation matrix of the residual time series has nearly zero entries outside its diagonal, which suggests that the time series are independent.

We now describe the implementation of the test and clustering algorithm on the data set. The significance level of the

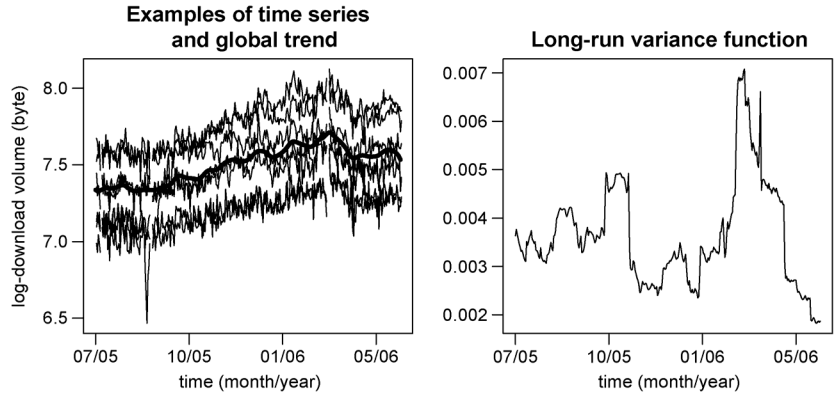


Fig. 2. Left panel: Examples of download volume time series (daily-totaled and log-transformed) and estimated global trend in thick line. Right panel: Estimated long-run variance function.

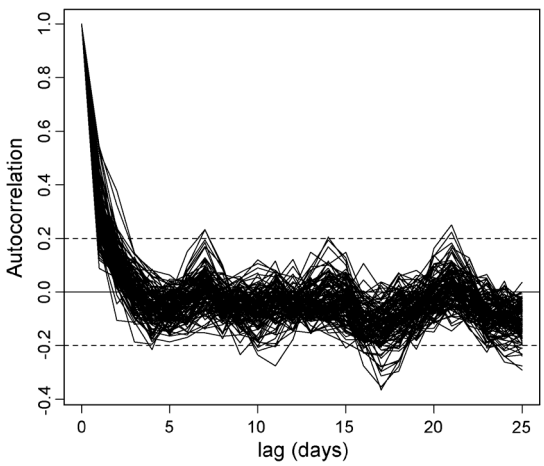


Fig. 3. Autocorrelation functions of the residual time series after a local linear fit with bandwidth $b = 10$ days.

TABLE II
SUMMARY OF THE CLUSTERS. THE CLUSTERS ARE MAXIMAL SETS OF AREA CODES FOR WHICH THE PARALLELISM ASSUMPTION IS RETAINED AT THE SIGNIFICANCE LEVEL 5%

Cluster Size	# clusters	Cum. prop. of N
29	1	22%
21	1	39%
20	1	54%
10	1	62%
7	1	67%
5	1	71%
3	2	76%
2	7	87%
1	17	100%

test is set to 5%. The local linear estimation of the trends is based on a truncated standard Gaussian density. The inverse covariance matrix $\mathbf{\Gamma}^{-1}$ is estimated as in Section IV-A. Specifically, after a pilot trend estimation using a bandwidth b such that $Tb = 5$ days, the "banding the inverse covariance matrix" technique (e.g. [23]) has been applied to the sample covariance matrix of residuals with a banding parameter $k = 6$ days. The final bandwidth b is obtained by minimizing the GCV score (21). The long-run variance function g is estimated as in Section IV-B based on the residuals of a local linear smoothing with $Tb = 10$ days. (A larger b is used to estimate the long-run variance function than for the trend estimation so as to make the estimate less sensitive to extreme observations.) The parameters $\tau = 0.04$ and $\varrho = 0.31$ are employed so that the estimate of $g(u), u \in [0, 1]$ utilizes about two weeks of data before and after a time point u and the autocovariances are truncated at lag $K_T = 4$. The previous values of τ and ρ have been seen to produce a smooth yet precise estimate, and the lag K_T has been determined by the fact that after lag 4, the autocovariance plots become dissimilar (a violation of the i.i.d. assumption on the error processes). Finally, the number n of units to remove at each step of the cluster search (see Section VI) is set to 3, a good compromise between search accuracy and computational speed of the algorithm.

The results of the analysis are presented in Tables II and III. Note that performing the parallelism test on the entire data set resulted in a p-value $< 10^{-16}$. Shifting our focus to clustering the time series, we observe that the four largest clusters found contain respectively 29, 21, 20, and 10 area codes. This represents a sizable proportion (62%) of the 129 area codes under study. In Fig. 4, the differences between these clusters are best perceived by visually fitting a regression line to the curves: although all curves have globally the same features, the slope of the fitted straight line varies largely and the obtained clusters are relatively homogeneous with respect to that slope. The examination of Table III reveals that there is no obvious spatial pattern in the clusters. It also shows that there is no systematic relation between the size of a cluster and its associated p-value. Overall, our statistical analysis shows that most area codes under study can be classified in a small number of profiles, or clusters, according to the parallelism patterns in their phone download activity. These strong similarities across area codes would deserve to be investigated in more detail as potentially, they could be exploited by phone companies to, e.g., better target their marketing strategies or improve the bandwidth allocation.

VIII. CONCLUSION

We have presented a test methodology for assessing the parallelism between trends of multiple time series. The physical dependence structure used in the paper can flexibly model nonstationary time series without having to specify generating mechanism or autocovariance functions. As by-products of the test,

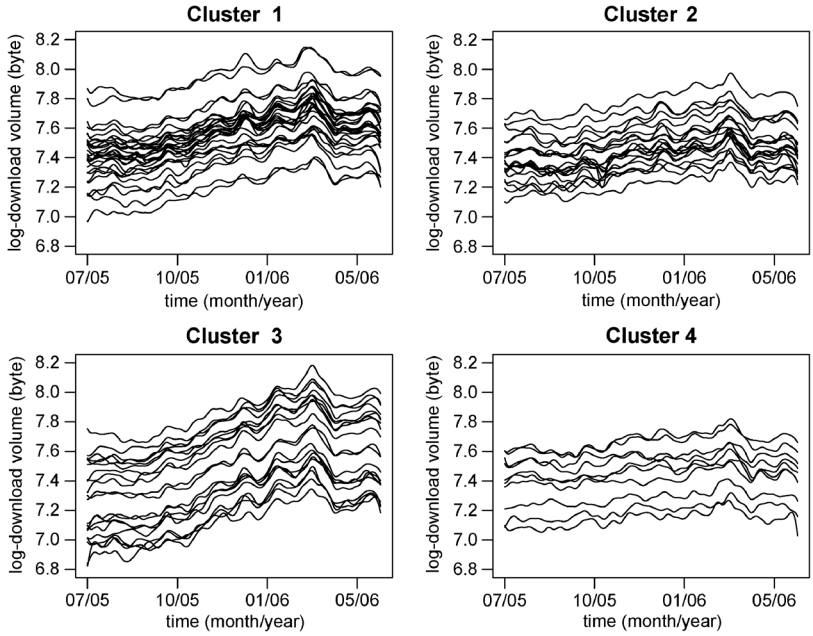


Fig. 4. Trends in the clusters of size $s \geq 10$. After a log-transform of the data, the trends are obtained by smoothing the time series with a bandwidth $b = 4$ days.

TABLE III
SPATIAL LOCATIONS OF CLUSTERS. FOR EACH AREA CODE, THE CORRESPONDING STATE IN THE U.S. IS DESIGNATED BY ITS POSTAL ABBREVIATION. ONLY CLUSTERS OF SIZE $s \geq 10$ ARE DISPLAYED

Cluster 1: size 29, p-value 0.095				
203 (CT)	219 (IN)	239 (FL)	301 (MD)	302 (DE)
321 (FL)	323 (CA)	484 (PA)	508 (MA)	513 (OH)
517 (MI)	540 (VA)	562 (CA)	603 (NH)	616 (MI)
617 (MA)	619 (CA)	646 (NY)	661 (CA)	703 (VA)
810 (MI)	813 (FL)	815 (IL)	856 (NJ)	859 (KY)
909 (CA)	941 (FL)	951 (CA)	978 (MA)	
Cluster 2: size 21, p-value 0.064				
209 (CA)	240 (MD)	510 (CA)	561 (FL)	586 (MI)
609 (NJ)	610 (PA)	626 (CA)	630 (IL)	631 (NY)
708 (IL)	714 (CA)	732 (NJ)	734 (MI)	772 (FL)
774 (MA)	781 (MA)	786 (FL)	818 (CA)	845 (NY)
908 (NJ)				
Cluster 3: size 20, p-value 0.151				
231 (MI)	269 (MI)	352 (FL)	386 (FL)	404 (GA)
407 (FL)	412 (PA)	419 (OH)	512 (TX)	704 (NC)
740 (OH)	773 (IL)	803 (SC)	816 (MO)	863 (FL)
864 (SC)	904 (FL)	919 (NC)	937 (OH)	989 (MI)
Cluster 4: size 10, p-value 0.071				
248 (MI)	516 (NY)	570 (PA)	571 (VA)	805 (CA)
808 (HI)	847 (IL)	914 (NY)	916 (CA)	917 (NY)

we have developed a method to estimate long-run variance functions of locally stationary processes and a simulation-based device to approximate distribution of statistics based on smoothed time series. Both these tools have shown good numerical performances in our simulations. They are of independent interest and could be used with profit in other statistical problems. A key assumption used to derive the theory of this paper is that the observed time series are independent from one another. A very interesting extension would be to allow for some form of dependence, e.g. to handle spatio-temporal data.

The paper also proposes an innovative method to cluster time series based on their parallelism. This method has at least two attractive features: first, it does not require to prespecify the number of clusters to be found, which guarantees the homogeneity of the clusters and allows atypical time series to be set apart; second, it readily provides significance levels for each cluster, thereby giving a quantitative sense of how strong the parallelism assumption holds. The implementation of this clustering method has given meaningful results with the cell phone download data. The algorithm is computationally fast as the individual trend functions and long-variance functions need being estimated only once, while the most computationally intensive step (Gaussian process simulation to approximate the distribution of the test statistic) is still manageable. The ideas used in this algorithm (greedy search, clustering based on individual contribution of sample units to test statistic) can be used to cluster time series according to other similarity measures than parallelism. An interesting direction of future research would be to compare the results of this type of clustering to the more conventional k-means or hierarchical approaches.

APPENDIX
TECHNICAL PROOFS

In the proofs, we denote by C a constant that is independent of N and T, and whose value may vary from place to place.

A. Proof of Theorem 1

The techniques for handling Case i) with large N and Case ii) with fixed N are different. For the former we apply the traditional Lindeberg–Feller CLT, while for the latter, we apply the m-dependence and martingale approximation techniques. For details, see Appendix Section A-1) and A-2), respectively.

We start by showing that under H_0, the test statistic $\widehat{\Delta}_{N,T}$ does not depend upon $\mu(\cdot)$ nor the c_i. To see this, introduce the weighted averages

$$\bar{w}_b(t) = T^{-1}\sum_{j=1}^{T} w_b(t, j/T).$$

Since $\sum_{t=1}^{T}[w_b(t,u) - \bar{w}_b(t)] = 0$, by (5), (8), and (9),

$$\begin{aligned}\hat{\mu}_i(u) - \hat{\mu}(u) - \hat{c}_i &= \sum_{t=1}^{T}[w_b(t,u) - \bar{w}_b(t)]\,(c_i + e_{it} - \bar{e}_{\cdot t}) \\ &= \sum_{t=1}^{T}[w_b(t,u) - \bar{w}_b(t)]\,(e_{it} - \bar{e}_{\cdot t}).\end{aligned}$$

1) Case i): $N \to \infty$: We shall prove the asymptotically equivalent form of (20)

$$Tb^{1/2}N^{-1/2}(\widehat{\Delta}_{N,T} - \mathbb{E}\widehat{\Delta}_{N,T}) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2 K_2^*\right). \tag{31}$$

To this end, we use the decomposition

$$\widehat{\Delta}_{N,T} - \mathbb{E}\widehat{\Delta}_{N,T} = \sum_{i=1}^{N}(A_i - \mathbb{E}A_i) - (R_N - \mathbb{E}R_N) \tag{32}$$

where

$$A_i = \int_0^1 \left(\sum_{t=1}^{T}(w_b(t,u) - \bar{w}_b(t))\,e_{it}\right)^2 du$$

$$\text{and} \quad R_N = N\int_0^1 \left(\sum_{t=1}^{T}(w_b(t,u) - \bar{w}_b(t))\,\bar{e}_{\cdot t}\right)^2 du,$$

and we show that asymptotically, $\sum_i A_i$ is normally distributed and R_N is negligible.

First, define

$$A_i^\circ = \int_0^1 \left(\sum_{t=1}^{T} w_b(t,u)e_{it}\right)^2 du.$$

By [26, Theorem 1], under the bandwidth conditions $Tb^{\frac{3}{2}} \to \infty$ and $b \to 0$ and the short-range dependence condition (19), we have

$$Tb^{1/2}\,(A_i^\circ - \mathbb{E}A_i^\circ) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2 K_2^*\right). \tag{33}$$

Observing that $A_1^\circ, \ldots, A_N^\circ$ are i.i.d., it results from (33) and the Lindeberg–Feller CLT that

$$\frac{Tb^{1/2}}{\sqrt{N}}\sum_{i=1}^{N}(A_i^\circ - \mathbb{E}A_i^\circ) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2 K_2^*\right). \tag{34}$$

We now show that $Tb^{\frac{1}{2}}N^{\frac{-1}{2}}\sum_{i=1}^{N}(A_i^\circ - A_i)$ is negligible as $N, T \to \infty$. Let $\dot{w}_b(t) = \int_0^1 w_b(t,u)du$,

$$J_i = \sum_{t=1}^{T}\bar{w}_b(t)e_{it} \quad \text{and} \quad \dot{J}_i = \sum_{t=1}^{T}\dot{w}_b(t)e_{it}.$$

Since $\max_t|\bar{w}_b(t)| = \mathcal{O}(T^{-1})$ and $\max_t|\dot{w}_b(t)| = \mathcal{O}(T^{-1})$, by Lemma 1 in [21], $|J_i|_4 = \mathcal{O}(T^{\frac{-1}{2}})$ and $\|\dot{J}_i\|_4 = \mathcal{O}(T^{\frac{-1}{2}})$. Hence,

$$\|A_i^\circ - A_i\|_2^2 = \left\|J_i^2 - 2J_i\dot{J}_i\right\|_2^2 = \mathcal{O}(T^{-2}) \tag{35}$$

and by the i.i.d. character of the $(A_i^\circ - A_i)$, one deduces that

$$\left\|Tb^{1/2}N^{-1/2}\sum_{i=1}^{N}(A_i^\circ - A_i)\right\|_2^2 = \mathcal{O}(b). \tag{36}$$

We proceed to study the remainder term $R_N - \mathbb{E}R_N$ in (32). By expanding R_N and using the i.i.d. character of the N time series, one easily finds that

$$R_N \overset{d}{=} A_1^\circ + J_1^2 - 2J_1\dot{J}_1 \tag{37}$$

where $\overset{d}{=}$ stands for equality in distribution. The terms in the above expansion have been studied before. More precisely, the relations (35) and (36) yield

$$\left\|Tb^{1/2}N^{-1/2}(R_N - \mathbb{E}R_N)\right\|_2^2 = \mathcal{O}(N^{-1} + N^{-1}b). \tag{38}$$

Putting together (32), (34), (35), and (38), one obtains the asymptotic normality (31).

2) Case ii): N is Fixed: Recall that $e_{it} = G(\frac{t}{T}; \mathcal{F}_{it})$. For $\zeta_{it}(u) = G(u; \mathcal{F}_{it})$, define

$$\tilde{\zeta}_{it}(u) = \mathbb{E}\left(\zeta_{it}(u)|\varepsilon_{i,t-m+1}, \varepsilon_{i,t-m+2}, \ldots, \varepsilon_{i,t}\right).$$

Then, the process $\{\tilde{\zeta}_{it}(u)\}_{t\in\mathbb{Z}}$ is m-dependent with long-run variance function g^* converging to g as $m \to \infty$. As in the proof of Theorem 1 in [26], we introduce the martingale difference

$$\begin{aligned}\tilde{D}_{i,t}^* &= \sum_{l=0}^{\infty}\mathbb{E}\left(\tilde{\zeta}_{i,t+l}(t/T)|\mathcal{F}_{it}\right) - \mathbb{E}\left(\tilde{\zeta}_{i,t+l}(t/T)|\mathcal{F}_{i,t-1}\right) \\ &= \sum_{l=0}^{m}\mathbb{E}\left(\tilde{\zeta}_{i,t+l}(t/T)|\mathcal{F}_{it}\right) - \mathbb{E}\left(\tilde{\zeta}_{i,t+l}(t/T)|\mathcal{F}_{i,t-1}\right).\end{aligned}$$

Observe that $\tilde{D}_{i,t}^*$, $1 \le t \le T$, are also m-dependent. Let $\tilde{D}_{i,t}^\dagger = \tilde{D}_{i,t}^* - \tilde{D}_{\cdot,t}^*$, where $\tilde{D}_{\cdot,t}^* = \sum_{i=1}^{N}\frac{\tilde{D}i,t^*}{N}$; let $(\sigma^*)^2 = \int_0^1 (g^*(u))^2 du$. By the argument of Theorem 1 in [26], to derive the asymptotic normality (20), it suffices to show that as $T \to \infty$,

$$\sum_{1\le t<t'\le T}\left[K^*\left(\frac{t-t'}{2Tb}\right)\right]^2\sum_{i=1}^{N}\sum_{i'=1}^{N}\mathbb{E}\left(\tilde{D}_{i,t}^\dagger\tilde{D}_{i',t}^\dagger\right)\mathbb{E}\left(\tilde{D}_{i,t'}^\dagger\tilde{D}_{i',t'}^\dagger\right) \times\frac{1}{T^2b(N-1)} \to K_2^*(\sigma^*)^2.$$

Since the $\tilde{D}^{\dagger}_{i,t}$, $i = 1,\dots,N$, are i.i.d., we see that $\mathbb{E}(\tilde{D}^{\dagger}_{i,t}\tilde{D}^{\dagger}_{i',t}) = g^*(t)\frac{(N-1)}{N}$ if $i = i'$ and $\mathbb{E}(\tilde{D}^{\dagger}_{i,t}\tilde{D}^{\dagger}_{i',t}) = -\frac{g^*(t)}{N}$ if $i \neq i'$. With a few manipulations, we then obtain

$$\sum_{i=1}^{N}\sum_{i'=1}^{N}\mathbb{E}\left(\tilde{D}^{\dagger}_{i,t}\tilde{D}^{\dagger}_{i',t}\right)\mathbb{E}\left(\tilde{D}^{\dagger}_{i,t'}\tilde{D}^{\dagger}_{i',t'}\right) = (N-1)g^*(t)g^*(t').$$

Furthermore, with the continuity of g^*, classic arguments for kernel smoothing show that

$$\frac{1}{T^2 b}\sum_{1\le t<t'\le T}\left[K^*\left(\frac{t-t'}{2Tb}\right)\right]^2 g^*(t)g^*(t') = K_2^*(\sigma^*)^2 + o(1),$$

the asymptotic normality (20) follows. □

B. Proof of Theorem 2

By (32) and the Cauchy–Schwarz inequality, we can write

$$\widehat{\Delta}_{N,T} = I_{N,T} + \sum_{i=1}^{N} A_i - R_N + \mathcal{O}_p\left(I_{N,T}^{1/2}\left(\sum_{i=1}^{N} A_i - R_N\right)^{1/2}\right)$$

where

$$I_{N,T} = \sum_{i=1}^{N}\int_0^1\left\{\sum_{t=1}^{T}\left[w_b(t,u) - \bar{w}_b(t)\right]\left[\mu_i(t/T) - \mu(t/T)\right]\right\}^2 du$$

and $\mu = N^{-1}\sum_{i=1}^{N}\mu_i$. By the approximation properties of local linear smoothers (see, for example, [27, p. 39, Prop. 1.13]), we obtain

$$I_{N,T} = \Delta_N + \mathcal{O}\left(N(b^2 + T^{-1})\right), \tag{39}$$

provided that the $\mu_i, i = 1,\dots,N$, have uniformly bounded second derivatives on [0, 1].

On the other hand, we know from (34), (36), and (38) that

$$Tb^{1/2}N^{-1/2}\left(\sum_{i=1}^{N}(A_i - \mathbb{E}A_i) - R_N\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2 K_2^*\right). \tag{40}$$

By the stochastic Lipschitz continuity (14) and the short-range dependence condition (19), and by properties of weight functions of local linear smoothers (see [27, p. 38, Lemma 1.3]) we also have $\mathbb{E}A_i = \mathcal{O}((Tb)^{-1})$. Hence,

$$\widehat{\Delta}_{N,T} = \Delta_N + \mathcal{O}(Nb^2) + \mathcal{O}(N/Tb) + o_p(N/Tb). \tag{41}$$

If $N^{-1}\Delta_N$ converges to 0 at a rate slower than $b^2 + \frac{1}{Tb}$, then $N^{-1}(b^2 + \frac{1}{Tb})\hat{\Delta}_{N,T} \to \infty$ in probability. Hence the test based on $\hat{\Delta}_{N,T}$ has unit asymptotic power. □

C. Proof of Theorem 3

Let $\hat{g}_i(u) = \sum_{k=-K_T}^{K_T}\hat{\gamma}_{ik}(u)$ be the estimated long-run variance based on $\{e_{it}\}_{t=1}^{T}$, where $K_T = \lfloor T\tau\varrho\rfloor$ is the truncation order and $\hat{\gamma}_{ik} = [|\mathcal{N}_\tau(u)| - |k|]^{-1}\sum_{t,t+k\in\mathcal{N}_\tau(u)} e_{it}e_{i,t+k}$ is the sample autocovariance at lag k. Since the cardinality $|N_\tau(u)|$ is of order $T\tau$, one sees that

$$\hat{g}_i(u) = \frac{1+\mathcal{O}(\varrho)}{|\mathcal{N}_\tau(u)|}\sum_{t\in\mathcal{N}_\tau(u)}\sum_{t'\in\mathcal{N}_\tau(u)} e_{it}e_{it'}\mathbb{1}_{\{|t-t'|\le K_T\}}. \tag{42}$$

By the argument of [21, Prop. 1], it can be shown that $\sup_{u\in[0,1]}\|\hat{g}_i(u) - \mathbb{E}\hat{g}_i(u)\|_2 = \mathcal{O}(\sqrt{\varrho})$v and by the i.i.d. property of the $\{e_{it}\}_{t=1}^{T}$, one deduces that

$$\sup_{u\in[0,1]}\|\hat{g}(u) - \mathbb{E}\hat{g}(u)\|_2 = \mathcal{O}(\sqrt{\varrho/N}). \tag{43}$$

The expectation $\mathbb{E}\hat{g}_i(u)$ can be used to approximate the truncation of $g(u)$ to order K_T thanks to the stochastic Lipschitz continuity (14) and the martingale decomposition of [28]. Specifically, let $\Gamma_2(k) = \sum_{j=0}^{\infty}\delta_2(j)\delta_2(j+k)$. Then for all $u\in[0,1]$ and $t,t'\in\mathcal{N}_\tau(u)$ such that $|t-t'|\le K_T$, it holds that

$$\left|\mathbb{E}(e_{it}e_{it'}) - \gamma_{|t-t'|}(t/T)\right| \le C\left(\Gamma_2\left(|t-t'|\right)\wedge(\tau\varrho)\right). \tag{44}$$

Moreover, we obtain after easy calculation that

$$\sum_{k=0}^{\infty}\left(\Gamma_2(k)\wedge(\tau\varrho)\right) = \mathcal{O}\left((\tau\varrho)^{\varsigma/(1+\varsigma)}\right). \tag{45}$$

Taking the expectation in (42) and adding terms so that the summation index set is $\{(t,t') : t\in\mathcal{N}_\tau(u), 1\le t'\le T, |t-t'|\le K_T\}$, it stems from (44) and (45) that

$$\begin{aligned}\mathbb{E}\hat{g}_i(u) &= \frac{1}{|\mathcal{N}_\tau(u)|}\sum_{t\in\mathcal{N}_\tau(u)}\sum_{k=-K_T}^{K_T}\mathbb{E}e_{it}e_{i,t+k} + \mathcal{O}(\varrho)\\ &= \frac{1}{|\mathcal{N}_\tau(u)|}\sum_{t\in\mathcal{N}_\tau(u)}\left[g(t/T) - 2\sum_{k=K_T}^{\infty}\gamma_k(t/T)\right]\\ &\quad + \mathcal{O}\left((\tau\varrho)^{\varsigma/(1+\varsigma)}\right) + \mathcal{O}(\varrho).\end{aligned} \tag{46}$$

In (46), a Taylor expansion of $g\in\mathcal{C}^2[0,1]$ at order 2 yields

$$\sup_{u\in[0,1]}\left|\frac{1}{|\mathcal{N}_\tau(u)|}\sum_{t\in\mathcal{N}_\tau(u)} g(t/T) - g(u)\right| = \mathcal{O}(\tau^2). \tag{47}$$

Also, the martingale decomposition of [28] can be applied to show that $\sup_{u\in[0,1]}|\gamma_k(u)| \le \Gamma_2(k)$, so that under the assumptions of Theorem 2,

$$\sup_{u\in[0,1]}\sum_{k=K_T}^{\infty}|\gamma_k(u)| = \mathcal{O}\left(\sum_{k=K_T}^{\infty}\delta_2(k)\right) = \mathcal{O}\left((T\tau\varrho)^{-\varsigma}\right). \tag{48}$$

Finally, to obtain (27), it suffices to note that $\mathbb{E}\hat{g}(u) = \mathbb{E}\hat{g}_i(u)$.

To derive (28), an easy calculation shows that

$$\begin{aligned}\tilde{g}_i(u) - \hat{g}_i(u) &= \frac{2}{|\mathcal{N}_\tau(u)|} \sum_{t\in\mathcal{N}_\tau(u)} \sum_{t'\in\mathcal{N}_\tau(u)} (\hat{e}_{it} - e_{it}) e_{it'} \mathbb{1}_{\{|t-t'|\le K_T\}} \\ &\quad + \frac{1}{|\mathcal{N}_\tau(u)|} \sum_{t\in\mathcal{N}_\tau(u)} \sum_{t'\in\mathcal{N}_\tau(u)} (\hat{e}_{it} - e_{it}) \\ &\quad \times (\hat{e}_{it'} - e_{it'}) \mathbb{1}_{\{|t-t'|\le K_T\}} \\ &:= I_{N,T}(u) + II_{N,T}(u).\end{aligned}$$

Since $\hat{e}_{it} - e_{it} = \beta_i(\frac{t}{T}) - \sum_{t''=1}^{T} \beta_i(\frac{t''}{T}) w_b(t'', \frac{t}{T}) - \sum_{t''=1}^{T} e_{it''} w_b(t'', \frac{t}{T})$, we have

$$\max_{t=1,\ldots,T} \|\hat{e}_{it} - e_{it}\|_p \le C(b^2 + T^{-1/2} b^{-1/2}).$$

Hence, by [26, Lemma A1], $\sup_{u\in[0,1]} \|I_{N,T}(u)\|_p \le C\iota$ and $\sup_{u\in[0,1]} \|II_{N,T}(u)\|_p \le C\iota^2$, (28) follows. □

ACKNOWLEDGMENT

The authors would like to thank the reviewers and an Associate Editor for their valuable suggestions and comments.

REFERENCES

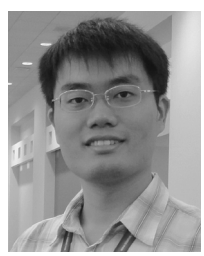
[1] S. G. Young and A. W. Bowman, "Non-parametric analysis of covariance," *Biometrics*, vol. 51, pp. 920–931, 1995.
[2] P. G. Gottschalk and J. R. Dunn, "Measuring parallelism, linearity, and relative potency in bioassay and immunoassay data," *J. Biopharm. Stat.*, vol. 15, no. 3, pp. 437–463, 2005.
[3] C. Park, A. Vaughan, J. Hannig, and K.-H. Kang, "SiZer analysis for the comparison of time series," *J. Stat. Plann. Inf.*, vol. 139, no. 12, pp. 3974–3988, 2009.
[4] S. Orbe, E. Ferreira, and J. Rodriguez-Poo, "Nonparametric estimation of time varying parameters under shape restrictions," *J. Econometr.*, vol. 126, no. 1, pp. 53–77, 2005.
[5] W. B. Wu and Z. Zhao, "Inference of trends in time series," *J. Roy. Stat. Soc. Ser. B*, vol. 69, pt. 3, pp. 391–410, 2007.
[6] W. Härdle and J. S. Marron, "Semiparametric comparison of regression curves," *Ann. Stat.*, vol. 18, no. 1, pp. 63–89, 1990.
[7] P. Hall and J. D. Hart, "Bootstrap test for difference between means in nonparametric regression," *J. Amer. Stat. Assoc.*, vol. 85, no. 412, pp. 1039–1049, 1990.
[8] E. King, J. D. Hart, and T. E. Wehrly, "Testing the equality of two regression curves using linear smoothers," *Stat. Probab. Lett.*, vol. 12, no. 3, pp. 239–247, 1991.
[9] P. Guo and A. J. Oyet, "On wavelet methods for testing equality of mean response curves," *Int. J. Wavelets Multiresolut. Inf. Process.*, vol. 7, no. 3, pp. 357–373, 2009.
[10] H. Dette and A. Munk, "Nonparametric comparison of several regression functions: Exact and asymptotic theory," *Ann. Stat.*, vol. 26, no. 6, pp. 2339–2368, 1998.
[11] N. Bissantz, H. Holzmann, and A. Munk, "Testing parametric assumptions on band- or time-limited signals under noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3796–3805, 2005.
[12] M. Pawlak and U. Stadtmüller, "Signal sampling and recovery under dependent errors," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2526–2541, 2007.
[13] D. Degras, "Simultaneous confidence bands for nonparametric regression with functional data," *Stat. Sinica*, vol. 21, pp. 1735–1765, 2011.
[14] F. Li, "Testing for the equality of two nonparametric regression curves with long memory errors," *Comm. Stat. Simulat. Comput.*, vol. 35, no. 3, pp. 621–643, 2006.
[15] J. Fan and S.-K. Lin, "Test of significance when data are curves," *J. Amer. Stat. Assoc.*, vol. 93, no. 443, pp. 1007–1021, 1998.
[16] M. A. Delgado, "Testing the equality of nonparametric regression curves," *Stat. Probab. Lett.*, vol. 17, no. 3, pp. 199–204, 1993.
[17] H. L. Koul and A. Schick, "Testing for the equality of two nonparametric regression curves," *J. Stat. Plann. Inf.*, vol. 65, no. 2, pp. 293–314, 1997.
[18] P. Lavergne, "An equality test across nonparametric regressions," *J. Econometr.*, vol. 103, no. 1–2, pp. 307–344, 2001.
[19] W. B. Wu, "Nonlinear system theory: Another look at dependence," *Proc. Natl. Acad. Sci.*, vol. 102, no. 40, pp. 14150–14154, 2005.
[20] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications*. London, U.K.: Chapman & Hall, 1996.
[21] W. Liu and W. B. Wu, "Asymptotics of spectral density estimates," *Econometr. Theory*, vol. 26, pp. 1218–1245, 2010.
[22] W. Härdle and E. Mammen, "Comparing nonparametric versus parametric regression fits," *Ann. Stat.*, vol. 21, no. 4, pp. 1926–1947, 1993.
[23] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Stat.*, vol. 36, no. 1, pp. 199–227, 2008.
[24] J. M. Vilar-Fernández, J. A. Vilar-Fernández, and W. González-Manteiga, "Bootstrap tests for nonparametric comparison of regression curves with dependent errors," *Test*, vol. 16, no. 1, pp. 123–144, 2007.
[25] W. B. Wu and Z. Zhou, "Gaussian approximations for non-stationary multiple time series," *Stat. Sinica*, vol. 21, pp. 1397–1413, 2011.
[26] T. Zhang and W. B. Wu, "Testing parametric assumptions of trends of a nonstationary time series," *Biometrika*, vol. 98, no. 3, pp. 599–614, 2011.
[27] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. New York: Springer, 2009.
[28] W. B. Wu, "Strong invariance principles for dependent random variables," *Ann. Prob.*, vol. 35, no. 6, pp. 2294–2320, 2007.

David Degras received the Ph.D. degree in statistics from the Université Paris 6, France, in 2007.

He is currently an Assistant Professor in the Department of Mathematical Sciences at DePaul University, Chicago, IL. His research interests include functional data analysis, survey sampling, dynamic systems and applications, and neuroimaging.

Zhiwei Xu received the Ph.D. degree in computer engineering from Florida Atlantic University, Boca Raton, in 2001.

He is currently an Assistant Professor in the Department of Computer and Information Science at the University of Michigan-Dearborn. His research interests include software metrics, quality modeling, software V&V, data mining, pattern recognition, artificial intelligence, cost modeling, and optimization.



Ting Zhang received the B.S. degree in mathematics and applied mathematics from Fudan University, China, in 2007.

He is currently working towards the Ph.D. degree in statistics at The University of Chicago, Chicago, IL. His research interests include nonparametric and semiparametric methods, model selection under parameter instability, locally stationary processes, Bayesian learning, and robust control problems in risk pricing.

Wei Biao Wu received the Ph.D. degree in statistics from the University of Michigan-Ann Arbor in 2001.

He is currently a Professor in the Department of Statistics at The University of Chicago, Chicago, IL. His research interests include probability theory, statistics, and econometrics. He is currently interested in estimating covariance matrices of temporally observed series.