A test of the long memory hypothesis based on self-similarity[*]

James Davidson   Dooruj Rambaccussing
University of Exeter   University of Dundee

Revised, January 2015, forthcoming in *Journal of Time Series Econometrics*

Abstract

This paper develops a new test of true versus spurious long memory, based on log-periodogram estimation of the long memory parameter using skip-sampled data. A correction factor is derived to overcome the bias in this estimator due to aliasing. The procedure is designed to be used in the context of a conventional test of significance of the long memory parameter, and a composite test procedure is described that has the properties of known asymptotic size and consistency. The test is implemented using the bootstrap, with the distribution under the null hypothesis being approximated using a dependent-sample bootstrap technique to approximate short-run dependence following fractional differencing. The properties of the test are investigated in a set of Monte Carlo experiments. The procedure is illustrated by applications to exchange rate volatility and dividend growth series.

1   Introduction

Estimation of the long memory parameter d by the method due to Geweke and Porter-Hudak (1983), or one of its variants, is a popular methodology in time series analysis. This estimator (henceforth, GPH) exploits fact that the autocovariances of a long memory process are nonsummable, and the spectral density f accordingly diverges at the origin at a particular rate, with $f(\lambda) = O(|\lambda|^{-2d})$ as $\lambda \to 0$. GPH estimates d by regressing the logarithms of the periodogram points in the neighbourhood of zero onto a suitable trend. However, except in very large samples this method has well-known limitations. As documented by Agiakloglou et al. (1993), the neglect of components of f representing short-run autocorrelation imply omitted terms in the regression, resulting in potentially substantial bias. In particular, the method is problematic as a basis for testing the null hypothesis of short memory, the case $d = 0$, since the conventional Wald statistic can be severely over-sized.

A simple illustration of this difficulty is provided by the observational equivalence between the fractionally integrated process $(1 - L)^d x_t = u_t$ with $d = 1$ and the autoregressive process $(1 - \phi L)x_t = u_t$ with $\phi = 1$. The ARFI model

$$(1 - \phi L)(1 - L)^d x_t = u_t$$

can exhibit a characteristically bimodal likelihood function when either of the parameters ϕ and d is close to unity in the process generating the sample. For every finite sample size, there exists a ϕ close enough to unity to bias the GPH estimator of d significantly, when its true value is zero.

It is desirable to have a means of distinguishing the cases of true d and spurious d, and goodness of fit criteria are an unreliable guide.

The approach explored in this paper is to devise a test with null and alternative interchanged. Recent research has highlighted the well-known property of self-similarity of hyperbolic decay processes under transformations such as periodic aggregation and periodic sub-sampling, otherwise known as skip-sampling. Chambers (1998) was the first to point out that if a long memory process is recorded at different rates, the rate of decay of the autocovariances is invariant to the rate of observation.

There are two ways to conceive of lowering the observation rate. Temporal aggregation means taking the sums of N successive observations to create the new sequence. This is the natural transformation for flow data, such that (for example) quarterly flows are each the sum of three successive monthly flows. Ohanissian et al. (2008) implement a test of long memory based on comparing log-periodogram estimates under different rates of temporal aggregation.

Skip-sampling, by contrast, means taking every Nth observation and discarding the remainder. This is the natural way of lowering the observation rate for stock or price data, although for the present purpose the nature of the observations is irrelevant, since the required properties of the skip-sampled series hold in all cases. Consider this transformation in the context of hyperbolic memory decay. Let the parameter δ index the rate of decay such that the autocovariance sequence $\{\gamma_j, j = 1, 2, \ldots\}$ of a stationary process satisfies

$$\gamma_j = O(j^{-\delta}) \tag{1.1}$$

for some $\delta > 0$. The hyperbolic memory class includes short memory processes having summable autocovariances, such that $\delta > 1$, and long memory processes where $\delta = 1 - 2d$ for $0 < d < \frac{1}{2}$, and hence $0 < \delta < 1$. It is immediately evident that, for any fixed, finite N,

$$\gamma_{Nj} = O(j^{-\delta}).$$

It follows that for the long memory class, the property of the spectral density near the origin should likewise be invariant to the sampling frequency.

This is in contrast to the case of exponential memory decay where $\gamma_j = o(j^{-\delta})$ for every finite δ, but there exists $\rho > 0$ such that

$$\gamma_j = O(e^{-\rho j}). \tag{1.2}$$

In this case, note that

$$\gamma_{jN} = O(e^{-\rho N j})$$

so that the memory decay parameter rises from ρ to ρN following skip-sampling. Since the estimator of (spurious) d in the exponential decay case is inevitably sensitive to the value of ρ, this suggests that comparing estimates under different rates of sampling might yield a useful test of the null hypothesis of long memory.

A range of nonlinear models, such as threshold autoregressive and Markov-switching processes, are often thought of as likely to be to be mistaken for long memory, since they can exhibit local patterns of apparent persistence, switches of local mean, for example or, in the case of ESTAR threshold models, unit root-like behaviour in the neighbourhood of the origin. As for the linear autoregressive model, the essential difference between these models and the long-memory case is that the serial dependence decays exponentially as the lag increases beyond a certain point, whereas long memory implies hyperbolic decay. Whether linear or nonlinear, stable difference equations of finite order necessarily exhibit exponential decay (see Gallant and White 1988, Davidson 1994) whereas unstable difference equations are nonstationary, featuring unit roots or explosive behaviour.

The class of cases of (1.1) with $\delta > 1$ count as instances of the alternative hypothesis for present purposes, because the autocovariances are summable. Models with this characteristic have not been significantly exploited in econometrics to date except in two rather special contexts, over-differenced fractional models (where $d < 0$ and there is the additional "anti-persistence" property of the autocovariances summing to zero) and stochastic volatility modelling. The FI-GARCH (Baillie et al. 1996) and HYGARCH (Davidson 2004) models are cases of the ARCH(∞) model where the lag weights in the conditional variance equation decline hyperbolically but are nonetheless summable. The co-moments of fourth order (when they exist) are likewise summable in these latter models. In the present case, by contrast, our null hypothesis is that of true long memory with $d > 0$.

This paper considers tests of the long memory hypothesis based on a comparison of the log-periodogram estimator of the d parameter in skip-sampled data with that from the original data. The test statistic is asymptotically standard Gaussian under the null hypothesis, given the usual assumptions of this literature (notably Gaussianity of the observations, see Robinson 1995, Hurvich et al., 1998). Convergence to the limit may be slow, and the formulation adopted depends on ancillary assumptions. The test is therefore implemented, for evaluation purposes, both as an asymptotic test and as a bootstrap test. We further recognise that the test is not consistent, for the rejection probabilities must be ultimately decreasing in sample size under the alternative hypothesis. However, we propose that the procedure be utilized as a component of a composite test, in combination with the Wald significance test on the fractional integration parameter in which the roles of null and alternative are reversed. The composite test can be formalized by the construction of a pseudo-p-value, and we show that this defines a test of the null of long memory that is both consistent and asymptotically correctly sized.

The paper is organized as follows. Section 2 reviews the important issue of aliasing in skip-sampled data, and its consequences for the form of the periodogram. Section 3 derives a bias-corrected form of the GPH estimator appropriate to skip-samples. Next, Section 4 describes the test procedure and derives the null asymptotic distribution of the statistic. Section 5 describes the implementation of the bootstrap version of the test. Section 6 describes the composite test, and Section 7 comments on the nonstationary case of the null hypothesis. Monte Carlo findings are reported in Section 8, Section 9 describes two contrasting applications, and Section 10 contains concluding comments. Some proofs are gathered in the appendix.

2   Aliasing

The distribution of the GPH estimator in skip-sampled data has been studied *inter alia* by Smith and Souza (2002, 2004) and Souza (2005). Skip-sampling induces a bias in the estimator due to the effect of aliasing on the form of the spectral density. For a comprehensive analysis of the aliasing phenomenon, see Hassler (2011). The essential result is that the spectral density of the skip-sampled data can be represented as an average of the spectral densities over the range of aliased frequencies.

Proposition 2.1 *If $\{y_t, t = 1, 2, \ldots\}$ is a discrete stationary stochastic process with spectral density f and $x_t = y_{tN}$ for $t = 1, 2, 3, \ldots$ and $N > 1$, the spectral density of the process x_t is*

$$f_N(\lambda) = \frac{1}{N} \sum_{j=0}^{N-1} f\left(\frac{\lambda + 2\pi j}{N}\right), \quad 0 \leq \lambda \leq \pi.$$

The straightforward proof is given in the appendix. Note that cycles of frequency λ/N in the original data become cycles of frequency λ in the skip-sampled data, and frequencies above π/N

are no longer identifiable. Hence, these contributions to the variance of the series are effectively aggregated with the identifiable frequencies.

In the fractionally integrated case where

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d}g(\lambda) = [2\sin(\lambda/2)]^{-2d} g(\lambda) \tag{2.1}$$

with $0 < g(0) < \infty$, we find that $f_N(\lambda)$ cannot be directly log-linearized in the GPH manner. What can be done, following the suggestion of Smith and Souza (2002), is to write

$$f_N(\lambda) = \frac{1}{N}\sum_{k=0}^{N-1} \left(2\sin\left(\frac{\lambda + 2\pi k}{2N}\right)\right)^{-2d} g\left(\frac{\lambda + 2\pi k}{N}\right) \tag{2.2}$$

$$= \left(2\sin\frac{\lambda}{2N}\right)^{-2d} g\left(\frac{\lambda}{N}\right) H_N(\lambda)$$

where

$$H_N(\lambda) = \frac{1}{N}\sum_{k=0}^{N-1} \left(\frac{\sin\left(\pi k/N\right)}{\tan(\lambda/2N)} + \cos(\pi k/N)\right)^{-2d} \frac{g\left((\lambda + 2\pi k)/N\right)}{g(\lambda/N)}. \tag{2.3}$$

There is, evidently, an omitted term $\log H_N$ in the log-periodogram regression in skip-sampled data, depending on d as well as λ. The omission of this term will be liable to produce a bias in the GPH regression, and its omission is not rendered negligible by taking frequencies close to the origin. Indeed, what is commonly observed is that estimates of $d > 0$ obtained from skip-sampled data are substantially closer to zero than those from the original data.

Remark

Note the implication for the standard analysis of a model such as (2.1), which is revealed to be specifically linked to the frequency of observation. Without this assumption, there is no reason to suppose that the function g does not also depend on d, nor that it is constant near the origin. In this light, the standard long memory analysis appears a little more fragile than is commonly taken for granted. Nonetheless, in this paper we shall work with the standard assumptions for the purposes of developing a test.

3   The Bias-corrected Estimator

The test we propose is based on the comparison of two narrow-band regression estimators of the memory parameter d, one based on the full sample, the other based on skip-sampling of the test series. As before, let N denote the periodicity of the skips. Skip-sampling is done by taking every Nth observation, so yielding a sample of size $[T/N]$ where $[z]$ denotes the largest integer below z. This can be done N times, by off-setting the initial observation, so that the N skip-samples can be represented as $\{x_{0t}\},\{x_{1t}\},\ldots, \{x_{N-1,t}\}$ where, for $k = 0,\ldots,N-1$,

$$x_{k1} = y_{k+1}$$
$$x_{k2} = y_{N+k+1}$$
$$x_{k3} = y_{2N+k+1}$$
$$\cdots$$
$$x_{kT} = y_{([T/N]-1)N+k+1}.$$

Each of these samples can be used to compute a modified log-periodogram estimator, which we denote $\hat{d}_{Nk}$ for $k = 0,\ldots,N-1$.

As shown in Section 2, the conventional GPH estimator applied to skip-sampled data is biased. To be more specific, it exhibits a bias different in character from the well-known case of data with short run dependence in the fractional differences, being present even if the original series is a pure fractional without short-run components. Moreover, the bias is not attenuated by choosing a narrow bandwidth. The logarithm of H_N in (2.3) is a missing term in the log-periodogram regression, and bias correction involves finding a computable surrogate for this function.

Expression (2.3) as a function of λ depends in the first place on the unknown d, and the natural approximation is to replace this with the asymptotically unbiased estimator $\hat{d}$. It also depends on the unknown spectral density component g evaluated at different points, and except in the case of the pure fractional model, the term $g\left(\lambda + 2\pi k\right)/N)/g(\lambda/N)$ in (2.3) varies with λ in general over the whole of the interval $[0, 2\pi]$, including points close to the origin. Approximating it by a constant, in the manner of dealing with $g(\lambda)$ in the narrow-band estimator, is therefore not an attractive option.

Possible methods for estimating this term include constructing a kernel estimator of g from the spectrum of the fractional differences. However, in this implementation we have adopted a semiparametric approach. Let the null hypothesis specify that the random sequence has a representation of the ARFI form

$$\phi(L)(1-L)^d x_t = u_t \tag{3.1}$$

where $\phi(L)$ is an invertible lag polynomial, of possibly infinite order, and $u_t \sim \text{NI}(0, \sigma^2)$ where 'NI' denotes independent Gaussian. To approximate $\phi(L)$ we use the the Durbin-Levinson algorithm to fit an autoregression of order $p_T = 0.6T^{1/3}$ to the fractional differences $(1-L)^{\hat{d}} x_t$, where $\hat{d}$ is the estimator of d based on the full sample. This yields an estimated lag polynomial $\hat{\phi}(L)$, and we then approximate $g(\lambda)$ by

$$\hat{g}(\lambda) = |\hat{\phi}(e^{-i\lambda})|^{-2}. \tag{3.2}$$

It suffices for our application that $\hat{g}(\lambda)$ converges in probability to $g(\lambda)$ pointwise in a neighbourhood of zero, and since $\hat{g}$ is a smooth differentiable function of the data this property should in fact hold in a wider class of processes than (3.1). Absolute summability of the autocovariances of the fractional differences should hold in processes for which log-periodogram regression has good properties. Since the estimators in question depend only on second moments, they will yield the same consistency properties if u_t in (3.1) is merely white noise. With caveats concerning invertibility, the Wold theorem therefore extends validity to the general covariance stationary case. The issues arising here are carefully analysed, in the bootstrap context, by Kreiss et al. (2011). They show that Gaussianity of the series is certainly sufficient and this is, in any case, an assumption adopted for our subsequent asymptotic analysis and imposed in our experiments. Of course, these considerations strictly relate to the case where d is known, and the largest source of error in finite samples will be due to the replacement of d by $\hat{d}$.

Letting $\lambda_j = 2\pi j/T$ as usual, the skip-sampled series consists of $[T/N]$ observations, and the frequencies at which the periodogram is evaluated are $\lambda_{Nj} = 2\pi Nj/T$ for $j = 1, \ldots, M_N$ where $M_N = [(T/N)^q]$, for $0 < q < 1$, represents the usual GPH bandwidth function of sample size. In practice q should be chosen according the the established prescriptions of the literature and, following Hurvich Deo and Brodsky (1998) (henceforth HDB), setting $q < 4/5$ ensures limiting Gaussianity of the estimator, with bias of $O(T^{2(q-1)})$. Let I_{Nk} denote the periodogram computed from the kth skip-sampled data set with period N, and let $\hat{H}_N(\lambda)$ denote the formula in (2.3) approximated as described, using the estimated parameters and the representation of the short-run spectral density in (3.2). The kth bias-corrected skip-sample estimator then takes the form

$$\hat{d}_{Nk} = \frac{\sum_{j=1}^{M_N}(X_{Nj} - \bar{X}_N)[\log I_{Nk}(\lambda_{Nj}) - \log \hat{H}_N(\lambda_{Nj})]}{\sum_{j=1}^{M_N}(X_{Nj} - \bar{X}_N)^2} \tag{3.3}$$

5

where $X_{Nj} = -2\log(2\sin\lambda_{Nj}/2)$. Provided N is treated as fixed and not linked to sample size note that $M_N = O(M)$ where $M = [T^q]$, and this is the assumption we maintain henceforth.

While the formula in (3.3) employs an estimator of the function g as a component of the aliasing correction, note that this estimator has not been included in the log-periodogram regression itself, and this remains a narrow-band estimator. Be careful to note that $\hat{g}$ depends on the narrow-band estimator of d based on the full sample, which is used to fractionally difference the data, and hence it does not provide a direct route to a broad-band estimation procedure of the type proposed by Moulines and Soulier (1999), for example.

4   The Skip-sampling Test

Letting the conventional GPH estimator based on the complete sample be denoted $\hat{d}$, the test statistic we consider is

$$\hat{\tau} = \frac{\hat{d} - \hat{d}_N}{\text{s.e.}(\hat{d} - \hat{d}_N)} \tag{4.1}$$

where $\hat{d}_N = N^{-1}\sum_{k=0}^{N-1}\hat{d}_{Nk}$. We use the signed statistic and perform a one-tailed test, on the assumption that the leading cases of the alternative will give rise to a smaller value of d in the skip-sampled data. Also note that, in view of the form of the estimator, using the average of the $\hat{d}_{Nk}$ estimates from the N skip samples is equivalent to adopting the average of the log-periodogram points across the N offset samples as regressand. This scheme makes the most efficient use of the available data.

When the sample is large enough, both the conventional GPH estimator $\hat{d}$ and the skip-sampled estimator $\hat{d}_N$ defined in (3.3) can be analysed using the techniques developed in HDB. These authors obtain their results from the following assumptions, which here relate to our null hypothesis under test.

Assumption 1 *The process $\{y_t, t = 1, 2, \ldots\}$ is stationary and Gaussian with the spectral density given in (2.1) with $0 < d < \frac{1}{2}$.*

Assumption 2 *$M \to \infty$ as $T \to \infty$ with $M/T \to 0$ and $(M\log M)/T \to 0$.*

Assumption 3 *$g'(0) = 0$, and $g''(\lambda)$, $g'''(\lambda)$ are bounded for all λ in a neighbourhood of zero.*

Letting $\varepsilon_{Nkj} = \log(I_{Nk}(\lambda_{Nj})/f(\lambda_{Nj}))$ there exists a function f^* such that (analogous to the expression in HDB page 42)

$$\hat{d}_{Nk} - d = \frac{1}{S_N}\sum_{j=1}^{M_N} a_{Nj}\log f_{Nj}^* + \frac{1}{S_N}\sum_{j=1}^{M_N} a_{Nj}\varepsilon_{Nkj}$$

where $a_{Nj} = X_{Nj} - \bar{X}_N$ and $S_N = \sum_{j=1}^{M_N} a_{Nj}^2 = O(M)$. Under our assumptions. the first right-hand side term is $o(1)$. In the case $N = 1$, such that there is no skip sampling and $\hat{d}_{Nk} = \hat{d}$, $f_{1j}^* = f_j^* = g(\lambda_j)$ and $\varepsilon_{10j} = \varepsilon_j$. In the cases with $N > 1$, on the other hand,

$$\log f_{Nj}^* = \log g\left(\frac{\lambda_{Nj}}{N}\right) - \log\left(\frac{\hat{H}(\lambda_{Nj})}{H(\lambda_{Nj})}\right).$$

Since H is twice-differentiable with respect to d and $\hat{d}$ is $M^{1/2}$-consistent under our assumptions, we can expand $\log\hat{H}(\lambda_{Nj})$ as

$$\log\hat{H}(\lambda_{Nj}) = \log H(\lambda_{Nj}) + \frac{H(\lambda_{Nj})'}{H(\lambda_{Nj})}(\hat{d} - d) + O_p(M^{-1}).$$

6

Then, using Lemma 1 of HDB, and letting

$$B_T(N, d) = \frac{1}{S_N} \sum_{j=1}^{M_N} a_{Nj} \frac{H(\lambda_{Nj})'}{H(\lambda_{Nj})}$$

we have

$$\hat{d}_{Nk} - d = -B_T(N, d)(\hat{d} - d) + \frac{1}{S_N} \sum_{j=1}^{M_N} a_{Nj} \varepsilon_{Nkj} + o_p(M^{-1/2}).$$

Note that the relevant properties of the random variables ε_{Nkj} extend from the full-sample to the skip-sampled case, specifically, that their distribution has finite second moments that asymptotically do not depend on nuisance parameters – see Lemmas 2 and 6-8 of HDB. Since the regressors are the same for each k, we further find

$$\hat{d}_N - d = -B_T(N, d)(\hat{d} - d) + \frac{1}{NS_N} \sum_{j=1}^{M_N} a_{Nj} \sum_{k=0}^{N-1} \varepsilon_{Nkj} + o_p(M^{-1/2}). \tag{4.2}$$

and hence

$$
\begin{aligned}
\hat{d} - \hat{d}_N &= [1 + B_T(N, d)](\hat{d} - d) - \frac{1}{NS_N} \sum_{j=1}^{M_N} a_{Nj} \sum_{k=0}^{N-1} \varepsilon_{Nkj} + o_p(M^{-1/2}) \\
&= \frac{1 + B_T(N, d)}{S} \sum_{j=1}^{M} a_j \varepsilon_j - \frac{1}{NS_N} \sum_{j=1}^{M_N} a_{Nj} \sum_{k=0}^{N-1} \varepsilon_{Nkj} + o_p(M^{-1/2}) \tag{4.3}
\end{aligned}
$$

where $a_j = X_j - \bar{X}$ and $S = \sum_{j=1}^{M} a_j^2$. In the appendix, we show the following.

Proposition 4.1 *For fixed finite N, and d such that $(1 - L)^d x_t$ is a weakly dependent process, $B_T(N, d)$ converges in probability to a finite nonstochastic limit $B(N, d)$.*

The next thing to note using further results from HDB is that, with $q < 4/5$,

$$\frac{1}{\sqrt{S}} \sum_{j=1}^{M} a_j \varepsilon_j \xrightarrow{D} \mathrm{N}(0, \pi^2/6).$$

Also note that $S_N^{-1/2} \sum_{j=1}^{M_N} a_{Nj} \varepsilon_{Nkj}$ has the same limit in distribution for each k, where $S_N = \sum_{j=1}^{M_N} a_{Nj}^2$, and so similarly,

$$\frac{1}{\sqrt{NS_N}} \sum_{j=1}^{M_N} a_{Nj} \sum_{k=0}^{N-1} \varepsilon_{Nkj} \xrightarrow{D} \mathrm{N}(0, \pi^2/6).$$

It follows that under these conditions,

$$M^{1/2}(\hat{d} - \hat{d}_N) \xrightarrow{D} \mathrm{N}(0, V) \tag{4.4}$$

where .

$$V = \left[\frac{(1 + B(N, d))^2}{\bar{S}} + \frac{N^q}{N\bar{S}_N} \right] \frac{\pi^2}{6} - 2 \frac{N^{q/2}(1 + B(N, d))}{\sqrt{\bar{S}}\sqrt{N\bar{S}_N}} C(q, N) + o_p(1)$$

N	4	6	8	10	12
q 0.5	0.350	0.245	0.183	0.156	0.129
0.55	0.336	0.234	0.178	0.143	0.124
0.6	0.328	0.226	0.173	0.142	0.118
0.65	0.322	0.220	0.167	0.139	0.120
0.7	0.311	0.213	0.165	0.092	0.103
0.75	0.302	0.278	0.118	0.121	0.131

Table 1: Numerical estimates of $C(n,q)$

where $\bar{S} = \lim_{T \to \infty} S/M$, $\bar{S}_N = \lim_{T \to \infty} S_N/M_N$, and

$$C(q, N) = \lim_{T \to \infty} \frac{\sum_{j=1}^{M} a_j \sum_{j'=1}^{M_N} a_{Nj'} \sum_{k=0}^{N-1} E(\varepsilon_j \varepsilon_{Nkj'})}{\sqrt{\bar{S}} \sqrt{N \bar{S}_N}}.$$

To derive a formula for $C(q, N)$ analytically would entail quite a challenging calculation, and we have circumvented the need for this by a numerical evaluation. Note that ε_j is the logarithm of the periodogram point of an independent Gaussian series (having $d = 0$ and g constant) whereas the $\varepsilon_{Nkj'}$ are the log-periodograms of the corresponding skip-sampled series. Therefore, $C(q, N)$ can be approximated as closely as desired, for given q and N, by a simulation based on a sufficiently large sample. The accuracy of the approximation can be monitored by computing the sample variances of the components at the same time, and checking how close these lie to their known asymptotic counterpart of $\pi^2/6$. We have performed the simulation with 200,000 replications in a sample size of 2000, with the results shown in Table 1. Denoting by $\hat{V}$ the variance formula computed using these approximations, replacing d by its full-sample GPH estimator, the test statistic is calculated as

$$\hat{\tau} = \frac{\hat{d} - \hat{d}_N}{(\hat{V}/M)^{1/2}}. \tag{4.5}$$

This statistic is used as the basis for a one-tailed test with rejections in the upper tail.

5   The Bootstrap Test

A difficulty with the semiparametric approach to estimation is the slow convergence to the asymptote, at the rate $M^{1/2}$ rather than $T^{1/2}$. The mean and variance approximations derived in the previous section are accordingly slow to improve, especially with the reduction in effective sample size following skip-sampling. This suggests that the bootstrap may have a useful role to play in implementing the test, while not overlooking that the parametric bootstrap is likewise dependent on slowly converging estimated parameters. Nonetheless, a comparison of the two procedures, asymptotic and bootstrap, may serve to triangulate the uncertainty.

The bootstrap distribution of the statistic has to be estimated by simulating the null hypothesis as a fractionally integrated process, while allowing for the possibility of short-run dependence of the fractional differences Given an estimator $\hat{d}$ of the fractional parameter and test statistic $\hat{\tau}$ computed from the sample, the calculation is performed as follows.

1. Compute the fractional differences $\hat{u}_t = (1-L)_+^{\hat{d}}(y_t - y_1)$ where $(1-L)_+^{\hat{d}} = \sum_{j=0}^{t} \hat{a}_j L^j$ and $a_0 = 1$, $a_j = a_{j-1}(j - \hat{d} - 1)/j$ for $j \geq 1$.

2. Repeat the following steps for $j = 1, ..., B$.

a) Draw a random sample $\hat{u}_{ij}^*, \ldots, \hat{u}_{Tj}^*$ from the distribution of $\hat{u}_1, \ldots, \hat{u}_T$ using a method that preserves the dependence structure; see Remark 1 below.

b) Construct the sequence

$$\hat{y}_{tj}^* = (1 - L)_+^{-\hat{d}} \hat{u}_{tj}^* + \hat{z}_{tj}, \ t = 1, \ldots, T$$

where $(1 - L)_+^{-\hat{d}} = \sum_{s=0}^{t} b_j L^j$ where $b_0 = 1$, $b_j = b_{j-1}(j + \hat{d} - 1)/j$ for $j \geq 1$, and $\hat{z}_{tj}$ is explained in Remark 2 below.

c) Compute the bootstrap test statistic $\hat{\tau}_j^*$ as in (4.5) for the sample $\hat{y}_{ij}^*, \ldots, \hat{y}_{Tj}^*$.

3. Compute the estimated p-value for the test as 0 if $\hat{\tau} > \hat{\tau}_{(B)}^*$ or else as

$$1 - \frac{\min\left\{j : \hat{\tau} \leq \hat{\tau}_{(j)}^*\right\}}{B},$$

where $\hat{\tau}_{(j)}^*$ is the jth order statistic for the bootstrap statistics $\hat{\tau}_1^*, \ldots, \hat{\tau}_B^*$.

Remarks

1. Methods for constructing the drawings $\hat{u}_{ij}^*, \ldots, \hat{u}_{Tj}^*$ include the stationary bootstrap of Politis and Romano (1994) and the sieve autoregression method of Bühlmann (1997). Note that the latter calculation is also used to obtain expression (3.2), and the same remarks apply regarding the validity of the sieve AR method in this context; see Kreiss et al. (2011).

2. The correction terms $\hat{z}_{tj}$ are constructed using Gaussian drawings and weights computed from the estimated parameters to have a covariance structure matching the components omitted through truncating the innovation sequence at 0. These replace the sample initial condition which has been truncated in step 1. The resulting sequence is approximately stationary for $|\hat{d}| < \frac{1}{2}$. If $\hat{d} \geq \frac{1}{2}$ the data are modelled in differences, replacing $\hat{d}$ by $\hat{d} - 1$, and the simulation is then integrated using the first observation for the initial condition. Nonstationary processes generated by this procedure converge after normalization to Type I Brownian motion. For details of the simulation procedure, see Davidson and Hashimzade (2009).

3. In practice, different estimators of d, employing different bandwidths in particular, might be used to compute the statistic and to implement the bootstrap resampling as in Step 1. Using a wider bandwidth in the former case could increase power by emphasizing bias under the alternative, while a different balance between bias and variance might be advantageous in achieving the best bootstrap distribution. Such choices can be guided by simulation experiments.

6 The composite test

Testing the degree of persistence of a time series is a problem that has attracted a degree of controversy, as documented by one of the present authors (Davidson 2009). This is one of a class of problems have been characterized by Dufour (1997) as 'ill-posed', and has close links with the testing frameworks critically analysed by Pötscher (2002) and Faust (1996, 1999), *inter alia*. Tests of the null hypothesis that the series has summable autocovariances – the 'I(0) hypothesis' – face a common difficulty for valid inference. This difficulty manifests itself in different ways

in different contexts, but the essential common feature is that the null hypothesis constitutes an open set in the parameter space. It follows that test power cannot exceed test size, where the latter is defined as the supremum of the rejection probabilities over the null set of the model space. While this problem extends to more general parameterizations it is most transparent in the case where the 'I(0)' property relates to the modulus of the maximal autoregressive root.[1] The null hypothesis is represented by the interval $[0,1)$ with its closure containing the leading case of the alternative.

Although the null and alternative are interchanged, the present case is clearly similar. The null hypothesis relating to the value of d is the open interval $(0, \infty)$, with its closure containing the cases of the alternative with $d = 0$. This is another situation where, under a literal interpretation, power cannot exceed size. The test is based on a comparison of two estimators of d, where under the alternative, one (the full-sample estimator) is expected to exhibit more bias than the other (the skip-sampled estimator) *as estimators of zero*. Since the estimators being compared are both consistent, albeit biased in finite samples, the test is evidently inconsistent. The probability of exceeding the rejection criteria under the alternative evidently cannot be monotone nondecreasing in sample size.

While the test might therefore appear of doubtful value in applications, this conclusion overlooks the context in which such a test might be applied. The question actually being posed, in most cases, is whether a 'significantly positive' estimator of d should be treated as a biased estimator of zero. If the significance test does *not* result in rejection, then we might on these grounds decide to reject the null hypothesis of long memory and either forego the skip-sampling test or, at least, to overlook a non-rejection in the latter test.

To formalize this idea, consider a composite test in which the skip-sampling test is performed in partnership with a one-tailed Wald test of the hypothesis $d \leq 0$ with alternative $d > 0$. With contamination by short-run positive autocorrelation, we anticipate possible over-rejection in this latter test. Non-rejection in the Wald test implies effective rejection of the null hypothesis of long memory, and there is, arguably, no need to proceed to the skip-sampling test. Here's how we can compute a composite test which delivers a p-value taking account of the outcome of the initial Wald test. Suppose that the Wald test delivers a p-value π_{1T} in a sample of size T, and the skip-sampling test a p-value π_{2T}. Consider the pseudo-p value calculated as

$$\hat{\pi}_{2T} = \pi_{2T}(1 - \pi_{1T})^{T^\delta} \tag{6.1}$$

for some $\delta > 0$.

Proposition 6.1 *The test "reject H_0 when $\hat{\pi}_{2T} < \alpha$" is consistent and asymptotically of size α.*

To prove the proposition, first consider the behaviour of the statistic $\hat{\pi}_{2T}$ in the case $d > 0$ (the null hypothesis). Let the Wald statistic be denoted t_T. The corresponding p-value π_{1T} is the area under the upper tail of the null distribution (standard normal) bounded by t_T. Since the Wald test is consistent with $t_T = O_p(T^{q/2})$, we have

$$\begin{aligned}
\pi_{1T} &= \frac{1}{\sqrt{2\pi}} \int_{t_T}^{\infty} e^{-x^2/2} dx \\
&= e^{-t_T^2/2} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-x^2/2 - t_T x} dx \\
&= O_p(e^{-T^q/2}).
\end{aligned}$$

---

[1] We assume the model is parameterized so that stable autoregressive roots lie inside the unit circle.

Hence, consider $\hat{\pi}_{2T}/\pi_{2T} = (1 - \pi_{1T})^{T^\delta}$, and note that for every $\delta < \infty$,

$$\begin{aligned}
\log(\hat{\pi}_{2T}/\pi_{2T}) &= T^\delta \log(1 - \pi_{1T}) \\
&= -T^\delta(\pi_{1T} + O(\pi_{1T}^2)) \\
&= o_p(1).
\end{aligned}$$

It follows that $\hat{\pi}_{2T}/\pi_{2T} \to 1$ in probability. Since the null hypothesis is true, $\hat{\tau}$ converges to N(0,1) in distribution according to (4.4). Therefore, π_{2T} is asymptotically uniformly distributed on the unit interval. By the indicated convergence in probability the composite test shares this property, and so rejects asymptotically with probability α in an α-level test.

Next, suppose that the null is false, with $d \le 0$. Recalling that the Wald test is one-sided, π_{1T} is asymptotically uniformly distributed on the unit interval in this case and so, in particular, $P(\pi_{1T} > 0) \to 1$. When $\delta > 0$, it follows that for any $\varepsilon > 0$,

$$P(\hat{\pi}_{2T} > \varepsilon) = P(\pi_{2T}(1 - \pi_{1T})^{T^\delta} > \varepsilon) \to 0$$

as $T \to \infty$, and the proposition is proved.

The convergence of $\hat{\pi}_{2T}$ to the uniform distribution under the null must be somewhat slower than that of π_{2T}, depending on the choice of δ. The smaller that δ is chosen, the nearer $\hat{\pi}_{2T}/\pi_{2T}$ is to unity in any given sample size and the smaller is the size distortion *ceteris paribus*, while not overlooking the fact that the test based on π_{2T} may itself exhibit size distortion in one direction or another, so that the net distortion in a given sample size is unpredictable. On the other hand, the larger δ is chosen, the more rapidly $\hat{\pi}_{2T}$ approaches 0 under the alternative. Hence, the choice of δ represents a trade-off of power against size.

Take care to note that the consistency of the composite test holds whether or not $\pi_{2T} \to 0$ under the alternative. To appreciate the contribution of the skip-sampling test, it may be helpful to envisage the "test" based on simply drawing a uniform random number from $[0, 1]$ at the second stage, instead of computing the quantile of the skip-sampling statistic. Proposition 6.1 holds also for this test! What we have done here is to give an alternative way of formalizing the properties of the Wald test. The tendency of this test to over-reject the conventional null hypothesis $d \le 0$, due to bias, is converted into a case of low finite-sample power to reject the hypothesis $d > 0$. However, the expectation is that the power of the composite test in finite samples is greater, to the extent that π_{2T} is distributed closer to zero than a uniform variate under the alternative. The simulation experiments reported in Section 8 show that such improvements, judged by the performance of the basic skip-sampling test, can be large.

We emphasize once again that the composite test does not need to be taken literally as an operational procedure. We can think of it as a formalization of the procedure of taking two test results into account in making a decision. If we cannot reject the hypothesis $d \le 0$ on the Wald test, we are unlikely to proceed to the second stage. If we do find d "signicantly positive" on conventional criteria, then we want to know how far this outcome might be attributable to bias, and the the skip-sampling test can in this case provide countervailing evidence.

7 The nonstationary case

As the observational equivalence issue raised in the introduction would lead us to predict, autoregressively generated series with a root in the stable region but close to unity characteristically yield an estimated d in the non-stationary range $\frac{1}{2} \le d \le 1$. However it is known (from, e.g. Velasco 1999, Kim and Phillips 2006) that log-periodogram regression in this range is consistent, and also asymptotically normal, under regularity conditions, for $d < \frac{3}{4}$. Our test should

exhibit similar characteristics in stationary and nonstationary cases of the null hypothesis, and this conjecture is borne out by the simulation experiments reported in Section 8.

In a well-known paper, Diebold and Inoue (2001) point out that in certain models exhibiting structural change, in which the frequency of change has a particular relation with sample size, there is the appearance of hyperbolic memory decay. In some of their examples, the processes in question are 'revealed' as really I(1) (stationary in differences) as T is extended with fixed parameters. To understand how the skip-sample test might behave in these cases, we must not overlook the fact that a unit root process, like a serially independent process, is technically a case of the null hypothesis. Both cases exhibit the invariance of memory to skip-sampling characteristic of fractional integration. Thus, a skip-sampled unit root remains a unit root. For this reason we should not expect the present test to have greatest power against local-to unity autoregressive alternatives. The natural approach, faced with a time series that does not exhibit mean reversion, might be to test for hyperbolic memory in the differences. Diebold and Inoue also propose examples in which processes appearing to show hyperbolic decay in a given sample size are 'revealed' as I(0) as T increases, and here our test should perform better. In particular, they consider a simple independent process subject to Markov-switching, which is one of the cases to be studied in the next section.

8 Monte Carlo experiments.

We present some experiments using three sample sizes, $T = 250$, 1000 and 5000, with 5000 replications in each case. Following preliminary investigations a bandwidth for the GPH estimator of $M = [T^{0.7}]$ was chosen to compute the tests, with a skip-sampling period of $N = 8$. A relatively wide bandwidth, emphasizing bias, is intended to optimize the performance of the test under the alternative. The skip-sampling periods $N = 4$ and $N = 12$ have also been tried, although the properties of the test do not appear very sensitive to this setting. These settings emerged as the best compromise in performance in null and alternative cases. The experiments returned both asymptotic and bootstrap p-values, using 300 bootstrap replications. The fractional differencing of the series prior to resampling has to be performed with an estimated d, as described in Section 5, and for this purpose a narrower bandwidth $M = [T^{0.55}]$ was used, to attenuate the bias. The simple bootstrap with independent resampling was used in the simulations of the pure fractional null hypothesis.

Table 2 shows the results obtained in nominal 5% tests for three cases of the null hypothesis. Under H_0, the data are generated as

$$y_t = (1 - L)_+^{-d} \varepsilon_t + z_t(d) \quad \varepsilon_t \sim \mathrm{NI}(0, 1)$$

where $z_t(d)$ is an independent Gaussian term generated by the method of Davidson and Hashimzade (2009), such that the sequence $\{y_t\}_{t=1}^{T}$ is stationary (see also Remark 2 of Section 5). The chosen values of d are shown in the column headings. The table entries show the proportion of replications in which the asymptotic and bootstrap p-values, respectively, fell below 0.05. The rows of the table show the performance in the three sample sizes of the basic skip-sampling test, and three cases of the composite test based on (6.1), with the values of δ shown in the first column of the table. The bootstrap p-values were computed using the algorithm from Section 5 using simple independent sampling to draw the fractional differences.

Under-sizing of the asymptotic test occurs in all sample sizes, and also in the bootstrap test in the larger samples, suggesting that the convergence is non-monotone.[2] If it is found surprising

---

[2] That this size distortion is wholly due to the replacement of true d by the estimated d was verified by conducting experiments using the former. In this case, the tests appear exact apart from experimental error.

		$d = 0.2$		$d = 0.4$		$d = 0.6$	
δ	T	Asymptotic	Bootstrap	Asymptotic	Bootstrap	Asymptotic	Bootstrap
Basic Test							
	250	0.009	0.080	0.003	0.091	0.013	0.064
	1000	0.008	0.086	0.010	0.058	0.023	0.040
	5000	0.001	0.108	0.020	0.040	0.027	0.026
Composite Test							
0.1	250	0.073	0.193	0.022	0.109	0.019	0.067
	1000	0.005	0.116	0.013	0.069	0.025	0.042
	5000	0.007	0.115	0.015	0.037	0.029	0.032
0.3	250	0.247	0.321	0.057	0.140	0.029	0.080
	1000	0.152	0.245	0.019	0.069	0.025	0.040
	5000	0.040	0.140	0.016	0.031	0.028	0.029
0.6	250	0.594	0.623	0.219	0.271	0.055	0.090
	1000	0.472	0.504	0.061	0.097	0.029	0.047
	5000	0.219	0.266	0.018	0.036	0.027	0.037

Table 2: Rejection rates in cases of the pure null (0.05 tests).

that these errors in rejection probability do not diminish more quickly, it is as well to remember that the components of the statistic depend upon as few as $[T/8]^{0.55}$ periodogram points, a mere 34 even in the case of $T = 5000$. The convergence to the asymptote is inevitably slow.

The issue of size distortion would clearly benefit from further study, and alternative estimators and bandwidths could certainly be considered. However, we note that under-rejection is a relatively benign problem provided the rejection rates under null and alternative differ sufficiently. Moreover, because the composite test yields a pseudo p-value that is always smaller than the bootstrap p-value, under-rejection is a desirable feature in the sense that the composite test is less prone to over-rejection under the null. Considering the alternative cases of δ in the composite test, the trade-off between power and size is evident here. The rate of over-rejection by the pseudo p-value under the null can be unacceptably large, but even with $\delta = 0.6$ this is a problem chiefly in small samples, or when d is close to zero. The latter results are not surprising, because when d is small, π_{1T} in (6.1) is on average closer to 1 then otherwise, and hence $\hat{\pi}_{2T}/\pi_{2T}$ smaller and over-rejection more acute. It is interesting that this effect persists in the largest sample considered but, again, it is important to remember that the slow rate of convergence. It is, if anything, more surprising to see how well the test can perform in modest sample sizes.

Next, in Tables 3 and 4 some cases of the alternative are shown. The performance of the simple and composite tests with the data generated as first-order autoregressions is reported in Table 3. This is the exponential decay model

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{NI}(0,1)$$

with $y_0 = 0$, for three different values of ϕ. Note that these are rejection rates, not power estimates, since the test sizes are uncorrected. With our compound null hypothesis, no consistent scheme for correcting rejection rates can be defined. This table therefore needs to be read in conjunction with Table 2. It's interesting to note that the rejection rate increases as ϕ is increased. This reflects the fact that the test has most power when autocorrelation is substantial but not hyperbolic. When the amount of autocorrelation is small, it is correspondingly difficult for the test to discriminate between exponential and hyperbolic decay, and this fact is reflected in the lower rejection rates observed.

		$\phi = 0.5$		$\phi = 0.7$		$\phi = 0.9$	
δ	T	Asymptotic	Bootstrap	Asymptotic	Bootstrap	Asymptotic	Bootstrap
Basic Test							
	250	0.010	0.336	0.133	0.513	0.280	0.393
	1000	0.010	0.382	0.306	0.713	0.914	0.885
	5000	0.005	0.375	0.369	0.742	1.00	1.00
Composite Test							
0.1	250	0.244	0.498	0.296	0.584	0.284	0.398
	1000	0.357	0.557	0.551	0.762	0.917	0.882
	5000	0.430	0.568	0.675	0.801	1.00	1.00
0.3	250	0.539	0.641	0.469	0.649	0.289	0.396
	1000	0.733	0.762	0.736	0.804	0.905	0.871
	5000	0.828	0.815	0.873	0.871	1.00	1.00
0.6	250	0.874	0.886	0.731	0.787	0.332	0.414
	1000	0.950	0.947	0.925	0.928	0.924	0.892
	5000	0.984	0.979	0.982	0.977	1.00	0.999

Table 3: Rejection rates, cases of the AR alternative (0.05 tests).

Table 4 shows rejection rates, in the basic test only, against a range of nonlinear I(0) processes. Here, we report the averages over replications of the log-periodogram estimates of d, side by side with the rejection frequencies by the bootstrap test, where we know in each case that the true d is zero. The models reported are as follows, where in each case $\varepsilon_t \sim \mathrm{NI}(0,1)$.

- "Bilinear" is of the form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1}\varepsilon_{t-1} + \phi_3 y_{t-1}\varepsilon_{t-2} + \varepsilon_t$$

  with $\phi_1 = 0.8$ and $\phi_2 = \phi_2 = 0.3$.

- "ESTAR" is the exponential self-exciting threshold AR case,

$$y_t = \alpha_1 y_{t-1}(1 - e^{(-\gamma y_{t-1}^2)}) + \alpha_2 y_{t-1} + \varepsilon_t,$$

  where $\alpha_1 = -1.5$, $\alpha_2 = 1$, $\gamma = 0.01$.

- "Markov Mean" is a model with Markov–switching intercepts. This takes the form

$$y_t = \alpha(S_t) + \varepsilon_t,$$

  where $\alpha(1) = 1$, $\alpha(2) = -1$ and $S_t = 1$ or 2 with $P(S_t = 1|S_{t-1} = 2) = P(S_t = 2|S_{t-1} = 1) = 0.05$.

- "Markov AR " is an autoregressive Markov-switching model,

$$y_t = \phi(S_t)y_{t-1} + \varepsilon_t,$$

  where $\phi(1) = 1.0$, $\phi(2) = 0.6$ and $P(S_t = 1|S_{t-1} = 2) = 0.03$ and $P(S_t = 2|S_{t-1} = 1) = 0.05$.

All of these models generate I(0) series, in the sense that their memory decay is ultimately exponential, but they have tended to give rise to highly biased GPH estimates, even in quite large samples.

	Bilinear		ESTAR		Markov Mean		Markov-AR	
T	p-val	$\hat{d}$	p-val	$\hat{d}$	p-val	$\hat{d}$	p-val	$\hat{d}$
250	0.720	0.343	0.415	0.503	0.138	0.457	0.292	0.653
1000	0.887	0.164	0.853	0.276	0.660	0.297	0.710	0.523
5000	0.960	0.052	0.993	0.099	0.995	0.125	0.994	0.309

Table 4: Bootstrap rejection rates in Nonlinear I(0) Alternatives (0.05 tests)

		Asymptotic	Bootstrap			
			Independent	Stationary		Sieve AR
d	T			Block Mean 5	Block Mean 10	
0.2	250	0.008	0.182	0.063	0.039	0.047
	1000	0.007	0.146	0.066	0.039	0.045
	5000	0.014	0.086	0.065	0.029	0.053
0.4	250	0.028	0.138	0.060	0.028	0.040
	1000	0.040	0.081	0.037	0.022	0.026
	5000	0.044	0.021	0.021	0.009	0.025
0.6	250	0.044	0.086	0.045	0.023	0.036
	1000	0.054	0.054	0.035	0.021	0.023
	5000	0.061	0.036	0.024	0.009	0.016

Table 5: Bootstrap rejection rates for ARFIMA(1,d,0) Models, $\phi = 0.3$ (0.05 tests)

Our final set of experiments examines rejection rates (basic test only) under what we might call the "contaminated null hypothesis", in other words, models in which the fractional differences of the process are autocorrelated. In such cases we have to consider resampling the differences using a bootstrap for dependent data. We consider two cases of the stationary bootstrap of Politis and Romano (1994) with exponential block length distributions with mean block lengths of 5 and 10 observations, and also the sieve-autoregressive method of Bühlmann (1997), where the lag length for the autoregression is chosen by the Akaike criterion up to a maximum of 10 lags. Including the asymptotic criterion, this makes for five test variants in all. Table 5 shows the results for models and sample sizes as in Table 2, but with an autoregressive component with $\phi = 0.3$. Thus, these are cases of the ARFIMA(1,d,0) class

$$(1 - \phi L)y_t = (1 - L)_+^{-d}\varepsilon_t + z_t(d), \quad \varepsilon_t \sim \mathrm{NI}(0,1)$$

The asymptotic test acquits itself relatively well here, and the independent bootstrap fails seriously only in the smallest sample. Under-rejection in the larger samples is again a feature of the findings, with the poorest performance delivered by the stationary bootstrap with the block-length 10.

9   Applications

We report two applications of the skip-sampling test. The first case considers a volatility measure for daily exchange rates of the British pound sterling against six currencies, in excess of 9,500 observations covering the period January 1975 to October 2012, (source: Bank of England). The measure in question is the logarithm of the absolute value of daily appreciation (log-change) augmented by 0.005. Taking logarithms normalizes the distribution by alleviating asymmetry and excess kurtosis, while adding the small constant overcomes the problem of days when zero change was recorded, so that the volatility measure is undefined.

	$\hat{d}$	Bootstrap p-values		
		Wald test	Skip-sample test	Bias test
Australian Dollar	0.532	0	0.789	0.013
Canadian Dollar	0.548	0	0.375	0.251
Danish Krone	0.471	0	0.883	0
Japanese Yen	0.414	0	0.987	0.278
New Zealand Dollar	0.523	0	0.977	0
US Dollar	0.429	0	0.319	0.144

Table 6: Tests for long memory in exchange rate volatility

Dividend growth	$\hat{d}$	Bootstrap p-values		
		Wald test	skip-sample test	Bias test
1946.1-2012.6	0.330	0.003	0	0.873
1871.2-2012.6	-0.068	1	0	0.575

Table 7: Tests for long memory in dividend growth

Table 6 reports in the first column the estimated d from GPH estimation with a bandwidth of $[T^{0.55}]$ where T is sample size. The succeeding columns show the bootstrap p-values for three tests: the usual Wald test (t-test) of significance of d, the skip-sampling test (skip period 8, bandwidth of $[T^{0.7}]$), and lastly the bias test of Davidson and Sibbertsen (2009) with a bandwidth of 0.92. The latter test has the null hypothesis of a pure fractional process, and tests for the presence of short-run autocorrelation in the fractional differences. The stationary block-bootstrap with a mean block-length of 5 was implemented with 299 bootstrap replications. Note that in these cases, given the significance test outcomes, the composite test could not return a result different from the simple skip-sampling test.

As can be seen, none of these skip-sampling tests leads to a rejection, so that the skip-sampling test reinforces the evidence from the Wald test that these series are long memory. The skip-sampling p-values tend to appear at the upper end of the unit interval, which is expected behaviour of the bootstrap in samples of this size, given the Monte Carlo findings. Even allowing for this distortion, however, the evidence in favour of the null hypothesis appears unequivocal.

Our second application is to the growth (log-change) in Shiller's S&P500 monthly real dividends series for January 1871–June 2012.[3] First, consider the sub-period starting in January 1946 (798 observations) with the results shown in the first row of the Table 7. The estimation and test settings are the same here as for the previous example and, once again, note that the composite test cannot return different findings. This result suggests that the long memory indicated by the Wald test is spurious. However, we have a more direct check on this finding in the present case, by extending the sample. The result with the full set of 1697 observations, starting in February 1871, is shown in the second row of Table 7, where the Wald test p-value falls emphatically in the non-rejection region of the 1-tailed test. This result shows how the biases in log-periodogram regression can persist in large samples, but also how the skip-sampling test offers the possibility of providing counter-evidence to this spurious significance.

---

[3]Posted by Robert Shiller at http://www.econ.yale.edu/~shiller/data.htm.

10   Conclusion

In this paper we have investigated the performance of a test for the null hypothesis of long memory, based on the self-similarity property of sequences with hyperbolic memory decay. The idea is to compare GPH log-periodogram estimators in original and skip-sampled versions of the data set. The aliasing phenomenon, which introduces an estimation bias in skip-samples, poses a problem for the implementation of this test, but a bias-corrected estimator permits the construction of an asymptotically pivotal statistic.

The use of a semiparametric method to construct the estimators and test statistic, with correspondingly slow convergence to the asymptote, inevitably poses a challenge for the implementation of the test, and the bootstrap variant of the test performs relatively well, in spite of being implemented using semi-parametric estimates of the null distribution. This combination of factors proves to pose a problem of under-rejection even in quite large samples, in the Monte Carlo evaluations. A bias reduction strategy such as the double bootstrap (Beran 1988) might alleviate this problem, at the cost of a large computational overhead, but in the settings where the test might be applied, under-rejection is a relatively benign problem, and the experiments indicate reasonable power properties. Alternative choices of test settings, such as GPH bandwidths and skip period, could prove helpful, although to research these must take us beyond the scope of this study. Notwithstanding these qualifications, the test may prove a useful addition to the arsenal of diagnostic procedures for long memory models, beside the bias test of Davidson and Sibbertsen (2009), which compares log-periodogram estimates with different bandwidths, and the aggregation test of Ohanissian et al. (2008).

Appendix

Proof of Proposition 2.1

Let γ_k denote the kth autocovariance, defined by the identity

$$\gamma_k = \int_0^{2\pi} \cos(k\omega) f(\omega) d\omega.$$

For the skip-sampled data with sampling period N, the autocovariances are γ_{Nk} where

$$\begin{aligned}
\gamma_{Nk} &= \int_0^{2\pi} \cos(Nk\omega) f(\omega) d\omega \\
&= \sum_{j=0}^{N-1} \int_{2\pi j/N}^{2\pi(j+1)/N} \cos(Nk\omega) f(\omega) d\omega \\
&= \sum_{j=0}^{N-1} \int_0^{2\pi/N} \cos(Nk\omega) f(\omega + 2\pi j/N) d\omega \\
&= \int_0^{\pi} \cos(k\lambda) f_N(\lambda) d\lambda.
\end{aligned}$$

where the third equality makes use of the fact that $\cos(Nk\omega) = \cos(Nk\omega + 2\pi j)$, and the fourth one makes the change of variable $\lambda = N\omega$ and the substitution

$$f_N(\lambda) = \frac{1}{N} \sum_{j=0}^{N-1} f((\lambda + 2\pi j)/N). \qquad \blacksquare$$

Proof of Proposition 4.1

Letting

$$A(\lambda, k, N) = \frac{\sin(\pi k/N)}{\tan(\lambda/2N)} + \cos(\pi k/N)$$

note first that

$$\frac{dH}{dd} = \frac{1}{N} \sum_{k=0}^{N-1} A(\lambda, k, N)^{-2d} \left[-2 \frac{g\left((\lambda + 2\pi k)/N\right)}{g(\lambda/N)} \log A(\lambda, k, N) \right.$$
$$\left. + \frac{d}{dd} \left(\frac{g\left((\lambda + 2\pi k)/N\right)}{g(\lambda/N)} \right) \right].$$

We obtain a formula for the derivative in the second term, and show that this is bounded in the limit. The terms of the form (3.2) depend on d because the data used to construct the sieve autoregressive estimates are the fractional differences of the measured data. Assume that p is fixed, and let $z_t = (1-L)^d x_t$ and so let $\mathbf{Z}_0$ $(T-p \times p)$ be the normalized data matrix whose columns are the vectors $\mathbf{z}_j = (z_{p+1-j}, \ldots, z_{T-j})'$ for $j = 1, \ldots, p$. Also, let $\mathbf{Z}_j$ for $j = 1, \ldots, p$ denote the matrix equal to $\mathbf{Z}_0$ except that the jth column has been replaced by $\mathbf{z}_0 = (z_{p+1}, \ldots, z_T)'$. Then, note that the coefficients $\hat{\phi}_j$ in the autoregression of order p can be written using Cramer's rule as

$$\hat{\phi}_j = \frac{|\mathbf{Z}_0' \mathbf{Z}_j|}{|\mathbf{Z}_0' \mathbf{Z}_0|}, \ j = 1, \ldots, p.$$

Let these elements define the $p + 1 \times 1$-vector $\hat{\boldsymbol{\phi}}$ by also putting $\hat{\phi}_0 = -1$.

Now, let $\boldsymbol{Q}(\theta)$ $(p + 1 \times p + 1)$ denote the real part of the Fourier matrix with elements $q_{rs} = \cos \theta(r - s)$ for $r, s = 0, \ldots, p$. Setting $\theta_1 = (\lambda + 2\pi k)/N$ and $\theta_2 = \lambda/N$, note that

$$\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} = \frac{|\hat{\phi}(e^{-i\theta_1})|^{-2}}{|\hat{\phi}(e^{-i\theta_2})|^{-2}} = \frac{\hat{\boldsymbol{\phi}}' \boldsymbol{Q}(\theta_2) \hat{\boldsymbol{\phi}}}{\hat{\boldsymbol{\phi}}' \boldsymbol{Q}(\theta_1) \hat{\boldsymbol{\phi}}} = \frac{\boldsymbol{b}' \boldsymbol{Q}(\theta_2) \boldsymbol{b}}{\boldsymbol{b}' \boldsymbol{Q}(\theta_1) \boldsymbol{b}}$$

where $\boldsymbol{b}$ is the $p + 1$-vector having elements $b_0 = -T^{-p} |\boldsymbol{Z}_0' \boldsymbol{Z}_0|$ and $b_j = T^{-p} |\boldsymbol{Z}_0' \boldsymbol{Z}_j|$ for $j = 1, \ldots, p$. In this notation we have

$$\frac{\mathrm{d}}{\mathrm{d}d} \left(\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} \right) = \left[\frac{\boldsymbol{b}'[\boldsymbol{Q}(\theta_2) + \boldsymbol{Q}'(\theta_2)]}{\boldsymbol{b}' \boldsymbol{Q}(\theta_1) \boldsymbol{b}} - \frac{\boldsymbol{b}' \boldsymbol{Q}(\theta_2) \boldsymbol{b} \boldsymbol{b}'[\boldsymbol{Q}(\theta_1) + \boldsymbol{Q}'(\theta_1)]}{[\boldsymbol{b}' \boldsymbol{Q}(\theta_1) \boldsymbol{b}]^2} \right] \frac{\mathrm{d}\boldsymbol{b}}{\mathrm{d}d}$$

and it remains to evaluate the second right-hand side factor.

Start with the elements of the $\boldsymbol{Z}_j$ matrices. Considering row t, let m denote the generic lag associated with a column of $\boldsymbol{Z}_j$. Using the argument from Tanaka (1999), Section 3.1, the derivatives with respect to d can be written as

$$\begin{aligned} \frac{\mathrm{d}z_{t-m}}{\mathrm{d}d} &= \frac{\mathrm{d}}{\mathrm{d}d}(1 - L)^d x_{t-m} \\ &= \log(1 - L)(1 - L)^d x_{t-m} \\ &= -\sum_{k=1}^{\infty} k^{-1} z_{t-m-k} \\ &= z_{t-m}^* \end{aligned}$$

where the last equality defines z_{t-m}^*. We have from Magnus and Neudecker (1988), p149, that for $j = 0, \ldots, p$,

$$\begin{aligned} \mathrm{d} \left| T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j \right| &= \left| T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j \right| . \mathrm{tr}(T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j)^{-1} T^{-1} \mathrm{d}(\boldsymbol{Z}_0' \boldsymbol{Z}_j) \\ &= \left| T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j \right| . \mathrm{tr}(T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j)^{-1} T^{-1} \left(\mathrm{d}\boldsymbol{Z}_0' \boldsymbol{Z}_j + \boldsymbol{Z}_0' \mathrm{d}\boldsymbol{Z}_j \right) \\ &= \left| T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j \right| . \mathrm{tr}(T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j)^{-1} \left(T^{-1} \boldsymbol{Z}_0^{*'} \boldsymbol{Z}_j + T^{-1} \boldsymbol{Z}_0' \boldsymbol{Z}_j^* \right) \mathrm{d}d \\ &= b_j^* \mathrm{d}d \end{aligned}$$

(defining b_j^*) where the $\boldsymbol{Z}_j^*$ denote the matrices with elements z_{t-m}^*, with the value of m defined as appropriate, according to the construction of $\boldsymbol{Z}_j$. Letting $\boldsymbol{b}^*$ denote the vector with elements $-b_0^*$ and b_j^*, for $j = 1, \ldots, p$, we now have the result

$$\frac{\mathrm{d}}{\mathrm{d}d} \left(\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} \right) = \frac{\boldsymbol{b}'[\boldsymbol{Q}(\theta_2) + \boldsymbol{Q}'(\theta_2)]\boldsymbol{b}^*}{\boldsymbol{b}' \boldsymbol{Q}(\theta_1) \boldsymbol{b}} - \frac{\boldsymbol{b}' \boldsymbol{Q}(\theta_2) \boldsymbol{b} \boldsymbol{b}'[\boldsymbol{Q}(\theta_1) + \boldsymbol{Q}'(\theta_1)]\boldsymbol{b}^*}{[\boldsymbol{b}' \boldsymbol{Q}(\theta_1) \boldsymbol{b}]^2}.$$

Since $\{z_t\}$ is a weakly dependent process by hypothesis, the process z_t^* is covariance stationary. It follows directly that, for every finite p, $\boldsymbol{b}^*$ converges in probability to a non-stochastic limit, depending on the autocovariances of $\{z_t\}$. From the fact that $\boldsymbol{b}$ converges in the same manner, and the Slutsky theorem, the proposition follows under the conditions stated.

Two simplifying assumptions have been made to reach this conclusion. The first is that z_t has been constructed as an infinite order moving average, whereas in practice the sums will be truncated, containing only the first $t - m$ terms. However, since the truncation affects at most a finite number of terms, this cannot change the value of the limit. Second, the lag length p has been assumed fixed. However, since z_t is a weakly dependent process by hypothesis, the autocovariances are summable and hence equal zero for lags exceeding some finite value. Letting p tend to infinity with T cannot change the distribution of $B(N, d)$ beyond some point, since the additional elements of $\boldsymbol{b}$ and $\boldsymbol{b}^*$ have sums converging to zero as p increases. ∎

References

Agiakloglou, C., P. Newbold and M. Wohar (1993) Bias in an estimator of the fractional difference parameter, *Journal of Time Series Analysis* 14, 235-46

Baillie, R. T., T. Bollerslev and H. O. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 74, 3-30

Beran, R. 1988, Prepivoting test statistics: a bootstrap view of asymptotic refinements, *Journal of the American Statistical Association* 83, 687-697.

Bühlmann, P. (1997) Sieve bootstrap for time series, *Bernoulli* 3, 123-148.

Chambers, M. J. (1998) Long Memory and Aggregation in Macroeconomic Time Series, *International Economic Review*, Vol. 39, No. 4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance (Nov., 1998), pp. 1053-1072

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press

Davidson, J. (2004) Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economics Statistics* 22, 1, 16-29.

Davidson, J. (2009) "When is a time series I(0)?". Chapter 13 of *The Methodology and Practice of Econometrics, a festschrift for David F. Hendry* eds. Jennifer Castle and Neil Shepherd, Oxford University Press.

Davidson, J. and N. Hashimzade (2009)Type I and type II fractional Brownian motions: a reconsideration" *Computational Statistics and Data Analysis* 53(6) 2089-2106

Davidson, J. and P. Sibbertsen (2009) Tests of Bias in Log-Periodogram Regression *Economics Letters* 102 (2009), 83-86.

Diebold, F. X. and A. Inoue (2001) "Long memory and regime switching", *Journal of Econometrics* 105, 131-159.

Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models, *Econometrica* 65 (6), 1365-1387.

Faust, J. (1996) Near observational equivalence and theoretical size problems with unit root tests, *Econometric Theory* 12 (4) 724-731.

Faust, J. (1999) Conventional confidence intervals for points on spectrum have confidence level zero, *Econometrica* 67 (3) 629-637.

Gallant, A. R. and White H. 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford.

Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221-237

Hassler (2011) Estimation of fractional integration under temporal aggregation, *Journal of Econometrics* 162, 240-247

Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of a long memory time series, *Journal of Time Series Analysis* 19, 19–46

Kim, C. S. and P. C. B. Phillips (2006) Log-periodogram regression: the nonstationary case. Cowles Foundation Discussion Paper 1587, Yale University.

Kreiss, J-P., E. Paparoditis and D. N. Politis (2011) "On the range of validity of the autoregressive sieve bootstrap", *Annals of Statistics*, 39, 2103-2130.

Magnus, J. R. and H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Chichester, John Wiley & Sons.

Moulines, E. and P. Soulier (1999) Broad band log-periodogram estimation of time series with long-range dependence. *Annals of Statistics* 27, 1415–1439

Ohanissian, A, J. R. Russell and R. S. Tsay et al. (2008) True or Spurious Long Memory? A New Test, *Journal of Business and Economic Statistics* 26, 2, 161-175

Politis, D N. and J P. Romano (1994) The stationary bootstrap *Journal of the American Statistical Association* 89, 1303-1313.

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 (3) 1035–1065.

Robinson, P. (1995) Log-periodogram regression of time series with long-range dependence. *Annals of Statistics* 23, 1048–1072.

Smith, J. and Souza, L. R, (2002) Bias in the memory parameter for different sampling rates, *International Journal of Forecasting* 18, 299-313

Smith, J. and Souza, L. R, (2004) Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: a Monte-Carlo study, *International Journal of Forecasting* 20, 487–502.

Souza, L. R. (2005) A note on Chambers's "long memory and aggregation in macroeconomic time series" *International Economic Review*, Vol. 46, No. 3, 1059-1062

Tanaka, K. (1999) The nonstationary fractional unit root. Econometric Theory 15, 549-582

Velasco, C. (1999) Non-Stationary Log-Periodogram Regression, *Journal of Econometrics* 91, 325-371.