The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Astill, Sam and Harvey, David I. and Leybourne, Stephen J. and Taylor, A. M. Robert (2014) Robust and powerful tests for nonlinear deterministic components. Oxford Bulletin of Economics and Statistics, 77 (6). pp. 780-799. ISSN 1468-0084

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/32658/1/nonlinear%20rev2.pdf

Robust and powerful tests for nonlinear deterministic components[*]

Sam Astill[a], David I. Harvey[b], Stephen J. Leybourne[b] and A.M. Robert Taylor[c]

a. Department of Economics, University of Warwick.

b. Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham.

c. Essex Business School, University of Essex.

May 2014

Abstract

We develop a test for the presence of nonlinear deterministic components in a univariate time series, approximated using a Fourier series expansion, designed to be asymptotically robust to the order of integration of the process and to any weak dependence present. Our approach is motivated by the Wald-based testing procedure of Harvey, Leybourne and Xiao (2010) [Journal of Time Series Analysis, vol. 31, p.379-391], but uses a function of an auxiliary unit root statistic to select between the asymptotic I(0) and I(1) critical values, rather than modifying the Wald test statistic as in Harvey *et al.*. We show that our proposed test has uniformly greater local asymptotic power than the test of Harvey *et al.* when the shocks are I(1), identical local asymptotic power when the shocks are I(0), and also improved finite sample properties. We also consider the issue of determining the number of Fourier frequencies used to specify any nonlinear deterministic components, evaluating the performance of algorithmic- and information criterion-based model selection procedures.

Keywords: Trend function testing; Robust tests; Fourier approximation.
JEL Classification: C22.

1   Introduction

The ability to detect the presence and magnitude of deterministic components in a financial or economic time series is of key importance when conducting empirical analysis, particularly for the purposes of forecasting and testing for a unit root. For example, in the latter case, failure to correctly specify a relevant deterministic component present in the data is known to result in non-similar and (usually) inconsistent tests. Moreover, the power of unit root tests to reject the null under the I(0) alternative when deterministic components are unnecessarily included in a model specification is also reduced.

Traditionally, attention in the literature has focused on a linear deterministic component, most often the case of a constant and/or linear trend. The possibility of breaks in such linear deterministics

---

has also received considerable attention, due in particular to the effect these breaks have on standard unit root and stationarity tests; see, *inter alia*, Perron (1998). However, as noted by Harvey, Leybourne and Xiao (2010) [HLX hereafter], these may not necessarily take the form of an instantaneous break in the trend function. They note that changes in economic aggregates are affected by the response of a large number of potentially heterogeneous individuals who are unlikely to respond instantaneously to shocks. As a consequence, they suggest that a smoothly evolving nonlinear deterministic component might provide a better approximation to the underlying deterministic component of these aggregates.

One possible method to capture such nonlinear behaviour is to approximate the deterministic component using a Fourier series expansion. This approach is explored by Becker *et al.* (2004) and is found to provide a good approximation for a variety of functions. Enders and Lee (2012) show that modelling the deterministic components of an economic time series via a Fourier function can approximate changes of various forms, such as a number of sharp breaks or deterministic smooth transitions, e.g. exponential or logistic smooth transtions (ESTR or LSTR). Becker *et al.* (2004) derive a test for the presence of nonlinear deterministic components using a Fourier expansion under the assumption that the shocks are I(0). Enders and Lee (2012) propose a corresponding test under the assumption that the data are I(1). The practical implementation of these tests is clearly problematic since both assume the order of integration of the data to be known. This results in a circular testing problem as we would need to know the order of integration of our data before performing the tests of either Becker *et al.* (2004) and Enders and Lee (2012); however, in order to perform a unit root or stationarity test we would need to specify the form of the deterministic component.

Motivated by this problem, and drawing on the robust linear trend tests of Vogelsang (1998), HLX suggest a test that is robust to the order of integration of the data, thereby eliminating this circular testing problem. This is achieved by using a composite statistic based around a Wald statistic (that has a well defined limit distribution for both I(0) and I(1) shocks) multiplied by a function of an auxiliary unit root test statistic. This function is specified such that when the shocks are I(0) it converges in probability to one, leaving the asymptotic distribution of the Wald statistic unaffected, but when the shocks are I(1), it converges to a well-defined limit distribution. Judicious choice of the precise function to be used then allows the asymptotic critical values of the composite test statistic to be lined up in the I(0) and I(1) environments, for a given significance level. This approach therefore yields tests that display correct asymptotic size for both I(0) and I(1) shocks.

Here we propose an alternative to the methodology of HLX. In our approach a function of an auxiliary unit root test statistic is used to select between the asymptotic I(0) and I(1) *critical values* for the Wald test, rather than creating a composite test via multiplication with the Wald statistic as in HLX. The motivation underlying this is that the presence of the multiplicative function of a unit root test statistic in the HLX procedure impacts negatively on the local asymptotic power when the shocks are I(1), relative to a test that compares the unmodified Wald statistic with its asymptotic I(1) critical value directly. In contrast, the approach considered in this paper always uses the Wald statistic without modification, and ensures that the correct asymptotic critical value is used, appropriate to the order of integration. We show that this new procedure achieves the same local asymptotic power

as the HLX test in the I(0) setting, while delivering (often substantial) local asymptotic power gains for I(1) shocks. The new approach also provides improved finite sample behaviour.

The paper is organized as follows. Section 2 outlines the model and the hypotheses of interest, with the HLX procedure reviewed in section 3. Section 4 presents our new approach, with the limiting null distribution and local asymptotic power of the tests detailed in section 5. Section 6 addresses issues related to the practical implementation of the new procedure, and section 7 reports results of Monte Carlo simulations to assess the finite sample properties of the proposed tests for a given number of frequencies in the Fourier expansion. In section 8 we examine the issue of selecting the number of frequencies to include in the Fourier expansion, and compare the performance of a selection algorithm based on the proposed tests with one based on the tests of HLX and also an information criterion approach. Section 9 reports results from an empirical study where the procedures considered in this paper are applied to a number of financial volatility indices, while section 10 concludes.

2   The Model and Testing Hypotheses

Following HLX, we consider a sample of T observations generated according to the following data generating process (DGP)

$$y_t = d_t + \sum_{f=1}^{n} \gamma_{1f,T} \sin\left(\frac{2\pi f t}{T}\right) + \sum_{f=1}^{n} \gamma_{2f,T} \cos\left(\frac{2\pi f t}{T}\right) + u_t, \quad t = 1, ..., T. \tag{1}$$

The nonlinear deterministic component of y_t is specified using a Fourier series expansion with the maximum number of frequencies contained in the expansion denoted by n, and $f \in Z^+$ denoting a particular frequency. A linear deterministic component is contained in d_t, the two leading cases of which are $d_t = \alpha$ (a constant) and $d_t = \alpha + \beta t$ (a constant and a linear trend).

We allow the stochastic component, u_t, to be either I(0) or I(1), satisfying either Assumption I(0) or Assumption I(1), respectively, below.

Assumption I(0)   *The stochastic process u_t is such that $u_t = v_t$, where the process $\{v_t\}$ satisfies $v_t = C(L)\varepsilon_t$, $C(L) := \sum_{i=0}^{\infty} C_i L^i$, $C(1)^2 > 0$ with $\sum_{i=0}^{\infty} i|C_i| < \infty$, and where ε_t is a mean zero i.i.d. sequence with $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and finite fourth moment. The long run variance of v_t is defined as $\omega_v^2 := \lim_{T\to\infty} T^{-1} E(\sum_{t=1}^{T} v_t)^2 = \sigma_\varepsilon^2 C(1)^2$.*

Assumption I(0) ensures that we can apply a Functional Central Limit Theorem (FCLT) to the partial sums of $\{v_t\}$; i.e., $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} v_t \xrightarrow{d} \omega_v W(r)$ where $\lfloor . \rfloor$ denotes the integer part of its argument, $\xrightarrow{d}$ denotes weak convergence and $W(r)$ is a standard Wiener process.[1]   When the process is I(0) we make the following assumption on the Fourier coefficients, $\gamma_{1f,T} := \gamma_{1f} \omega_v T^{-1/2}$, $\gamma_{2f,T} := \gamma_{2f} \omega_v T^{-1/2}$, $f = 1, ..., n$. The $T^{-1/2}$-scalings on the Fourier coefficients are analytical devices that provide the appropriate Pitman drifts for the local asymptotic power analysis that follows. The scaling by ω_v is a

---

[1] We maintain the assumption that ε_t is i.i.d. for consistency with the assumptions in HLX, however, the asymptotic results that follow would continue to hold if we relax this assumption to allow ε_t to be a martingale difference sequence.

convenience device that allows the long run variance to be factored out of the limit distributions that arise for the test statistics outlined in this paper.

Assumption I(1) *The stochastic process u_t is such that $\Delta u_t = v_t$, where the process $\{v_t\}$ is as defined in Assumption I(0).*

When the process is I(1) we assume that the Fourier coefficients satisfy, $\gamma_{1f,T} := \gamma_{1f}\omega_v T^{1/2}$, $\gamma_{2f,T} := \gamma_{2f}\omega_v T^{1/2}$, $f = 1, ..., n$, with the $T^{1/2}$-scalings now representing the appropriate Pitman drifts.

In the context of (1), when testing for the presence of nonlinear deterministic components our null and alternative hypotheses are given by

$$H_0 : \gamma_{1f,T} = \gamma_{2f,T} = 0, \quad f = 1, ..., n$$
$$H_1 : \text{at least one of } \gamma_{1f,T}, \gamma_{2f,T} \neq 0, \quad f = 1, ..., n.$$

Under H_0 the deterministic component in (1) reduces to d_t, while H_1 specifies that some form of nonlinear deterministic component is present in the data. The formulation of H_1 allows for the case where $\gamma_{1f,T} = \gamma_{2f,T} = 0$ for some but not all frequencies $f = 1, ..., n$.

3 The HLX Test

In order to construct tests that are robust to whether the shocks are I(0) or I(1), HLX first consider the partially summed counterpart to regression (1)

$$z_t = \sum_{s=1}^{t} d_s + \sum_{f=1}^{n} \gamma_{1f,T} \sum_{s=1}^{t} \sin\left(\frac{2\pi f s}{T}\right) + \sum_{f=1}^{n} \gamma_{2f,T} \sum_{s=1}^{t} \cos\left(\frac{2\pi f s}{T}\right) + \eta_t \tag{2}$$

where $z_t := \sum_{s=1}^{t} y_s$ and $\eta_t := \sum_{s=1}^{t} u_s$. They then consider a scaled Wald statistic for H_0 against H_1 based on (2), of the form $SW_0^n := (RSS_R - RSS_U)/RSS_U$, where RSS_R denotes the residual sum of squares from OLS regression of z_t on $\sum_{s=1}^{t} d_s$, and RSS_U denotes the residual sum of squares from OLS estimation of the unrestricted regression (2). The notation SW_0^n indicates we are testing the null of zero frequencies against the alternative of up to n frequencies.

The limit distribution of SW_0^n depends on whether Assumption I(0) or Assumption I(1) holds, but crucially it is well defined in each case. Consequently, HLX propose a Vogelsang (1998)-type modification, basing their recommended test on the modified statistic $MW_0^n := SW_0^n \exp(-b_\xi |DF|^{-1})$, where b_ξ is a finite positive constant and DF is the Dickey-Fuller t-statistic applied to the residuals $\hat{u}_t$ obtained from OLS estimation of (1), i.e. $DF = (\hat{\phi} - 1)/s.e.(\hat{\phi})$ with $\hat{\phi}$ obtained from the OLS regression

$$\hat{u}_t = \phi \hat{u}_{t-1} + \sum_{i=0}^{p} \zeta_i \Delta \hat{u}_{t-i} + e_t. \tag{3}$$

The $|DF|^{-1}$ statistic has the property that it is $O_p(1)$ when the shocks are I(1), since DF has a limit distribution (U, say), and converges in probability to zero when the shocks are I(0). Defining

the limiting null distributions of SW_0^n under I(0) and I(1) shocks as D_0 and D_1, respectively, it then follows that, under H_0, $MW_0^n \xrightarrow{d} D_0$ under Assumption I(0) and $MW_0^n \xrightarrow{d} D_1 \exp(-b_\xi |U|^{-1})$ under Assumption I(1). The constant b_ξ is chosen such that, for a given significance level ξ, the MW_0^n test has the same asymptotic critical value under I(0) and I(1) shocks. As a result, the MW_0^n test will have the correct size asymptotically regardless of whether Assumption I(0) or Assumption I(1) holds.

4   An Alternative Robust Test

Our proposed testing procedure involves utilising the same scaled Wald statistic as HLX, namely SW_0^n, but we propose using an auxiliary unit root statistic (denoted by J) to switch between the critical values appropriate for SW_0^n under Assumption I(0) and Assumption I(1), i.e. critical values from D_0 and D_1, respectively, rather than modifying the test statistic itself with the multiplicative term $\exp(-b_\xi |DF|^{-1})$ as in HLX. Such an approach has the advantage that it leads to a test with local asymptotic power identical to the tests of HLX when the shocks are I(0), but with greater local asymptotic power when the shocks are I(1) due to the removal of the influence of the auxiliary unit root test statistic on the asymptotic distribution of the test in the I(1) case.

Denoting by $cv_{0,\xi}$ and $cv_{1,\xi}$ the ξ-level critical values from D_0 and D_1, respectively, we wish to compare SW_0^n with $cv_{0,\xi}$ when the shocks are I(0), and SW_0^n with $cv_{1,\xi}$ when the shocks are I(1). To achieve this, we consider the following adaptive critical value

$$cv_\xi^\lambda := \lambda_{J,\xi} cv_{0,\xi} + (1 - \lambda_{J,\xi}) cv_{1,\xi}, \qquad \lambda_{J,\xi} := \exp(-\tau_\xi T^\delta J) \tag{4}$$

with τ_ξ and δ positive constants. We require a value of δ and a statistic J such that $T^\delta J$ converges in probability to zero under Assumption I(0) but diverges to positive infinity under Assumption I(1). Then, when the shocks are I(0), $\lambda_{J,\xi} \xrightarrow{p} 1$ and $cv_\xi^\lambda \xrightarrow{p} cv_{0,\xi}$, and when the shocks are I(1), $\lambda_{J,\xi} \xrightarrow{p} 0$ and $cv_\xi^\lambda \xrightarrow{p} cv_{1,\xi}$. This ensures that the correct critical value is used asymptotically, yielding a robust testing approach. Note that although the constant τ_ξ will play a role in the finite sample behaviour of $\lambda_{J,\xi}$, it is asymptotically irrelevant. In what follows, we denote by ASW_0^n our proposed test that compares the statistic SW_0^n with the adaptive critical value cv_ξ^λ, given suitable choices of δ and J that satisfy our requirements on $T^\delta J$.

5   Asymptotic Results

Defining $\boldsymbol{\gamma} := [\gamma_{11}, ..., \gamma_{1n}, \gamma_{21}, ..., \gamma_{2n}]'$, $X_k(r) := [m_{11}(r), ..., m_{1k}(r), m_{21}(r), ..., m_{2k}(r)]'$, and $H_k(r) := [F(r), \sin(2\pi r), ..., \sin(2\pi k r), \cos(2\pi r), ..., \cos(2\pi k r)]'$, where $F(r) := 1$ if $d_t = \alpha$, $F(r) := (1, r)$ if $d_t = \alpha + \beta t$, $m_{1f}(r) := \frac{1}{2\pi f}(1 - \cos(2\pi f r))$ and $m_{2f}(r) := \frac{1}{2\pi f} \sin(2\pi f r)$, the following large sample results obtain (see HLX for the limits of MW_0^n; the limits for SW_0^n follow in a straightforward way). If y_t is generated by (1) under H_1, then

(a) Under Assumption I(0), $SW_0^n, MW_0^n \xrightarrow{d} \dfrac{\int_0^1 L_R(r, \boldsymbol{\gamma})^2 dr}{\int_0^1 L_U(r)^2} - 1 =: D_0(\boldsymbol{\gamma})$

where $L_R(r, \boldsymbol{\gamma})$ denotes the continuous time residuals from the projection of $\boldsymbol{\gamma}' X_n(r) + W(r)$ onto the space spanned by $\int_0^r F(s)ds$, and $L_U(r)$ denotes the continuous time residuals from the projection of $W(r)$ onto the space spanned by $[\int_0^r F(s), X_n(r)']$; and

$$(b) \text{ Under Assumption I(1),} \quad SW_0^n \xrightarrow{d} \frac{\int_0^1 N_R(r, \boldsymbol{\gamma})^2 dr}{\int_0^1 N_U(r)^2} - 1 =: D_1(\boldsymbol{\gamma})$$

$$MW_0^n \xrightarrow{d} D_1(\boldsymbol{\gamma}) \exp(-b_\xi |U|^{-1})$$

where $N_R(r, \boldsymbol{\gamma})$ denotes the continuous time residuals from the projection of $\boldsymbol{\gamma}' X_n(r) + \int_0^r W(s)ds$ onto the space spanned by $\int_0^r F(s)ds$, $N_U(r)$ denotes the continuous time residuals from the projection of $\int_0^r W(s)ds$ onto the space spanned by $[\int_0^r F(s), X_n(r)']$, and $U := (K(1)^2 - K(0)^2 - 1)/(2\sqrt{\int_0^1 K(r)^2 dr})$, with $K(r)$ the continuous time residuals from the projection of $W(r)$ onto the space spanned by $H_n(r)'$.

The asymptotic null distributions of SW_0^n and MW_0^n obtain by setting $\boldsymbol{\gamma} = 0$ in the foregoing representations, so that, linking with the notation of the previous section, $D_0 \equiv D_0(\mathbf{0})$ and $D_1 \equiv D_1(\mathbf{0})$. Note that D_0 and D_1 are functions of n, and thus the limit distribution of SW_0^n depends on the choice of n. The asymptotic critical values of the with-constant SW_0^n ($d_t = \alpha$) and with-trend SW_0^n ($d_t = \alpha + \beta t$) statistics, for $n = 1, 2, 3$ and for both I(0) and I(1) shocks, are given in Table 1. These were obtained from direct simulation of these limiting distributions, with the Wiener process approximated using $NIID(0, 1)$ random variates, and with the integrals approximated by normalized sums based on 1,000 steps. Here and throughout the paper, all Monte Carlo simulations were conducted in Gauss 9.0 using 50,000 replications. The critical values reported for the I(0) case are also the limit critical values of MW_0^n, and coincide with the values reported in Table 1 of HLX.

Under Assumption I(0), both the MW_0^n and ASW_0^n tests share the same asymptotic distribution and, as such, will possess identical local asymptotic power functions. Under Assumption I(1), however, their local asymptotic power functions differ. Henceforth, we will concentrate attention on the with-constant ASW_0^n test to facilitate direct comparison with the results presented in HLX. Corresponding results for the with-trend variant are available on request. Figure 1 reports the local asymptotic power (at the nominal 0.05 significance level) under both Assumption I(0) and Assumption I(1) for $n = 1, 2, 3$, with the results again obtained from direct simulation of the limiting distributions. Due to the multi-parameter nature of the testing problem we present results under the alternative that $\gamma_{11} = ... = \gamma_{1n} = \gamma_{21} = ... = \gamma_{2n} = \gamma$. Panels (a), (c) and (e) confirm that the local asymptotic powers of both tests are identical when the shocks are I(0). However, when the shocks are I(1), Panels (b), (d) and (f) reveal a distinct power ordering between the two test procedures across all values of n. The new ASW_0^n test has uniformly superior local asymptotic power to the MW_0^n test, with the potential power gains being quite substantial. Indeed, we find that the maximum power gains across γ afforded by using ASW_0^n rather than MW_0^n are 0.28, 0.30 and 0.31 for $n = 1$, $n = 2$ and $n = 3$, respectively.

6 Practical Implementation of the Test Procedure

To implement the ASW_0^n test, we require choices for δ and J which satisfy the conditions on $T^\delta J$ given below (4). We followed HLX and experimented with functions of the Dickey-Fuller-type statistic, DF, and also a Breitung (2002)-type variance ratio unit root statistic

$$B := s_u^{-2} T^{-3} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \hat{u}_s \right)^2 \tag{5}$$

where $s_u^2 := T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$, with $\hat{u}_t$ the residuals obtained from OLS estimation of (1). While all choices of δ and J that satisfy the required conditions on $T^\delta J$ result in asymptotically equivalent test procedures, we found the best overall finite sample behaviour was achieved when setting $\delta = 1/2$ and $J = B$. With these choices, following results in Breitung (2002), under Assumption I(0) we find $B = O_p(T^{-1})$ and so $T^{1/2} J = O_p(T^{-1/2})$, while under Assumption I(1), $B = O_p(1)$ with $T^{1/2} J = O_p(T^{1/2})$, clearly satisfying the required conditions on $T^\delta J$. We recommend these settings for δ and J in the implementation of the test.

While the adaptive critical value cv_ξ^λ of (4) delivers an appropriate critical value for SW_0^n asymptotically, we also consider a finite sample adjustment that proves to be beneficial in controlling the new test's size in small samples. Results from unreported simulations showed that the procedure outlined thus far has a tendency to exhibit finite sample over-size when the shocks are I(1), and finite sample under-size when the shocks are I(0). We therefore consider a modification designed to inflate (deflate) the finite sample critical value in the I(1) (I(0)) case. Specifically, we consider the following adjusted adaptive critical value in place of (4)

$$cv_\xi^{\lambda,adj} := \lambda_{J,\xi}(1 - \kappa_\xi \lambda_{J,\xi} T^{-1/2}) cv_{0,\xi} + (1 - \lambda_{J,\xi})\{1 + \kappa_\xi(1 - \lambda_{J,\xi}) T^{-1/2}\} cv_{1,\xi} \tag{6}$$

where $\kappa_\xi > 0$. Note that this has no effect asymptotically as $\kappa_\xi \lambda_{J,\xi} T^{-1/2} \xrightarrow{p} 0$ under I(0) and $\kappa_\xi(1 - \lambda_{J,\xi}) T^{-1/2} \xrightarrow{p} 0$ under I(1).

Although κ_ξ and τ_ξ in (6) and (4) have no impact on the asymptotic behaviour of ASW_0^n, specific values of these parameters are required to implement the test in practice, and the choice of these values affects the test's finite sample properties. We calibrated these choices according to a set of unreported size and power simulations for $n = 1, 2, 3$. As a starting point, for a given n, we restricted attention to the set of pairings of (κ_ξ, τ_ξ) which delivered a correct empirical size of exactly ξ (for each of $\xi = 0.10$, 0.05 and 0.01) for the pure unit root case $\Delta y_t = \varepsilon_t \sim NIID(0,1)$, $y_1 = \varepsilon_1$, and for the sample size $T = 300$. Among such pairings, increases in τ_ξ (with corresponding decreases in κ_ξ) were found to reduce the degree of any over-size displayed in stationary scenarios, but at the cost of decreased finite sample power. We first chose (κ_ξ, τ_ξ) such that the test also had empirical size equal to ξ when $y_t = \varepsilon_t \sim NIID(0,1)$ and $T = 300$. In some cases, however, we found that this choice led to significant size distortions for moving average ε_t; in such cases, we selected a (κ_ξ, τ_ξ) pairing with a larger τ_ξ (and therefore smaller κ_ξ) to reduce the size distortions. Specifically, we chose a pairing such that, for $T = 300$, the empirical size lay below 0.07 for nominal 0.05-level tests (and below 0.14 and

0.014 for nominal 0.10- and 0.01-level tests, respectively) across the ARMA(1,1) and ARIMA(0,1,1) simulation settings considered in section 7.1 below. We advocate tolerating this modest amount of potential over-size in order to preserve decent finite sample power levels. The chosen pairings are given in Table 2, and we recommend these settings for practical applications of the test.

7   Finite Sample Simulations

7.1   Empirical Size

In this section we consider the finite sample size behaviour of the MW_0^n and ASW_0^n tests, focusing on the case $n = 1$. We generate data according to the following DGP, which allows for stationary ARMA(1,1) and integrated ARIMA(0,1,1) shocks: $y_t = \phi y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$, $t = 2, ..., T$, with $u_1 = \varepsilon_1$ and $\varepsilon_t \sim NIID(0,1)$. Table 3 reports the empirical sizes of nominal 0.05-level MW_0^1 and ASW_0^1 tests for the sample sizes $T = \{150, 300\}$ and serial correlation parameters $\phi = \{0, 0.5, 0.7, 0.9, 0.95, 1\}$ and $\theta = \{-0.5, 0, 0.5\}$. When calculating the DF unit root test statistic required for the MW_0^1 procedure, the lag truncation parameter p in (3) was chosen using the modified Akaike information criterion (MAIC) of Ng and Perron (2001) with $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$, and using the modification of Perron and Qu (2007), as outlined in HLX.

The finite sample sizes of the two tests follow broadly similar patterns across the different serial correlation parameter settings. Both are close to nominal size for I(1) shocks ($\phi = 1$), apart from some over-size observed for ASW_0^1 when $T = 150$. At the other extreme, when $\phi = 0$, we see that in the absence of moving average components ($\theta = 0$), the with-constant ASW_0^1 test is approximately correctly sized and the with-trend ASW_0^1 test is a little under-sized, while some modest size distortions are observed when $\theta \neq 0$. On the other hand, the MW_0^1 tests are severely under-sized in all cases when $\phi = 0$. For stationary but autocorrelated DGPs ($0 < \phi < 1$) all tests can be substantially under-sized in finite samples. As was observed by HLX, for some values of ϕ the tests are more under-sized for $T = 300$ than for $T = 150$; unreported simulations confirm, however, that this phenomenon eventually vanishes for much larger sample sizes, in line with our asymptotic results. The results of Table 3 show that the new ASW_0^n tests are generally conservative, and are therefore unlikely to spuriously signal the presence of nonlinear deterministic components when they are in fact absent. Furthermore, while under-size is apparent for stationary shocks, the degree of this downward size distortion is less marked for the ASW_0^n tests than for the MW_0^n tests.

7.2   Empirical Power

To examine the finite sample power properties of the tests, we generate data according to the DGP

$$y_t = \gamma \sum_{f=1}^{n} \sin\left(\frac{2\pi f t}{T}\right) + \gamma \sum_{f=1}^{n} \cos\left(\frac{2\pi f t}{T}\right) + u_t, \quad t = 1, ..., T \tag{7}$$

$$u_t = \phi u_{t-1} + \varepsilon_t, \quad t = 2, ..., T \tag{8}$$

with $n = 1$, $u_1 = \varepsilon_1$ and $\varepsilon_t \sim NIID(0, 1)$. Figure 2 presents power curves for nominal 0.05-level with-constant MW_0^1 and ASW_0^1 tests, for $T = 150$ and $\phi = \{0, 0.5, 0.7, 0.9, 0.95, 1\}$. The power curves were computed using a grid of 50 steps of values for γ from to 0 to $\gamma_{\max}$, with $\gamma_{\max} = \{2, 4, 8, 20, 20, 100\}$, corresponding to the six values of ϕ considered.

Figures 2(f) and 3(f) show that when the shocks are I(1), the power of the new ASW_0^1 test is clearly superior to that of MW_0^1, in line with the asymptotic power results in Figure 1. It is reassuring to see that the power gains observed in the limit are also manifest in finite samples, with quite substantial power advantages available through use of ASW_0^1 in the I(1) setting. When the shocks are I(0), the two tests are asymptotically equivalent, although it is to be expected that their power properties will differ in finite samples, particularly given the differential finite sample size results discussed above. Indeed, Figures 2(a)-2(e) show that differences between the ASW_0^1 and MW_0^1 power curves do occur. The main observation is that ASW_0^1 generally outperforms MW_0^1 under I(0) shocks. The power gains associated with ASW_0^1 are most marked when $\phi = 0$, where it is evident that the under-size associated with MW_0^1 in this case has a detrimental impact on power relative to the better sized ASW_0^1 test. A similar, albeit less exaggerated, pattern is seen when $\phi = 0.5$, while there is little to choose between the two tests when $\phi = 0.7$ and $\phi = 0.9$. Finally, when $\phi = 0.95$, relative power gains are again displayed by ASW_0^1. In summary, the ASW_0^1 test offers valuable improvements in finite sample power relative to MW_0^1, both in the case of I(1) shocks (as would be expected) and also in the case of I(0) shocks where the tests are asymptotically equivalent.

As suggested by a referee, an alternative to using $\hat{u}_t$ in the construction of B would be to instead use residuals obtained from (1) but with the null hypothesis imposed, i.e. to use residuals from OLS estimation of $y_t = d_t + u_t$, $t = 1, ..., T$. It can be shown that $cv_\xi^{\lambda, adj}$ computed using B based on such restricted residuals has exactly the same asymptotic properties as $cv_\xi^{\lambda, adj}$ in (6) under Assumption I(0) and Assumption I(1), both under the null and under the respective local alternative hypotheses. Unreported simulations show that this approach tends to be susceptible to a greater degree of under-size than our suggested procedure, while neither procedure dominates the other across ϕ in terms of finite sample power. As a result, we do not pursue this alternate approach further here.

8   Determining The Number Of Frequencies

The analysis in the previous section assumed that the true maximum number of frequencies, n, was known. In practice, however, this setting is unknown and must be specified by the practitioner. Unreported simulation results show that incorrectly specifying the number of frequencies, n, to include in the testing procedure has a detrimental effect on the power of all of the tests considered in this paper. For instance, if we choose to perform a test for at most $n + 1$ frequencies when the true number of frequencies is n we will sacrifice power due to over-specification in the test procedure; similarly, under-specification of n can result in tests with very low power.

HLX attempt to overcome this problem by presenting an algorithm for determining the number

of frequencies to include in their testing procedure. This algorithm can also be applied to the new testing approach proposed in this paper. As part of the algorithm, HLX develop a test of the null of at most $m-1$ frequencies versus m frequencies, i.e., within the context of (1),

$$H_0^* : \ \gamma_{1m,T} = \gamma_{2m,T} = 0; \quad \gamma_{1f,T}, \gamma_{2f,T}, \ f = 1,...,m-1 \text{ unrestricted}$$
$$H_1^* : \ \text{at least one of } \gamma_{1m,T}, \gamma_{2m,T} \neq 0$$

They recommend the test $MW_{m-1}^m := SW_{m-1}^m \exp(-b_\xi |DF|^{-1})$, where $SW_{m-1}^m := (RSS_R - RSS_U)/RSS_U$ with RSS_R and RSS_U the restricted and unrestricted residual sums of squares from OLS estimation of (2) with n replaced by $m-1$ and m, respectively, and where DF is the Dickey-Fuller t-statistic applied to the OLS residuals from estimation of (1) with n replaced by m.

The new approach proposed in this paper can also be used to construct a test of H_0^* against H_1^*. Specifically, we adopt the same SW_{m-1}^m statistic that appears as a component in MW_{m-1}^m, and compare this statistic with the adjusted adaptive critical value $cv_\xi^{\lambda,adj}$ taking the form of (6), where $\lambda_{J,\xi}$ is as defined in (4) with $\delta = 1/2$ and $J = B$ where B now denotes the statistic in (5) with $\hat{u}_t$ being the residuals from OLS estimation of (1) with n replaced by m.

The following large sample results for SW_{m-1}^m can be obtained directly from HLX. When y_t is generated by (1) under H_1^*,

$$\text{(a) Under Assumption I(0),} \quad SW_{m-1}^m \xrightarrow{d} \frac{\int_0^1 L_R^*(r, \boldsymbol{\gamma}^*)^2 dr}{\int_0^1 L_U^*(r)^2} - 1 =: D_0^*(\boldsymbol{\gamma}^*)$$

where $\boldsymbol{\gamma}^* := [\gamma_{1m}, \gamma_{2m}]$, $L_R^*(r, \boldsymbol{\gamma}^*)$ denotes the continuous time residuals from the projection of $\gamma_{1m} m_{1m}(r) + \gamma_{2m} m_{2m}(r) + W(r)$ onto the space spanned by $\left[\int_0^r F(s)ds, X_{m-1}(r)' \right]$, and $L_U^*(r)$ denotes the continuous time residuals from the projection of $W(r)$ onto the space spanned by $\left[\int_0^r F(s)ds, X_m(r)' \right]$; and

$$\text{(b) Under Assumption I(1),} \quad SW_{m-1}^m \xrightarrow{d} \frac{\int_0^1 N_R^*(r, \boldsymbol{\gamma}^*)^2 dr}{\int_0^1 N_U^*(r)^2} - 1 =: D_1^*(\boldsymbol{\gamma}^*)$$

where $N_R^*(r, \boldsymbol{\gamma}^*)$ denotes the continuous time residuals from the projection of $\gamma_{1m} m_{1m}(r) + \gamma_{2m} m_{2m}(r) + \int_0^r W(s)ds$ onto the space spanned by $\left[\int_0^r F(s)ds, X_{m-1}(r)' \right]$, and $N_U^*(r)$ denotes the continuous time residuals from the projection of $\int_0^r W(s)ds$ onto the space spanned by $\left[\int_0^r F(s)ds, X_m(r)' \right]$.

The asymptotic distributions under the null H_0^* follow by setting $\boldsymbol{\gamma}^* = 0$ in the above limits. Asymptotic critical values for SW_{m-1}^m for $m = 2, 3$ under both Assumption I(0) and Assumption I(1) are given in Table 1. We also determined suitable values for τ_ξ and κ_ξ to be used in constructing $cv_\xi^{\lambda,adj}$, and these recommended values are given in Table 2. Hereafter we denote by ASW_{m-1}^m the new tests that compare SW_{m-1}^m with $cv_\xi^{\lambda,adj}$.

Following the HLX algorithm, the ASW_0^n and ASW_{m-1}^m tests can now be used to determine the number of frequencies, n. Given an assumption on the largest possible value of n, $n_{\max}$, we first conduct the tests ASW_0^i, $i = 1, ..., n_{\max}$, and if none of these tests reject we conclude $n = 0$. If any of the tests do reject, we identify the largest value of i for which the null is rejected and set m to this value. If $m = 1$ it is concluded that $n = 1$; otherwise if $m > 1$ we conclude that m is the largest

10

value that n might take. We then consider ASW_0^{m-1}, and if this test fails to reject we set $n = m$. If, however, ASW_0^{m-1} does reject we then perform the ASW_{m-1}^m test; then if this test rejects we conclude $n = m$, otherwise we reduce m by one and repeat the loop. For a diagrammatic representation of the algorithm, see Figure 3 of HLX.

A natural alternative approach to identifying the number of frequencies present in a series would be a standard model selection criterion. As such we also consider a Bayesian information criterion (BIC) approach, which for the constant case ($d_t = \alpha$) is based on the following two regressions

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{f=1}^{n} \gamma_{1f,T} \sin\left(\frac{2\pi f t}{T}\right) + \sum_{f=1}^{n} \gamma_{2f,T} \cos\left(\frac{2\pi f t}{T}\right) + \sum_{i=0}^{k} c_i \Delta y_{t-i} + e_t \qquad (9)$$

$$\Delta y_t = \sum_{f=1}^{n} \gamma_{1f,T}^* \sin\left(\frac{2\pi f t}{T}\right) + \sum_{f=1}^{n} \gamma_{2f,T}^* \cos\left(\frac{2\pi f t}{T}\right) + \sum_{i=0}^{k} c_i^* \Delta y_{t-i} + e_t^*. \qquad (10)$$

Consider selecting n by minimising the BIC across $n = \{0, 1, ..., n_{\max}\}$ and $k = \{0, 1, ..., k_{\max}\}$ for a given regression. BIC based on (9) will be appropriate when the shocks are I(0), while BIC based on (10) will be appropriate for I(1) shocks. Consequently, we propose minimising the BIC jointly over both regressions (9) and (10), and it this procedure with which we will compare the performance of the HLX-type algorithm below.

We now assess, by means of Monte Carlo simulation, the relative performance of the HLX algorithm based on MW_0^n and MW_{m-1}^m (which we refer to as MW), the same algorithmic approach based on ASW_0^n and ASW_{m-1}^m (which we denote by ASW), and the BIC procedure outlined above. Note that all tests were performed for the with-constant variants and were conducted at the nominal 0.05-level. Data were generated according to (7)-(8) with $n = 2$. We compute the frequency with which each procedure correctly selects $n = 2$, along with the frequencies with which each of the incorrect selections of $n = 0, 1, 3$ are made. The choices $n_{\max} = 3$ and $k_{\max} = 4$ are adopted in all cases, and we consider experiments with $T = 150$, $\phi = \{0, 0.5, 0.7, 0.9, 0.95, 1\}$, as before. For each value we report results for four different values of γ, including the case $\gamma = 0$. Results for serially correlated ε_t were found to be similar to those for i.i.d. innovations and are available upon request.

The results are reported in Table 4. Comparing first the two algorithmic approaches, we observe that ASW generally outperforms MW, in line with the superior finite sample testing properties that the constituent tests involved in ASW display. Due to the inherent multiple testing issues with the algorithms, when $\gamma = 0$ we see that both ASW and MW select a non-zero value of n with frequency greater than the nominal level, less so in the case of MW due to its lower finite sample size. However, as might be expected from the results of section 7, when $\gamma > 0$ the ASW approach generally outperforms the MW approach in terms of the frequency with which $n = 2$ is selected. Indeed, the improvements offered by ASW over the original MW are quite substantial in a number of cases, particularly for modest values of γ. While MW can outperform ASW, such gains are always relatively minor, and tend to occur in situations where both procedures select $n = 2$ with high probability.

Turning our attention to a comparison of ASW with the BIC approach, it can be seen that when $\gamma = 0$, BIC selects a value of n greater than zero with roughly the same frequency as was seen for

ASW. In some cases, differences are seen, but neither procedure dominates the other overall when there are truly no nonlinear deterministic components present. When $\gamma > 0$, we find that neither *ASW* nor BIC dominates the other across the different γ magnitudes considered. While the performance of the two methods for determining the number of frequencies can be quite different for any given DGP, these differences do not follow a systematic pattern across all ϕ or γ; as such, it is difficult to argue for a particular ranking between *ASW* and BIC. What is clear, however, is that both procedures offer improvements relative to the *MW* approach of HLX.

The previous simulation study has examined the ability of the proposed algorithms and BIC procedure to correctly specify n when the deterministic component of the series is exactly approximated by a Fourier series expansion. It is also important to investigate how useful these approaches are in approximating other forms of nonlinear deterministic components. To that end we generated data according to $y_t = d_t + u_t$, $t = 1,...,T$, with u_t as in (8) for $\phi = \{0, 0.5, 1\}$, and with d_t specified as either a mid-sample ESTR, i.e. $d_t = \gamma[1 - \exp(-0.1(t - 0.5T)^2)]$, or as a mid-sample LSTR, i.e. $d_t = \gamma/[1 + \exp(0.1(t - 0.5T))]$. For a sample size of $T = 150$ we compute the power of the methods to detect nonlinear deterministic elements, measured as the percentage of replications for which the *ASW* algorithm or BIC selects $n > 0$ frequencies. Table 5 reports, for a range of γ magnitudes, the power of with-constant *ASW* and BIC (results are omitted for *MW* as its performance was uniformly worse than *ASW*). For ESTR deterministics, both *ASW* and BIC have power that is increasing in γ, indicating that the Fourier approximation works reasonably well in terms of modelling the ESTR-type nonlinearities. Of the two procedures, BIC offers higher power than *ASW*, particularly when $\phi > 0$.

In the case of LSTR deterministics, however, the powers of both with-constant *ASW* and BIC are typically decreasing in γ (with the exception of BIC when $\phi = 1$), so while we again find that BIC outperforms *ASW* here, neither procedure performs well across all values of ϕ considered. This feature arises since the LSTR component involves a relatively slow transition from one level to another, which is not well modelled by the Fourier terms of frequency $f = 1, 2, 3$ that are included in the fitted unrestricted model (this approximation error may also be magnified in the *ASW* procedure due to the statistics being based on the partially summed regression (2)). To capture these low frequency movements, one would need to incorporate Fourier terms with lower frequency than one, e.g. $f = 0.5$, in the approximation. Of course, such low frequency Fourier terms can themselves be reasonably well approximated by a linear trend term, hence we also report in Table 5 results obtained from application of the with-trend version of *ASW* to the LSTR data generating processes. We see that for the cases where the with-constant *ASW* suffered from very low power, the with-trend variant has decent power which is also now increasing in the magnitude of γ. We also see in Table 5 that with-trend *ASW* also delivers some power improvements over with-constant *ASW* in the ESTR simulations, particularly when $\phi = 1$. It appears, therefore, that when low frequency changes are present in the data, the with-trend variant of *ASW* is a potentially more robust approach for detecting deterministic nonlinearities, as the underlying fitted models provide a superior approximation to the true deterministic component.

9   Empirical Application

There has been much interest in the financial literature in modelling the volatility of economic in-
dicators, most notably financial indicators such as stock market indices. While early work by, *inter
alia*, Poterba and Summers (1986), French *et al.* (1987) and Schwert and Seguin (1990) concentrated
on modelling volatility in stock market indices using linear methods, more recent work by Cao and
Tsay (1992) attempts to use nonlinear methods to analyse such indices using threshold autoregressive
and nonlinear GARCH and EGARCH models. Such methods assume that any observed nonlinearity
in the volatility indices is stochastic rather than deterministic. It is of interest, however, to assess
whether we can detect nonlinear behaviour in the *deterministic* components of such volatility indices.

To that end we collected daily data on six volatility indices for the one year period ending 18
March 2013. Five of these indices measure the volatility of a particular stock market index: the Dow
Jones Industrial Average Volatility Index (Ticker: VXD), the NASDAQ-100 Volatility Index (VXN),
the S&P 500 Volatility Index (VIX), the S&P 100 Volatility Index (VXO) and the FTSE 100 Volatility
Index (VFTSE). The final series considered is the EuroCurrency Volatility Index (EVZ), which is an
index of the volatility of the US\$/EUR exchange rate. We applied the with-constant ASW and MW
algorithms described in section 8 (with $n_{\max} = 3$), implemented at both the 0.05 and 0.10 significance
levels, to each series. We also applied the BIC selection procedure as a point of comparison. The
number of frequencies, $\hat{n}$, detected by these methods are reported in Table 6. In all cases, some form
of nonlinear deterministic behaviour is detected in the volatility indices, suggesting a consistent body
of evidence for nonlinear behaviour in the deterministic components of these financial volatility series.
The pattern of results observed in Table 6 is consistent with the finite sample simulations presented
in sections 7 and 8. In all cases, for a given significance level, ASW chooses $\hat{n}$ to be greater than or
equal to that chosen by MW; indeed there are three series for which MW does not find any evidence
of nonlinearity at the 0.05-level, while ASW selects $\hat{n} = 2$. In comparison with the BIC procedure, we
find that ASW finds more evidence for deterministic nonlinearity, with BIC selecting $\hat{n} = 0$ for four
of the series, and never identifying a greater number of frequencies than ASW.

As a measure of the underlying persistence in each series, Table 6 also reports (in parentheses)
the estimate of ϕ obtained from the DF regression (3) used in the MW_0^n statistic, where n is set to
the corresponding number of frequencies listed in Table 6. These suggest that the series are highly
persistent around a nonlinear deterministic component, but it is unclear whether the stochastic com-
ponents would be best modelled by stationary or unit root processes. This highlights the advantages
of the robust procedures considered in this paper, as we do not need to take a stand on the integration
properties of the data. Interestingly, fitting only a constant to the data results in estimates of ϕ very
close to unity, suggesting that failure to specify the nonlinear deterministic components of these series
could well lead to the inference that they contain a unit root. Note also that since the with-constant
ASW procedure always selects $\hat{n} > 0$, it does not appear that problematic low frequency movements
are a feature of these time series.

The fitted deterministic components associated with the corresponding values of $\hat{n}$ obtained by

ASW are presented alongside the original and detrended series in Figure 3. Note that for the EVZ series we plot the fitted and detrended series for $\hat{n} = 3$ rather than $\hat{n} = 1$. We see that the deterministic components detected by *ASW* appear to fit the data rather well, with the apparent nonlinear behaviour in the series well approximated by the Fourier frequency representation in most cases.

10 Conclusions

We have proposed a new approach to testing for the presence of nonlinear deterministic components in an economic time series designed to be robust to the order of integration of the data and to any weak dependence present. The recommended approach involves using a scaled Wald statistic that has well defined, but different, limit distributions depending on whether the stochastic component of the series is I(0) or I(1). Robustness is achieved by using an adaptive critical value constructed so that the appropriate I(0) or I(1) critical value is selected asymptotically. This delivers a test with the same asymptotic properties as the test of HLX in an I(0) environment, but with worthwhile local asymptotic power gains in an I(1) setting. Monte Carlo simulations suggest that the new test also has superior overall finite sample size and power properties to the HLX test for both I(0) and I(1) shocks. The proposed testing procedure can be employed in the algorithm of HLX to determine the number of frequencies used to model a series that contains potential nonlinear deterministic components. We have also considered a model selection approach based on the BIC which was also shown to outperform the original HLX procedure. Our empirical application to financial market volatility indices over the last year supports the notion that the new procedures offer benefits over extant methods.

References

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.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. Journal of Econometrics 108, 343-363.

Cao, C.Q. and Tsay, R.S. (1992). Nonlinear time-series analysis of stock volatilities. Journal of Applied Econometrics 7, 165-185.

Enders, W. and Lee, J. (2012). A unit root test using a Fourier series to approximate smooth breaks. Oxford Bulletin of Economics and Statistics 74, 574-599.

French, K.R., Schwert, G.W. and Stambaugh, R. (1987). Expected stock returns and volatility. Journal of Financial Economics 19, 3-29.

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.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

Perron, P. (1998). Trends and random walks in macroeconomic time series: further evidence from a new approach. Journal of Economic Dynamics and Control 12, 297-332.

Perron, P. and Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. Economics Letters 94, 12-19.

Poterba, J.M. and Summers, L.H. (1986). The persistence of volatility and stock market fluctuaions. American Economic Review 76, 1143-1151.

Schwert, G.W. and Seguin, P.J. (1990). Heteroscedasticity in stock returns. Journal of Finance 45, 1129-1155.

Vogelsang, T.J. (1998). Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123-148.

Table 1. Asymptotic ξ-level critical values for SW_0^n and SW_{m-1}^m statistics

	$n=1$			$n=2$			$n=3$		
ξ	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
	With-constant SW_0^n								
$cv_{0,\xi}$	5.268	7.439	13.370	9.337	12.754	21.495	13.259	17.702	29.159
$cv_{1,\xi}$	50.631	76.222	167.163	193.162	274.977	565.278	475.264	680.520	1300.226
	With-trend SW_0^n								
$cv_{0,\xi}$	2.677	3.708	6.370	4.685	6.124	9.921	6.550	8.468	13.308
$cv_{1,\xi}$	30.601	47.342	103.816	100.248	150.228	295.168	227.125	329.713	644.469

	$m=2$			$m=3$		
ξ	0.10	0.05	0.01	0.10	0.05	0.01
	With-constant SW_{m-1}^m					
$cv_{0,\xi}$	1.779	2.406	4.007	1.038	1.393	2.252
$cv_{1,\xi}$	9.829	14.138	27.101	4.636	6.437	12.147
	With-trend SW_{m-1}^m					
$cv_{0,\xi}$	1.310	1.774	2.904	0.856	1.138	1.849
$cv_{1,\xi}$	8.556	12.624	24.795	4.591	6.513	12.293

Table 2. τ_ξ and κ_ξ values for ASW_0^n and ASW_{m-1}^m tests

	$n=1$			$n=2$			$n=3$		
ξ	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
	With-constant ASW_0^n								
τ_ξ	14.6	17.5	26.0	23.0	29.5	50.0	32.0	41.0	90.0
κ_ξ	5.8	6.9	8.5	10.5	11.2	8.6	14.3	14.5	6.5
	With-trend ASW_0^n								
τ_ξ	43.2	50.0	63.0	79.0	110.0	180.0	151.0	210.0	325.0
κ_ξ	9.5	9.8	10.5	14.4	8.3	5.0	11.2	6.4	3.0

	$m=2$			$m=3$		
ξ	0.10	0.05	0.01	0.10	0.05	0.01
	With-constant ASW_{m-1}^m					
τ_ξ	32.4	37.0	53.9	47.7	57.0	80.0
κ_ξ	4.9	5.6	7.4	4.4	5.2	6.8
	With-trend ASW_{m-1}^m					
τ_ξ	90.0	105.0	160.0	140.0	195.0	315.0
κ_ξ	9.9	8.6	5.9	10.6	6.0	2.1

Table 3. Finite sample size of MW_0^1 and ASW_0^1 tests

| | | With-constant tests | | | | With-trend tests | | | |
| | | $T = 150$ | | $T = 300$ | | $T = 150$ | | $T = 300$ | |
ϕ	θ	MW_0^1	ASW_0^1	MW_0^1	ASW_0^1	MW_0^1	ASW_0^1	MW_0^1	ASW_0^1
0.00	-0.5	0.006	0.033	0.010	0.034	0.002	0.010	0.004	0.014
	0.0	0.009	0.051	0.012	0.050	0.004	0.026	0.007	0.032
	0.5	0.004	0.057	0.011	0.063	0.001	0.053	0.006	0.068
0.50	-0.5	0.003	0.018	0.004	0.018	0.000	0.003	0.001	0.004
	0.0	0.007	0.025	0.008	0.024	0.002	0.007	0.003	0.007
	0.5	0.009	0.051	0.012	0.050	0.004	0.026	0.007	0.032
0.70	-0.5	0.002	0.014	0.003	0.011	0.000	0.002	0.001	0.002
	0.0	0.006	0.018	0.006	0.014	0.002	0.004	0.002	0.003
	0.5	0.008	0.040	0.006	0.033	0.004	0.015	0.003	0.013
0.90	-0.5	0.010	0.022	0.005	0.009	0.002	0.005	0.001	0.001
	0.0	0.015	0.025	0.007	0.010	0.005	0.006	0.002	0.002
	0.5	0.015	0.045	0.007	0.018	0.006	0.017	0.002	0.005
0.95	-0.5	0.024	0.041	0.011	0.015	0.008	0.014	0.003	0.003
	0.0	0.030	0.044	0.013	0.016	0.013	0.016	0.005	0.004
	0.5	0.031	0.069	0.014	0.025	0.015	0.032	0.005	0.008
1.00	-0.5	0.039	0.058	0.040	0.049	0.038	0.057	0.039	0.049
	0.0	0.047	0.060	0.047	0.050	0.047	0.061	0.047	0.050
	0.5	0.049	0.076	0.048	0.058	0.052	0.091	0.049	0.063

Table 4. Number of frequencies selected by with-constant MW and ASW algorithms and BIC: $T = 150$

| | | MW | | | | ASW | | | | BIC | | | |
ϕ	γ	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 0$	$n = 1$	$n = 2$	$n = 3$
0.00	0.00	0.989	0.009	0.001	0.000	0.933	0.050	0.014	0.004	0.991	0.009	0.000	0.000
	0.30	0.419	0.017	0.501	0.064	0.208	0.043	0.679	0.070	0.281	0.144	0.565	0.009
	0.60	0.051	0.000	0.882	0.067	0.014	0.000	0.931	0.055	0.001	0.000	0.984	0.014
	1.20	0.003	0.000	0.976	0.022	0.000	0.000	0.950	0.050	0.000	0.000	0.986	0.014
0.50	0.00	0.993	0.007	0.001	0.000	0.971	0.024	0.004	0.001	0.987	0.012	0.001	0.000
	0.70	0.408	0.004	0.528	0.060	0.334	0.006	0.578	0.082	0.274	0.051	0.656	0.020
	1.40	0.030	0.000	0.917	0.052	0.032	0.000	0.906	0.062	0.009	0.000	0.964	0.027
	2.80	0.001	0.000	0.982	0.017	0.000	0.000	0.955	0.045	0.000	0.000	0.973	0.027
0.70	0.00	0.993	0.006	0.001	0.000	0.978	0.017	0.004	0.001	0.981	0.016	0.002	0.001
	1.20	0.450	0.002	0.481	0.067	0.383	0.002	0.514	0.101	0.433	0.013	0.524	0.029
	2.40	0.032	0.000	0.915	0.053	0.040	0.000	0.881	0.078	0.032	0.000	0.919	0.049
	4.80	0.001	0.000	0.977	0.023	0.000	0.000	0.949	0.051	0.000	0.000	0.949	0.051
0.90	0.00	0.977	0.014	0.005	0.003	0.952	0.023	0.013	0.011	0.943	0.037	0.013	0.007
	3.20	0.349	0.000	0.510	0.141	0.248	0.000	0.577	0.175	0.277	0.000	0.596	0.127
	6.40	0.038	0.000	0.861	0.101	0.013	0.000	0.877	0.110	0.000	0.000	0.836	0.164
	12.80	0.002	0.000	0.935	0.063	0.000	0.000	0.909	0.091	0.000	0.000	0.836	0.164
0.95	0.00	0.939	0.028	0.018	0.016	0.896	0.040	0.033	0.031	0.906	0.051	0.025	0.018
	4.00	0.364	0.000	0.469	0.167	0.260	0.000	0.552	0.188	0.145	0.000	0.659	0.195
	8.00	0.063	0.000	0.799	0.138	0.016	0.000	0.855	0.129	0.000	0.000	0.781	0.218
	16.00	0.007	0.000	0.904	0.089	0.000	0.000	0.892	0.108	0.000	0.000	0.782	0.218
1.00	0.00	0.900	0.041	0.030	0.029	0.876	0.050	0.039	0.035	0.883	0.064	0.031	0.021
	10.00	0.369	0.000	0.515	0.116	0.234	0.000	0.667	0.099	0.000	0.000	0.814	0.186
	20.00	0.209	0.000	0.690	0.101	0.032	0.000	0.903	0.065	0.000	0.000	0.814	0.186
	40.00	0.116	0.000	0.796	0.088	0.000	0.000	0.944	0.056	0.000	0.000	0.814	0.186

Table 5. Power of ASW algorithm and BIC to detect exponential and logistic smooth transitions: $T = 150$

		ESTR				LSTR		
ϕ	γ	ASW_c	BIC	ASW_t	γ	ASW_c	BIC	ASW_t
0.0	1.0	0.508	0.526	0.520	2.0	0.484	0.969	0.263
	1.5	0.806	0.929	0.837	4.0	0.055	0.428	0.687
	2.0	0.936	0.997	0.958	6.0	0.001	0.034	0.895
0.5	3.0	0.588	0.893	0.660	10.0	0.001	0.042	0.474
	4.0	0.797	0.990	0.803	15.0	0.000	0.005	0.765
	5.0	0.906	0.997	0.922	20.0	0.000	0.026	0.921
1.0	25.0	0.405	0.730	0.868	25.0	0.163	0.332	0.312
	50.0	0.701	0.999	1.000	50.0	0.057	0.922	0.698
	75.0	0.872	1.000	1.000	75.0	0.011	0.993	0.474

Note: ASW_c and ASW_t denote with-constant and with-trend ASW, respectively.

Table 6. Number of frequencies selected by with-constant MW and ASW algorithms and BIC for volatility indices

Volatility index	$MW(5\%)$	$MW(10\%)$	$ASW(5\%)$	$ASW(10\%)$	BIC
DJIA (VXD)	0 (0.92)	2 (0.76)	2 (0.76)	2 (0.76)	0 (0.92)
NASDAQ-100 (VXN)	0 (0.94)	2 (0.82)	2 (0.82)	2 (0.82)	0 (0.94)
S&P 500 (VIX)	0 (0.92)	2 (0.76)	2 (0.76)	2 (0.76)	2 (0.76)
EuroCurrency (EVZ)	1 (0.90)	1 (0.90)	1 (0.90)	3 (0.85)	0 (0.97)
S&P 100 (VXO)	2 (0.74)	2 (0.74)	2 (0.74)	2 (0.74)	2 (0.74)
FTSE 100 (VFTSE)	2 (0.81)	2 (0.81)	2 (0.81)	2 (0.81)	0 (0.95)

Note: Values of $\hat{\phi}$ are given in parentheses.
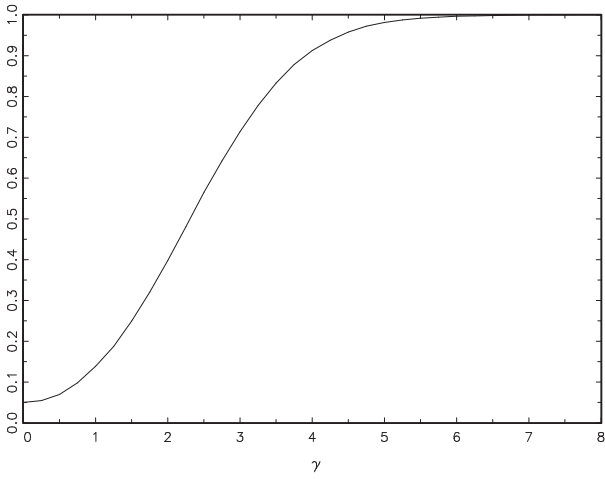
(a) I(0) shocks, $n = 1$
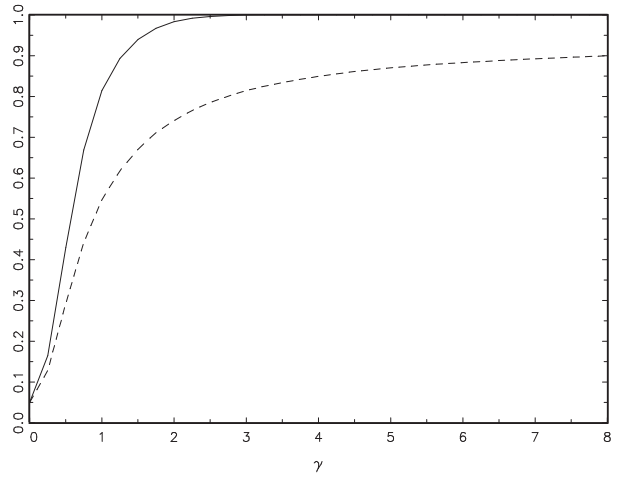
(b) I(1) shocks, $n = 1$
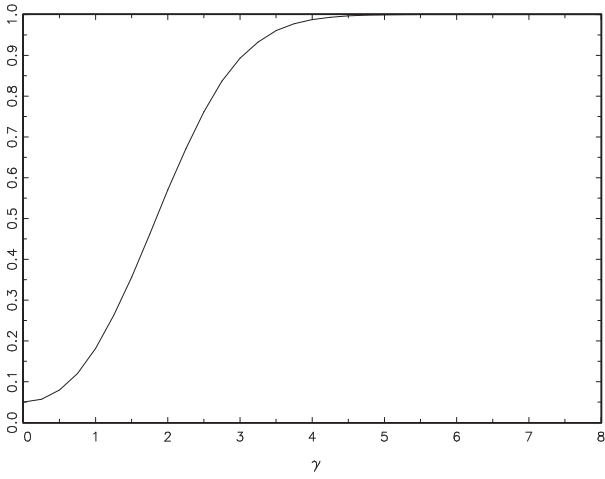
(c) I(0) shocks, $n = 2$
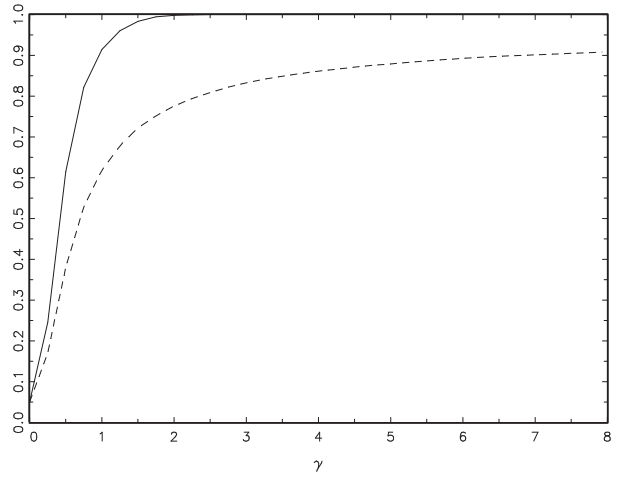
(d) I(1) shocks, $n = 2$
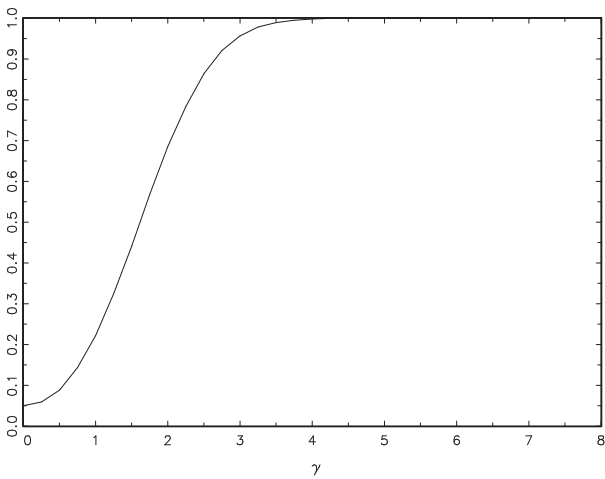
(e) I(0) shocks, $n = 3$
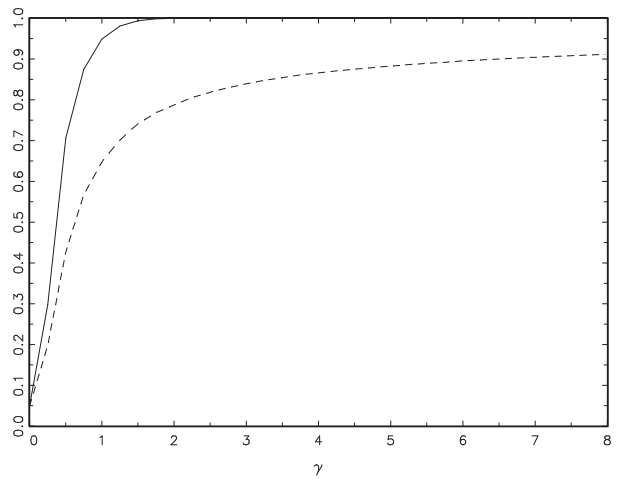
(f) I(1) shocks, $n = 3$

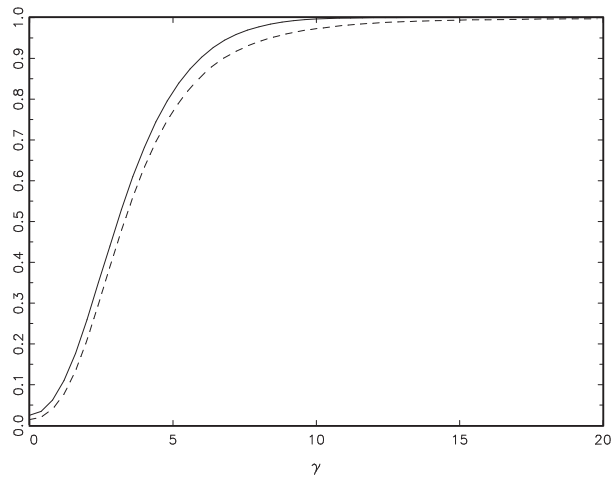Figure 1. Local asymptotic power of with-constant tests: MW_0^n: - - - , ASW_0^n: ——
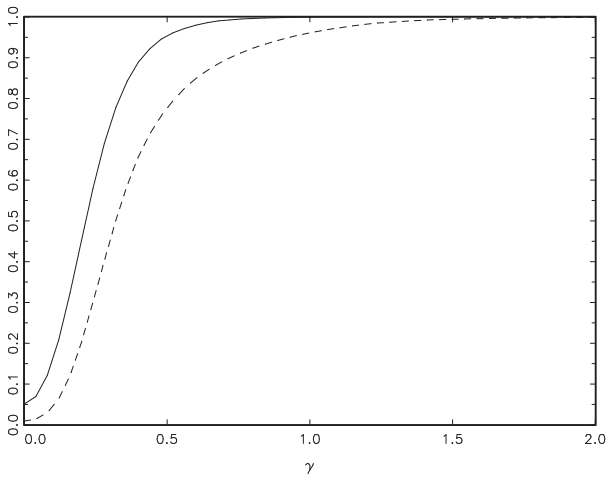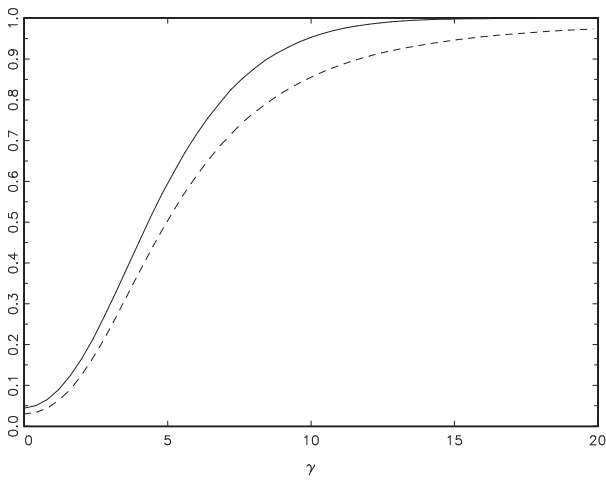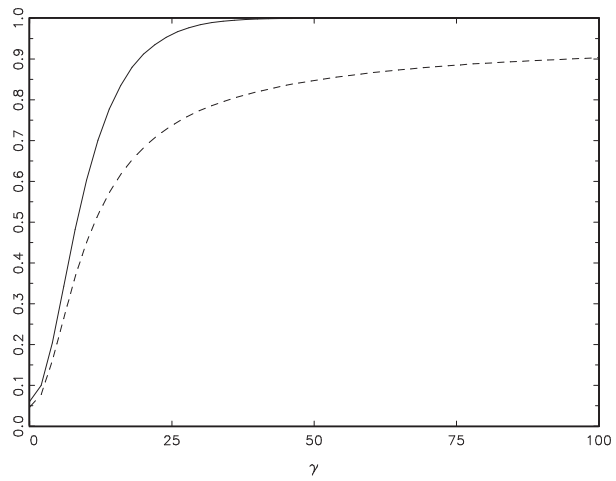
(a) $\phi = 0.00$
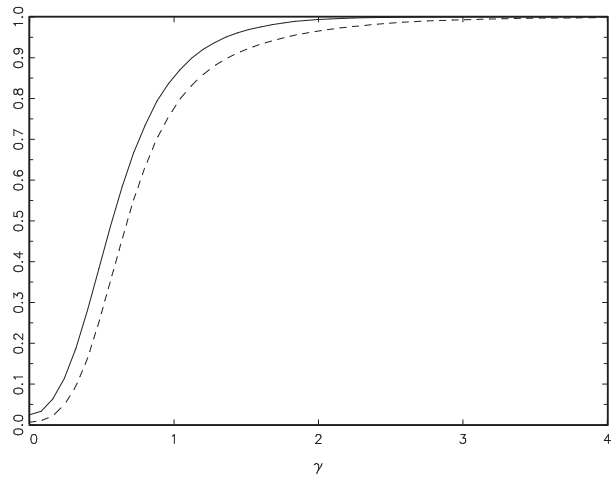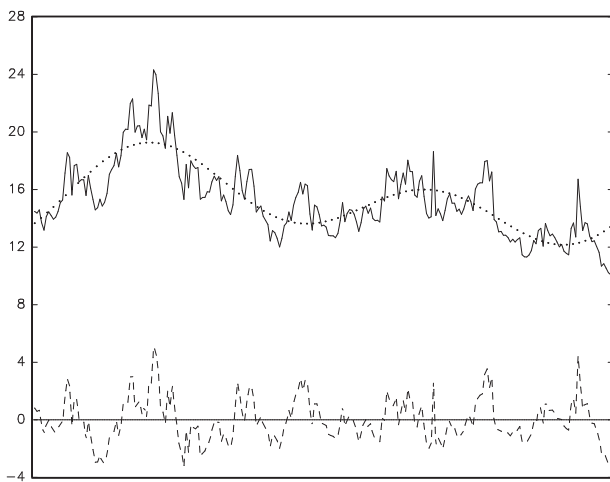
(b) $\phi = 0.50$
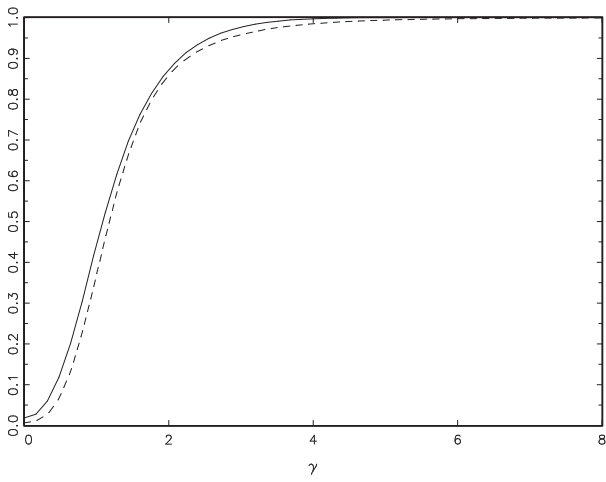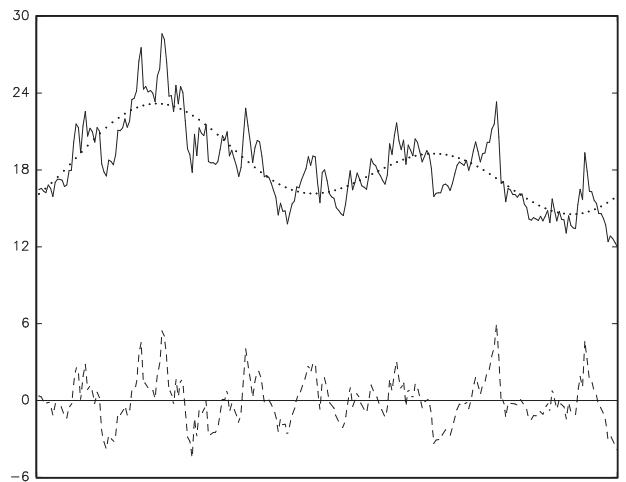
(c) $\phi = 0.70$
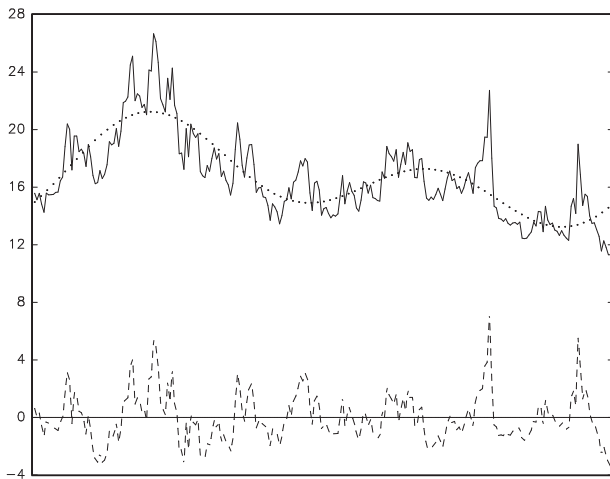
(d) $\phi = 0.90$

(e) $\phi = 0.95$

(f) $\phi = 1.00$

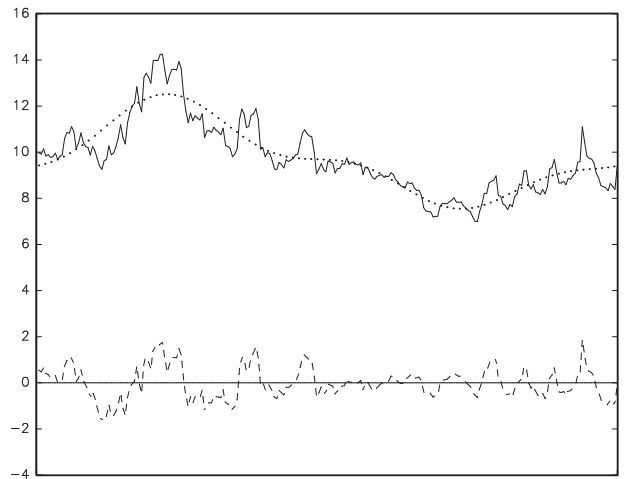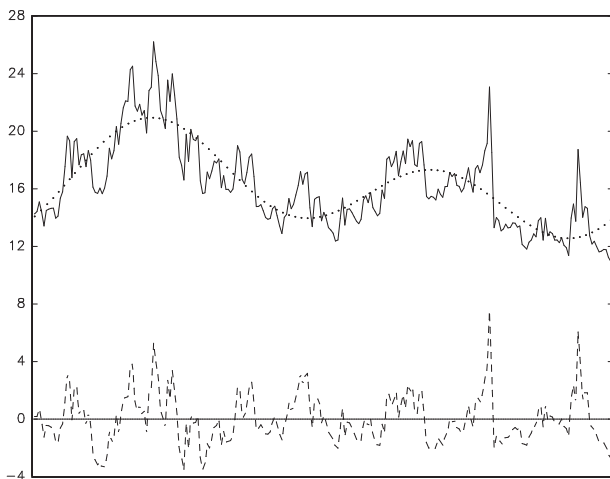Figure 2. Finite sample power of with-constant tests: $T = 150$, $n = 1$; MW_0^n: - - - , ASW_0^n: ——
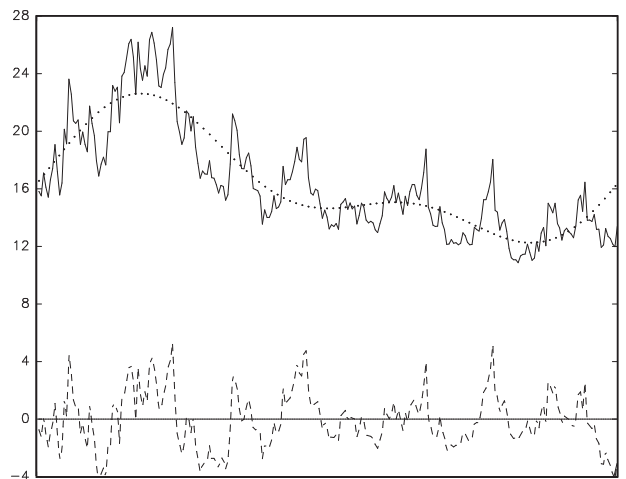
Figure 3. Nonlinear deterministic components detected in volatility indices:
Index: —— , Fitted deterministic: · · · , Detrended series: - - -

21