Trigonometric Trend Regressions of Unknown Frequencies with Stationary or Integrated Noise*

Pierre Perron† **Mototsugu Shintani**‡ **Tomoyoshi Yabu**§

This version: January 31, 2020

Abstract

We propose a new procedure to select the unknown frequencies of a trigonometric function, a problem first investigated by Anderson (1971) under the assumption of serially uncorrelated noise. We extend the analysis to general linear processes without the prior knowledge of a stationary or integrated model allowing the frequencies to be unknown. We provide a consistent model selection procedure. We first show that if we estimate a model with fewer frequencies than those in the correct model, the estimates converge to a subset of the frequencies in the correct model. This opens the way to a consistent model selection strategy based on a specific to general procedure that tests whether additional frequencies are needed. This is achieved using tests based on the feasible "super efficient" (under unit root noise) Generalized Least Squares estimator of Perron, Shintani and Yabu (2017) who assumed the frequencies to be known. We show that the limiting distributions of our test statistics are the same for both cases about the noise function. Simulation results confirm that our frequency selection procedure works well with sample sizes typically available in practice. We illustrate the usefulness of our method via applications to unemployment rates and global temperature series.

JEL Classification Number: C22.

Keywords: Cyclical trends, median-unbiased estimator, nonlinear trends, super-efficient estimator, unit root.

---

*We thank Bertille Antoine and Takeshi Hoshikawa for helpful comments and Francsico Estrada for providing some data used. We also acknowledge the helpful comments from seminar and conference participants at Kindai University, Kyoto University and the 4th Hitotsubashi Summer Institute (HSI2018).

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02458; e-mail: perron@bu.edu.

‡Faculty of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; e-mail: shintani@e.u-tokyo.ac.jp.

§Department of Business and Commerce, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-0073, Japan; e-mail: tyabu@fbc.keio.ac.jp.

1 Introduction

We consider the issue of selecting trigonometric terms when the trend of a univariate time series is periodic whether with a trending component or with a fixed mean. This problem was first investigated by Anderson (1971) under the assumption of a serially uncorrelated random noise component. We generalize his results to provide a consistent model selection procedure robust to serial correlation in the noise of unknown form allowing $I(0)$, i.e., stationary, or $I(1)$, i.e., integrated process with an autoregressive root on the unit circle. The goal is to determine whether there is a need to consider a simple linear trend or a more general nonlinear one. The main issue is that the limiting distributions of statistics to test for the presence of nonlinearities in the trend depends on the order of integration which is also unknown. On the other hand, testing whether the noise component is $I(0)$ or $I(1)$ depends on the nature of the deterministic trend (e.g., Perron, 1989, 1990, for the cases of abrupt structural changes in slope or level). In particular, if the trend is misspecified, unit root (and other misspecification) tests, will lose power and can be outright inconsistent (e.g., Perron, 1988, Campbell and Perron, 1991). Hence, we are faced with a circular problem and what is needed is a procedure to test for nonlinearity that is robust to the possibilities of $I(0)$ or $I(1)$ noise components. We circumvent this circular problem using the feasible GLS estimator proposed by Perron, Shintani and Yabu (2017); henceforth PSY, following work by Perron and Yabu (2009a,b), which uses a so-called super-efficient estimator of the sum of the autoregressive coefficients when the errors are $I(1)$, though they required the assumption of known frequencies. We first show that if we estimate a model with fewer (or the same number of) frequencies than those in the correct model, our suggested estimates converge to a subset (or all) of the frequencies in the correct model. This opens the way to a consistent model selection strategy based on a specific to general procedure that tests whether additional frequencies are needed. The problem is to obtain tests that have the same limit distribution under both $I(0)$ and $I(1)$ noise. We propose two tests, mean and sup-type tests, that achieve this goal. We show that the limiting distributions of our proposed test statistics are the same for both $I(0)$ and $I(1)$ noise even if the frequencies in the trigonometric function are unknown. This contrasts with the results of tests for structural breaks at unknown dates (Perron and Yabu, 2009b).

The use of trigonometric functions is motivated by the observed regular fluctuations in time series data which are referred to as cyclical trends by Anderson (1971). In principle, frequencies in cyclical trends can be selected by testing zero restrictions on the amplitudes, namely, the coefficients on trigonometric functions. However, unlike the case of a polynomial trend function where there is a clear ordering of the degrees of the polynomials, there is no a priori meaningful ordering of the frequencies. Furthermore, as emphasized by Davies (1987), the frequencies are not identified under the null hypothesis of no cyclical component. Anderson (1971) circumvents this problem in a way that is similar to the approach typically used for problems related to structural breaks (e.g., Andrews, 1993, and Andrews and Ploberger, 1994). Our procedure generalizes Anderson's (1971) approach to cover a broader class of time series data and offers new procedures that provide reliable inference in practice whether the noise component is $I(0)$ or $I(1)$, a feature that can be especially useful if one wants to assess the nature of the noise component using some unit root test (or other misspecification test).

Trigonometric functions have also been used to represent the seasonal, trend and irregular components (e.g., Harvey, 1993). Gao and Tsay (2019) used this approach with a prespecified seasonal frequency when analyzing high-dimensional multivariate time series with the noise components allowed to be $I(0)$ or $I(1)$. An alternative interpretation is to use trigonometric functions as a base function for Fourier approximations to model general nonlinear trend functions. For example, Becker et al. (2004) used this approach to approximate the time-varying coefficients in a regression model with the aim of testing for parameter constancy, while Becker et al. (2006) used it to capture an unknown form of structural break in their test for stationarity. In both cases, the noise is $I(0)$ and the number of frequencies is assumed known while which frequencies to include is done via some tests. Trigonometric functions have also been used in the context of unit root tests; e.g., Enders and Lee (2012a, 2012b) and Rodrigues and Taylor (2012). Testing for trigonometric trend allowing both $I(0)$ and $I(1)$ noise have been investigated by Harvey et al. (2010), Astill et al. (2015) and PSY assuming a known set of frequencies, or using some sequential procedure to determine the maximum frequency.

The paper is structured as follows. Section 2 presents the trigonometric trend model, the statistical procedures and the theoretical results about the consistent selection of the frequencies. Section 3 assesses the adequacy of our proposed methods via simulations. Section 4 illustrates their usefulness via applications to unemployment rates and aggregate temperatures series. An appendix contains the proofs of the results stated in the text.

2 The Trigonometric Trend Model and the Statistical Procedures

We describe the model in Section 2.1 focusing first on the AR(1) case for clarity of exposition. Sections 2.2 contains the main theoretical results about the consistent selection of the frequencies. Section 2.3 discusses the modifications needed for general linear processes in the noise component via an autoregressive approximation and to allow for a trending component.

2.1 The model

Following Anderson (1971), we consider a scalar random variable y_t be generated by:

$$y_t = \beta_0 + \sum_{j=1}^{m} \{\gamma_{1j} \sin(2\pi k_j t/T) + \gamma_{2j} \cos(2\pi k_j t/T)\} + u_t \tag{1}$$

where $u_t = \alpha u_{t-1} + e_t, -1 < \alpha \leq 1$ for $t = 1, ..., T$ with $\mathbb{K}_m = (k_1, \ldots, k_m) \in \Lambda = \{1, ..., n\}$, n is the upper bound on the frequencies, e_t is a martingale difference sequence with respect to the sigma-field $\mathcal{F}_t$ generated by $\{e_{t-s}, s \geq 0\}$, i.e., $E(e_t|\mathcal{F}_{t-1}) = 0$, with $E(e_t^2) = \sigma^2$ and $E(e_t^4) < \infty$. Also, $u_0 = O_p(1)$. For now we focus on the AR(1) case with $-1 < \alpha \leq 1$, so that both stationary, $I(0)$ with $|\alpha| < 1$, and integrated, $I(1)$ with $\alpha = 1$, processes are allowed. The validity of our methods with a more general structure for u_t will be discussed later. Note that the indices k_j are nonnegative integers for $j = 1, ..., m$ where m is the total number of frequencies, which can take values in a proper subset of all the integers from 1 to n provided $m < n$. For example, when $m = 2$ and $n = 3$, $\mathbb{K}_2 = (k_1, k_2)$ can be either $(1, 2)$, $(1, 3)$ or $(2, 3)$. In general, the number of possible choices of $\mathbb{K}_m$ is ${}_nC_m$.

If the set of frequencies $\mathbb{K}_m = (k_1, \ldots, k_m)$ is known, the model can be estimated by using the following feasible GLS (FGLS) estimator even if the order of integration of u_t is unknown. Let $x_t = (1, \sin(2\pi k_1 t/T), \cos(2\pi k_1 t/T), ..., \sin(2\pi k_m t/T), \cos(2\pi k_m t/T))'$ and $\Psi = (\beta_0, \gamma_{11}, \gamma_{21}, ..., \gamma_{1m}, \gamma_{2m})' = (\beta_0, \gamma')'$, $y = (y_1, ..., y_T)'$ and $X = (x_1, ..., x_T)'$. Typically, $\hat{\alpha} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2$ where $\hat{u}_t = y_t - x_t' \hat{\Psi}_{OLS}$ and $\hat{\Psi}_{OLS} = (X'X)^{-1} X'y$, can be used as an estimate of α required for a Cochrane and Orcutt (1949) type transformation. Instead, we follow PSY and use $\hat{\alpha}_S = \hat{\alpha}$ if $|\hat{\alpha} - 1| > T^{-1/2}$ and 1 otherwise. The estimator $\hat{\alpha}_S$ is often referred to as super-efficient as it converges to 1 at a rate faster than T when $\alpha = 1$. As explained in PSY, in order to have a limit distribution identical under both the $I(0)$ and $I(1)$ cases, it is necessary to use the FGLS estimator $\hat{\Psi}$ of Prais and Winsten (1954) which minimizes $SSR_T(\mathbb{K}_m) = \sum_{t=1}^{T} (\widetilde{y}_t - \widetilde{x}_t' \hat{\Psi})^2$, where $\widetilde{x}_t' = (1 - \hat{\alpha}_S L) x_t'$, $\widetilde{y}_t = (1 - \hat{\alpha}_S L) y_t$ for $t = 2, ..., T$, and $\widetilde{x}_1' = (1 - \hat{\alpha}_S^2)^{1/2} x_1'$, $\widetilde{y}_1 = (1 - \hat{\alpha}_S^2)^{1/2} y_1$. Alternatively, in matrix format, $\hat{\Psi} = (\widetilde{X}'\widetilde{X})^{-} \widetilde{X}'\widetilde{y}$ where $\widetilde{X} = (\widetilde{x}_1, ..., \widetilde{x}_T)'$ is a $T \times (2m + 1)$ matrix of transformed data and the $T \times 1$ vector $\widetilde{y} = (\widetilde{y}_1, ..., \widetilde{y}_T)'$ is similarly defined. Here, $(\widetilde{X}'\widetilde{X})^{-}$ is the generalized inverse of $\widetilde{X}'\widetilde{X}$, which is needed since the first column of $\widetilde{X}$ is asymptotically zero when $\alpha = 1$.

This poses no problem since we do not make inference about the constant β_0. For known $\mathbb{K}_m$, PSY considered testing the absence of a cyclical trend, $H_0 : \gamma_{1j} = \gamma_{2j} = 0$ for all j, against the alternative $H_1 : \gamma_{1j} \neq 0$ or $\gamma_{2j} \neq 0$ for some j with the Wald statistic: $W_{\widehat{\gamma}}(\mathbb{K}_m) = \hat{\Psi}' R' [s^2 R(\widetilde{X}'\widetilde{X})^{-} R']^{-1} R\hat{\Psi}$, where $R = [0 : I_{2m}]$ is a $2m \times (2m+1)$, $s^2 = T^{-1} \sum_{t=1}^{T} \hat{v}_t^2$ and $\hat{v}_t = \widetilde{y}_t - \widetilde{x}_t' \hat{\Psi}$. Their results imply that $W_{\widehat{\gamma}}(k)$ converges to a chi-square distribution with $2m$ degrees of freedom, $\chi^2(2m)$, for both the $I(0)$ and $I(1)$ cases, i.e.,

$$\begin{aligned} W_{\widehat{\gamma}}(\mathbb{K}_m) \quad &\Rightarrow \quad [R(\int_0^1 G(r)G(r)^{'}dr)^{-} \int_0^1 G(r)dW(r)]'[R(\int_0^1 G(r)G(r)^{'}dr)^{-} R']^{-1} \qquad (2) \\ &\quad \times [R(\int_0^1 G(r)G(r)^{'}dr)^{-} \int_0^1 G(r)dW(r)] \equiv A(\mathbb{K}_m) =^d \chi^2(2m) \end{aligned}$$

where $G(r) = F(r) = [1, \sin(2\pi k_1 r), \cos(2\pi k_1 r), ..., \sin(2\pi k_m r), \cos(2\pi k_m r)]'$ if $|\alpha| < 1$ and $G(r) = Q(r) = [0, 2\pi k_1 \cos(2\pi k_1 r), -2\pi k_1 \sin(2\pi k_1 r), ..., 2\pi k_m \cos(2\pi k_m r), -2\pi k_m \sin(2\pi k_m r)]$ if $\alpha = 1$. Below, we consider related statistical analyses when $\mathbb{K}_m = (k_1, \ldots, k_m)$ is unknown.

2.2 Consistent selection of the frequencies

We use m_0 as the true number of frequencies and m as the number included in the estimated model. Correct specification occurs when $m = m_0$, while the model is underspecified if $m < m_0$. The estimates of the frequencies is given by $(\hat{k}_1, \ldots, \hat{k}_m) = \arg\min_{\Lambda} SSR_T(\mathbb{K}_m)$.

Theorem 1 *Let y_t be generated by (1) with $\gamma_{1j} = \gamma_{1j0}$ and $\gamma_{2j} = \gamma_{2j0}$ if $|\alpha| < 1$; $\gamma_{1j} = \gamma_{1j0} T^h$ and $\gamma_{2j} = \gamma_{2j0} T^h$ $(h > 1/2)$ if $\alpha = 1$. Then, as $T \to \infty$, a) if $m = m_0$: $(\hat{k}_1, \ldots, \hat{k}_m) \xrightarrow{p} (k_1, \ldots, k_{m_0})$; b) if $m < m_0$: for all $j = 1, \ldots, m$, $P[\hat{k}_j \notin \{k_1, k_2, \ldots, k_{m_0}\}] \to 0$.*

The conditions of the theorem state that the coefficients γ_{1j} and γ_{2j} increase with T when $\alpha = 1$. This is a needed theoretical device to ensure consistent estimates of the frequencies since, when $\alpha = 1$, the coefficient estimates are not consistent for fixed values. Hence, the results should be viewed as applying to large values when $\alpha = 1$, though they are still relevant for moderate values as shown in the simulations of Section 3. The results first state that $(\hat{k}_1, \ldots, \hat{k}_m)$ are consistent estimates of $(k_1, \ldots, k_{m_0})$ for the correctly-specified case $(m = m_0)$. What is more relevant to develop a consistent model selection procedure is part (b), which states that when the fitted model is underspecified, the estimates nevertheless converge to some of the true frequencies, those that minimize the overall SSR. This is what is needed to develop a sequential procedure to consistently estimate the frequencies. This result is akin to that in the structural change literature; i.e., when estimating a structural change model with a number of changes lower than the true value, one still obtains consistent estimates of a subset of the true break dates; see Bai and Perron (1998) and Bai (1997). We next consider the construction of the relevant sequential test for consistent model selection.

Consider a model with ℓ frequencies included in (1) where $0 \leq \ell \leq n-1$ so that either γ_{1j} or γ_{2j} is non-zero for all $j = 1, \ldots, \ell$, and the values k_j's for $j = 1, \ldots, \ell$ are known (possibly via some prior testing procedure). We wish to test the null hypothesis (H_0) of no trigonometric trend at some additional frequency $k_{\ell+1}$ $(H_0 : \gamma_{1,\ell+1} = \gamma_{2,\ell+1} = 0)$ against the alternative hypothesis $H_1 : \gamma_{1,\ell+1} \neq 0$ or $\gamma_{2,\ell+1} \neq 0$ at the unknown frequency $k_{\ell+1}$. In this setup, a parameter is not identified under the null hypothesis, namely $k_{\ell+1}$. Following the structural change literature (e.g., Andrews, 1993, and Andrews and Ploberger, 1994), we consider two tests: sup-$W(\ell+1) = \sup_{k_{\ell+1} \in \Lambda_{-\ell}} W_{\hat{\gamma}_{\ell+1}}(k_{\ell+1})$ and $Mean$-$W(\ell+1) = (n-\ell)^{-1} \sum_{k_{\ell+1} \in \Lambda_{-\ell}} W_{\hat{\gamma}_{\ell+1}}(k_{\ell+1})$, where $\Lambda_{-\ell} = \Lambda \backslash \{k_1, \ldots, k_\ell\}$ and $W_{\hat{\gamma}_{\ell+1}}(k_{\ell+1}) = \hat{\Psi}' R' [s^2 R (\widetilde{X}' \widetilde{X} X)^{-} R']^{-1} R \hat{\Psi}$, with $\widetilde{X} = (\widetilde{x}_1, ..., \widetilde{x}_T)'$ a $T \times \{2(\ell+1)+1\}$ matrix of transformed data whose t^{th}-row is given by $\widetilde{x}_t' = (1 - \hat{\alpha}_S L) x_t'$ for $t = 2, ..., T$ and $\widetilde{x}_1' = (1 - \hat{\alpha}_S^2)^{1/2} x_1'$, $x_t = (1, \sin(2\pi k_1 t/T), \cos(2\pi k_1 t/T), ..., \sin(2\pi k_{\ell+1} t/T), \cos(2\pi k_{\ell+1} t/T))'$ and $R = [0_{2 \times (2\ell+1)} : I_2]$ is a $2 \times \{2(\ell+1)+1\}$ restriction matrix. A simple modification of Theorem 1 of PSY yields: $W_{\widehat{\gamma}_{\ell+1}}(k_{\ell+1}) \Rightarrow A(k_{\ell+1})$ as defined by (2). Hence, $W_{\widehat{\gamma}_{\ell+1}}(k_{\ell+1})$ has the same $\chi^2(2)$ distribution under both $I(0)$ and $I(1)$ errors. We show that the same holds for the sup-W and $Mean$-W tests, even though the functionals $F(r)$ and $Q(r)$ are different.

Theorem 2 *Under H_0 specified above and for $-1 < \alpha \leq 1$:* sup-$W(\ell+1) \Rightarrow \sup_{k_{\ell+1}\in\Lambda_{-\ell}} g(k_{\ell+1}) \equiv \sup_{j\in(1,\ldots,n-\ell)} A_j$, *$Mean$-$W(\ell+1) \Rightarrow (n-\ell)^{-1}\sum_{k_{\ell+1}\in\Lambda_{-\ell}} g(k_{\ell+1}) \equiv (n-\ell)^{-1}\sum_{j=1}^{n-\ell} A_j$ where*

$$g(k_j) = \frac{[\int_0^1 \sin(2\pi k_j r)dW(r)]^2}{\int_0^1 \sin(2\pi k_j r)^2 dr} + \frac{[\int_0^1 \cos(2\pi k_j r)dW(r)]^2}{\int_0^1 \cos(2\pi k_j r)^2 dr}$$

and $A_j =^d \chi^2(2)$.

It should be noted that the limit distribution depends only on $n-\ell$. The critical values, valid for both the $I(0)$ and $I(1)$ cases, are provided in Table 1 for $n-\ell = 1, ..., 5$. These were obtained via simulations with one million replications. When $n-\ell = 1$, (the known frequency case) the critical values correspond to those of a $\chi^2(2)$, but are still reported for convenience. Theorem 2 allows the following sequential procedure to estimate the number and nature of the frequencies: 1) starting with $\ell = 0$, use the sup-$W(\ell+1)$ or $Mean$-$W(\ell+1)$ test; if the null is not rejected, conclude for a model with ℓ frequencies estimated as $(\hat{k}_1, \ldots, \hat{k}_\ell) = \arg\min_\Lambda SSR_T(\mathbb{K}_\ell)$; 2) if the null is rejected, update ℓ to $\ell+1$ and repeat until a non-rejection occurs or the maximal allowed value $\ell = n-1$ is attained. This procedure will result in a consistent model selection if the size of the test converges to 0 slowly enough for the tests to be consistent. This generalizes Anderson's (1971) procedure valid with uncorrelated errors. An alternative approach would use a general-to-specific procedure by running a trigonometric trend regression using the largest allowed value n for the number of frequencies and keeping only the components that are significant according to a Wald test applied to a pair of coefficients for each frequency, again using the FGLS procedure. This approach can be justified using the asymptotic results of PSY. However, the simulations reported below show that our proposed sequential procedure is preferable.

2.3 Extensions to general linear processes and to trending series

The assumption of AR(1) errors is restrictive but can easily be extended to allow general linear processes assuming that u_t is generated by one of the following two structures: a) $I(0)$ noise: $u_t = C(L)e_t$, $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$, $0 < |C(1)| < \infty$; b) $I(1)$ noise: $\Delta u_t = D(L)e_t$, $D(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i|d_i| < \infty$, $0 < |D(1)| < \infty$. As before e_t is a martingale difference sequence (with $E(e_t^2) = \sigma^2$) and $u_0 = O_p(1)$. Then, $\hat{\alpha}_S$ is computed based on the regression $\hat{u}_t = \alpha \hat{u}_{t-1} + \sum_{i=1}^{p_T} a_i^* \Delta \hat{u}_{t-i} + e_{pt}$ where $p_T \to \infty$ and $p_T^3/T \to 0$ as $T \to \infty$, with p_T allowed to be in the range (0,int$[12(T/100)^{1/4}]$). Let $\hat{e}_{pt}$ be the estimated residuals, then, s^2 in the Wald statistic $W_{\widehat{\gamma}}(\mathbb{K}_m)$ needs to be replaced by (see PSY for details):

$$\hat{\omega}^2 = \begin{cases} (T - p_T)^{-1} \sum_{t=p_T+1}^{T} \hat{e}_{pt}^2 & \text{if} \quad |\hat{\alpha} - 1| > T^{-1/2} \\ T^{-1} \sum_{t=1}^{T} \hat{v}_t^2 + T^{-1} \sum_{j=1}^{T-1} w(j, m_T) \sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j} & \text{if} \quad |\hat{\alpha} - 1| \leq T^{-1/2} \end{cases} \quad (3)$$

where $w(j, m_T)$ is a weight function with bandwidth m_T. We use Andrews' (1991) automatic selection procedure for m_T along with the quadratic spectral window. When the sample size is small and α is near one, the OLS estimator $\hat{\alpha}$ is known to be biased downward. Using Roy and Fuller's (2001) 'upper-biased' corrected estimator turns out to be very effective in reducing the bias without changing any of the asymptotic results reported. The exact form of the estimate is described in the appendix. See Roy et al. (2004), Perron and Yabu (2012), and PSY for more details.

Theorem 2 showed the equivalence of the sup-$W(\ell+1)$ and $Mean$-$W(\ell+1)$ between the $I(0)$ and $I(1)$ cases for model (1) for which a linear time trend is not included. However, if β_0 is replaced by $\beta_0 + \beta_1 t$, thereby allowing trending series, the equivalence no longer holds and the limit distributions are different for the cases with $I(0)$ and $I(1)$ noise. For the $I(1)$ case, the limiting distributions of the test statistics will be the same as for the constant only case so that the critical values are those in Table 1. For the $I(0)$ case, the limiting distribution differs because the correlation of the linear trend and the sine function is not zero. Still, the two sets of critical values are very similar; see the supplement. Hence, one can safely use the critical values in Table 1 expecting the same size in large samples.

3 Monte Carlo Experiments

We now present simulation results to assess the performance of our methods in finite samples typically available in practice. To investigate the methods to select the true frequencies based on minimizing the SSR of the FGLS estimator, labelled the FGLS method (see Theorem 1), we first generate data from a process with a single frequency $k_1 \in \{1, 2, 3, 4, 5\}$:

$$y_t = \gamma\{\sin(2\pi k_1 t/T) + \cos(2\pi k_1 t/T)\} + u_t \tag{4}$$

where $u_t = \alpha u_{t-1} + e_t$, $\gamma \geq 0$, $e_t \sim i.i.d.\ N(0, 1)$ and $u_0 = 0$. We set $\alpha = 1.0$ and 0.8 and consider various values of γ; $T = 200$ and we use 10,000 replications [1]. The FGLS method is applied without assuming an AR(1) structure, we allow for an unknown AR(p) process with p selected using the BIC and employ the bias correction method described in the appendix. We compare its performance with two alternative procedures. The first is to select k_1 by minimizing the SSR from using OLS applied to (1) with one frequency. Such a procedure has been considered in Becker et al. (2006) and is labelled as the OLS method [2]. The second is to select k_1 by minimizing the SSR of the following Dickey-Fuller (1979) type regression

$$\Delta y_t = \beta_0 + \rho y_{t-1} + \textstyle\sum_{i=1}^{p_T} \Delta y_{t-i} + \gamma_1 \sin(2\pi k_1 t/T) + \gamma_2 \cos(2\pi k_1 t/T) + e_t, \tag{5}$$

considered by Enders and Lee (2012b) and labelled as the DF method (with p_T selected using BIC). The OLS and DF methods have been used to select the trigonometric components to properly specify the trend function prior to performing unit root or stationarity tests. The

[1] While the results for the model without linear trend are reported here, we also conducted experiments for the linear trend case and obtained similar results. The additional results are available in the supplement.

[2] No lagged variables were included for the OLS estimator as in Becker et al. (2016). We computed their test using an estimate of the long-run variance to account for serial correlation. The results were similar.

DF method is based on assuming $I(1)$ errors to obtain the relevant limit distribution. It is conservative under $I(0)$ errors. On the other hand, the OLS method relies on critical values derived assuming $I(0)$ errors. Our results will show our procedure, valid under both $I(1)$ and $I(0)$ errors, to perform at least as well, in general, than both the OLS and DF methods.

Table 2 shows the proportions of selecting the true k_1 for each values of k_1 from 1 to 5, and the average of the correctly selected proportions among all five cases. Note that when $\gamma = 0$, k_1 is not identified; hence one would expect the estimates to be randomly distributed across permissible values of k_1 so that each is selected about 20% of the time. First, note that the proportion of correctly selected frequencies increases with γ in all cases. Second, when $\alpha = 1.0$, using the DF and FGLS methods yields more accurate selections in general compared to using the OLS method. When $\gamma = 0$, the DF (resp., OLS) method unreasonably selects $k_1 = 1$ in 47% (resp., 79%) of the cases, while the FGLS method selects all frequencies equally 20% of the time. The OLS method performs very well when $k_1 = 1$ but relatively badly for other values. The FGLS method performs best overall when $k_1 > 1$, though it is inferior to DF when $k_1 = 1$. Third, when $\alpha = 0.8$, all three procedures perform equally well. In summary, the FGLS method strikes the best balance in selecting the true frequency when the number is known. The OLS procedure tends to select incorrect frequencies when $\alpha = 1.0$, so one should be careful when using it as a prior step to perform unit root tests.

The second set of simulation (Figures 1 and 2) pertains to the evaluation of the performance of the sup-W and $Mean$-W tests to estimate the frequencies to be included for each step of the sequential procedure. We consider testing the null hypothesis of no trigonometric trend ($m_0 = \ell = 0$) against the alternative of one frequency ($m_0 = \ell + 1 = 1$) since this is frequently used in practice. We set the number of candidate frequencies at $n = 5$. We compare the properties of our tests to two others available in the literature. The first is a sup-F test, similar to our sup-W, based on the OLS estimates of (1) considered by Anderson (1971) and Becker et al. (2006), labelled sup-OLS. Since it is based on the limiting distribution assuming an $I(0)$ noise, it is not expected to perform well when the noise is $I(1)$. Enders and Lee (2012b) considered a sup-F test based on the DF regression (5), labelled sup-DF. The test is based on the limiting distribution obtained under the assumption of $I(1)$ noise and, hence, it is not expected to be robust to $I(0)$ noise (Enders and Lee, 2012a, considered a similar test for the model with a linear trend). To evaluate the exact size of the tests, we generate data from (4) setting $\gamma = 0$ for various values of α. We use $10,000$ replications and critical values for nominal 5% size tests. We report results for $T = 200$ and 500.

The empirical size results are presented Figure 1 as a function of α varying from 0 to 1. The exact size of the sup-OLS test quickly becomes very large as α gets further away from 0 and is therefore not reliable in the presence of serial correlation in the noise. In contrast, the sup-DF test is significantly undersized over the entire stationary region for α and achieves an exact size close to 5% only when α is very close to 1, even when $T = 500$. Among the two tests we propose, the rejection frequencies of the $Mean$-W is, in general, closer to the 5% nominal size, especially when $T = 200$. The $Mean$-W test is conservative when α is near to but not equal to 1, while in the same region the sup-W is slightly liberal. Both tests have roughly the correct size when $\alpha = 1$ or α is far from 1, especially when $T = 500$. Overall, the results show that our tests have reasonable size properties for both the $I(0)$ and $I(1)$ cases, and have much better properties than the sup-DF or sup-OLS procedures.

To evaluate the power of the tests, we henceforth omit the sup-OLS test given its very bad size properties. Instead, as a benchmark, we include the infeasible GLS-based supremum Wald test that uses the true value of α, labelled as sup-GLS, which is expected to provide an upper bound on the power function. The data are generated from (4) setting $\gamma \geq 0$ and $k_1 = 2$, which is assumed to be unknown when constructing the tests. We are interested in the power performance of the tests for the unit root ($\alpha = 1.0$) and stationary ($\alpha = 0.8$) cases.

We set $\gamma = \gamma_0 T^{1/2}$ for the unit root case, while we set $\gamma = \gamma_0 T^{-1/2}$ for the stationary case. This is done in order to obtain results that pertain to the local asymptotic power of the tests. The results are presented in Figure 2. When $\alpha = 1.0$, the power functions of sup-W and sup-GLS are almost indistinguishable. The power of the $Mean$-W test is slightly lower but still clearly dominates that of the sup-DF even in the unit root case. In summary, amongst the feasible tests, the highest power is achieved with sup-W followed by $Mean$-W.

Finally, we investigate the performance of sequential procedures to select the correct combination of frequencies using our proposed sequential strategy based on either the sup-W or $Mean$-W tests. Sequential procedures using other tests, such as sup OLS or sup DF are not included. For purposes of comparisons, we also consider the selection frequency using a general-to-specific procedure when the full model with all possible frequencies are estimated by FGLS but keeping only the frequencies for which the zero restriction on a pair of coefficients (on the sine and cosine trend functions with a common frequency) is rejected using a Wald test at the 5% significance level. We consider two data generating processes. The first uses data generated from (4) with $\gamma \geq 0$, $k_1 = 2$, $\gamma = \gamma_0 T^{1/2}$ for the unit root case and $\gamma = \gamma_0 T^{-1/2}$ for the stationary case. We report the proportions of correctly selecting a single frequency with $k_1 = 2$ when the maximum number of frequency is set at $n = 5$, with the total number of frequencies assumed unknown. If more than one frequency are selected, it is not counted as a success or correct specification, even if the frequency 2 is included. Figure 3 reports the proportions of success with $T = 500$ for $\alpha = 1.0$ and 0.8 obtained from $10,000$ replications. Note that by construction the procedures involve a type I error so that the proportion of successes cannot reach 100% reflecting the fact that overfitting cannot be avoided when the size of the test is fixed, though this problem can be remedied in large samples adopting a size that decreases as T increases at some appropriate rate. For both $I(1)$ and $I(0)$ cases, the general-to-specific procedure performs well when γ is small, but the suggested specific-to-general sequential procedure performs markedly better when γ becomes larger. The nonincreasing power of the general-to-specific procedure for large values of γ comes from the fact that the procedure tends to over-specify the total number of frequencies. Note that when γ is relatively small, the cost of neglecting nonlinear trend components is relatively minor. For these reasons, the specific-to-general sequential procedure is preferable. We also considered a second experiment with data generated by

$$y_t = \gamma\{\sin(2\pi k_1 t/T) + \cos(2\pi k_1 t/T) + \sin(2\pi k_2 t/T) + \cos(2\pi k_2 t/T)\} + u_t \qquad (6)$$

where $u_t = \alpha u_{t-1} + e_t$, $\gamma \geq 0$ with $k_1 = 2$ and $k_2 = 3$. All other settings are the same as before. Figure 4 shows the proportions of success, correct specification, with $T = 500$ for $\alpha = 1.0$ and 0.8. Overall, the results are similar to those for the single frequency model.

4 Empirical Applications

To illustrate the usefulness of our procedure, we consider first the estimation of cyclical trends in unemployment rate series for the G7 countries. The monthly harmonized unemployment rate series for Canada, France, Germany, Italy, Japan, UK and US were obtained from the OECD database and a logarithmic transformation was applied. The sample period is from January 1983 to July 2015 for all countries except Germany for which it is from January 1991 to July 2015. The extracted nonlinear trend components may be interpreted as estimates of the natural rates of unemployment, a conceptual unobserved component used in macroeconomic policy. The unemployment gap defined as the deviations from the natural rate is countercyclical and expected to be stationary. However, in a competing hysteresis hypothesis on unemployment argued by Blanchard and Summers (1987), deviations from the natural rates can be highly persistent and possibly permanent. Hence, to permit these competing views it is important to allow for both $I(0)$ and $I(1)$ error components when estimating the nonlinear trend functions. We set $n = 5$ as the maximum number of frequencies. We first determine whether to include a linear trend term using the asymptotic results of PSY for the FGLS estimator of a model with all possible frequencies included (a 10% size is used). We then estimate the frequencies of the trigonometric trend function using the sequential procedure based on the $Mean\text{-}W(\ell+1)$ test with a 5% size with a maximal value $n = 5$. The results are presented in Table 3(a), while Figure 5 shows the time series plots of the unemployment rate series along with the estimated cyclical trends. The results show clear statistically significant evidence of nonlinear trends for all seven countries. The linear trend component is absent, except for Italy and the UK. Overall, the selected fitted non-linear trend captures very well the low-frequency variations in the series.

As a second application, we consider the estimation of a nonlinear trend for annual global, northern and southern hemispheres temperature series from 1850 to 2010. The data used are from the HadCRUT3 database (http://www.metoffice.gov.uk/hadobs/hadcrut3/), which were also used in Estrada et al. (2013a,b) and PSY. The data, graphed in Figure 6, suggest that World War II and the Great Crash contributed to the mid-20th century cooling via important reductions in CO2 emissions. There is also a marked increase in the growth rates of temperatures near 1960, marking the start of sustained global warming. Since the mid-90s, reductions in the emission of chlorofluorocarbons and methane contributed to the so-called "hiatus", a slowdown in the growth rate of temperatures. As discussed in PSY, it is also important to allow for both cases with a $I(0)$ or $I(1)$ noise when estimating the nonlinear trend components. Again, we use the sequential procedure based on the $Mean\text{-}W(\ell+1)$ test with a maximal value $n = 5$ and a 5% size to determine the frequencies. For all three series, we include a linear trend term in the regression based on the procedure of PSY described above. The results are presented in Table 3(b) and the fitted trend functions are shown in Figure 6. For the global and southern hemisphere temperature series, $\mathbb{K}_2 = (k_1, k_2) = (1, 3)$

is selected. For the northern hemisphere, $\mathbb{K}_1 = k_1 = 1$ is selected. Our results suggest that, even with a small number of frequencies, the trigonometric trend model captures well the main features of the climate trend, namely the slowdown in growth during the 40s-mid-50s, the change in growth following 1960 and a slowdown in growth in the post mid-90s. A decrease is also present in the late 19th century for global and southern hemisphere.

5 Conclusions

We proposed a new consistent model selection procedure for a trigonometric trend function. The problem of unknown frequencies in such a cyclical trend was first investigated by Anderson (1971) under the assumption of serially uncorrelated errors. We extended the result in important ways by allowing for general linear processes approximated by a sequence of autoregressions without the prior knowledge as to whether the error term is stationary or contains an autoregressive unit root. This was achieved by modifying the test statistic proposed by Perron, Shintani and Yabu (2017) which requires the assumption of a known set of frequencies. We showed that the limiting distribution of our test statistics is the same for both the stationary and unit root cases even if the frequencies are unknown. Simulation results confirmed that our procedure works well with sample sizes typically available in practice. We illustrated the usefulness of our method via applications to international data on unemployment rates and to global temperature series.

References

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, Wiley, New York: New York.

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817-858.

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383-1414.

Astill, S., Harvey, D. I., Leybourne, S. J., and Taylor, A. M. R. (2015), "Robust and Powerful Tests for Nonlinear Deterministic Components," *Oxford Bulletin of Economics and Statistics*, 77, 780–799.

Bai, J. (1997), "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315-352.

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

Becker, R., Enders, W., and Hurn, S. (2004), "A General Test for Time Dependence in Parameters," *Journal of Applied Econometrics*, 19, 899–906.

Becker, R., Enders, W., and Lee, J. (2006), "A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks," *Journal of Time Series Analysis*, 27, 381–409.

Blanchard, O. J., and Summers, L. H. (1987), "Hysteresis in Unemployment," *European Economic Review*, 31, 288–295.

Campbell, J. Y. and Perron, P. (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots', in *NBER Macroeconomics Annual 1991*, 6, MIT Press, 141-220.

Cochrane, D., and Orcutt, G. H. (1949), "Application of Least Squares Regression to Relationships Containing Auto-correlated Error Terms," *Journal of the American Statistical Association*, 44, 32–61.

Davies, R. B. (1987), "Hypothesis Testing When a Nuisance Parameter is Only Identified Under the Alternative," *Biometrika*, 47, 33–43.

Dickey, D. A. and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Enders, W., and Lee, J. (2012a), "A Unit Root Test Using a Fourier Series to Approximate Smooth Breaks," *Oxford Bulletin of Economics and Statistics*, 74, 574–599.

Enders, W., and Lee, J. (2012b), "The Flexible Fourier Form and Dickey-Fuller Type Unit Root Tests," *Economics Letters*, 117, 196–199.

Estrada, F., Perron, P., Gay-García, C., and Martínez-López, B. (2013a), "A Time-Series Analysis of the 20th Century Climate Simulations Produced for the IPCC's Fourth Assessment Report," *PloS one*, 8(3), e60017.

Estrada, F., Perron, P., and Martínez-López, B. (2013b), "Statistically Derived Contributions of Diverse Human Influences to Twentieth-century Temperature Changes," *Nature Geoscience*, 6, 1050–1055.

Gao, Z., and Tsay R. S. (2019), "A Structural-factor Approach to Modeling High-dimensional Time Series and Space-time Data," *Journal of Time Series Analysis*, 40, 343-362.

Harvey, A. C. (1993), *Time Series Models*, MIT Press, Cambridge, MA.

Harvey, D. I., Leybourne, S. J., and Xiao, L. (2010), "Testing for Nonlinear Deterministic Components When the Order of Integration is Unknown," *Journal of Time Series Analysis*, 31, 379–391.

Perron, P. (1988), "Trends and Random Walks in Macroeconomic Time Series: Further Evidence From a New Approach", *Journal of Economic Dynamics and Control*, 297-332.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica*, 57, 1361-1401.

Perron, P. (1990), "Testing for a Unit Root in a Time Series With a Changing Mean", *Journal of Business and Economic Statistics*, 8, 153-162.

Perron, P., Shintani, M., and Yabu, T. (2017), "Testing for Flexible Nonlinear Trends with an Integrated or Stationary Noise Component," *Oxford Bulletin of Economics and Statistics*, 79, 822–850.

Perron, P. and Yabu, T. (2009a), "Estimating Deterministic Trends With an Integrated or Stationary Noise Component," *Journal of Econometrics*, 151, 56-69.

Perron, P., and Yabu, T. (2009b), "Testing for Shifts in Trend with an Integrated or Stationary Noise Component," *Journal of Business and Economic Statistics*, 27, 369–396.

Perron, P., and Yabu, T. (2012), "Testing for Trend in the Presence of Autoregressive Error: A Comment," *Journal of the American Statistical Association*, 107, 844.

Prais, S. J., and Winsten, C. B. (1954), "Trend Estimators and Serial Correlation," Cowles Foundation Discussion Paper 383.

Rodrigues, P. M. M., and Taylor, A. M. R. (2012), "The Flexible Fourier Form and Local Generalised Least Squares De-trended Unit Root Tests," *Oxford Bulletin of Economics and Statistics*, 74, 736–759.

Roy, A., Falk, B., and Fuller, W. A. (2004), "Testing for Trend in the Presence of Autoregressive Error," *Journal of the American Statistical Association*, 99, 1082–1091.

Roy, A., and Fuller, W. A. (2001), "Estimation for Autoregressive Time Series with a Root Near 1," *Journal of Business and Economic Statistics*, 19, 482–493.

Appendix (a): Proofs of the Theoretical Results

Proof of Theorem 1: Without loss of generality, we only consider model (1) with a maximum of two frequencies, setting $\beta_0 = 0$ and using only the sine components, i.e.,

$$y_t = \gamma_{1T} \sin(2\pi k_1 t/T) + \gamma_{2T} \sin(2\pi k_2 t/T) + u_t \tag{A.1}$$

We show that, when the regression model has only one frequency, the estimated frequency $\hat{k} = \arg\min_{\Lambda} SSR_T(\mathbb{K}_1)$ for $\Lambda = \{1, ..., n\}$ converges to either k_1 or k_2 under the conditions of Theorem 1. Without loss of generality, we let $|\gamma_{10}| > |\gamma_{20}|$ if $|\alpha| < 1$ and $|\gamma_{10} k_1| > |\gamma_{20} k_2|$ if $\alpha = 1$. We show that if $|\alpha| < 1$, $T^{-1}[SSR_T(k) - SSR_T(k_1)]$ converges to a positive value for any $k \neq k_1$ and likewise for $T^{1-2h}[SSR_T(k) - SSR_T(k_1)]$ when $\alpha = 1$. Therefore, the selected frequency is a consistent estimate of the frequency k_1, arbitrarily selected as the one which minimizes the overall SSR. This result can easily be generalized by adding the cosine functions and allowing the number of frequencies to be m. Note that if a one frequency model holds, i.e., (A.1) with $\gamma_{20} = 0$, the model is correctly specified. Theorem 1(b) implies that if the true number of frequencies is one, the selected frequency obtained by minimizing the SSR is consistent. This is a special case of Theorem 1(a), hence we only prove part (b).

Stationary Case ($|\alpha| < 1$): The OLS estimate of α converges to a weighted average of α and 1, i.e., $\hat{\alpha} \xrightarrow{p} \alpha_i^* \equiv \alpha\lambda_i + (1 - \lambda_i)$, where $\lambda_1 = \omega^2/(\omega^2 + \gamma_{20}^2/2)$ with $\omega^2 = \sigma^2/(1 - \alpha^2)$ if $k = k_1$; $\lambda_2 = \omega^2/(\omega^2 + \gamma_{10}^2/2)$ if $k = k_2$; $\lambda_3 = \omega^2/(\omega^2 + (\gamma_{10}^2 + \gamma_{20}^2)/2)$ if $k \notin \{k_1, k_2\}$. In large samples, $\alpha < \hat{\alpha} < 1$ and thus $\hat{\alpha}_S = \hat{\alpha}$. Also,

$$T^{-1} SSR_T(k) \xrightarrow{p} \begin{cases} SSR(k_1) = \sigma^2 + (\alpha - \alpha_1^*)^2\omega^2 + (1 - \alpha_1^*)^2\gamma_{20}^2/2, \text{ if } k = k_1 \\ SSR(k_2) = \sigma^2 + (\alpha - \alpha_2^*)^2\omega^2 + (1 - \alpha_2^*)^2\gamma_{10}^2/2, \text{ if } k = k_2 \\ SSR(k) = \sigma^2 + (\alpha - \alpha_3^*)^2\omega^2 + (1 - \alpha_3^*)^2(\gamma_{10}^2 + \gamma_{20}^2)/2, \text{ if } k \notin \{k_1, k_2\} \end{cases}$$

and, hence, $SSR(k)$ has a unique minimum at k_1. For the case $k = k_1$,

$$\hat{\alpha} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2$$

$$
\begin{aligned}
&= \frac{T^{-1}\sum_{t=2}^{T}(\gamma_{20}\sin(2\pi k_2 t/T)+u_t)(\gamma_{20}\sin(2\pi k_2(t-1)/T)+u_{t-1})}{T^{-1}\sum_{t=2}^{T}(\gamma_{20}\sin(2\pi k_2(t-1)/T)+u_{t-1})^2}+o_p(1) \\
&= \frac{T^{-1}\sum_{t=2}^{T}u_t u_{t-1}+\gamma_{20}^2 T^{-1}\sum_{t=2}^{T}\sin(2\pi k_2 t/T)\sin(2\pi k_2(t-1)/T)}{T^{-1}\sum_{t=2}^{T}u_{t-1}^2+\gamma_{20}^2 T^{-1}\sum_{t=2}^{T}\sin^2(2\pi k_2(t-1)/T)}+o_p(1)
\end{aligned}
$$

$$
\xrightarrow{p} [(\alpha\omega^2+\gamma_{20}^2/2)/(\omega^2+\gamma_{20}^2/2)] = \alpha\lambda_1+(1-\lambda_1)
$$

using the results of $T^{-1}\sum_{t=2}^{T}\sin^2(2\pi k_2 t/T)\to 1/2$ and $T^{-1}\sum_{t=2}^{T}u_{t-1}^2\xrightarrow{p}\omega^2$. Also,

$$
\begin{aligned}
& T^{-1}SSR_T(k_1) \\
&= T^{-1}\sum_{t=2}^{T}(\gamma_{20}(\sin(2\pi k_2 t/T)-\alpha_1^*\sin(2\pi k_2(t-1)/T))+(\alpha-\alpha_1^*)u_{t-1}+e_t)^2+o_p(1) \\
&= T^{-1}\sum_{t=2}^{T}e_t^2+(\alpha-\alpha_1^*)^2 T^{-1}\sum_{t=2}^{T}u_{t-1}^2 \\
&\quad +\gamma_{20}^2 T^{-1}\sum_{t=2}^{T}(\sin(2\pi k_2 t/T)-\alpha_1^*\sin(2\pi k_2(t-1)/T))^2+o_p(1)
\end{aligned}
$$

$$
\xrightarrow{p} SSR(k_1)=\sigma^2+(\alpha-\alpha_1^*)^2\omega^2+(1-\alpha_1^*)^2\gamma_{20}^2/2
$$

using $T^{-1}\sum_{t=2}^{T}e_t{}^2\xrightarrow{p}\sigma^2$ and $T^{-1}\sum_{t=2}^{T}(\sin(2\pi k_2 t/T)-\alpha_1^*\sin(2\pi k_2(t-1)/T))^2\to(1-\alpha_1^*)^2/2$. We can similarly derive the convergence results of $\hat{\alpha}$ and $T^{-1}SSR_T(k)$ for the case of $k=k_2$ and $k\notin\{k_1,k_2\}$. Next, we show that $SSR(k)$ has a unique minimum at $k=k_1$. We have,

$$
\begin{aligned}
SSR(k_2)-SSR(k_1) &= [(\alpha-\alpha_2^*)^2\omega^2+(1-\alpha_2^*)^2\gamma_{10}^2/2]-[(\alpha-\alpha_1^*)^2\omega^2+(1-\alpha_1^*)^2\gamma_{20}^2/2] \\
&= (1-\alpha)^2\left\{[(1-\lambda_2)^2\omega^2+\lambda_2^2\gamma_{10}^2/2]-[(1-\lambda_1)^2\omega^2+\lambda_1^2\gamma_{20}^2/2]\right\} \\
&= (1-\alpha)^2\omega^2\left\{(\gamma_{10}^2/2)/(\omega^2+\gamma_{10}^2/2)-(\gamma_{20}^2/2)/(\omega^2+\gamma_{20}^2/2)\right\} \\
&= [(1-\alpha)^2\omega^4(\gamma_{10}^2-\gamma_{20}^2))/(2(\omega^2+\gamma_{10}^2/2)(\omega^2+\gamma_{20}^2/2)]>0
\end{aligned}
$$

using $\alpha-\alpha_i^*=(1-\lambda_i)(\alpha-1)$ and $1-\alpha_i^*=\lambda_i(1-\alpha)$. Similarly, for $k\notin\{k_1,k_2\}$,

$$
\begin{aligned}
SSR(k)-SSR(k_1) &= [(\alpha-\alpha_3^*)^2\omega^2+(1-\alpha_3^*)^2(\gamma_{10}^2+\gamma_{20}^2)/2]-[(\alpha-\alpha_1^*)^2\omega^2+(1-\alpha_1^*)^2\gamma_{20}^2/2] \\
&= (1-\alpha)^2\left\{[(1-\lambda_3)^2\omega^2+\lambda_3^2(\gamma_{10}^2+\gamma_{20}^2)/2]-[(1-\lambda_1)^2\omega^2+\lambda_1^2\gamma_{20}^2/2]\right\}
\end{aligned}
$$

$$= (1-\alpha)^2\omega^2 \left\{ \frac{(\gamma_{10}^2+\gamma_{20}^2)/2}{\omega^2+(\gamma_{10}^2+\gamma_{20}^2)/2} - \frac{\gamma_{20}^2/2}{\omega^2+\gamma_{20}^2/2} \right\}$$
$$= [(1-\alpha)^2\omega^4\gamma_{10}^2/(2(\omega^2+(\gamma_{10}^2+\gamma_{20}^2)/2)(\omega^2+\gamma_{20}^2/2)] > 0.$$

Therefore, $SSR(k)$ has an unique minimum at k_1. **Unit Root Case (**$\alpha = 1$**)**: Denote the OLS estimate of γ by $\hat{\gamma}_{OLS}$. When $k = k_1$,

$$T(\hat{\alpha}-1) = \frac{T^{-2h}\sum_{t=2}^{T}(-(\hat{\gamma}_{OLS}-\gamma_{1T})\Delta\sin(2\pi k_1 t/T) + \gamma_{2T}\Delta\sin(2\pi k_2 t/T) + e_t)\widehat{u}_{t-1}}{T^{-2h-1}\sum_{t=2}^{T}\widehat{u}_{t-1}^2}$$

and thus $T(\hat{\alpha}-1) \xrightarrow{p} 0$ since $T^{-2h-1}\sum_{t=2}^{T}\hat{u}_{t-1}^2 = O_p(1)$ and all the terms in the numerator are $o_p(1)$. $T(\hat{\alpha}-1) \xrightarrow{p} 0$ leads to $T(\hat{\alpha}_S-1) \xrightarrow{p} 0$. Similarly, we can show $T(\hat{\alpha}-1) \xrightarrow{p} 0$ and thus $T(\hat{\alpha}_S-1) \xrightarrow{p} 0$ for the case with $k = k_2$ and $k \notin \{k_1, k_2\}$. For $SSR_T(k)$, we have

$$T^{1-2h}[SSR_T(k_2) - SSR_T(k_1)]$$
$$= \gamma_{10}^2 T \sum_{t=2}^{T}\Delta\sin^2(2\pi k_1 t/T) - \gamma_{20}^2 T\sum_{t=2}^{T}\Delta\sin^2(2\pi k_2 t/T) + o_p(1)$$
$$\xrightarrow{p} 2\pi^2(k_1^2\gamma_{10}^2 - k_2^2\gamma_{20}^2) > 0,$$

and

$$T^{1-2h}[SSR_T(k) - SSR_T(k_1)] = \gamma_{10}^2 T\sum_{t=2}^{T}\Delta\sin^2(2\pi k_1 t/T) + o_p(1)$$
$$\xrightarrow{p} 2\pi^2(k_1^2\gamma_{10}^2) > 0, \text{ if } k \notin \{k_1, k_2\}.$$

Therefore, the estimated frequency is consistent for k_1.

Proof of Theorem 2: Without loss of generality, we consider the single frequency model

$$y_t = \beta_0 + \gamma_{11}\sin(2\pi kt/T) + \gamma_{21}\cos(2\pi kt/T) + u_t \tag{A.2}$$

Generalizing the result to a multiple frequency model is straightforward. If $\alpha = 1$,

$$W_{\hat{\gamma}}(k) \Rightarrow [R(\textstyle\int_0^1 G_{1k}(r)G_{1k}'dr)^{-1}\int_0^1 G_{1k}(r)dW(r)]'[R(\int_0^1 G_{1k}(r)G_{1k}(r)'dr)^{-1}R']^{-1}$$
$$\times[R(\textstyle\int_0^1 G_{1k}(r)G_{1k}(r)'dr)^{-1}\int_0^1 G_{1k}(r)dW(r)] =^d \chi^2(2).$$

Substituting $G_{1k}(r) = [0, 2\pi k\cos(2\pi kr), -2\pi k\sin(2\pi kr)]'$ yields

$$\begin{aligned}
&\int_0^1 G_{1k}(r)G_{1k}(r)' dr \\
= &\begin{bmatrix} 0 & 0 & 0 \\ 0 & (2\pi k)^2 2\int_0^1 \cos(2\pi kr)^2 dr & -2\pi k\int_0^1 \sin(2\pi kr)\cos(2\pi kr)dr \\ 0 & -2\pi k\int_0^1 \sin(2\pi kr)\cos(2\pi kr)dr & (2\pi k)^2 \int_0^1 \sin(2\pi kr)^2 dr \end{bmatrix} \\
= &(2\pi k)^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & \int_0^1 \cos(2\pi kr)^2 dr & 0 \\ 0 & 0 & \int_0^1 \sin(2\pi kr)^2 dr \end{bmatrix}
\end{aligned}$$

and

$$\int_0^1 G_{1k}(r)dW(r) = 2\pi k \begin{pmatrix} 0 & \int_0^1 \cos(2\pi kr)dW(r) & -\int_0^1 \sin(2\pi kr)dW(r) \end{pmatrix}'.$$

Therefore,

$$W_{\hat{\gamma}}(k) \Longrightarrow g(k) = \frac{\left[\int_0^1 \sin(2\pi kr)dW(r)\right]^2}{\int_0^1 \sin(2\pi kr)^2 dr} + \frac{\left[\int_0^1 \cos(2\pi kr)dW(r)\right]^2}{\int_0^1 \cos(2\pi kr)^2 dr}.$$

Note that $\int_0^1 \sin(2\pi kr)dW(r)/[\int_0^1 \sin(2\pi kr)^2 dr]^{1/2}$ and $\int_0^1 \cos(2\pi kr)dW(r)/[\int_0^1 \cos(2\pi kr)^2 dr]^{1/2}$ are independent standard normal random variables. Therefore, the limiting distributions of sup-W and $Mean$-W are those of the supremum of a chi-square random variable with 2 degrees of freedom, respectively. On the other hand, if $|\alpha| < 1$

$$\begin{aligned}
W_{\hat{\gamma}}(k) \;\Rightarrow\; & [R(\textstyle\int_0^1 G_{0k}(r)G_{0k}(r)' dr)^{-1}\int_0^1 G_{0k}(r)dW(r)]'[R(\int_0^1 G_{0k}(r)G_{0k}(r)' dr)^{-1}R']^{-1} \\
& \times[R(\textstyle\int_0^1 G_{0k}(r)G_{0k}(r)' dr)^{-1}\int_0^1 G_{0k}(r)dW(r)] =^d \chi^2(2).
\end{aligned}$$

Substituting $G_{0k}(r) = [1, \sin(2\pi kr), \cos(2\pi kr)]'$ yields

$$\begin{aligned}
&\int_0^1 G_{0k}(r)G_{0k}(r)' dr \\
= &\begin{bmatrix} 1 & \int_0^1 \sin(2\pi kr)dr & \int_0^1 \cos(2\pi kr)dr \\ \int_0^1 \sin(2\pi kr)dr & \int_0^1 \sin(2\pi kr)^2 dr & \int_0^1 \sin(2\pi kr)\cos(2\pi kr)dr \\ \int_0^1 \cos(2\pi kr)dr & \int_0^1 \sin(2\pi kr)\cos(2\pi kr)dr & \int_0^1 \cos(2\pi kr)^2 dr \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \int_0^1 \sin(2\pi kr)^2 dr & 0 \\ 0 & 0 & \int_0^1 \cos(2\pi kr)^2 dr \end{bmatrix}$$

and

$$\int_0^1 G_{0k}(r)dW(r) = (\ 1 \quad \int_0^1 \sin(2\pi kr)dW(r) \quad \int_0^1 \cos(2\pi kr)dW(r)\ )'.$$

Therefore, $W_{\hat{\gamma}}(k) \Rightarrow g(k)$. This is the same functional form for the distribution as the one in the $I(1)$ case. This result implies sup-W and $Mean$-W have the same limiting distribution.

Appendix (b): Roy and Fuller's (2001) bias corrected estimator

It is well known that the OLS estimate of α is biased downward especially when α is near one. Hence, in many cases no truncation may apply when some would be desirable. We adopt the bias correction proposed by Roy and Fuller (2001). We consider here the one based on the OLS estimate (Roy et al., 2004). It is a function of a unit root test, namely the t-ratio $\hat{\tau} = (\hat{\alpha} - 1)/\hat{\sigma}_\alpha$, where $\hat{\alpha}$ is the OLS estimate and $\hat{\sigma}_\alpha$ its standard deviation. The bias-corrected estimate is given by

$$\hat{\alpha}_M = \hat{\alpha} + C(\hat{\tau})\hat{\sigma}_\alpha, \tag{A.3}$$

$$C(\hat{\tau}) = \begin{cases} -\hat{\tau} & \text{if} \quad \hat{\tau} > \tau_{pct} \\ I_p T^{-1}\hat{\tau} - (1+r)[\hat{\tau} + c_2(\hat{\tau} + a)]^{-1} & \text{if} \quad -a < \hat{\tau} \le \tau_{pct} \\ I_p T^{-1}\hat{\tau} - (1+r)\hat{\tau}^{-1} & \text{if} \quad -c_1^{1/2} < \hat{\tau} \le -a \\ 0 & \text{if} \quad \hat{\tau} \le -c_1^{1/2} \end{cases}$$

where $c_1 = (1 + r)T$ with r the number of parameters estimated in the trend function; $r = 1+2m$ for the constant only case and $r = 2+2m$ when a constant and trend are included, with m the number of frequencies included. Also, $c_2 = [(1 + r)T - \tau_{pct}^2(I_p + T)][\tau_{pct}(a + \tau_{pct})(I_p + T)]^{-1}$, a is some constant and τ_{pct} is a percentile of the limit distribution of $\hat{\tau}$ when

$\alpha = 1$. Also, $I_p = [(p+1)/2]$ where p is the order of the autoregressive process considered for the noise component. The parameters for which specific values need to be selected are τ_{pct} and a. Based on extensive simulation experiments, we selected $a = 10$ since it leads to tests with better properties, and for τ_{pct} we use $\tau_{0.85}$, the upper biased version whose values are presented in Table A-1. Hence, our suggested procedure involves the following steps: 1) Detrend the data by OLS to obtain residuals, say $\hat{u}_t$; 2) Estimate an $AR(p)$ for $\hat{u}_t$ yielding the estimate $\hat{\alpha}$ and the t-ratio $\hat{\tau}$; 3) Use $\hat{\alpha}$ and $\hat{\tau}$ to get the Roy and Fuller (2001) biased corrected estimates $\hat{\alpha}_M$; 4) Apply the truncation $\hat{\alpha}_{MS} = \hat{\alpha}_M$ if $|\hat{\alpha}_M - 1| > T^{-1/2}$ and 1 otherwise; 5) Apply the GLS procedure of Prais and Winsten (1954) with $\hat{\alpha}_{MS}$ to obtain the estimates of the coefficients of the trend and the estimate of the variance of the residuals and construct the standard Wald-statistic $W_{\widehat{\gamma}}(k)$; 6) Construct the sup-$W(\ell+1)$ or $Mean$-$W(\ell+1)$ tests as described in the text.

Using the biased corrected versions $\hat{\alpha}_M$, instead of the OLS estimates, does not change anything to the stated large sample results (Theorems 1 and 2). All that is needed for these asymptotic results to hold is that $T(\hat{\alpha}_M - 1) = O_p(1)$ when $\alpha = 1$, and $T^{1/2}(\hat{\alpha}_M - \alpha) \xrightarrow{d} N(0, 1-\alpha^2)$ when $|\alpha| < 1$. These conditions are satisfied. Roy et al. (2004) and Perron and Yabu (2009a) use a similar bias correction based on a weighted symmetric least-squares estimator of α instead of the OLS estimator. Both lead to tests with similar properties. However, note that the test proposed by Roy et al. (2004) has very different sizes in the I(0) and I(1) cases; see Perron and Yabu (2012) for details.

Table 1: Critical values of the sup-W and *Mean*-W tests with a constant

	$n-\ell$								
	1	2		3		4		5	
	$\chi^2(2)$	sup-W	$Mean$-W	sup-W	$Mean$-W	sup-W	$Mean$-W	sup-W	$Mean$-W
0.90	4.61	5.94	3.89	6.73	3.54	7.32	3.35	7.75	3.20
0.95	5.99	7.35	4.75	8.14	4.19	8.74	3.88	9.18	3.66
0.99	9.21	10.59	6.64	11.41	5.60	12.00	5.02	12.45	4.65

Note: The critical values can also be applied to the case with trending series; see the Supplement. They are the same in the I(1) case and nearly so in the I(0) case.

Table 2: Proportion of correct frequency selection with a constant, $T = 200$

(a) $\alpha = 1$

	OLS						DF						FGLS					
$\gamma \setminus k_1=$	1	2	3	4	5	Average	1	2	3	4	5	Average	1	2	3	4	5	Average
0	0.79	0.15	0.04	0.01	0.00	0.20	0.47	0.20	0.14	0.11	0.10	0.20	0.22	0.20	0.21	0.19	0.19	0.20
1	0.80	0.20	0.08	0.04	0.03	0.23	0.49	0.28	0.30	0.38	0.49	0.39	0.24	0.30	0.41	0.54	0.66	0.43
2	0.83	0.31	0.21	0.17	0.16	0.34	0.55	0.50	0.69	0.85	0.93	0.71	0.31	0.53	0.79	0.93	0.99	0.71
3	0.86	0.46	0.39	0.38	0.37	0.49	0.64	0.75	0.93	0.96	0.95	0.85	0.42	0.79	0.97	1.00	1.00	0.84
4	0.90	0.60	0.58	0.57	0.57	0.64	0.73	0.91	0.98	0.95	0.94	0.90	0.54	0.93	1.00	1.00	1.00	0.90
5	0.93	0.73	0.72	0.72	0.72	0.77	0.82	0.98	0.97	0.94	0.95	0.93	0.68	0.99	1.00	1.00	1.00	0.93

(b) $\alpha = 0.8$

	OLS						DF						FGLS					
$\gamma \setminus k_1 =$	1	2	3	4	5	Average	1	2	3	4	5	Average	1	2	3	4	5	Average
0.00	0.25	0.23	0.20	0.17	0.13	0.20	0.21	0.21	0.20	0.19	0.18	0.20	0.19	0.21	0.21	0.20	0.19	0.20
0.25	0.33	0.31	0.28	0.25	0.22	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.26	0.28	0.29	0.29	0.29	0.28
0.50	0.52	0.49	0.48	0.45	0.43	0.47	0.46	0.45	0.47	0.48	0.50	0.47	0.43	0.47	0.49	0.50	0.53	0.48
0.75	0.72	0.72	0.72	0.70	0.70	0.71	0.65	0.67	0.70	0.73	0.76	0.70	0.63	0.69	0.72	0.75	0.79	0.72
1.00	0.89	0.88	0.88	0.88	0.89	0.88	0.83	0.84	0.87	0.90	0.92	0.87	0.82	0.86	0.89	0.91	0.93	0.88
1.25	0.97	0.96	0.97	0.97	0.97	0.97	0.93	0.94	0.96	0.98	0.99	0.96	0.92	0.95	0.97	0.98	0.99	0.96

Table 3: Empirical applications to unemployment rates and temperature series

	Sample period	Model	Selected frequencies using the Mean-W test
(a) Unemployment rate series			
Canada	1983.1-2015.7	Constant	(2,4,5)
France	1983.1-2015.7	Constant	(1,3,5)
Germany	1991.1-2015.7	Constant	(1,2,3,4)
Italy	1983.1-2015.7	Trend	(1,2)
Japan	1983.1-2015.7	Constant	(1,4)
United Kingdom	1983.1-2015.7	Trend	(2,4,5)
United States	1983.1-2015.7	Constant	(1,2,3,4,5)
(b) Temperature series			
Global	1856-2010	Trend	(1,3)
Northern	1856-2010	Trend	(1)
Southern	1850-2010	Trend	(1,3)

Figure 1: Finite sample size

(a) $T = 200$ **(b) $T = 500$**

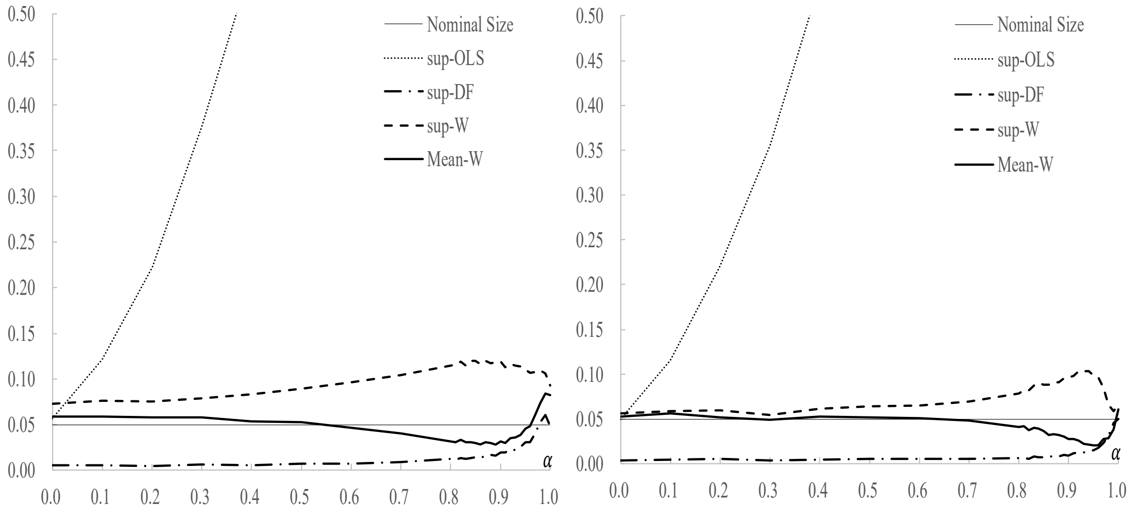


Figure 2: Finite sample power ($k_1 = 2,\ T = 500$)

(a) $\alpha = 1$ **(b) $\alpha = 0.8$**

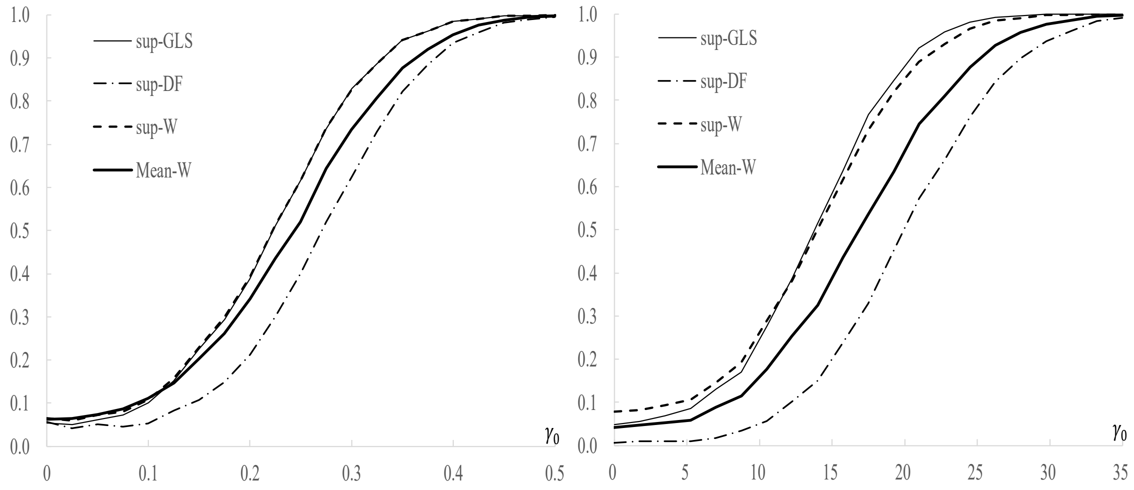


Figure 3: Probability of $k_1 = 2$ being selected ($T = 500$)

(a) $\alpha = 1$ **(b) $\alpha = 0.8$**

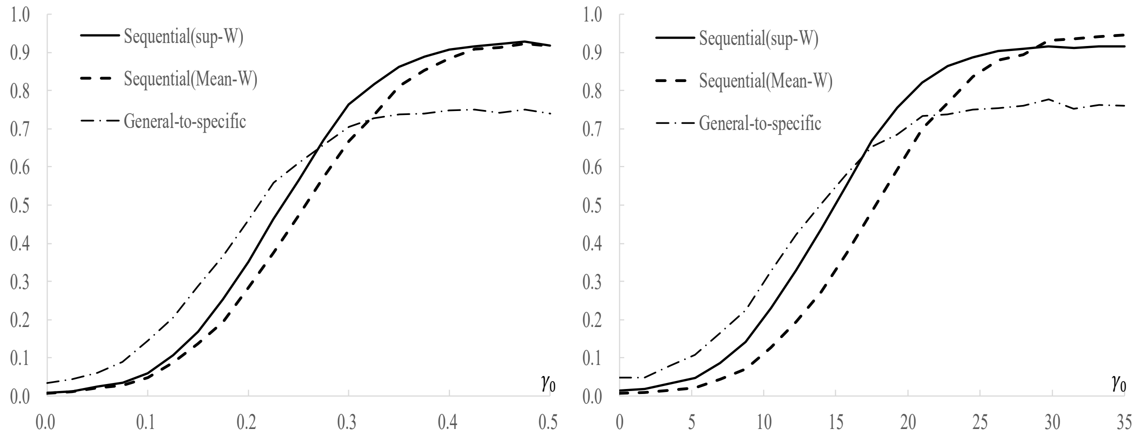


Figure 4: Probability of $k_1 = 2$ and $k_2 = 3$ being selected ($T = 500$)

(a) $\alpha = 1$ **(b) $\alpha = 0.8$**

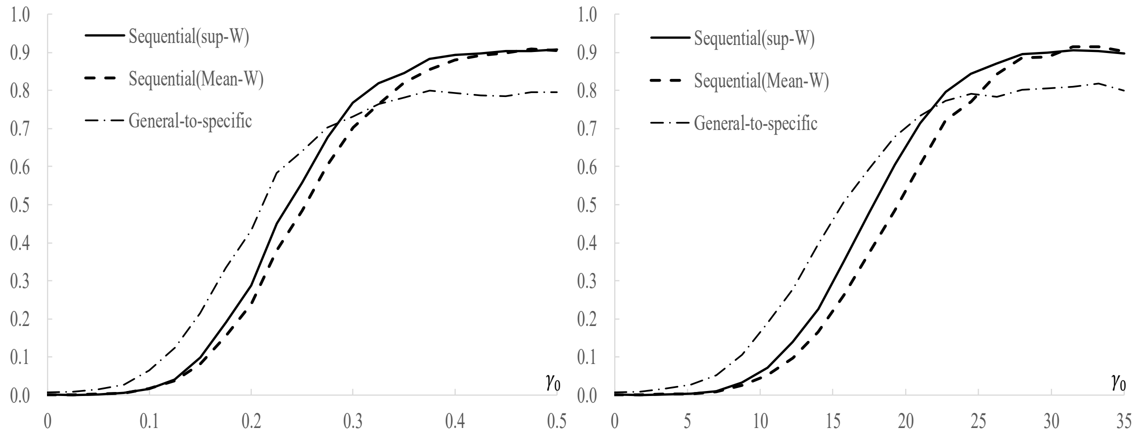
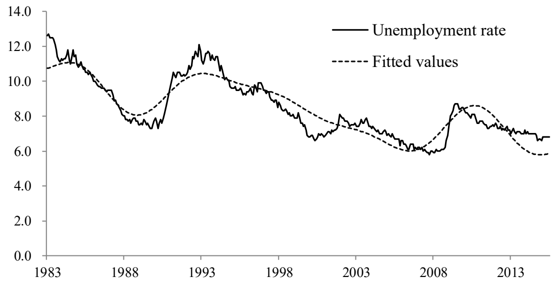
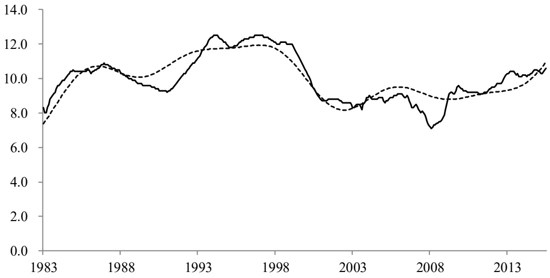


Figure 5: Monthly unemployment rate and fitted trigonometric trend for the G7 countries

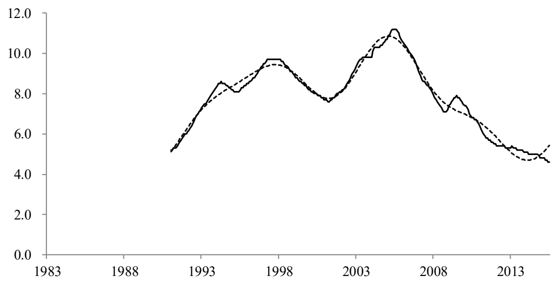
(a) Canada



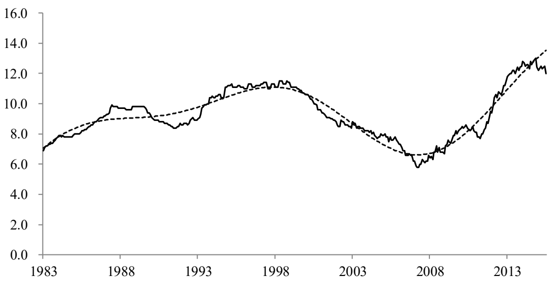
(b) France



(c) Germany

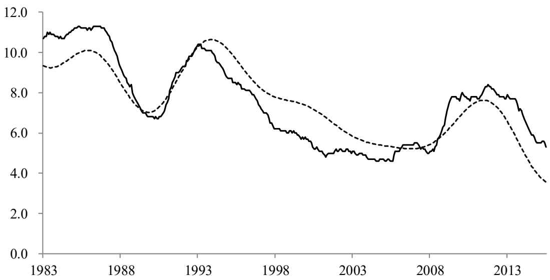


(d) Italy



(e) Japan

(f) United Kingdom



(g) United States

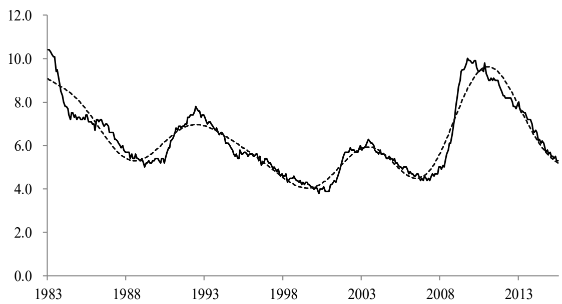
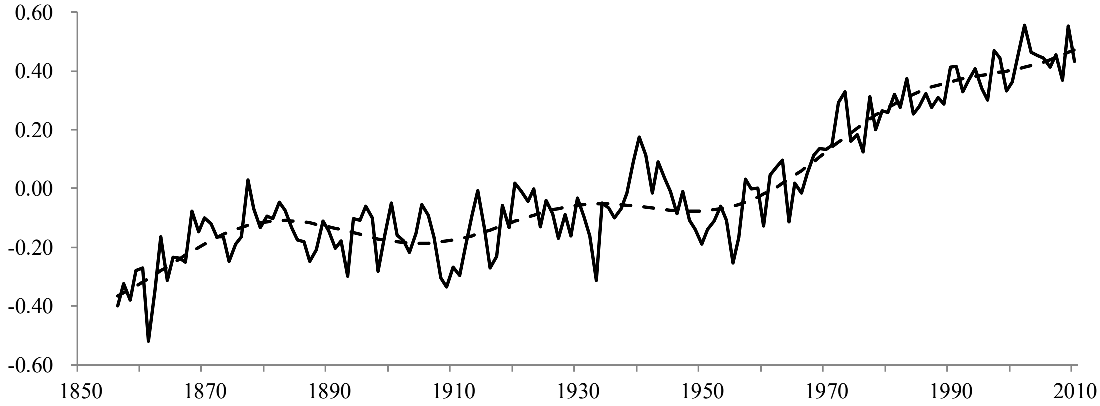
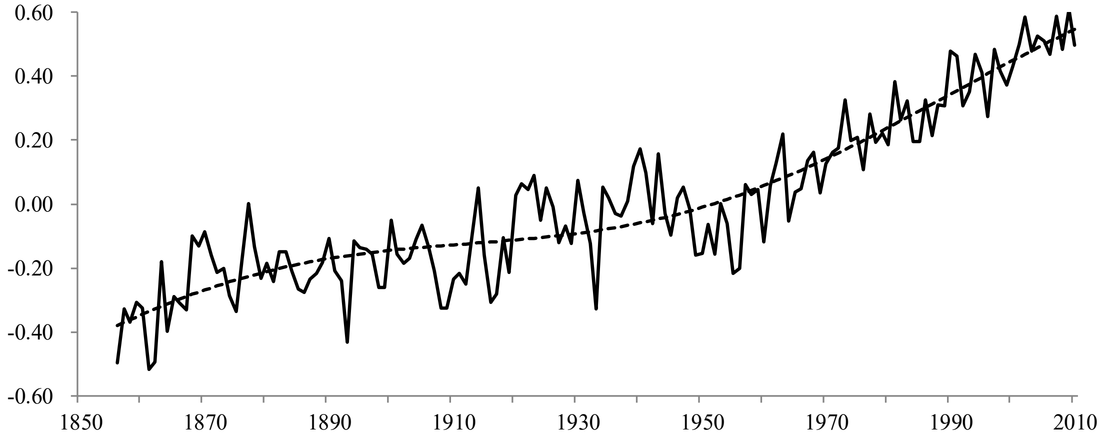


Figure 6: Temperature series and fitted trigonometric trend

(a) Global



(b) North Hemisphere



(c) South Hemisphere

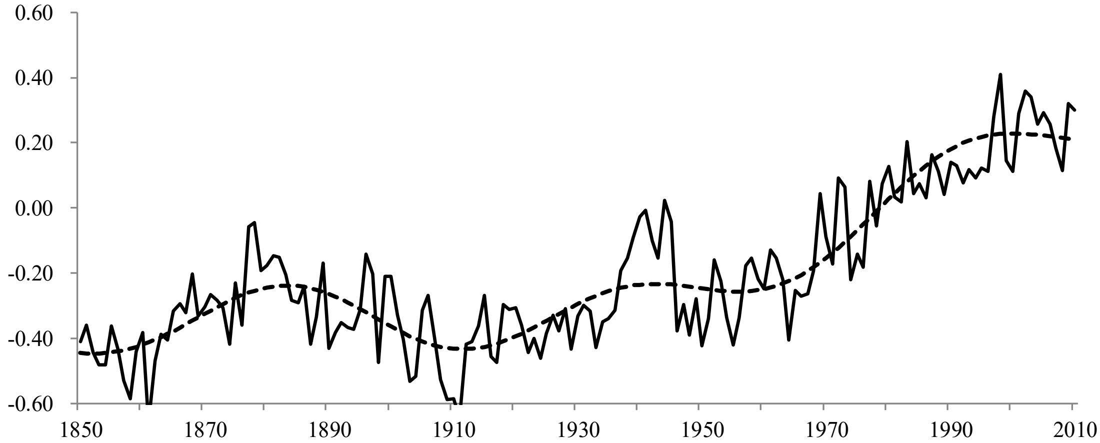


Table A-1: Values of $\tau_{0.85}$

1) With a constant only

a) Two frequencies (k_1, k_2)

$k_1 \setminus k_2$	2	3	4	5
1	-3.93	-3.63	-3.47	-3.39
2		-2.89	-2.78	-2.74
3			-2.58	-2.55
4				-2.49

b) Three frequencies (k_1, k_2, k_3)

k_1	$k_2 \setminus k_3$	3	4	5
1	2	-4.47	-4.28	-4.15
1	3		-3.91	-3.79
1	4			-3.61
2	3		-3.07	-3.02
2	4			-2.90
3	4			-2.67

c) Four frequencies (k_1, k_2, k_3, k_4)

k_1	$k_2 \setminus k_3, k_4$	3,4	3,5	4,5
1	2	-5.00	-4.84	-4.60
1	3			-4.16
2	3			-3.20

d) Five frequencies

$(k_1, k_2, k_3, k_4, k_5) = 1,2,3,4,5$

-5.48

2) With a constant and time trend

a) Two frequencies (k_1, k_2)

$k_1 \setminus k_2$	2	3	4	5
1	-4.51	-4.30	-4.15	-4.04
2		-3.90	-3.72	-3.64
3			-3.44	-3.36
4				-3.22

b) Three frequencies (k_1, k_2, k_3)

k_1	$k_2 \setminus k_3$	3	4	5
1	2	-5.11	-4.95	-4.84
1	3		-4.71	-4.59
1	4			-4.40
2	3		-4.28	-4.15
2	4			-3.94
3	4			-3.60

c) Four frequencies (k_1, k_2, k_3, k_4)

k_1	$k_2 \setminus k_3, k_4$	3,4	3,5	4,5
1	2	-5.63	-5.52	-5.34
1	3			-5.09
2	3			-4.61

d) Five frequencies

$(k_1, k_2, k_3, k_4, k_5) = 1,2,3,4,5$

-6.10