

Powerful Self-normalizing Tests for Stationarity against the Alternative of a Unit Root*

Uwe Hassler[†]

Goethe-Universität Frankfurt

Mehdi Hosseinkouchack

EBS Universität

October 12, 2022

Abstract

We propose a family of tests for stationarity against a local unit root. It builds on the Karhunen-Loève (KL) expansions of the limiting CUSUM process under the null hypothesis and a local alternative. The variance ratio type statistic \mathcal{VR}_q is a ratio of quadratic forms of q weighted Gaussian sums such that the nuisance long-run variance cancels asymptotically without having to be estimated. Asymptotic critical values and local power functions can be calculated by standard numerical means, and power grows with q . However, Monte Carlo experiments show that q may not be too large in finite samples to obtain tests with correct size under the null. Balancing size and power results in a superior performance compared to the classic KPSS test.

JEL classification C12 (hypothesis testing), C22 (time series)

Keywords KPSS, Karhunen-Loève, local alternative.

1 Introduction

Stationarity of economic and financial time series is an underlying assumption of many models, and more specifically integration of order 0, $I(0)$, is a prerequisite of many econometric techniques. For this reason the so-called KPSS test by Kwiatkowski, Phillips,

* **Acknowledgements:** Earlier versions of this paper were presented at the 13th International Conference on Computational and Financial Econometrics (London, 2019), the 2nd Italian Workshop of Econometrics and Empirical Economics (Venice, 2020). We are grateful to Ye Lu, Morten Nielsen and Martin Wagner for helpful comments. We further thank an anonymous referee and Isaac Miller (Editor) for thoughtful and very helpful suggestions.

[†]**Corresponding author:** Statistics and Econometric Methods, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, email: hassler@wiwi.uni-frankfurt.de.

Schmidt, and Shin (1992) enjoys great popularity in empirical economics. It relies on consistent estimation of the so-called long-run variance under the null hypothesis, which is a notoriously difficult issue; for a recent treatment in a continuous time framework under high frequency, see Lu and Park (2019) and Jiang, Lu, and Park (2020).

To circumvent long-run variance estimation one may call on the principle of self-normalization or scale invariance, see the low-frequency stationary test (LFST) by Müller and Watson (2008). It relies on the likelihood ratio statistic of weighted averages of the data characterized by different covariance matrices under the null hypothesis and the local alternative. The alternative assumes a so-called local level model, testing for $I(0)$ against a local random walk. The long-run variance cancels from the likelihood ratio statistic that hence enjoys the property of self-normalization, at least asymptotically. Critical values have been determined by simulation.

In this paper we adopt the local level model and the idea of computing ratios of weighted averages. However, we follow the proposal by Hassler and Hosseinkouchack (2019) and use weighting schemes obtained from the respective Karhunen-Loëve (KL) expansion. To this end we derive the eigenstructure of the autocovariance kernels of the respective limiting processes under different assumptions on the deterministic component in Proposition 1 through 3. For testing we consider only one of two factors of the likelihood ratio statistic corresponding to a kind of variance ratio of quadratic forms (\mathcal{VR}_q say) in (asymptotically) normal variates, which allows to compute critical values and local power function analytically from Hassler and Hosseinkouchack (2019, Thm. 1). Here, q is the number of weighted averages included in the statistic, which provides a family of tests, each having asymptotically correct size for any finite q , while asymptotic power grows with q . Monte Carlo experiments, however, show that q may not be too large to control the size of the test under the null hypothesis in small samples. A choice of $q = 25$ results in a procedure that outperforms KPSS and LFST in terms of size and power for many realistic combinations of sample size and strength of persistence under the alternative.

The rest of the paper is organized as follows. Section 2 becomes precise on the models and assumptions and briefly reviews KPSS. Section 3 introduces our tests as self-normalizing alternatives to KPSS. Asymptotic and finite sample power is addressed in Section 4, while some conclusions are drawn in the final section. Mathematical proofs are relegated to the Appendix.

A word on notation before we begin. Weak convergence in the Skorohod space of cadlag functions is denoted by \Rightarrow as the sample size T goes off to infinity. Integrals are taken from 0 to 1, unless indicated otherwise. $\lfloor x \rfloor$ returns the largest integer smaller than or equal to some real x , $x > 0$. And sinh and cosh denote the hyperbolic sine and cosine, respectively.

2 KPSS

We adopt the local level model maintained by Müller and Watson (2008),

$$y_t = d_t + \frac{\sigma_\eta}{T} \sum_{j=1}^t \eta_j + e_t, \quad t = 1, \dots, T, \quad (1)$$

with the null hypothesis $\sigma_\eta = 0$ and the (local) alternative $\sigma_\eta > 0$. We assume three models $M0$ through $M2$, where $M0$ stands for the case without deterministics ($d_t = 0$), while $M1$ and $M2$ allow for a constant term and a linear time trend, respectively, $d_t = \mu$ and $\mu + \delta t$. The least squares (LS) residuals under the null hypothesis are denoted accordingly, $\hat{u}_{t,0} = y_t$ and $\hat{u}_{t,1} = y_t - \bar{y}$ for $M0$ and $M1$, respectively, and $\hat{u}_{t,2} = y_t - \hat{\mu} - \hat{\delta}t$ for $M2$. The KPSS statistics build on the partial sums under the null hypothesis,

$$S_{t,k} = \sum_{j=1}^t \hat{u}_{j,k}, \quad k = 0, 1, 2, \quad (2)$$

and become

$$\mathcal{KPSS}_k = \frac{\sum_{t=1}^T S_{t,k}^2}{T^2 \hat{\omega}_{e,k}^2}, \quad k = 1, 2. \quad (3)$$

Here, $\hat{\omega}_{e,k}^2$ is a consistent estimator of the long-run variance from Assumption 1. The limiting distributions and critical values are discussed in Kwiatkowski et al. (1992). For the rest of the paper, we will work under assumptions parallel to Kwiatkowski et al. (1992) and Müller and Watson (2008).

Assumption 1 Let $\{e_t, \eta_t\}'$ be a zero mean vector process with

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \begin{pmatrix} e_t \\ \eta_t \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_e W_e(r) \\ W_\eta(r) \end{pmatrix} \quad \text{for } r \in [0, 1] \text{ as } T \rightarrow \infty,$$

where W_e and W_η are standard Wiener processes independent of each other, and $0 < \omega_e^2 < \infty$.

Kwiatkowski et al. (1992) did not consider model $M0$. It is, however, straightforward under $\sigma_\eta = 0$ with Assumption 1 that

$$\mathcal{KPSS}_0 = \frac{\sum_{t=1}^T (\sum_{j=1}^t \hat{u}_{j,0})^2}{T^2 \hat{\omega}_{e,0}^2} \Rightarrow \int_0^1 W_e^2(r) dr;$$

critical values are available from MacNeill (1978, Table 2) and Nyblom (1989, Table 1).

KPSS builds on the Lagrange Multiplier (LM) principle, see Tanaka (1983), and becomes

locally best invariant under Gaussianity and absence of serial correlation, while no uniformly most powerful tests exist, see Nyblom and Mäkeläinen (1983) and similarly Nabeya and Tanaka (1988, Sect. 4.1) for the case without deterministics. Under serial correlation, KPSS is plagued by the long-run variance ω_e^2 that has to be removed. Of course, this is a standard problem often labeled HAC (heteroskedasticity and autocorrelation consistent) variance estimation in econometrics. The vast majority of HAC estimators falls into the class of kernel estimators, see the pioneering papers by Newey and West (1987) and Andrews (1991), cf. also Liu and Wu (2010). All these estimators are subject to the critique by Müller (2007, p. 1338) “[...] that any consistent long-run variance estimator is necessarily a discontinuous function [...], i.e. sample paths [...] that are close in the sup norm do not in general lead to similar long-run variance estimates.”¹ To circumvent this non-robust estimation we call on the principle of self-normalization or scale invariance and discuss ratio tests.

3 Variance ratio tests

3.1 Limiting processes

Let us stick to the normalized partial sum process from the residuals computed under H_0 :

$$x_{\lfloor rT \rfloor, k} := \frac{S_{\lfloor rT \rfloor, k}}{\sqrt{T}} = \frac{\sum_{j=1}^{\lfloor rT \rfloor} \widehat{u}_{j,k}}{\sqrt{T}}, \quad k = 0, 1, 2.$$

Under Assumption 1, it then holds for $M0$ that

$$x_{\lfloor rT \rfloor, 0} \Rightarrow \omega_e X_{0,\theta}(r), \quad X_{0,\theta}(r) := W_e(r) + \theta \int_0^r W_\eta(s) ds \text{ with } \theta := \frac{\sigma_\eta}{\omega_e}. \quad (4)$$

For $M1$, $x_{\lfloor rT \rfloor, 1}$ converges to the corresponding tied down process: $x_{\lfloor rT \rfloor, 1} \Rightarrow \omega_e (X_{0,\theta}(r) - r X_{0,\theta}(1))$ with θ defined in (4). This can be rewritten as

$$X_{1,\theta}(r) := X_{0,\theta}(r) - r X_{0,\theta}(1) = W_e(r) - r W_e(1) + \theta \left(\int_0^r W_\eta(s) ds - r \int_0^1 W_\eta(s) ds \right).$$

¹The estimation of the long-run variance (being proportional to the spectrum of a stationary process at the origin) has been classified as an “ill-posed” problem before. Faust (1996, Prop. 2) argued that any bounded confidence interval for ω_e^2 has necessarily a coverage probability of zero if the spectrum is not sufficiently smooth over the parameter space, since convergence of a pointwise consistent estimator is not uniform; see Pötscher (2002, Thm. 4.3) for a reinforcement.

Finally, under $M2$ one obtains $x_{\lfloor rT \rfloor, 2} \Rightarrow \omega_e X_{2,\theta}(r)$ where

$$X_{2,\theta}(r) := X_{0,\theta}(r) + (2r - 3r^2) X_{0,\theta}(1) + 6(r^2 - r) \int_0^1 X_{0,\theta}(a) da ,$$

with θ as before. Of course, $X_{2,\theta}$ parallels the so-called second-level Brownian bridge from Kwiatkowski et al. (1992, eq. (16)).

Due to the independence maintained under Assumption 1, the autocovariance kernel of $X_{0,\theta}$ is

$$K_0(s, t; \theta) = \min(s, t) + \theta^2 K(s, t) ,$$

where we define $K(s, t)$ as the kernel of the integrated component in (4):

$$K(s, t) := \frac{\min^2(s, t)(3 \max(s, t) - \min(s, t))}{6} .$$

The autocovariance kernel of $X_{1,\theta}$ becomes

$$K_1(s, t; \theta) = K_0(s, t; 0) - st + \theta^2 [K(s, t) - sK(1, t) - tK(s, 1) + stK(1, 1)] .$$

For $X_{2,\theta}$ one obtains

$$\begin{aligned} K_2(s, t; \theta) &= K_1(s, t; 0) - 3st(1-s)(1-t) \\ &\quad + \theta^2 \left[\frac{1}{3}g_1(s)g_1(t) + \frac{25}{4}g_2(s)g_2(t) + \frac{5}{4}g_1(s)g_2(t) + \frac{5}{4}g_1(t)g_2(s) \right. \\ &\quad \left. - \frac{1}{4}g_2(t)g_4(s) - \frac{1}{4}g_2(s)g_4(t) - \frac{1}{6}g_1(s)g_3(t) - \frac{1}{6}g_1(t)g_3(s) + k(s, t) \right] , \end{aligned}$$

where $g_1(s) = 4s - 3s^2$, $g_2(s) = s^2 - s$, $g_3(s) = 3s^2 - s^3$, $g_4(s) = 6s^2 - s^4$.

Testing for the null hypothesis $\sigma_\eta = 0$ in model $M0$ through $M2$ amounts to testing for $\theta = \theta_0 = 0$ in $X_{k,\theta}(r)$, $k = 0, 1, 2$. The following tests are directed against specific alternatives, $\theta = \theta_1 > 0$. In order to denote these two values, we will write θ_h with $h \in \{0, 1\}$.

3.2 Test statistics

Obviously, $X_{0,\theta}(r)$ and $X_{1,\theta}(r)$ share the additive structure maintained in Hassler and Hosseinkouchack (2019, Ass. 2), and one can show that this holds true for $X_{2,\theta}(r)$, too. Hence, we can follow their route and build statistics from the eigenstructures of the kernels. They consist of the eigenfunctions $f_{j,\theta_h,k}(s)$ and eigenvalues $\lambda_{j,\theta_h,k}$ under the null and the alternative, $h \in \{0, 1\}$. They are the nontrivial solutions to the Fredholm integral

equation ($j = 1, 2, \dots$)

$$f_{j,\theta_h,k}(t) = \lambda_{j,\theta_h,k} \int_0^1 K_k(s, t; \theta_h) f_{j,\theta_h,k}(s) ds.$$

The eigenvalues are the zeros of the Fredholm determinant $D_{\theta_h,k}(\lambda)$ defined as

$$D_{\theta_h,k}(\lambda) = \lim_{T \rightarrow \infty} \det \left(I_T - \frac{\lambda}{T} \left[K_k \left(\frac{j}{T}, \frac{\ell}{T}; \theta_h \right) \right]_{j,\ell=1,\dots,T} \right),$$

where I_T denotes the identity matrix.

The eigenstructure provides the ingredients for the KL expansion,

$$X_{k,\theta}(r) = \sum_{j=1}^{\infty} \frac{f_{j,\theta,k}(r)}{\lambda_{j,\theta,k}^{1/2}} Z_j,$$

where $\{Z_j\}$ is a sequence of independent standard normal variates. In general, such infinite expansions are not unique, see e.g. the discussion in Phillips (1998). But a truncated KL expansion,

$$X_{k,\theta}(r) = \sum_{j=1}^q \frac{f_{j,\theta,k}(r)}{\lambda_{j,\theta,k}^{1/2}} Z_j + error_{k,\theta}(r),$$

minimizes the total mean squared error for any q . This suggests to compute q weighted sums of the normalized partial sum process under the null and under the alternative:

$$X_{j,\theta_h,k} := \sum_{t=1}^T \left[\int_{(t-1)/T}^{t/T} f_{j,\theta_h,k}(s) ds \right] \frac{S_{t,k}}{\sqrt{T}}, \quad j = 1, \dots, q, \quad k = 0, 1, 2. \quad (5)$$

Due to the continuous mapping theorem it holds that

$$X_{j,\theta_h,k} \Rightarrow \omega_e \int_0^1 f_{j,\theta_h,k}(s) X_{k,\theta_h}(s) ds.$$

The long-run variance cancels from the following variance ratio type statistics along the lines of Hassler and Hosseinkouchack (2019):

$$\mathcal{VR}_{k,q} = \frac{\sum_{j=1}^q \lambda_{j,\theta_0,k} X_{j,\theta_0,k}^2}{\sum_{j=1}^q \lambda_{j,\theta_1,k} X_{j,\theta_1,k}^2}, \quad k = 0, 1, 2. \quad (6)$$

In the limit, one obtains quadratic forms of normal variates independent of a nuisance scaling parameter, say $\mathcal{L}_k(\theta) = \mathcal{L}_{k,q,\theta_0}(\theta)/\mathcal{L}_{k,q,\theta_1}(\theta)$. Of course, $\mathcal{L}_k(\theta)$ depend on q , too, but this is suppressed for convenience. The cumulative distribution function can be computed by numerical means, see Hassler and Hosseinkouchack (2019, Thm. 1), which allows to

compute critical values as well as asymptotic power functions.

Remark 1 Computation of the variance ratio type statistic requires a choice of q . Notice that q does not affect the asymptotic size of the test. The number q does not parallel the bandwidth when computing the KPSS statistic, which has to grow with the sample size to ensure consistent long-run variance estimation. We rather propose a family of asymptotically valid tests for any finite q . Asymptotically, the choice of q affects only power, see Section 4.1. In finite samples, however, it is correct that q has to be small relative to the sample size, see Section 4.2.

Let $\ell_q(\alpha)$ denote the $(1 - \alpha)$ -quantiles of the null distribution $\mathcal{L}_k(\theta_0)$; they are the critical values when testing at level α . The 5% critical values will be determined jointly with θ_1 such that the asymptotic power equals 50% at θ_1 . Note that critical values and θ_1 depend on k , $k = 0, 1, 2$, as well as on q , which we suppress for notational convenience. These same values for θ_1 are employed to determine critical values at level 1% and 10%, too. A selection is provided in Table 1. The eigenstructures that are required to compute the test statistics are given next.

Table 1: Critical values $\ell_q(\alpha)$ for model Mk , $k = 0, 1, 2$.

		$q = 10$	$q = 15$	$q = 25$	$q = 50$	$q = 100$
$k = 0$	$\ell_q(0.01)$	2.3628	1.7128	1.3608	1.1606	1.0760
	$\ell_q(0.05)$	1.8156	1.4447	1.2317	1.1052	1.0502
	$\ell_q(0.10)$	1.6268	1.3460	1.1819	1.0831	1.0398
	θ_1	5.2583	4.7661	4.4503	4.2477	4.1555
$k = 1$	$\ell_q(0.01)$	3.0622	1.9622	1.4589	1.1975	1.0922
	$\ell_q(0.05)$	2.3833	1.6627	1.3226	1.1409	1.0662
	$\ell_q(0.10)$	2.1282	1.5456	1.2674	1.1173	1.0553
	θ_1	10.7880	9.1443	8.2355	7.6971	7.4618
$k = 2$	$\ell_q(0.01)$	5.3490	2.4653	1.6187	1.2511	1.1145
	$\ell_q(0.05)$	4.1877	2.0936	1.4683	1.1920	1.0880
	$\ell_q(0.10)$	3.7225	1.9402	1.4045	1.1664	1.0763
	θ_1	25.0919	17.7328	14.7969	13.2726	12.6451

Note: Values for θ_1 are all calculated such that the tests have a rejection rate of 50% at a 5% nominal size when the alternative hypothesis is $\theta = \theta_1$.

3.3 Eigenstructure

As in Table 1 we will suppress for simplicity that the eigenstructures depend on the model tested, $k = 0, 1, 2$. We begin with model $M0$ without intercept.

Proposition 1 For the eigenstructure of $K_0(s, t; \theta)$ from M0 it holds that the eigenvalues are the positive roots of $D_\theta(\lambda) = 0$ with

$$D_\theta(\lambda) = \frac{2\theta^2 + (2\theta^2 + \lambda) \cos \mu_1 \cosh \mu_2 - \mu_1 \mu_2 \sin \mu_1 \sinh \mu_2}{4\theta^2 + \lambda}$$

where $\mu_1 = \mu_1(\lambda; \theta)$ and $\mu_2 = \mu_2(\lambda; \theta)$ are short for

$$\mu_1 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} + \lambda}{2}} \quad \text{and} \quad \mu_2 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} - \lambda}{2}}.$$

The eigenfunctions are given as

$$f(t) = \alpha_2 (\alpha_1 \cos(\mu_1 t) + \sin(\mu_1 t)) + \alpha_3 \exp(\mu_2 t) + \alpha_4 \exp(-\mu_2 t),$$

with α_1 through α_4 defined in the Appendix.

Proof: See Appendix.

Note that $\lim_{\theta \rightarrow 0} \mu_1 = \sqrt{\lambda}$ and $\lim_{\theta \rightarrow 0} \mu_2 = 0$ such that $\lim_{\theta \rightarrow 0} D_\theta(\lambda) = \cos \sqrt{\lambda}$ with $\lambda_j = (j - 1/2)^2 \pi^2$, which reproduces of course the well known eigenvalues of the standard Wiener process. Further, careful expansions of α_1 through α_4 around $\theta = 0$ show that $\lim_{\theta \rightarrow 0} f(\lambda) = \sqrt{2} \sin(\sqrt{\lambda}t)$.

Next, we turn to the model with intercept.

Proposition 2 For the eigenstructure of $K_1(s, t; \theta)$ from M1 it holds that the eigenvalues are the positive roots of $D_\theta(\lambda) = 0$ with

$$D_\theta(\lambda) = \frac{\sin \mu_1 \sinh \mu_2}{\mu_1 \mu_2},$$

where $\mu_1 = \mu_1(\lambda; \theta)$ and $\mu_2 = \mu_2(\lambda; \theta)$ are from Proposition 1. The eigenfunctions are given as

$$f(t) = \sqrt{2} \sin(\mu_1 t).$$

Proof: See Appendix.

Note that $\lim_{\theta \rightarrow 0} D(\lambda) = \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}}$ with $\lambda_j = j^2 \pi^2$, and $\lim_{\theta \rightarrow 0} f(t) = \sqrt{2} \sin(j\pi t)$, which of course reproduces the case of a Brownian bridge, see e.g. Phillips (1998, p. 1303).

Remark 2 Somehow surprisingly, the case with intercept is simpler than the case without. Note that for $k = 1$, μ_1 has to be a multiple of π , such that $f(t) = \sqrt{2} \sin(j\pi t)$ irrespective

of the value of θ . Hence, the weights in (5) are identical, $f_{j,\theta_0,1}(s) = f_{j,\theta_1,1}(s)$, which is not the case for $k = 0$. Further, the eigenvalues required for (6) become for $k = 1$ simply $\lambda_{j,\theta_h,1} = \frac{j^4\pi^4}{j^2\pi^2+\theta_h^2}$.

Finally, we turn to the case with intercept and a linear time trend.

Proposition 3 *For the eigenstructure of $K_2(s, t; \theta)$ from M2 it holds that the eigenvalues are the positive roots of $D_\theta(\lambda) = 0$ with*

$$D_\theta(\lambda) = \frac{144(2 - 2\cos\mu_1 - \mu_1\sin\mu_1)(2 - 2\cosh\mu_2 + \mu_2\sinh\mu_2)}{\mu_1^4\mu_2^4}$$

where $\mu_1 = \mu_1(\lambda; \theta)$ and $\mu_2 = \mu_2(\lambda; \theta)$ are from Proposition 1. The eigenfunctions are given as

$$f(t) = \begin{cases} \sqrt{2}(\cot\frac{\mu_1}{2}\cos(\mu_1 t) + \sin(\mu_1 t) - \cot\frac{\mu_1}{2}) & \text{if } \mu_1 \cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2} = 0, \\ \sqrt{2}\sin(\mu_1 t) & \text{if } \sin\frac{\mu_1}{2} = 0. \end{cases}$$

Proof: See Appendix.

By expansions of $x\sinh x$ and $2 - 2\cosh x$ it is not hard to verify that $\lim_{\theta \rightarrow 0} D_\theta(\lambda) = \frac{12}{\lambda^2} (2 - 2\cos\sqrt{\lambda} - \sqrt{\lambda}\sin\sqrt{\lambda})$, which is the result for the second-level Brownian bridge by Nabeya and Tanaka (1988, Thm. 6).

Remark 3 *The computation of the test statistic from (6) is facilitated by the following considerations. Note that for $k = 2$, μ_1 solves $\sin\frac{\mu_1}{2}(\mu_1 \cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2}) = 0$. Let us call these solutions $\mu_{1,j}$, where $\mu_{1,j}$ is either an even multiple of π or it solves $\mu_1 \cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2} = 0$. Since the eigenfunctions are just a function of μ_1 , we have again that $f_{j,\theta_0,2}(s) = f_{j,\theta_1,2}(s)$ implying that the weights in (5) are identical. Further, it holds that $\lambda_{j,\theta_h,2} = \frac{\mu_{1,j}^4}{\mu_{1,j}^2 + \theta_h^2}$.*

4 Power

4.1 Asymptotic

Local power of $\mathcal{VR}_{k,q}$ can be computed along the lines of Hassler and Hosseinkouchack (2019, Thm. 1). Moreover, using Tanaka (1996, Thm. 5.11) we obtain the limiting local power function of the KPSS tests. Since it holds that

$$\mathcal{KPSS}_k \Rightarrow \int_0^1 (X_{k,\theta}(r))^2 dr,$$

the characteristic function $\phi_{KPSS,k}$ of this limit is given by

$$\phi_{KPSS,k}(t) = \frac{1}{\sqrt{D_{k,\theta}(2it)}},$$

where $D_{k,\theta}(\cdot)$ are the Fredholm determinants from Proposition 2 and 3 for $k = 1$ and $k = 2$, respectively. Using Lévy's inversion theorem it is only a numerical problem to compute distribution functions, see Figure 1 and 2. While the KPSS tests are locally best (having the steepest slope of the power function at the null), they are beaten for large enough q as we move away from the null hypothesis.

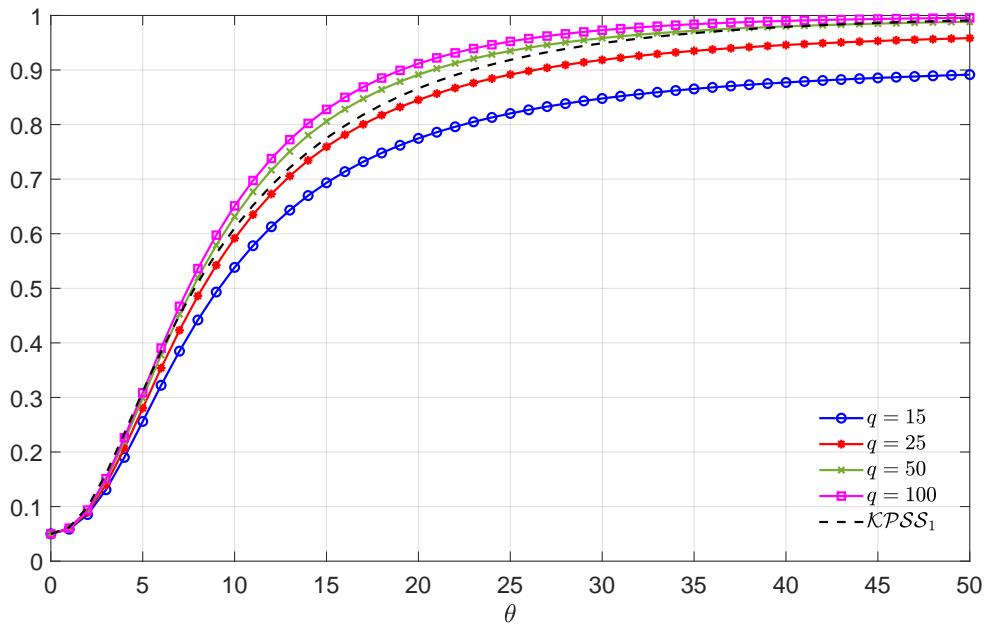


Figure 1: Limiting power for model with constant ($k = 1$) at size 5%.

4.2 Experimental evidence

All computer experiments were performed with MATLAB and rejection frequencies are computed from 10^4 replications when testing at nominal level of 5%. The errors from models $M0$ through $M2$ are autoregressive of order 1, $e_t = \phi e_{t-1} + \varepsilon_t$, and independent of the random walk. Both white noise innovations are standard normal, i.e.

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \text{i.i.}\mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

When a long-run variance has to be estimated, we use the Bartlett (or triangular) window popularized in econometrics by Newey and West (1987) and Kwiatkowski et al. (1992)

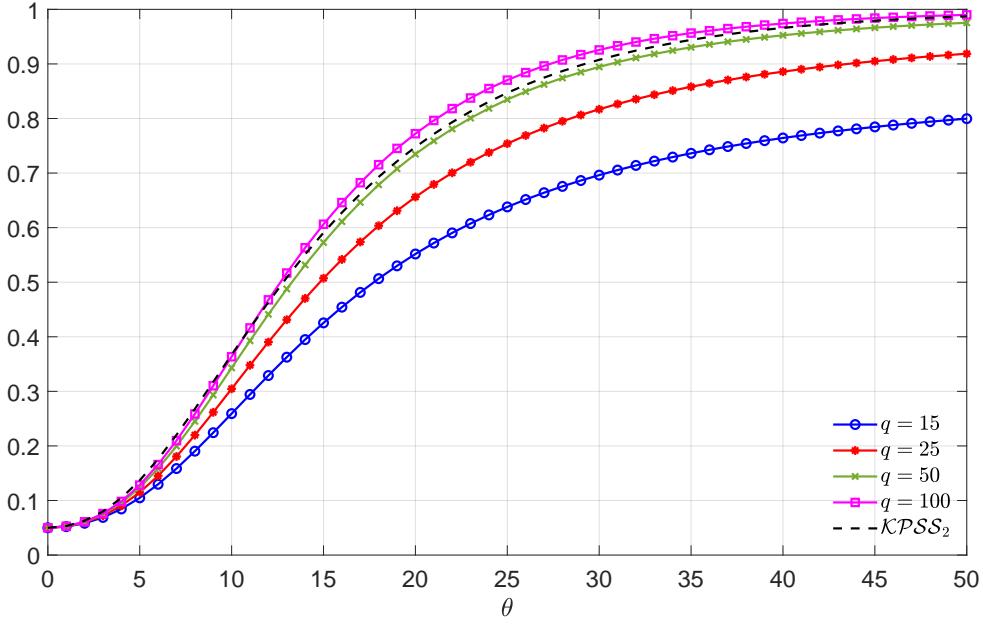


Figure 2: Limiting power for model with trend ($k = 2$) at size 5%.

with data driven bandwidth selection according to Andrews (1991, eq. (6.2)), which builds on the optimal rate. Sample sizes are $T \in \{100, 200, 500, 1000\}$. We only report results for the empirically most relevant case $k = 1$ with intercept. Evidence for $k = 0$ and $k = 2$ is available upon request. In particular, the case of detrending produces very similar findings in term of ranking of the tests and choice of q , only that the power functions are generally flatter, as one would expect from Figure 1 and 2.

For completeness, we also include the LFST test by Müller and Watson (2008). For the model with intercept ($k = 1$) they suggested weighted sums of the residuals, where the weights are derived from the autocovariance kernel of a demeaned Wiener process. Hence,

$$Y_{j,k} := \sum_{t=1}^T \left[\int_{(t-1)/T}^{t/T} \sqrt{2} \cos(\pi j s) ds \right] \widehat{u}_{t,1}, \quad j = 1, \dots, 13, \quad (7)$$

where the number of weights is restricted to 13, since Müller and Watson (2008) were only interested in lower than business cycle variability. Again, the nuisance parameter ω_e^2 cancels from the statistic asymptotically,

$$\mathcal{L}\mathcal{F}\mathcal{S}\mathcal{T}_1 := \frac{\sum_{j=1}^{13} Y_{j,k}^2}{\sum_{j=1}^{13} \frac{Y_{j,k}^2}{1+(g_1/j\pi)^2}}, \quad (8)$$

where $g_1 = 10$.

From Table 2 we learn that for $\phi = 0$, all tests are correctly sized or even under-sized

Table 2: Experimental size at level 5% for $M1$.

ϕ	$\mathcal{KPS}\mathcal{S}_1$	$\mathcal{LFS}\mathcal{T}_1$	$\mathcal{VR}_{1,15}$	$\mathcal{VR}_{1,25}$	$\mathcal{VR}_{1,50}$	$\mathcal{VR}_{1,100}$
$T = 100$						
-0.75	0.58	5.55	4.45	3.91	1.21	0.00
-0.50	1.30	5.56	4.34	3.52	1.21	0.00
-0.25	2.69	5.78	4.54	4.05	1.95	0.01
0.00	4.74	5.89	4.60	4.59	3.69	1.08
0.25	6.75	6.99	5.82	6.19	9.36	12.29
0.50	9.73	8.55	7.61	11.95	29.62	53.86
0.75	15.94	17.08	17.94	37.54	75.64	94.87
$T = 200$						
-0.75	1.59	5.78	4.68	4.76	3.84	1.30
-0.50	2.89	5.95	4.95	4.75	3.90	1.32
-0.25	3.35	6.35	5.50	4.66	4.35	1.96
0.00	4.72	5.89	4.84	4.66	4.42	3.46
0.25	6.41	6.44	5.16	5.38	6.54	10.21
0.50	7.39	6.26	5.28	6.52	12.16	30.31
0.75	12.67	9.82	9.05	16.59	42.68	80.34
$T = 500$						
-0.75	2.61	6.44	5.05	5.16	4.93	3.99
-0.50	3.35	6.28	4.90	5.08	4.83	4.35
-0.25	3.71	6.16	5.00	5.39	4.87	4.37
0.00	5.11	6.03	4.73	4.84	4.65	4.58
0.25	6.00	5.83	5.17	5.06	5.51	6.11
0.50	6.87	6.34	5.43	5.41	6.78	10.56
0.75	10.69	6.67	5.55	6.94	12.88	33.80
$T = 1000$						
-0.75	2.50	5.88	4.79	4.92	4.97	4.81
-0.50	3.04	5.95	5.08	5.01	5.30	5.12
-0.25	3.76	5.53	4.35	4.53	4.46	4.30
0.00	5.02	5.78	4.72	4.71	4.86	4.91
0.25	5.54	6.11	5.18	5.18	5.00	5.46
0.50	6.18	6.03	5.14	5.28	5.45	6.65
0.75	8.48	6.60	5.48	5.55	7.21	13.61

Note: The data are generated under H_0 with $e_t = \phi e_{t-1} + \varepsilon_t$.

in small samples. For $\phi > 0$, the case of particular interest, we find: \mathcal{KPSS}_1 is mildly oversized even in larger samples and outperformed by \mathcal{LFST}_1 . The size-distortion of $\mathcal{VR}_{1,q}$ is growing with q . $\mathcal{VR}_{1,15}$ and $\mathcal{VR}_{1,25}$ do not display notable distortions, being less distorted than \mathcal{KPSS}_1 , while $\mathcal{VR}_{1,50}$ and $\mathcal{VR}_{1,100}$ are distorted when the sample size is too small or the autocorrelation too strong. Generally, the closer ϕ is to a unit root the smaller has q to be (for fixed T) in order to control size.

For a power analysis we focus on $T = 500$ and $\phi \in \{0, 0.5\}$. For a fair comparison, size-adjusted power is presented in Figure 3. For $\phi = 0$, $\mathcal{VR}_{1,50}$ is more powerful than $\mathcal{VR}_{1,25}$, and both dominate \mathcal{KPSS}_1 , which is more powerful than \mathcal{LFST}_1 . The identical ranking is found for $\phi = 0.5$, only that the power curves are closer.

We conclude that the improved size performance of \mathcal{LFST}_1 relative to \mathcal{KPSS}_1 comes at a price in terms of power, while $\mathcal{VR}_{1,25}$ has similar size properties to \mathcal{LFST}_1 and is more powerful than \mathcal{KPSS}_1 at the same time.

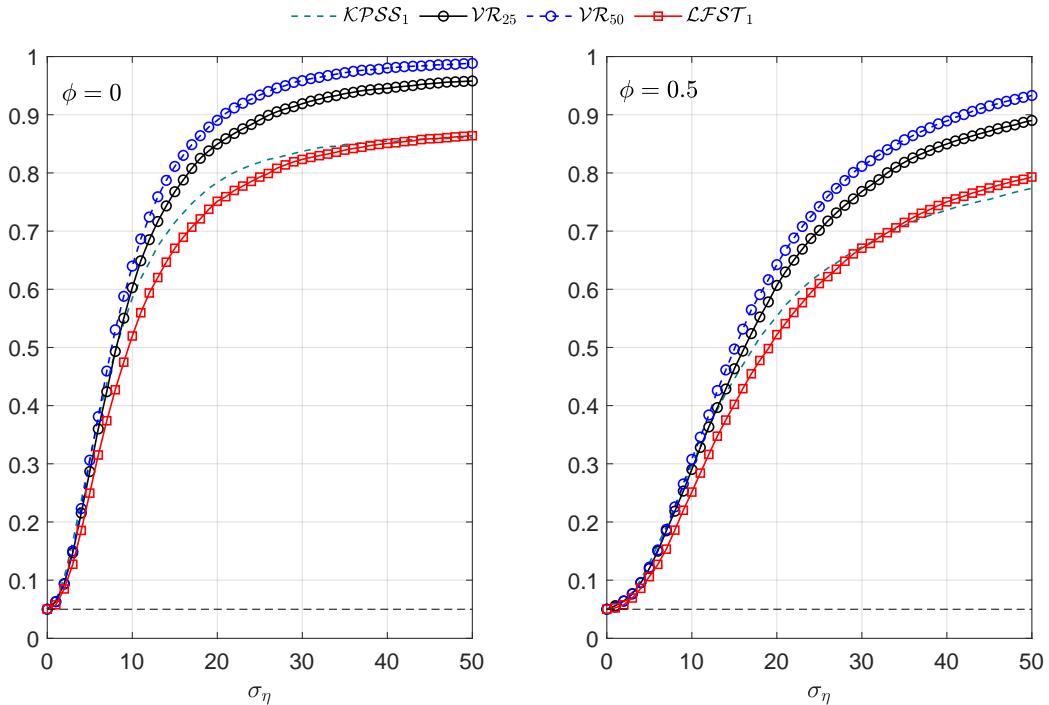


Figure 3: Size-adjusted power for $T = 500$ with demeaning; the process is from (1) with $e_t = \phi e_{t-1} + \varepsilon_t$.

5 Summary

We introduced a family of tests for stationarity against local alternatives of a random walk. It builds on variance ratios \mathcal{VR}_q , where q is the number of weighted sums of

CUSUM processes under three considered models: without deterministics, with a constant term, and with a linear time trend. The long-run variance as nuisance parameter cancels asymptotically from the ratios, which makes our procedure self-normalizing. The limit distribution involves quadratic forms of normal variates, and critical values have been computed and are provided in Table 1, and local power functions are available, too.

Asymptotically, q does not affect the size of the test, while the limiting power grows with q . Monte Carlo experiments, however, show that q may not be too large relative to a finite sample size and relative to the persistence in the stationary process under the null in order to obtain a test with correct size. While $q = 100$ or $q = 50$ may result in oversized tests, $q = 25$ balances size distortion and power over a wide range of autocorrelation and sample size. \mathcal{VR}_{25} is superior to the classic competitor by Kwiatkowski et al. (1992), not only in terms of power but also in terms of size distortion.

Appendix

First through fourth derivatives of f are denoted as f' , f'' , f''' and $f^{(4)}$.

Proof of Proposition 1

The integral equation associated with $K_0(s, t; \theta)$ is

$$f(t) = \lambda \int_0^1 f(s) K_0(s, t; \theta) ds,$$

which is equivalent to

$$f^{(4)}(t) = -\lambda f''(t) + \lambda \theta^2 f(t),$$

with $f(0) = 0$, $f''(1) = -\lambda f(1)$, $f''(0) = -(\lambda + \theta^2) f(0)$, and $f'''(1) = -\lambda f'(1)$. The solution of the latter differential equation reads

$$f(t) = c_1 \cos(\mu_1 t) + c_2 \sin(\mu_1 t) + c_3 \exp(\mu_2 t) + c_4 \exp(-\mu_2 t),$$

where

$$\mu_1 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} + \lambda}{2}} \quad \text{and} \quad \mu_2 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} - \lambda}{2}}.$$

Using the boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda) = \frac{2\theta^2 + (2\theta^2 + \lambda) \cos \mu_1 \cosh \mu_2 - \mu_1 \mu_2 \sin \mu_1 \sinh \mu_2}{4\theta^2 + \lambda}.$$

We can verify the following conditions:

$$\begin{aligned}
\lim_{\lambda \rightarrow 0} D_\theta(\lambda) &= 1, \\
\int K_0(t, t; \theta) dt &= \frac{1}{2} + \frac{1}{12}\theta^2 = -\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda), \\
\int \int K_0^2(s, t; \theta) ds dt &= \frac{1}{6} + \frac{11}{180}\theta^2 + \frac{11}{1680}\theta^4 \\
&= -\lim_{\lambda \rightarrow 0} \frac{\partial^2}{\partial \lambda^2} D_\theta(\lambda) + \left(\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda) \right)^2, \\
\lim_{R \rightarrow \infty} \frac{\log \log |D_\theta(Re^{i\varphi})|}{\log R} &< 1 \text{ (due to hyperbolic behavior)}.
\end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_0(s, t; \theta)$.

Now, using the first three boundary conditions together with $\int_0^1 f(t)^2 dt = 1$ we obtain

$$f(t) = \alpha_2 (\alpha_1 \cos(\mu_1 t) + \sin(\mu_1 t) + \alpha_3 \exp(\mu_2 t) + \alpha_4 \exp(-\mu_2 t)),$$

where

$$\begin{aligned}
\alpha_1 &= \frac{(\mu_2^2 - 2\theta^2)(1 + \cos \mu_1 \cosh \mu_2)}{\mu_1^2 \cosh \mu_2 \sin \mu_1 + \mu_1 \mu_2 \cos \mu_1 \sinh \mu_2}, \\
\alpha_2 &= \left(\frac{e^{-2\mu_2}}{4\mu_1 \mu_2 (\mu_1^2 + \mu_2^2)} \right)^{-1/2} \times \\
&\quad [2\alpha_4^2 (-1 + e^{2\mu_2}) \mu_1 (\mu_1^2 + \mu_2^2) + 8\alpha_4 e^{2\mu_2} \mu_1 \mu_2 (\mu_1 + \alpha_3 \mu_1^2 + \mu_2 (\alpha_1 + \alpha_3 \mu_2)) \\
&\quad + 2e^{2\mu_2} (4\alpha_3 \mu_1 \mu_2 (\mu_1 - \alpha_1 \mu_2) + \alpha_3^2 (-1 + e^{2\mu_2}) \mu_1 (\mu_1^2 + \mu_2^2) + (\alpha_1 + \mu_1 + \alpha_1^2 \mu_1) \mu_2 (\mu_1^2 + \mu_2^2)) \\
&\quad - 8\alpha_4 e^{\mu_2} \mu_2 \mu_1 ((\mu_1 + \alpha_1 \mu_2) \cos \mu_1 + (-\alpha_1 \mu_1 + \mu_2) \sin \mu_1) \\
&\quad + 8\alpha_3 e^{3\mu_2} \mu_2 \mu_1 ((-\mu_1 + \alpha_1 \mu_2) \cos \mu_1 + (\alpha_1 \mu_1 + \mu_2) \sin \mu_1) \\
&\quad + e^{2\mu_2} \mu_2 (\mu_1^2 + \mu_2^2) (-2\alpha_1 \cos(2\mu_1) + (-1 + \alpha_1^2) \sin(2\mu_1))]^{-1/2}, \\
\alpha_3 &= \frac{e^{\mu_2} \mu_1 (-\lambda + \mu_1^2) \mu_2 (\theta^2 + \lambda + \mu_2^2) + \mu_1 (\theta^2 - \lambda + \mu_1^2) (\lambda + \mu_2^2) (\mu_2 \cos \mu_1 - \mu_1 \sin \mu_1)}{\mu_2 (\lambda + \mu_2^2) (\theta^2 + \lambda + \mu_2^2) ((-1 + e^{2\mu_2}) \mu_2 \cos \mu_2 + (1 + e^{2\mu_2}) \mu_1 \sin \mu_1)}, \\
\alpha_4 &= \frac{e^{\mu_2} \mu_1 (-(\lambda - \mu_1^2) \mu_2 (\theta^2 + \lambda + \mu_2^2) + e^{\mu_2} (\theta^2 + \lambda - \mu_1^2) (\lambda + \mu_2^2) (\mu_2 \cos \mu_1 + \mu_1 \sin \mu_1))}{\mu_2 (\lambda + \mu_2^2) (\theta^2 + \lambda + \mu_2^2) ((-1 + e^{2\mu_2}) \mu_2 \cos \mu_1 + (1 + e^{2\mu_2}) \mu_1 \sin \mu_1)}.
\end{aligned}$$

Hence, the proof is complete.

Proof of Proposition 2

The Fredholm integral equation for $K_1(s, t; \theta)$ is equivalent to

$$f^{(4)}(t) = -\lambda f''(t) + \lambda \theta^2 f(t),$$

with boundary conditions $f(0) = 0$, $f(1) = 0$, $f''(0) = -\lambda f(0)$, $f''(1) = -\lambda f(1)$. The solution to the latter differential equation reads

$$f(t) = c_1 \cos(\mu_1 t) + c_2 \sin(\mu_1 t) + c_3 \exp(\mu_2 t) + c_4 \exp(-\mu_2 t),$$

where μ_1 and μ_2 are as defined under the proof of Proposition 1. Using the boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda) = \frac{\sin \mu_1 \sinh \mu_2}{\mu_1 \mu_2}.$$

We can verify the following conditions:

$$\begin{aligned} \lim_{\lambda \rightarrow 0} D_\theta(\lambda) &= 1, \\ \int K_1(t, t; \theta) dt &= \frac{1}{6} + \frac{1}{90}\theta^2, = -\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda), \\ \int \int K_1^2(s, t; \theta) ds dt &= \frac{1}{90} + \frac{2}{945}\theta^2 + \frac{1}{9450}\theta^4 \\ &= -\lim_{\lambda \rightarrow 0} \frac{\partial^2}{\partial \lambda^2} D_\theta(\lambda) + \left(\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda) \right)^2, \\ \lim_{R \rightarrow \infty} \frac{\log \log |D_\theta(Re^{i\varphi})|}{\log R} &< 1 \text{ (due to hyperbolic behavior).} \end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_1(s, t; \theta)$.

Further, using the first three boundary conditions together with $\int_0^1 f^2(t) dt = 1$ we obtain

$$f(t) = \sqrt{2} \sin(\mu_1 t),$$

as required to complete the proof.

Proof of Proposition 3

The Fredholm integral equation for $K_2(s, t; \theta)$ is equivalent to

$$\begin{aligned} f^{(4)}(t) &= -\lambda f''(t) + \lambda \theta^2 f(t) - 6\theta^2 \lambda a, \\ a &= \int_0^1 s(1-s) f(s) ds, \end{aligned}$$

with boundary conditions $f(0) = 0$, $f(1) = 0$ and

$$\begin{aligned} f''(0) &= -\lambda f(0) + \lambda \int_0^1 \frac{1}{10}(-1+s)(-5s^2\theta^2 + 5s^3\theta^2 + s(-60+\theta^2))f(s)ds, \\ f'''(0) &= -\lambda f'(0) + \lambda\theta^2 \int_0^1 (1-3s)(s-1)f(s)ds, \\ f'''(1) &= -\lambda f'(1) + \lambda\theta^2 \int_0^1 (3s-2)s f(s)ds. \end{aligned}$$

The solution to the latter differential equation reads

$$f(t) = c_1 \cos(\mu_1 t) + c_2 \sin(\mu_1 t) + c_3 \exp(\mu_2 t) + c_4 \exp(-\mu_2 t) + 6a.$$

Using the first boundary condition, $f(0) = 0$, we eliminate a from the rest of boundary conditions. Using the new boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda) = -\frac{144(-2+2\cos\mu_1+\mu_1\sin\mu_1)(2-2\cosh\mu_2+\mu_2\sinh\mu_2)}{\mu_1^4\mu_2^4}.$$

We can verify the following conditions:

$$\begin{aligned} \lim_{\lambda \rightarrow 0} D_\theta(\lambda) &= 1, \\ \int K_2(t, t; \theta) dt &= \frac{1}{15} + \frac{11}{12600}\theta^2 = -\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda), \\ \int \int K_2^2(s, t; \theta) ds dt &= \frac{11}{12600} + \frac{1}{27000}\theta^2 + \frac{509}{1164240000}\theta^4 \\ &= -\lim_{\lambda \rightarrow 0} \frac{\partial^2}{\partial \lambda^2} D_\theta(\lambda) + \left(\lim_{\lambda \rightarrow 0} \frac{\partial}{\partial \lambda} D_\theta(\lambda) \right)^2, \\ \lim_{R \rightarrow \infty} \frac{\log \log |D_\theta(Re^{i\varphi})|}{\log R} &< 1 \text{ (due to hyperbolic behavior).} \end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_2(s, t; \theta)$.

Finally using the boundary conditions together with $\int_0^1 f^2(t)dt = 1$ we obtain

$$f(t) = \begin{cases} \sqrt{2} \left(\cot \frac{\mu_1}{2} \cos(\mu_1 t) + \sin(\mu_1 t) - \cot \frac{\mu_1}{2} \right) & \text{if } \mu_1 \cos \frac{\mu_1}{2} - 2 \sin \frac{\mu_1}{2} = 0, \\ \sqrt{2} \sin(\mu_1 t) & \text{if } \sin \frac{\mu_1}{2} = 0. \end{cases}$$

as required to complete the proof.

References

- Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Faust, J. (1996). Near observational equivalence and theoretical size problems with unit root tests. *Econometric Theory* 12, 724–731.
- Hassler, U. and M. Hosseinkouchack (2019). Ratio tests under limiting normality. *Econometric Reviews* 38, 793–813.
- Jiang, B., Y. Lu, and J. Y. Park (2020). Testing for stationarity at high frequency. *Journal of Econometrics* 215, 341 – 374.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.
- Liu, W. and W. B. Wu (2010). Asymptotics of spectral density estimates. *Econometric Theory* 26, 1218–1245.
- Lu, Y. and J. Y. Park (2019). Estimation of longrun variance of continuous time stochastic process using discrete sample. *Journal of Econometrics* 210, 236 – 267.
- MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics* 6, 422–433.
- Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics* 141, 1331–1352.
- Müller, U. K. and M. W. Watson (2008). Testing models of low-frequency variability. *Econometrica* 76, 979–1016.
- Nabeya, S. (2001). Approximation to the limiting distribution of t- and F-statistics in testing for seasonal unit roots. *Econometric Theory* 17, 711–737.
- Nabeya, S. and K. Tanaka (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics* 16, 218–235.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

- Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.
- Nyblom, J. and T. Mäkeläinen (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association* 78, 856–864.
- Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
- Pötscher, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70, 1035–1065.
- Tanaka, K. (1983). Non-normality of the Lagrange multiplier statistic for testing the constancy of regression coefficients. *Econometrica* 51, 1577–1582.
- Tanaka, K. (1996). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley.