UNIVERSITY *of York*

Discussion Papers in Economics

No. 15/14

Autocorrelation robust inference using the Daniell kernel with fixed bandwidth

Javier Hualde and Fabrizio Iacone

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Autocorrelation robust inference using the Daniell kernel with fixed bandwidth

Javier Hualde
Universidad Pública de Navarra

Fabrizio Iacone*
University of York

August 20, 2015

Abstract

We consider alternative asymptotics for frequency domain estimates of the long run variance, in which the bandwidth is kept fixed. For a weakly dependent process, this does not yield a consistent estimate of the long run variance, but the standardized mean has t limit distribution, which, for any given bandwidth, appears to be more precise than the traditional Gaussian limit . In presence of fractionally integrated data, the limit distribution of the estimate is not standard, and we derive critical values for the standardized mean for various bandwidths. Again, we find that this asymptotic result provides a better approximation than other proposals like the Memory Autocorrelation Consistent (MAC) estimate. In multivariate set up, fixed bandwidth asymptotics may be also used to provide a characterization to the limit distribution of estimates of cointegrating parameter which differs substantially from the conventional Narrow Band asymptotics.

Keywords: long run variance estimation, long memory, large-m and fixed-m asymptotic theory.

JEL classification: C32

*Corresponding author: Fabrizio Iacone, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK.

1. INTRODUCTION

We consider inference for the mean of a time series with autocorrelation of unspecified nature. This problem has been widely studied for the case of weakly dependent processes (whose spectral density is finite and nonzero), where robust inference can be obtained by standardizing the sample average by the long run variance: this is usually unknown, but it can be consistently estimated by a range of techniques based either on weighted autocovariances or on weighted periodograms, see for example Priestley (1981).

In the present paper we emphasize the use of frequency domain techniques. In particular, the simplest frequency domain estimate of the long run variance can be obtained by direct averaging periodograms evaluated at the first m Fourier frequencies (which corresponds to using the Daniell kernel). The parameter m is known as bandwidth and, when discussing the limiting properties of this estimate, it is routinely assumed that $m \to \infty$, although at a rate slower than for the sample size T, so that the band is degenerating to 0. Throughout, we will denote the assumption $m \to \infty$ as the large-m approach. Here, we will consider instead an alternative approach to derive the asymptotic properties in which m is kept fixed. This is motivated by the fact that in any practical situation a finite m is used and, therefore, considering m fixed might yield a better asymptotic approximation. This approach will be denoted as fixed-m.

Related to this discussion, it is well known that frequency domain estimates may be approximated as time domain, and vice-versa. Actually, the fixed-m approach can be seen as a frequency domain analogue to the fixed-b approach for time domain estimates of the long run variance, see, e.g., Kiefer and Vogelsang (2002, 2005), Jansson (2004) and Sun, Phillips and Jin (2008). In fact, our approach is inspired by their work, in the sense that we advocate using the fixed-m limit to characterize more accurately the the finite sample properties of the standardized mean. Related works include also Sun (2013), who employed a non-parametric series method for estimating the long run variance which projects the observable series onto orthonormal basis functions. This leads to test statistics with standard limiting distributions under fixed-smoothing asymptotics, which correspond to using a fixed number of basis functions.

Despite these analogies, however, the fixed-m and fixed-b limits are different and, as we show in the paper, in the leading case of weakly dependent processes, one advantage of the fixed-m approach is that it does not require the simulation of a limit distribution, as opposed to what is necessary for fixed-b asymptotics. Note that even if consistency of the estimate of the long run variance cannot be justified under the fixed-m approach, for the particular weakly dependent case, the standardized mean has a t_{2m} limit distribution. We find in a Monte Carlo exercise that, in finite samples, for any given bandwidth, this limit distribution provides a better approximation to the distribution of the standardized sample mean than the Gaussian limit that is obtained using conventional asymptotics.

We allow in the paper for processes with general types of dependence, including the so-called long memory and antipersistent processes. In those cases, the spectral density at frequency zero is unbounded or 0, depending on whether the memory parameter is positive or negative, respectively. In this sense, our approach is close to Müller (2014), who also emphasized a frequency-domain perspective, but, instead of focusing on fractional processes, generated dependence by means of an autoregressive root close to unity. With large-m asymptotics, a semiparametric standardization of the test statistic is obtained using a Memory Autocorrelation Consistent (MAC) estimate of the variance of the sample mean, as in Robinson (2005). In comparison to the MAC standardization, which requires estimation of the memory of the process, the fixed-m asymptotics automatically adjust the rate of convergence of the sample mean to the fractional Brownian motion, to the rate of convergence of the estimate of the long run variance. This property is often called self-normalization, and, in the context of fractionally integrated processes, it has been considered by Shao (2011) and McElroy and Politis (2012, 2013), who employed a fixed-b approach to characterize the limiting behaviour of their time domain estimate of the long run variance. Under fractional integration (with positive or negative memory), the fixed-m limit of the standardized mean is not standard and we derive critical values for various bandwidths. We also find in a Monte Carlo exercise that this type of asymptotics provides a better approximation than with the feasible MAC standardization.

Further extensions of this approach are possible: in multivariate set up, it seems particularly interesting to consider using fixed bandwidth asymptotics to characterize the limit distribution for the estimate of the cointegrating parameter, in alternative to the conventional Narrow Band asymptotics.

The following section presents the standardized mean and discusses its large-m and fixed-m limits. In Section 3, we compare the large-m and fixed-m limiting approximations to the sampling distribution of the long run variance by means of a Monte Carlo experiment; in the same section, we also compare the large-m and fixed-m inference for weakly dependent and fractionally integrated processes. In Section 4, we conclude. The proofs of the theorems are given in the Appendix.

2. LARGE-m AND FIXED-m LIMITS OF THE STANDARDIZED MEAN

We consider the time series $x_1, ..., x_T$, observed from the stationary process

$$x_t := \mu + u_t,$$

where $E(u_t) = 0$ and u_t may be a weakly dependent process or, alternatively, fractionally integrated. We characterize this in Assumptions 1 and 2 below: we first introduce a weakly dependent process, and then we integrate it, so u_t is, in general, fractionally integrated.

Assumption 1. η_t is a linear process satisfying $\eta_t = A(L)\varepsilon_t := \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ where L is the usual lag operator. The weights $\{A_j\}$ are such that $A(1)^2 > 0$ and $\sum_{l=0}^{\infty} l\,|A_l| < \infty$ and ε_t is an independent, identically distributed (i.i.d.) sequence with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = 1$. We denote the long run variance as $\sigma^2 := A(1)^2$.

Assumption 2. Let $\Delta_t^{(\delta)} := \Gamma(t+\delta)/(\Gamma(\delta)\Gamma(t+1))$, $\Gamma(\cdot)$ denoting the Gamma function, such that $\Gamma(0) := \infty$ and $\Gamma(0)/\Gamma(0) := 1$. The process u_t is such that $u_t := \sum_{s=-\infty}^{t} \Delta_{t-s}^{(\delta)} \eta_s$, $\delta \in (-1/2, 1/2)$.

We consider inference on μ when the parametric structure of u_t is not known. In this case, the sample mean

$$\overline{x} := \frac{1}{T}\sum_{t=1}^{T} x_t$$

is a natural estimate of μ and, if u_t is weakly dependent, that is $u_t = \eta_t$ (or $\delta = 0$) and regularity conditions are met, inference on $\overline{x}$ can be based on the Central Limit Theorem (CLT)

$$\sqrt{T}(\overline{x} - \mu) \to_d N\left(0, \sigma^2\right).$$

Inference on μ could then be conducted using the standardization

$$\sqrt{T}(\overline{x} - \mu)/\sigma \to_d N(0,1). \tag{1}$$

Of course, in practice σ^2 is unknown, but a large number of semiparametric techniques are available to estimate it consistently, see, e.g., Priestley (1981). Letting

$$w_x(\lambda) := \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} x_t e^{i\lambda t}$$

be the Fourier transform of x_t, and the periodogram $I(\lambda) := |w_x(\lambda)|^2$, the Daniell kernel provides a very simple estimate of σ^2,

$$\widehat{\sigma}^2 := 2\pi \frac{1}{m}\sum_{j=1}^{m} I(\lambda_j), \text{ where } \lambda_j := \frac{2\pi j}{T}, \tag{2}$$

and m is known as the bandwidth. Feasible inference is then conducted using the statistic

$$\tau := \sqrt{T}\left(\overline{x}-\mu\right)/\sqrt{\widehat{\sigma}^2}. \tag{3}$$

When $m \to \infty$, $m/T \to 0$ and given other regularity conditions, the estimate $\widehat{\sigma}^2$ is consistent, and it can be substituted in (1) without altering the limit. We refer to this assumption as large-m asymptotics.

For a general u_t such that δ is not necessarily 0 and denoting by $f(\lambda)$ the spectral density of u_t, then, as $T \to \infty$,

$$Var\left(T^{1/2-\delta}\overline{x}\right) \to Gp(\delta),$$

where

$$\begin{aligned} p(\delta) &:= \begin{cases} 2\frac{\Gamma(1-2\delta)\sin(\pi\delta)}{\delta(1+2\delta)} \text{ if } \delta \neq 0, \\ 2\pi \text{ if } \delta = 0, \end{cases} \\ G &:= \lim_{\lambda\to 0} |\lambda|^{2\delta} f(\lambda). \end{aligned}$$

Let $\widehat{\delta}$ be a consistent estimate of δ: the feasible MAC standardized sample mean is

$$\theta = T^{1/2-\widehat{\delta}} \frac{(\overline{x}-\mu)}{\widehat{G} \times p\left(\widehat{\delta}\right)}, \tag{4}$$

where

$$\widehat{G} = \frac{1}{J}\sum_{j=1}^{J} \lambda_j^{2\widehat{\delta}} I(\lambda_j). \tag{5}$$

Under regularity conditions, which include $J \to \infty$, $J/T \to 0$, $\widehat{\delta} = \delta + o_p(\ln T)$, then $\theta \to_d N(0,1)$. The application of large-m asymptotics then requires the preliminary estimation of the order of integration and the construction of a specific test statistic for each value of δ. In contrast, in fixed-m asymptotics, we can still consider the statistic (3).

To derive the fixed-m asymptotic distribution of the estimate $\widehat{\sigma}^2$, we need to strenghten the moment conditions on ε_t.

Assumption 3. There is q such that $E\left(|\varepsilon_t|^q\right) < \infty$ with $q \geq \max\left(2, 2/(1+2\delta)\right)$.

Remark 1. Under Assumptions 1, 2 and 3, we have the Functional Central Limit

Theorem (FCLT) for fractional process: for $r \in [0,1]$, as $T \to \infty$,

$$\frac{1}{T^{1/2+\delta}} \sum_{t=1}^{\lfloor rT \rfloor} u_t \Rightarrow \Sigma_\delta W_{\delta+1}(r), \tag{6}$$

where $W_{\delta+1}(r)$ is a Type I fractional Brownian motion, as defined in Mandelbrot and Van Ness (1968),

$$\Sigma_\delta^2 := \sigma^2 \frac{\Gamma(1-2\delta)}{(1+2\delta)\Gamma(1+\delta)\Gamma(1-\delta)},$$

and $\lfloor \cdot \rfloor$ denotes integer part. For further details see Theorems 2.1 and 2.2 of Wang, Lin and Gulati (2003).

When $\delta = 0$, the process is weakly dependent and the limit (6) encompasses the standard convergence to the standard Brownian motion, $W(r)$, given by

$$\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} u_t \Rightarrow \sigma W(r),$$

as in Phillips and Solo (1992).

Let

$$\begin{aligned} \widehat{W}_{\delta+1}(r) &:= W_{\delta+1}(r) - rW_{\delta+1}(1), \\ Q_\delta(j) &:= \left\{ \left(2\pi j \int_0^1 \sin(2\pi j r) \widehat{W}_{\delta+1}(r)\, dr \right)^2 + \left(2\pi j \int_0^1 \cos(2\pi j r) \widehat{W}_{\delta+1}(r)\, dr \right)^2 \right\}. \end{aligned}$$

Our key result is Lemma 1 below.

Lemma 1. Under Assumption 1, 2 and 3, for fixed $j = 1, ..., m$, as $T \to \infty$,

$$T^{-2\delta} 2\pi I(\lambda_j) \Rightarrow \Sigma_\delta^2 Q_\delta(j). \tag{7}$$

By Lemma 1 and (6) we can establish the following theorem.

Theorem 2. Under Assumption 1, 2 and 3, for fixed m, as $T \to \infty$,

$$\tau \Rightarrow \frac{W_{\delta+1}(1)}{\sqrt{\frac{1}{m}\sum_{j=1}^m Q_\delta(j)}}.$$

When $\delta = 0$, it is well known that, under regularity conditions, the joint distribution of $I(\lambda_j)$, $j = 1, ..., m$, converges to that of m independent $2^{-1} f(0) \chi_2^2$ variates (see, e.g., Theorem 5.2.6 of Brillinger, 2001). Then, by the continuous mapping theorem and exploiting also the asymptotic independence of $I(0)$ and $I(\lambda_j)$, $j = 1, ..., m$, (see, e.g., Theorem 4.4.1 of Brillinger, 2001) it is straightforward to derive

Corollary 3. Under Assumption 1, 2 and $\delta = 0$, for fixed m, as $T \to \infty$,

$$\tau \to_d t_{2m}.$$

Remark 2. In related settings, the t limiting distribution has been already posed by Sun (2013) (Theorem 3.1) and Müller (2014, p.314).

Remark 3. Asymptotics for the Fourier transforms of possibly fractionally integrated processes are also given in Chen and Hurvich (2003) and Lahiri (2003). We find that (7) provides an easier way to simulate quantiles for the limit distribution of τ when $\delta \neq 0$.

Remark 4 Theorem 1 provides results for Type I fractionally processes. Results for Type II fractionally integrated processes may be obtained in the same way, simply replacing the type of fractional Brownian motion in the limit.

Remark 5. It is well known that weighted covariance estimates and spectral estimators are very similar, and it is interesting to discuss this more in detail. Denote the sample covariance of x_t,

$$\begin{aligned} c(l) \quad &: \quad = T^{-1} \textstyle\sum_{t=1}^{T-l} (x_t - \overline{x})(x_{t+l} - \overline{x}), \text{ for } l \geq 0, \\ &: \quad = T^{-1} \textstyle\sum_{t=1-l}^{T} (x_t - \overline{x})(x_{t+l} - \overline{x}), \text{ for } l < 0, \end{aligned}$$

the weighted covariance estimate of σ^2 is

$$\widehat{\sigma}_{WC} := \textstyle\sum_{l=-T+1}^{T-1} k(l/M)\, c(l), \tag{8}$$

where, for the Daniell kernel,

$$k(x) := \frac{\sin(\pi x)}{\pi x}, \tag{9}$$

and M is a bandwidth parameter such that $1 \leq M \leq T$. Then, for $K_M(\lambda) := (2\pi)^{-1} \sum_{|l|<T} k(l/M)\, e^{-il\lambda}$, the weighted covariance estimate has frequency domain representation

$$\int_{-\pi}^{\pi} K_M(\lambda)\, I^*(\lambda)\, d\lambda, \tag{10}$$

where $I^*(\lambda)$ is the periodogram of $x_t - \overline{x}$, and, in the case of the Daniell kernel, the spectral window $K_M(\lambda)$ takes value $M/(2\pi)$ when $-\pi/M \leq \lambda \leq \pi/M$ and 0 otherwise. Thus for $m = \lfloor T/(2M) \rfloor$, the estimate (2) is an approximation of (10) when (9) is used. Regularity conditions for consistent estimation of σ^2 for weakly autocorrelated series include $M \to \infty$ but $M/T \to 0$ as $T \to \infty$ for (8), or $m \to \infty$ but $m/T \to 0$ as

$T \to \infty$ for (2). Kiefer and Vogelsang (2005) on the other hand consider the alternative assumption for (8) $M = \lfloor bT \rfloor$, for a fixed b parameter, and derive the asymptotics under that assumption (denoted fixed-b versus the standard small-b approach, where it is assumed $b \to 0$). Noting that $M = \lfloor bT \rfloor$ in (10) corresponds in the approximation (2) to taking m as $\lfloor 1/(2b) \rfloor$, our fixed-m assumption is then closely related to the fixed-b assumption of Kiefer and Vogelsang (2005). However, despite the similarities, the weighted covariance and the weighted periodogram estimates are not equal, and Theorem 2 shows that under fixed-b and fixed-m asymptotics the limits are also different. It is worth noticing that, when $\delta = 0$, the fixed-b weighted covariance estimate of the long run variance is biased, whereas the average periodogram estimate is asymptotically unbiased. One clear advantage of using the Daniell kernel average periodogram estimate (2) is that the limit distribution in Corollary 3 is well known, so that the simulation of critical values is not necessary, as opposed to when weighted covariance estimates are used.

Remark 6. In a multivariate set up, fixed bandwidth asymptotics may be used to characterize the limit distribution in alternative to standard Narrow Band asymptotics of Robinson and Marinucci (2001). In view of the limit (7), we conjecture that, keeping m fixed, the Narrow Band least squares attains the fastest rate among those of first stage estimators (see Hualde and Iacone, 2012). Moreover, it would be possible to characterize the limit distribution of the estimate even for combinations of the orders of integration not considered in Robinson and Marinucci (2001).

In many cases, it is possible to assume that the series is weakly autocorrelated, and Corollary 3 then gives a well known limit distribution. If this is not the case, however, the quantiles of τ must be simulated. We simulated the limit distribution of τ approximating integrals with summations over 10,000 steps. We simulated the distribution for $\delta \in \{-0.49, -0.4, -0.3, ..., 0.3, 0.4, 0.49\}$ and $m \in \{1, 2, ..., 16\}$: for each case, we repeated the simulation 10,000 times. In Table 1 we report the values cv such that $P(|\tau| > cv) = 0.05$. Notice that we also simulated the distribution for $\delta = 0$: in that case, the simulation is not necessary, but we made for completeness to compare the simulated values with the quantiles from the t_{2m} distribution.

3. FINITE SAMPLE PERFORMANCE

We analyse the reliability of the fixed-m approximation in a range of Monte Carlo exercises. First, we verify the limits in Lemma 1, Theorem 2 and Corollary 3 in a large sample. In the second simulation, we consider the case of weakly autocorrelation, and compare the asymptotics from Theorem 2 with the Gaussian limit in finite samples. In the third exercise, we do not assume knowledge of δ anymore, and we compare the τ

statistic with a feasible MAC standardized sample mean.

3.1. Finite sample approximations to $Q_\delta(j)$ and t_{2m}

In the first exercise, we simulate the periodograms and the limit distributions $Q_\delta(j)$ in Lemma 1, also comparing them with the χ_2^2 limit. In detail, we simulate the periodogram $I(\lambda_j)$ both in case u_t is an i.i.d. $N(0,1)$ (Nid$(0,1)$) process, and when it is an AR(1), $u_t = \eta_t = \phi\eta_{t-1} + \varepsilon_t$, for ε_t Nid$(0,1)$ and $\phi = 0.5$. In all cases, we set $T = 500$ in the computation of the periodogram, with 5,000 repetitions, and we simulated $Q_\delta(j)$ approximating integrals with summations over 1,000 steps, with 50,000 repetitions. In all cases, we used $\delta = 0$ so that we were also able to compare these two distributions with the χ_2^2 limit (appropriately scaled). Results in Figure 1 are for $j = 1$ and $j = 3$, and are strongly supportive of the approximation in Lemma 1.

In the second part of this exercise, we used the same artificial data to simulate the τ statistic with $m = 1$ and $m = 3$, to compare it with the t_{2m} and standard normal distributions: the results are in Figure 2, and they convincingly show that the approximation from Corollary 3 is more appropriate than using the standard normal.

3.2. Fixed-m robust inference in presence of weakly autocorrelated errors.

In the second exercise, we consider a weakly autocorrelated series, and we study if the fixed-m limit allows for a more precise inference than the standard, large-m asymptotic. Again we simulate an AR(1) process $u_t = \eta_t = \phi\eta_{t-1} + \varepsilon_t$, for ε_t Nid$(0,1)$ and $\phi = 0.5$, but in this experiment we limit the sample size to $T = 64$ or $T = 256$ only. We repeated the experiment 10,000 times. For each case, we compute the τ statistic and we count the frequency with which the realization $|\tau|$ is above the 95th percentile: this is of course like testing the null hypothesis on μ with theoretical size 5% and using both the t_{2m} and the $N(0,1)$ distributions, so we refer to these as rejection frequencies.

The results of the simulation are in Table 2: we find that in all cases using the t_{2m} the empirical size is closer to 5%, so we recommend using this distribution in testing. Note that this is the case even with relatively large values for m, for example $m = \sqrt{T}$.

The second interesting finding from our experiment is that the size is more precisely approximated the smaller is m: this is not a serious problem when $T = 256$, but it is important in the experiment having $T = 64$, thus suggests that a low value for m should be considered, when the sample size is small. It is worth noting that both results are consistent with findings in Kiefer and Vogelsang (2005) for fixed-b estimates.

For the second part of the exercise, we consider the alternative problem of distinguishing between a series having $E(x_t) = \mu$ and $E(x_t) = \mu + \beta$ for $|\beta| > 0$ in finite samples by means of the τ statistic. To avoid any size distortion that may be due to autocorrelation, we generated $u_t = \varepsilon_t$, for ε_t Nid$(0,1)$, and considered $\beta = \kappa T^{-1/2}$ for

$\kappa \in \{0, 1, 2, ..., 9\}$ to study the local power for $T = 64$ and $T = 256$.

Results are in Figure 3: these are for $m = 1$, $m = 3$ and $m = 5$ when $T = 64$, and for $m = 1$, $m = 4$ and $m = 16$ when $T = 256$, and 10,000 replications; in both plots we also include $m =inf$ for the case in which the correct $\sigma^2 = 1$ and limit normal distribution are used, as this indicates the maximum power that can be attained in this exercise. Again our results are consistent with Kiefer and Vogelsang (2005) in the sense that a larger bandwidth m is associated with better power: taken together with Table 2, these confirm that there is a trade off between size and power, although it seems that this is mostly limited to the smallest sample, as for $T = 256$ the size correction comes with nearly any power loss.

3.3. Fixed-m robust inference in presence of fractionally integrated errors.

In our third exercise, we allow for fractionally integrated processes as well. In this case, it is not realistic to assume knowledge of δ and, in line with our semiparametric approach, we use the local Whittle estimate. We refer to Robinson (1995) for a formula and references, but we note that this estimate is based on smoothing the periodogram on the lowest frequencies, thus requiring a bandwidth: we refer to the bandwidth used in the local Whittle estimation of δ as J. Simulation results, such as in Abadir, Distaso and Giraitis (2009), recommend $J = \lfloor T^{0.65} \rfloor$, so this is the one we adopt. For simplicity, this is also the bandwidth we use in the computation of $\widehat{G}$ (see (5)) to be used in the feasible MAC standardized mean (4), θ: Abadir, Distaso and Giraitis (2009) do find that a larger bandwidth ($\lfloor T^{0.8} \rfloor$) may give a slightly superior estimate of G in mean squared error sense (while keeping $J = \lfloor T^{0.65} \rfloor$ in the estimation of δ), but the advantage seems limited and, given that the authors also recommend to experiment with shorter bandwidths anyway, using $J = \lfloor T^{0.65} \rfloor$ in the whole formula seems convenient.

We simulated a Gaussian ARFIMA(1,δ,0), $u_t = \Delta^{-\delta}\eta_t$ with $\eta_t = \phi\eta_{t-1} + \varepsilon_t$, ε_t Nid$(0, 1)$, with $\phi = -0.5$, 0, 0.5, and $\delta = -0.3$, 0, 0.3, and we computed the τ and θ statistics (for the τ statistic, we used $m = 1$, ..., $m = 16$). We considered samples of $T = 64, 256, 1,024$, and for each combination we simulate 10,000 repetitions. For the MAC standardized mean, we compute the infeasible statistic $\overline{\theta}$, where the correct δ was used and the feasible one θ as in (4). As mentioned before, the computation of τ does not require knowledge of δ, so τ is feasible. However, the fixed-b limit of τ depends on δ , so we will provide results for the unfeasible and feasible cases, where τ is compared to critical values calculated from the true and estimated δ, respectively. Note that the feasible case consists on a plug-in approach for obtaining critical values (as in, e.g., McElroy and Politis, 2012, 2013)

A summary of the results is in Tables 3-12. We present first results for $\delta = 0$

and $\phi = 0$, to better appreciate the effect of estimating δ. This causes a relevant size distortion, especially in the smallest samples and for larger the bandwidths m, and, of course, for the feasible θ, for which the asymptotic approximation is worse. The first conclusion that we draw from this experiment is then that the advantage of using fixed-m asymptotic is more important if fractional integration is also considered.

In the reminder of the experiment we analyse the effect of altering δ or ϕ. In summary, changing δ does not alter much the empirical size, but changing ϕ has a more relevant effect, with $\phi > 0$ reducing the empirical size (for given T and δ), and $\phi < 0$ increasing it. We conjecture that these results may be due to the lower order bias in the estimation of δ caused by the weak autocorrelation in η_t. When $\phi > 0$, the estimate $\widehat{\delta}$ is affected by a lower order bias with positive sign: as the quantiles in Table 1 are increasing in δ, this positive bias results in incorrectly selecting for the test based in the τ statistic a critical value that is higher than it would be optimal, thus distorting the size of the τ test towards 0. The size distortion for the MAC normalized sample mean is also very relevant: Wright (1998) showed that the usual Wald statistic, which is designed to cater for $\delta = 0$, converges to 0 in presence of positive fractional integration, and we conjecture that this argument also applies to the lower order bias induced by a positive weak autocorrelation. Of course these arguments are reversed in presence of negative autocorrelation, and this is actually the most important reason for concern, as it may generate a relevant size increase, as we indeed find in the tables. With respect to this, it is encouraging to see that the size is still reasonable if fixed-m asymptotics instead of large-m (and MAC standardization) are used, especially when m is kept small.

4. CONCLUSION

We analysed inference for the mean of a time series with autocorrelation of unspecified nature and that may be fractionally integrated. We considered a standardization of the sample mean obtained averaging the first m periodograms, as it usually done for the estimation of the long run variance using the Daniell kernel, and we derive a different type of asymptotics, in which m is kept fixed. We refer to these as fixed-m asymptotics. When the time series is weakly autocorrelated, the sample mean standardized in this way has limit t_{2m} distribution; when $\delta \neq 0$ the test statistic is self-normalizing, thus avoiding the need of estimating δ, as it is otherwise required when large m asymptotics are applied. The limit distribution when $\delta \neq 0$ depends on δ and it must be tabulated, and we provide quantiles for it. We compare fixed-m and large-m asymptotics in a Monte Carlo experiment, finding that the former allow for a better approximation of the size in finite samples. We also find that, if fractional integration is allowed for, then large-m asymptotics are more heavily affected by the size distortion in the estimation of

δ, thus making the application of fixed-m asymptotics even more preferable. As for the choice of m, we find that the size is better approximated the smaller is m, but also that there is a trade-off between size precision and power, as larger m are associated with bigger power.

Appendix.

Proof of Lemma 1.

Let

$$\widehat{x}_t = x_t - \overline{x}.$$

From (6),

$$\begin{aligned} &\frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}\widehat{x}_t = \frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}(u_t-\overline{u}) = \frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}u_t - \frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}\overline{u} \\ = &\frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}u_t - \frac{\lfloor rT\rfloor}{T^{1/2+\delta}}\frac{1}{T}\sum_{t=1}^{T}u_t = \frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}u_t - \frac{\lfloor rT\rfloor}{T}\frac{1}{T^{1/2+\delta}}\sum_{t=1}^{T}u_t \\ \Rightarrow\ &\Sigma_\delta\left(W_{\delta+1}(r) - rW_{\delta+1}(1)\right). \end{aligned}$$

Noting that for $j=\pm1,....,\pm m$, $\sum_{t=1}^{T} e^{i\lambda_j t}=0$, for $j=1,...,m$, it also holds that

$$w_x(\lambda_j) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\widehat{x}_t e^{i\lambda_j t},$$

and, using $e^{ix}=\cos x + i\sin x$,

$$\begin{aligned} I(\lambda_j) &= \frac{1}{2\pi T}\left(\sum_{t=1}^{T}\widehat{x}_t(\cos\lambda_j t + i\sin\lambda_j t)\right)\left(\sum_{s=1}^{T}\widehat{x}_s(\cos\lambda_j s - i\sin\lambda_j s)\right) \\ &= \frac{1}{2\pi T}\left\{\left(\sum_{t=1}^{T}\widehat{x}_t\cos\lambda_j t\right)^2 + \left(\sum_{s=1}^{T}\widehat{x}_s\sin\lambda_j s\right)^2\right\}. \end{aligned}$$

By summation by parts,

$$\begin{aligned} 2\pi I(\lambda_j) &= \left(\sum_{t=1}^{T-1}(\cos\lambda_j(t+1)-\cos\lambda_j t)\frac{1}{\sqrt{T}}\sum_{s=1}^{t}\widehat{x}_s\right)^2 \\ &\quad+\left(\sum_{t=1}^{T-1}(\sin\lambda_j(t+1)-\sin\lambda_j t)\frac{1}{\sqrt{T}}\sum_{s=1}^{t}\widehat{x}_s\right)^2, \end{aligned}$$

where we used $\sum_{s=1}^{T} \widehat{x}_s = 0$. Next, by the mean value theorem

$$\begin{aligned}
\cos \lambda_j (t+1) &= \cos \lambda_j t - \lambda_j \sin \lambda_j t + O\left(\lambda_j^2\right), \\
\sin \lambda_j (t+1) &= \sin \lambda_j t + \lambda_j \cos \lambda_j t + O\left(\lambda_j^2\right),
\end{aligned}$$

so,

$$\begin{aligned}
T^{-2\delta} 2\pi I(\lambda_j) &= \left(\sum_{t=1}^{T-1} \left(-\frac{2\pi j}{T} \sin\left(2\pi j \frac{t}{T}\right) + O\left(\lambda_j^2\right) \right) \frac{1}{T^{1/2+\delta}} \sum_{s=1}^{t} \widehat{x}_s \right)^2 \\
&\quad + \left(\sum_{t=1}^{T-1} \left(\frac{2\pi j}{T} \cos\left(2\pi j \frac{t}{T}\right) + O\left(\lambda_j^2\right) \right) \frac{1}{T^{1/2+\delta}} \sum_{s=1}^{t} \widehat{x}_s \right)^2,
\end{aligned}$$

and, as $T \to \infty$,

$$T^{-2\delta} 2\pi I(\lambda_j) \Rightarrow \Sigma_\delta^2 \left\{ \left(2\pi j \int_0^1 \sin(2\pi j r) \widehat{W}_{\delta+1}(r)\, dr \right)^2 + \left(2\pi j \int_0^1 \cos(2\pi j r) \widehat{W}_{\delta+1}(r)\, dr \right)^2 \right\},$$

noting that the contribution due to the $O\left(\lambda_j^2\right)$ terms is of smaller order.

Proof of Theorem 2.

Theorem 2 follows from Lemma 1, (6) and the continuous mapping theorem.

REFERENCES

Abadir, K.M., W. Distaso and L. Giraitis, 2009. Two estimators of the long-run variance: beyond short memory. *Journal of Econometrics* 150, 56-70.

Brillinger, D.R., 2001. *Time series. Data analysis and theory.* SIAM, Philadelphia.

Chen, W.W., and C.M. Hurvich, 2003. Estimating fractional cointegration in the presence of polynomial trends, *Journal of Econometrics* 117, 95-121.

Hualde, J., and F. Iacone, 2012. First stage estimation of fractional cointegration, *Journal of Time Series Econometrics* 4, Issue 1, Article 2.

Jansson, M., 2004. The error in rejection probability of simple autocorrelation robust tests, *Econometrica* 72, 937-946.

Kiefer, N.M., and T.J. Vogelsang, 2002. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size, *Econometric Theory* 18, 1350-1366.

Kiefer, N.M., and T.J. Vogelsang, 2005. A new asymptotic theory for heteroskedasticity - autocorrelation robust tests. *Econometric Theory* 21, 1130-64.

Lahiri, S.N., 2003. A necessary and sufficient condition for asymptotic independence of discrete Fourier Transforms under short- and long-range dependence, *Annals of Statistics* 31, 613-641.

Mandelbrot, B.B., and J.W. Van Ness, 1968. Fractional Brownian motion, fractional noise and applications, *SIAM Review* 10, 422-437.

McElroy, T., and D.N. Politis, 2012. Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory, *Econometric Theory* 28, 471-481.

McElroy, T., and D.N. Politis, 2013. Distribution theory for the studentized mean for long, short, and negative memory time series, *Journal of Econometrics* 177, 60-74.

Müller, U.K., 2014. HAC corrections for strongly autocorrelated time series, *Journal of Business and Economic Statistics* 32, 311-322.

Phillips, P.C.B., and V. Solo, 1992. Asymptotics for Linear Processes, *The Annals of Statistics* 20, 971-1001

Priestley, M.B., 1981. *Spectral Analysis and Time Series*, Academic Press, London.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630-1661.

Robinson, P.M., 2005. Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* 21, 171 180.

Robinson, P.M., and D. Marinucci, 2001. Narrow band analysis of nonstationary processes, *Annals of Statistics* 29, 947-86.

Shao, X., 2011. A simple test of changes in mean in the possible presence of long-range dependence, *Journal of Time Series Analysis* 32, 598–606.

Sun, Y., 2013. A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator, *Econometrics Reviews* 16, 1-26.

Sun, Y., P.C.B. Phillips and S. Jin, 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica* 76, 175-194.

Wang, Q.,Y.X. Lin, and C.M. Gulati, 2003. Asymptotics for general fractionally integrated processes with applications to unit root tests, *Econometric Theory* 19, 143-164.

Wright, J.H., 1998. Testing for a structural break at unknown date with long memory disturbances. *Journal of Time Series Analysis* 19, 369–376.

Table 1. Upper 5% quantile for $|\tau|$ for $m \in \{1, ..., 16\}$

$\delta \setminus m$	1	2	3	4	5	6	7	8
-0.49	2.725	1.646	1.332	1.185	1.076	0.996	0.941	0.895
-0.4	2.891	1.710	1.392	1.221	1.110	1.037	0.977	0.929
-0.3	3.136	1.854	1.527	1.334	1.241	1.169	1.107	1.055
-0.2	3.425	2.090	1.728	1.559	1.462	1.394	1.332	1.291
-0.1	3.849	2.401	2.040	1.886	1.790	1.719	1.680	1.644
0	4.421	2.848	2.470	2.310	2.237	2.204	2.168	2.135
0.1	5.171	3.411	3.063	2.955	2.888	2.877	2.857	2.850
0.2	6.289	4.325	3.961	3.913	3.888	3.874	3.940	3.938
0.3	8.086	5.724	5.397	5.413	5.401	5.472	5.612	5.719
0.4	12.358	8.748	8.442	8.689	8.893	9.173	9.427	9.645
0.49	41.485	30.391	30.180	31.282	32.339	33.676	34.777	35.990

$\delta \setminus m$	9	10	11	12	13	14	15	16
-0.49	0.854	0.821	0.787	0.762	0.735	0.711	0.692	0.672
-0.4	0.885	0.853	0.824	0.797	0.774	0.752	0.733	0.714
-0.3	1.022	0.984	0.955	0.928	0.910	0.888	0.876	0.855
-0.2	1.266	1.228	1.199	1.172	1.153	1.141	1.117	1.109
-0.1	1.625	1.597	1.571	1.547	1.530	1.505	1.501	1.482
0	2.114	2.105	2.097	2.087	2.066	2.070	2.067	2.061
0.1	2.843	2.855	2.858	2.857	2.872	2.886	2.909	2.904
0.2	3.973	4.007	4.030	4.059	4.104	4.148	4.197	4.229
0.3	5.792	5.885	6.008	6.125	6.181	6.289	6.364	6.468
0.4	9.847	10.114	10.369	10.532	10.737	10.968	11.170	11.339
0.49	37.112	38.193	39.272	40.218	41.149	42.089	42.989	43.799

Table 2. Rejection frequencies for different asymptotics

T	m	1	2	3	4	5	6	7	8
64	t_{2m}	0.052	0.053	0.056	0.063	0.069	0.076	0.084	0.093
64	$N(0,1)$	0.188	0.126	0.112	0.104	0.104	0.107	0.111	0.116
256	t_{2m}	0.049	0.047	0.049	0.047	0.048	0.051	0.050	0.051
256	$N(0,1)$	0.186	0.118	0.095	0.084	0.076	0.073	0.069	0.068
	m	9	10	11	12	13	14	15	16
64	t_{2m}	0.099	0.107	0.116	0.124	0.132	0.141	0.150	0.157
64	$N(0,1)$	0.120	0.129	0.136	0.144	0.151	0.158	0.164	0.172
256	t_{2m}	0.050	0.052	0.052	0.054	0.054	0.056	0.058	0.058
256	$N(0,1)$	0.067	0.064	0.066	0.067	0.066	0.067	0.068	0.067

Table 3. Rejection frequencies for the unfeasible case, $\delta = 0$, $\phi = 0$

T	m	1	2	3	4	5	6	7	8
64		0.049	0.048	0.048	0.051	0.051	0.050	0.050	0.049
256		0.048	0.043	0.046	0.047	0.046	0.047	0.046	0.046
1,024		0.047	0.050	0.048	0.049	0.047	0.048	0.046	0.048
	m	9	10	11	12	13	14	15	16
64		0.049	0.050	0.049	0.050	0.049	0.048	0.049	0.048
256		0.046	0.044	0.044	0.044	0.046	0.045	0.045	0.046
1,024		0.047	0.045	0.044	0.044	0.046	0.044	0.044	0.044

T	$\overline{\theta}$
64	0.061
256	0.051
1024	0.051

Table 4. Rejection frequencies for the feasible case, $\delta = 0$, $\phi = 0$

T	m	1	2	3	4	5	6	7	8
64		0.068	0.084	0.094	0.102	0.109	0.112	0.114	0.116
256		0.055	0.060	0.067	0.072	0.074	0.075	0.075	0.077
1,024		0.050	0.056	0.059	0.062	0.062	0.063	0.063	0.063
	m	9	10	11	12	13	14	15	16
64		0.118	0.119	0.118	0.119	0.120	0.120	0.121	0.125
256		0.078	0.080	0.080	0.080	0.081	0.081	0.082	0.082
1,024		0.063	0.063	0.064	0.063	0.065	0.064	0.063	0.065

T	θ
64	0.124
256	0.096
1024	0.076

Table 5. Rejection frequencies for the unfeasible case, $\delta = -0.3$, $\phi = -0.5$

T	m	1	2	3	4	5	6	7	8
64		0.046	0.051	0.050	0.056	0.056	0.054	0.053	0.052
256		0.048	0.048	0.048	0.052	0.050	0.050	0.051	0.053
1024		0.046	0.052	0.047	0.053	0.052	0.051	0.051	0.055

	m	9	10	11	12	13	14	15	16
64		0.052	0.051	0.049	0.047	0.045	0.042	0.038	0.036
256		0.050	0.052	0.053	0.054	0.052	0.052	0.051	0.052
1024		0.052	0.053	0.052	0.054	0.053	0.053	0.052	0.053

T	$\overline{\theta}$
64	0.044
256	0.048
1024	0.048

Table 6. Rejection frequencies for the feasible case, $\delta = -0.3$, $\phi = -0.5$

T	m	1	2	3	4	5	6	7	8
64		0.054	0.065	0.066	0.067	0.066	0.064	0.063	0.060
256		0.053	0.055	0.056	0.059	0.057	0.058	0.059	0.058
$1,024$		0.049	0.055	0.051	0.055	0.056	0.055	0.056	0.057

	m	9	10	11	12	13	14	15	16
64		0.060	0.057	0.055	0.054	0.049	0.044	0.041	0.041
256		0.057	0.058	0.059	0.059	0.058	0.060	0.061	0.060
$1,024$		0.057	0.057	0.055	0.056	0.056	0.057	0.055	0.057

T	θ
64	0.022
256	0.040
1024	0.052

Table 7. Rejection frequencies for the unfeasible case, $\delta = -0.3$, $\phi = 0.5$

T	m	1	2	3	4	5	6	7	8
64		0.048	0.049	0.051	0.060	0.066	0.071	0.080	0.091
256		0.049	0.045	0.043	0.046	0.043	0.045	0.046	0.047
1024		0.046	0.049	0.046	0.048	0.047	0.047	0.045	0.048

	m	9	10	11	12	13	14	15	16
64		0.098	0.113	0.123	0.135	0.144	0.156	0.165	0.176
256		0.045	0.047	0.049	0.051	0.051	0.054	0.052	0.055
1024		0.047	0.047	0.046	0.047	0.046	0.046	0.045	0.045

T	$\overline{\theta}$
64	0.128
256	0.081
1024	0.062

Table 8. Rejection frequencies for the feasible case, $\delta = -0.3$, $\phi = 0.5$

T	m	1	2	3	4	5	6	7	8
64		0.036	0.034	0.031	0.030	0.028	0.027	0.027	0.027
256		0.042	0.032	0.030	0.028	0.025	0.025	0.023	0.024
1024		0.043	0.043	0.039	0.037	0.034	0.033	0.035	0.033
	m	9	10	11	12	13	14	15	16
64		0.027	0.027	0.028	0.029	0.030	0.032	0.033	0.036
256		0.021	0.020	0.020	0.018	0.018	0.018	0.018	0.018
1024		0.033	0.032	0.030	0.030	0.029	0.028	0.028	0.029

T	θ
64	0.031
256	0.022
1024	0.026

Table 9. Rejection frequencies for the unfeasible case, $\delta = 0.3$, $\phi = -0.5$

T	m	1	2	3	4	5	6	7	8
64		0.049	0.047	0.050	0.049	0.051	0.051	0.048	0.047
256		0.047	0.044	0.044	0.045	0.048	0.048	0.048	0.046
$1,024$		0.051	0.053	0.051	0.053	0.053	0.053	0.050	0.051
	m	9	10	11	12	13	14	15	16
64		0.047	0.045	0.041	0.040	0.039	0.038	0.036	0.034
256		0.048	0.047	0.046	0.046	0.046	0.046	0.046	0.046
$1,024$		0.052	0.052	0.051	0.051	0.052	0.051	0.052	0.051

T	$\overline{\theta}$
64	0.040
256	0.041
1024	0.049

Table 10. Rejection frequencies for the feasible case, $\delta = 0.3$, $\phi = -0.5$

T	m	1	2	3	4	5	6	7	8
64		0.111	0.160	0.186	0.202	0.211	0.217	0.222	0.226
256		0.068	0.090	0.099	0.107	0.112	0.113	0.115	0.117
1024		0.054	0.060	0.061	0.062	0.064	0.063	0.063	0.063
	m	9	10	11	12	13	14	15	16
64		0.229	0.231	0.232	0.232	0.233	0.231	0.231	0.233
256		0.118	0.119	0.120	0.122	0.123	0.124	0.124	0.125
1024		0.065	0.065	0.064	0.064	0.066	0.065	0.065	0.065

T	θ
64	0.244
256	0.145
1024	0.101

Table 11. Rejection frequencies for the unfeasible case, $\delta = 0.3$, $\phi = 0.5$

T	m	1	2	3	4	5	6	7	8
64		0.051	0.052	0.055	0.059	0.066	0.072	0.075	0.081
256		0.044	0.046	0.046	0.046	0.050	0.050	0.049	0.049
1,024		0.049	0.053	0.052	0.054	0.054	0.056	0.054	0.052
	m	9	10	11	12	13	14	15	16
64		0.088	0.095	0.098	0.103	0.110	0.113	0.119	0.121
256		0.050	0.051	0.050	0.050	0.051	0.052	0.052	0.052
1,024		0.053	0.053	0.052	0.051	0.053	0.053	0.052	0.052

T	$\overline{\theta}$
64	0.154
256	0.097
1024	0.073

Table 12. Rejection frequencies for the feasible case, $\delta = 0.3$, $\phi = 0.5$

T	m	1	2	3	4	5	6	7	8
64		0.025	0.030	0.031	0.030	0.031	0.031	0.031	0.031
256		0.023	0.023	0.023	0.022	0.023	0.023	0.022	0.022
1,024		0.034	0.033	0.031	0.031	0.030	0.030	0.028	0.027
	m	9	10	11	12	13	14	15	16
64		0.032	0.032	0.032	0.033	0.033	0.034	0.034	0.035
256		0.021	0.021	0.021	0.020	0.020	0.019	0.019	0.019
1,024		0.027	0.026	0.025	0.024	0.024	0.024	0.024	0.024

T	θ
64	0.038
256	0.026
1024	0.031

Figure 1(a) Periodogram Finite sample and Limit distributions, i.i.d., j=1

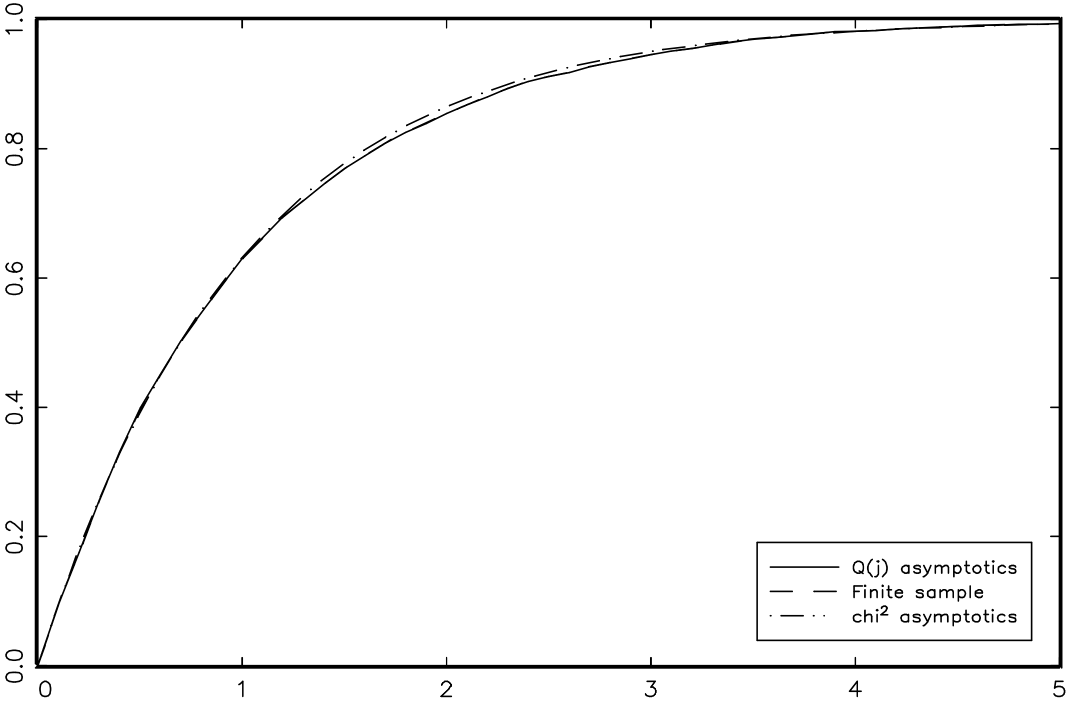

Figure 1(b) Periodogram Finite sample and Limit distributions, i.i.d., j=3

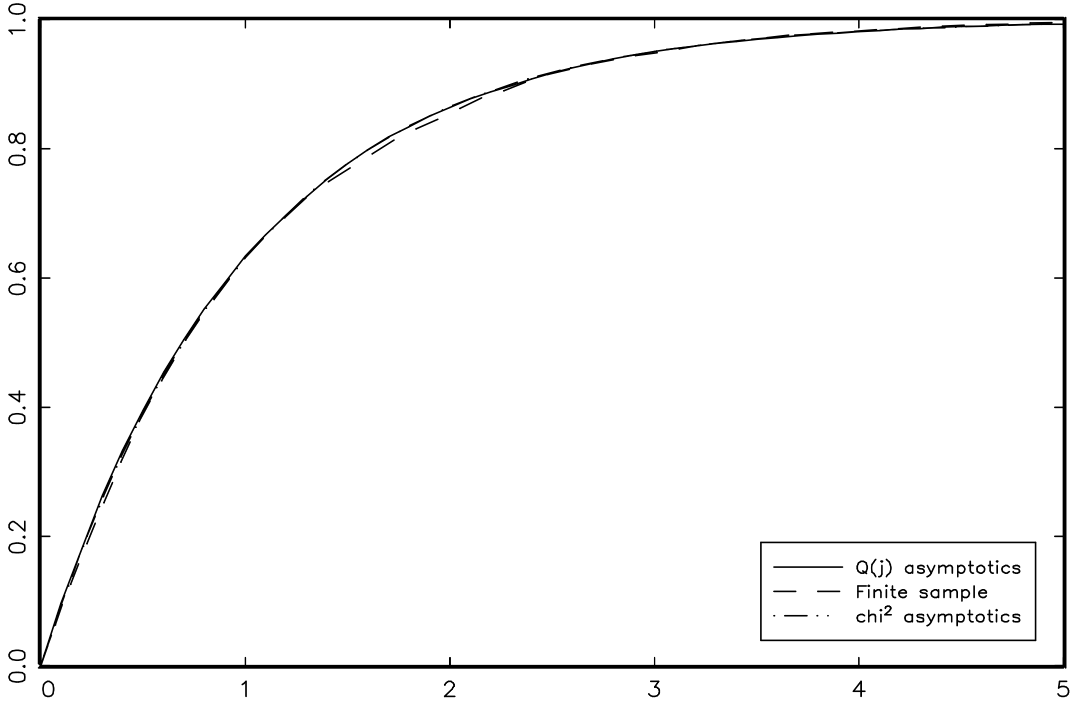

Figure 1(c) Periodogram Finite sample and Limit distributions, AR(1), j=1

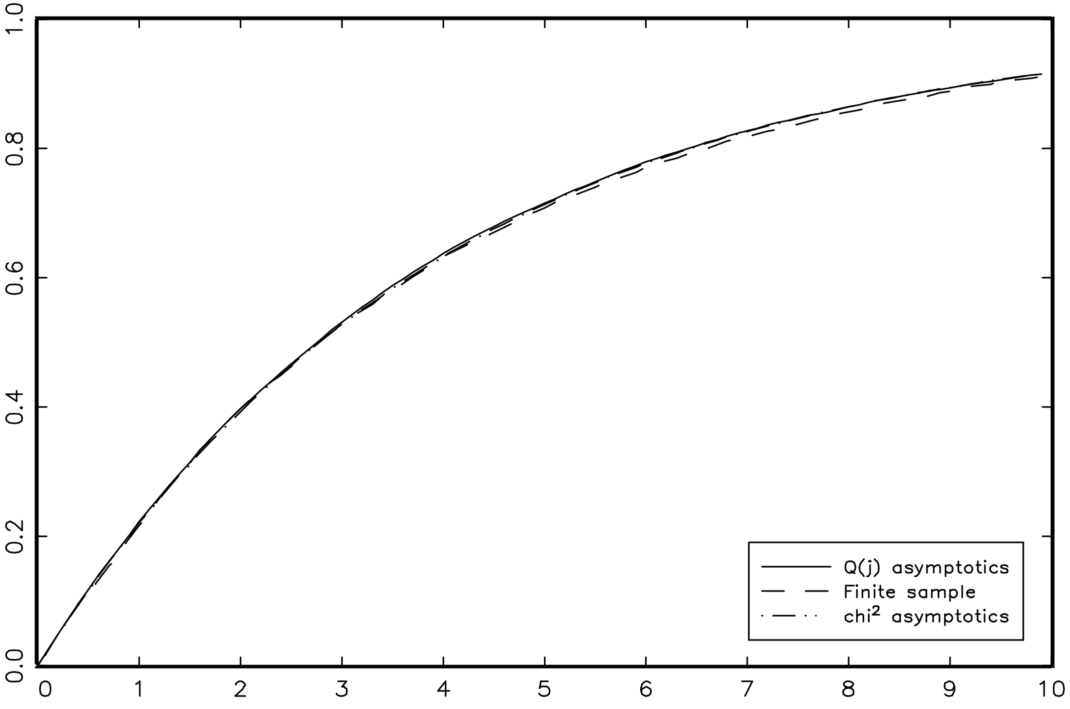

Figure 1(d) Periodogram Finite sample and Limit distributions, AR(1), j=3

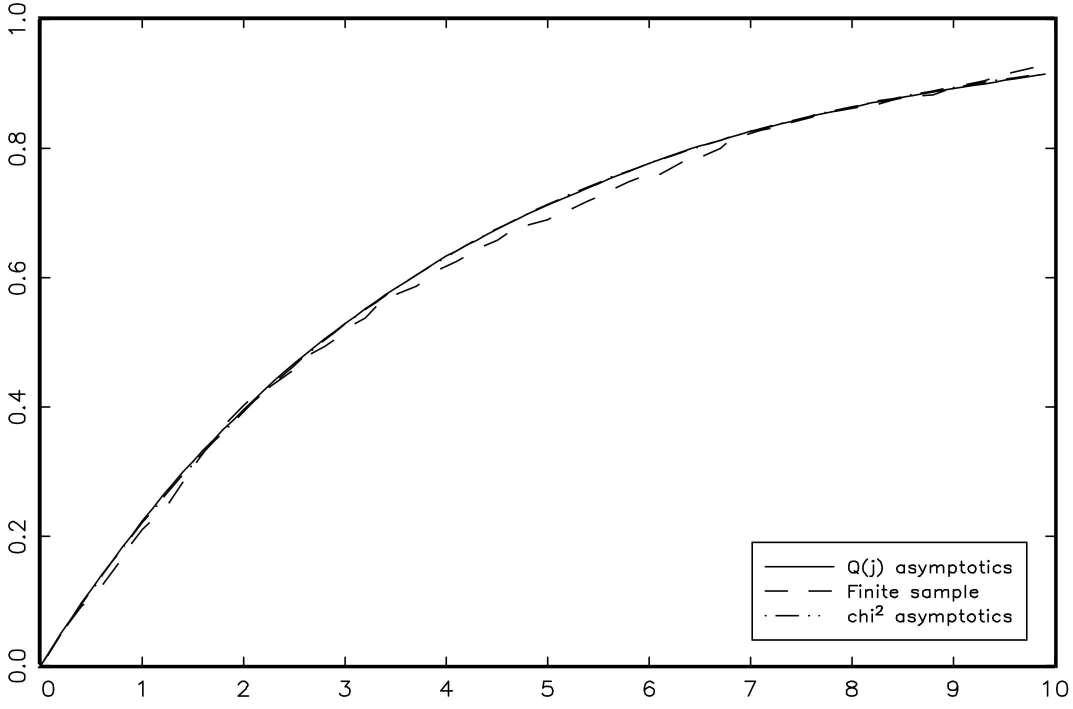

Figure 2(a) τ statistic Finite sample and Limit distributions, i.i.d., m=1

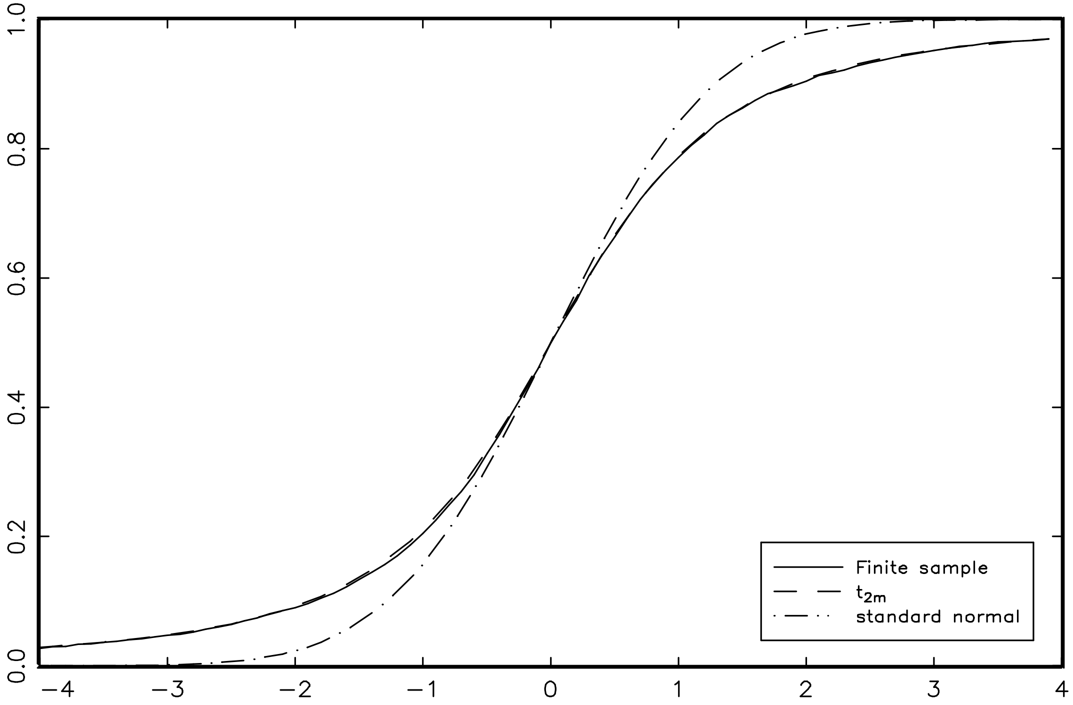

Figure 2(b) τ statistic Finite sample and Limit distributions, i.i.d., m=3

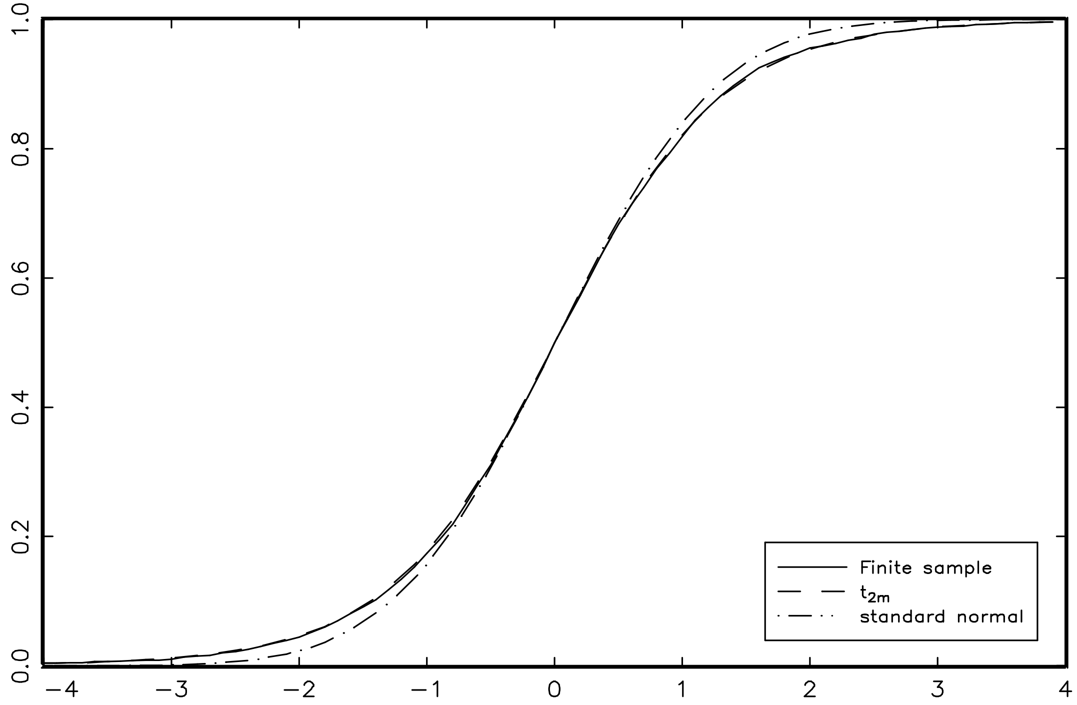

Figure 2(c) τ statistic Finite sample and Limit distributions, AR(1), m=1

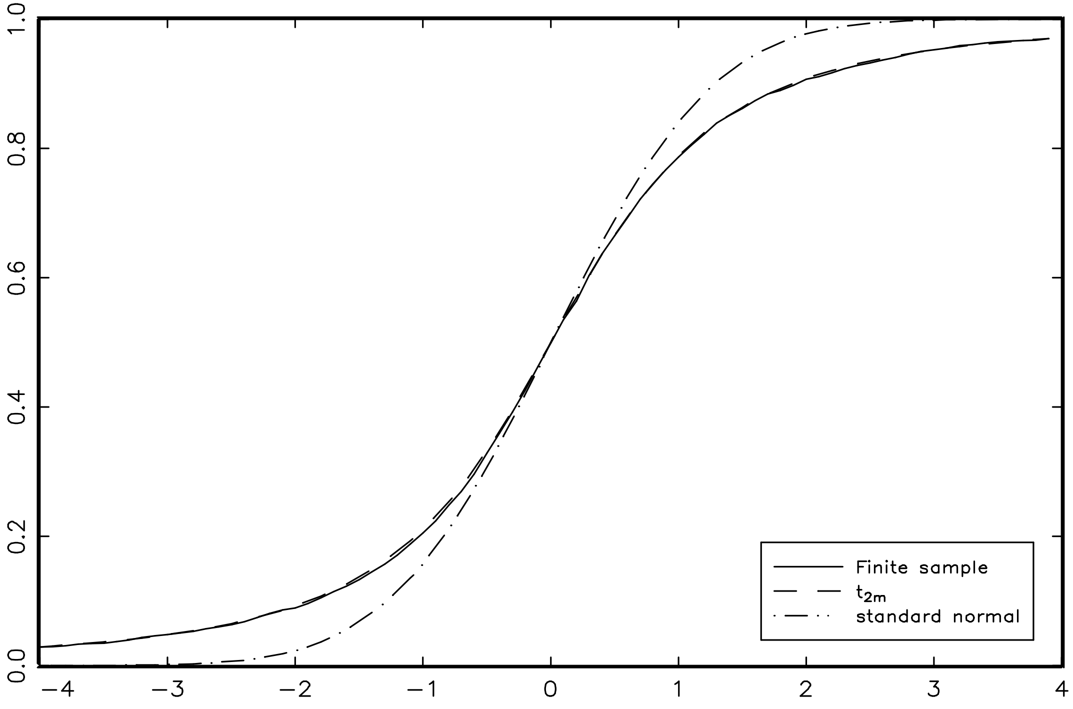

Figure 2(d) τ statistic Finite sample and Limit distributions, AR(1), m=3

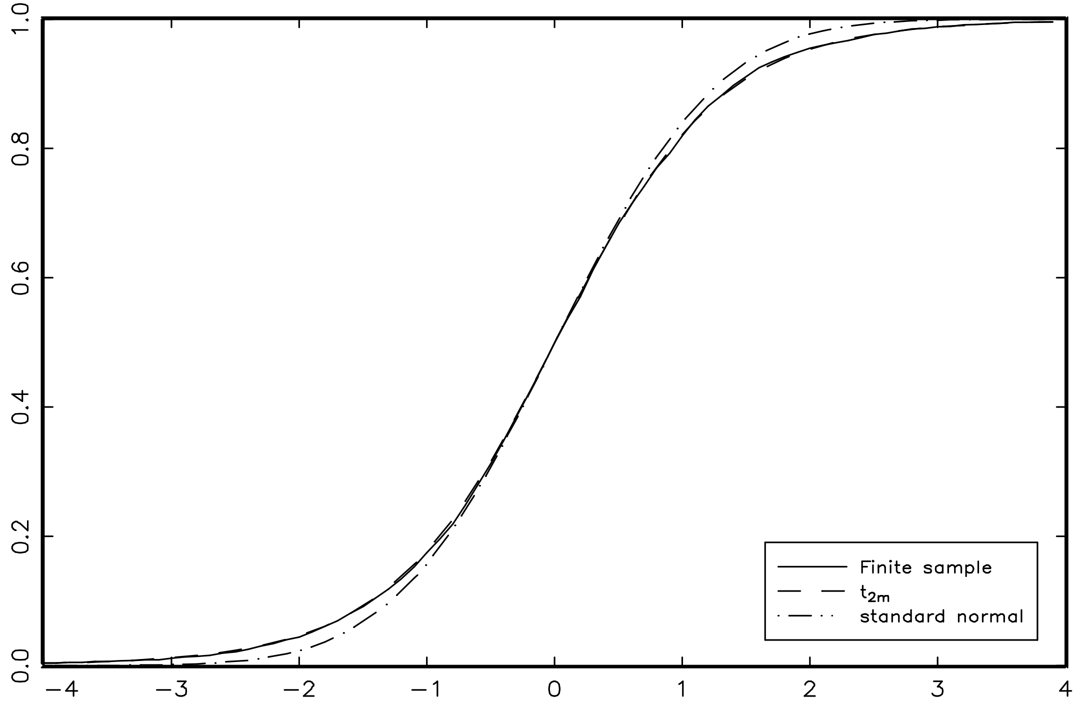

Figure 3(a) τ statistic Finite sample power, i.i.d.

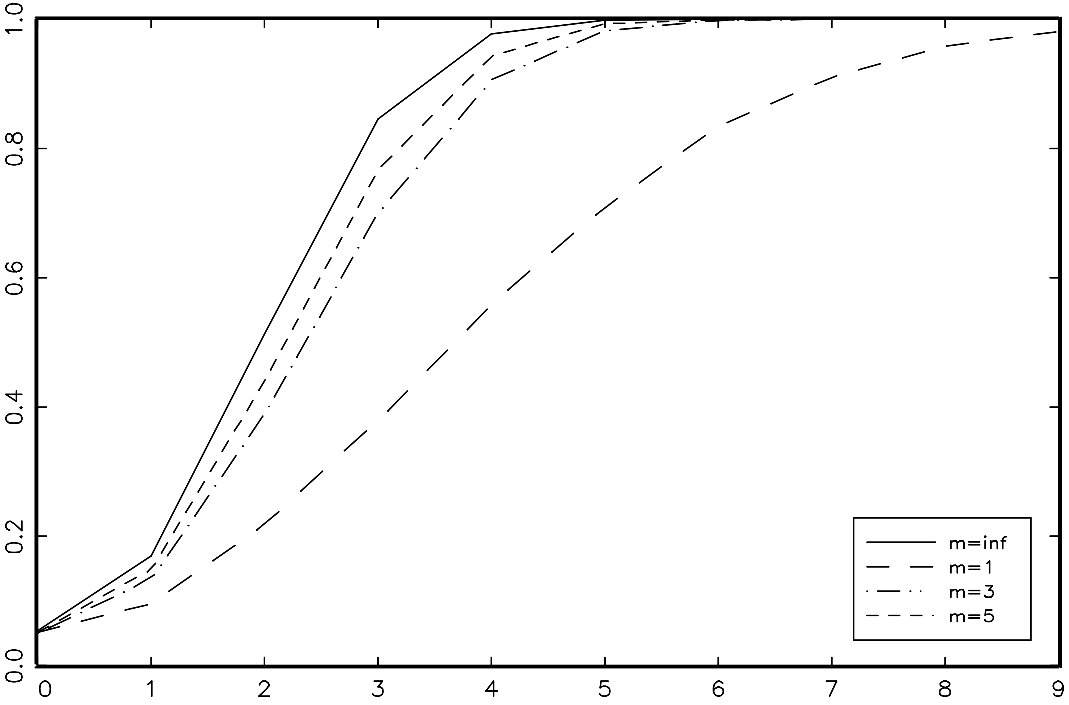

Figure 3(b) τ statistic Finite sample power, i.i.d.