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

James Davidson and Dooruj Rambaccussing
Department of Economics
University of Exeter Business School
Rennes Drive
Exeter EX4 4PU
U.K.

October 2011

Abstract

This paper develops a new test of true versus spurious long memory, based on log-peri- odogram 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 test is imple- mented using the bootstrap, with the distribution under the null hypothesis approximated using the sieve-autoregression to approximate short-run dependence following fractional differencing. The procedure is designed to be used in the context of rejection in a conventional test of significance of the long memory parameter, and in this context the test is consistent. Its properties are investigated in a set of Monte Carlo experiments.

1 Introduction

Semiparametric estimation of long memory is a popular methodology in time series analysis. When the autocovariances of a process are nonsummable, it is well known that the spectral density $f$ diverges at the origin with

\[ f(\lambda) = O(|\lambda|^{-2d}) \quad (1.1) \]

as $\lambda \to 0$, and this characteristic provides the basis for semiparametric estimation of the long memory parameter $d$, which is zero in the summable case. The log-periodogram regression originally proposed by Geweke and Porter-Hudak (1983, henceforth GPH) is a well-established implementation of this idea. The convenient fractional differencing representation of long memory (Granger and Joyeux 1980, Hosking 1981) in which

\[ f(\lambda) = |1 - e^{-i\lambda}|^{-2d}g(\lambda) \quad (1.2) \]

where $g$ is bounded at the origin, is often adopted. Since $|1 - e^{-i\lambda}|^2 = \lambda^2 + O(\lambda^4)$ there is little loss of generality in assuming (1.2) provided the form of $g$ is otherwise unspecified.

An inherent problem with this approach to investigating time series is that the form of the spectrum at the origin defines long-run – truthfully, infinite run – behaviour, whereas observed samples are finite. There are time series models for which the GPH estimator will return large and "significant" values of $d$ in finite samples, in spite of the fact that the autocovariances are summable. 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 $\phi = 1$. Indeed, the ARFI model

\[ (1 - \phi L)(1 - L)^d x_t = u_t \]
is well-known to exhibit a characteristically bimodal likelihood function, when either of the parameters $\phi$ and $d$ is close to unity in the process generating the sample. For every finite sample size, there exists a $\phi$ large enough 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 evidently an unreliable guide.

Recent research has highlighted the well-known property of self-similarity of hyperbolic decay processes under rescaling transformations, such as periodic aggregation and periodic subsampling, 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 in the context of 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 $n$th 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 these in the context of hyperbolic memory decay. Let the parameter $\delta$ index the rate of decay such that the autocovariance sequence of a stationary process satisfies

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

for some $\delta > 0$. The hyperbolic memory class includes short memory processes having summable autocovariances, such that $\delta > 1$, and the long memory class 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 at 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 $\delta$, but there exists $\rho > 0$ such that

$$\gamma_j = O(e^{-\rho j}),$$

In this case, note that

$$\gamma_{nj} = O(e^{-\rho nj})$$

so that the memory decay parameter rises from $\rho$ to $\rho n$ following skip-sampling. Since the estimator of (spurious) $d$ in the exponential decay case is inevitably sensitive to the value of $\rho$, 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 ESTAR, SETAR and Markov-switching processes are often thought of as likely to be mistaken for long memory, since they can exhibit local patterns of apparent persistence, switches of local mean, for example, or unit root-like behaviour in the neighbourhood of the origin. As in the case of the linear autoregressive model, the essential difference between these latter 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.
Note that the class of cases of (1.3) with $1 \leq \delta < \infty$ count as instances of the alternative hypothesis for present purposes. Models of this sort do not seem to have been significantly exploited to date in econometrics, except in the rather special contexts of 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 FIGARCH (Baillie et al. 1996) and HYGARCH (Davidson 2005) models are cases of the ARCH($\infty$) 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 of true long memory.

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 we have in mind, the simple difference of the two estimators, is asymptotically Gaussian under the usual assumptions of this literature (notably Gaussianity of the observations, see Robinson 1995, Hurvich et al., 1998) although with unknown variance. We develop a bootstrap procedure that should be asymptotically correctly sized, at least for processes that are linear under the null hypothesis.

The rest of 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 describes the procedure for the bootstrap test. Section 4 derives the bias-corrected estimator for skip-samples needed to implement the test. Section 5 discusses the null asymptotic distribution of the statistic, and Section 6 considers the appropriate usage of the test and its properties under the alternative. Our Monte Carlo findings are reported in Section 7 and Section 8 concludes the paper.

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 \( \{x_t, t = 1, 2, \ldots\} \) is a discrete stochastic process with spectral density $f$ and $y_t = x_{nt}$ for $t = 1, 2, 3, \ldots$ and $n > 1$, the spectral density of the process $y_t$ is

$$f_n(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} f \left( \frac{\lambda + 2\pi j}{n} \right).$$

The straightforward proof is given in the appendix. Note that cycles of frequency $\lambda/n$ in the original data become cycles of frequency $\lambda$ in the skip-sampled data, and frequencies above $\pi/n$ are no longer identifiable. Hence, these contributions to the variance of the series are effectively aggregated with the identifiable frequencies.

Applying this formula to the fractionally integrated case

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

(2.1)
where $0 < g(0) < \infty$, we find that $f_n(\lambda)$ does not admit to direct log-linearization in the GPH manner. What can be done, following the suggestion of Smith and Souza (2002), is to write

$$f_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} \left( 2 \sin \left( \frac{\lambda + 2\pi k}{2n} \right) \right)^{-2d} g \left( \frac{\lambda + 2\pi k}{n} \right)$$

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

where

$$H_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} \left( \frac{\cos(\lambda/2n)}{\sin(\lambda/2n)} \sin(\pi k/n) + \cos(\pi k/n) \right)^{-2d} g \left( \frac{(\lambda + 2\pi k)/n}{g(\lambda/n)} \right).$$

There is, evidently, an omitted term $\log H_n$ in the log-periodogram regression in skip-sampled data, depending on $d$ as well as $\lambda$. 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 test

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. Let $n$ denote the periodicity of the skips. Skip-sampling is done by taking every $n$th observation, so yielding a sample of size $\lceil T/n \rceil$ where $\lceil z \rceil$ denotes the largest integer below $z$. This can be done $n$ times, by off-setting the initial observation, so that the $n$ skip-samples are \(\{x_{0}\}, \{x_{1}\}, \ldots, \{x_{n-1}\}\), where for $k = 0, \ldots, n - 1$,

\[\begin{align*}
  x_{k1} &= y_{k+1} \\
  x_{k2} &= y_{n+k+1} \\
  x_{k3} &= y_{2n+k+1} \\
  \vdots \\
  x_{kT} &= y_{(\lceil T/n \rceil - 1)n+k+1}.
\end{align*}\]

Let the conventional GPH estimator, based on the complete sample, be denoted $\hat{d}$ and let modified log-periodogram estimators, with a bias correction to be explained in greater detail in Section 4 below, be denoted $\hat{d}_{nk}$ for $k = 0, \ldots, n - 1$. The test statistic is defined as

$$\tau = \hat{d} - \hat{d}_{n}$$

where $\hat{d}_{n} = n^{-1} \sum_{k=0}^{n-1} \hat{d}_{nk}$.

The test is implemented as a bootstrap test where the distribution under the null hypothesis is simulated as a fractionally integrated process, allowing for the possibility of linear short-run
dependence of the fractional differences. Thus, linearity is the chief restriction on the class of
models included in the null hypothesis. The sieve-autoregression procedure (Bühlmann, 1997) is
used to model this dependence in the bootstrap draws. Given \( \hat{d} \) and \( \tau \) as in (3.1), computed from
the observed sample, the steps leading to the computation of a p-value for comparison with the
chosen significance level are as follows.

1. Compute the fractional differences \( \hat{u}_t = (1 - L)^{\hat{d}} y_t \).

2. Fit an autoregression of order \( p_T \) for \( \hat{u}_t \) using the Durbin-Levinson algorithm, where \( p_T \)
is chosen to optimize the Akaike criterion subject to \( p_T \leq 0.6T^{1/3} \). Let \( \hat{\phi}(L) \hat{u}_t \),
\( t = p_T + 1, \ldots, T \) denote the residuals from this model.

3. Repeat the following steps for \( j = 1, \ldots, B \).
   
   (a) Draw a random sample \( \hat{\varepsilon}^*_1, \ldots, \hat{\varepsilon}^*_T \) with replacement from the distribution
   \( P(\hat{\varepsilon}^*_t = \hat{\varepsilon}_t - \bar{\varepsilon}) = 1/(T - p_T) \), \( t = 1, \ldots, T - p_T \), where
   \( \bar{\varepsilon} = (T - p_T)^{-1} \sum_{t=p_T+1}^T \hat{\varepsilon}_t \).
   
   (b) Generate the bootstrap data sample as
   \( \hat{y}^*_t = (1 - L)^{-\hat{d}} \hat{\phi}(L)^{-1} \hat{\varepsilon}^*_t 1_{\{t \geq 1\}} + \hat{z}_{tj} \), \( t = 1, \ldots, T \)
   where \( \hat{\varepsilon}^*_t 1_{\{t \geq 1\}} = 0 \) for \( t < 1 \) and the sequence \( \hat{z}_{1j}, \ldots, \hat{z}_{Tj} \) is generated independently
to simulate the contribution of the presample shocks.
   
   (c) Compute the bootstrap statistic \( \tau^*_j \) as in (3.1) for the sample \( \hat{y}^*_1, \ldots, \hat{y}^*_T \).

4. Compute the estimated p-value for the test as 0 if \( \tau > \tau^*_B \) or else
   \[
   1 - \frac{\min\{j : \tau \leq \tau^*_j\}}{B}
   \]
   where \( \tau^*_j \) is the \( j \)th order statistic for the bootstrap statistics \( \tau^*_1, \ldots, \tau^*_B \).

Remarks

1. We use the signed test statistic and hence do 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.

2. In view of the form of the estimator, the use of the average of the \( d \) estimates from the \( n \) skip
samples is equivalent to adopting the average of the log-periodogram points as regressand.
This scheme makes the most efficient use of the available data.

3. The correction terms \( \hat{z}_{jt} \) are constructed using Gaussian drawings and weights constructed
from the estimated parameters to have a covariance structure matching the components
omitted through truncating the innovation sequence at 0. The resulting sequence is
approximately stationary for \( |\hat{d}| < 1/2 \). If \( \hat{d} \geq 1/2 \) 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).
4. The use of the AR sieve bootstrap to reproduce the limiting distribution of the statistic appears valid provided that the spectral density of the fractional differences satisfies the smoothness conditions specified in Kreiss et al. (2011); actual linearity is not necessary. Since the data are first fractionally differenced using an estimator that is consistent but converges more slowly than $\sqrt{T}$, the convergence of the bootstrap distribution is correspondingly slow. This fact does not appear to contradict the convergence itself.

4 The Bias-corrected estimator

The construction of the estimators $\hat{d}_{nk}$ used in the test is a key issue. As shown in Section 2 the conventional GPH estimator applied to skip-sampled data is biased. The magnitude of the bias depends on $d$ as well as the form of the short-run dependence, and this bias is not attenuated by a narrow bandwidth. Since the test is to be implemented with the bootstrap, bias might not be regarded as critical. An asymptotically correctly sized test is assured provided the bootstrap replications reproduce the distribution of $\tau$ under the null hypothesis. However, there are two main reasons why the mean of the statistic is in fact an important issue.

First, the dependence of the mean on nuisance parameters implies that the null distribution is not asymptotically pivotal. As is well known (see e.g. Horowitz 2000) this has the effect of increasing the order of magnitude in $T$ of the error in rejection probability (ERP). This is especially important because of the relatively slow rates of convergence of the narrow band estimators employed here. The bias-corrected test has mean zero asymptotically, still exhibiting some dependence on the unknown $d$, but of $O_p(M^{1/2})$, where $M$ is the GPH bandwidth, instead of $O_p(1)$.

Second, the test is designed to determine whether "large" (nominally significant) estimates of the $d$ parameter are spurious or consistent for the rate of hyperbolic memory. The power of the test depends on the skip-sampling estimates lying significantly closer to zero under alternatives than the full-sample estimates. When the null is false, however, the long-memory component of the bias term that would be present under the null hypothesis is absent. Hence, the test comparison becomes a matter of comparing one bias with another. Test power could be correspondingly poor.

Bias correction involves finding a computable surrogate for $H_n$ in (2.3), and there are two issues to be considered. The expression depends in the first place on the unknown parameter $d$, and in the second place on the unknown spectral density component $g(\lambda)$. Since for the purposes of the test the estimates from the skip samples are to be compared with those from the full sample, the natural approximation is to replace $d$ with the asymptotically unbiased estimator $\hat{d}$.

Except in the case of the pure fractional difference, the term $g(\lambda + 2\pi j)/n) / g(\lambda/2)$ will in general vary with $\lambda$ over the whole of the interval $[0, 2\pi]$, including points close to the origin. Approximating it by a constant is therefore not an attractive option, notwithstanding that this is the approach for dealing with $g(\lambda/2)$ in the narrow-band estimator. One option would be to construct a kernel estimator of $g$ from the spectrum of the fractional differences. However, the method chosen here is semiparametric. In view of the fact that the sieve autoregression is to be used in any case to simulate the data for the purposes of the bootstrap test, a natural approach is to make use of these fitted parameters, and so approximate $g(\lambda)$ by

$$ \hat{g}(\lambda) = |\hat{\phi}(e^{-i\lambda})|^2. \quad (4.1) $$

Essentially, this option trades the requirement to assume a linear process under the null hypothesis for the greater efficiency afforded by the autoregressive parameterization.

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. Let \( I_{nk} \) denote the periodogram computed from the \( k \)th skip-sampled data set, and let \( \hat{H}_n(\lambda) \) denote formula in (2.3) approximated as described, using the estimated parameters and the representation of the short-run spectral density in (4.1). The bias-corrected skip-sample estimator then takes the form.

\[
\hat{d}_{nk} = \frac{\sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \log I_{nk}(\lambda_{nj}) - \log \hat{H}_n(\lambda_{nj})}{\sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n)^2}
\]  

(4.2)

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.

5 Asymptotic distribution of the statistic

Let the null hypothesis specify that the random sequence is stationary and Gaussian, having a Wold representation of the form

\[ x_t = (1 - L)^{-d} \theta(L) \varepsilon_t \]

where \( \theta(L) \) is an invertible lag polynomial of potentially infinite order and \( \varepsilon_t \sim \text{NID}(0, \sigma^2) \). Since \( \theta(L) \) is arbitrary apart from having summable coefficients this representation is very general, restricting only the tail of the lag distribution. It suffices that the moving average coefficients are of \( O(j^{d-1}) \) as \( j \to \infty \).

When the sample is large enough, both the conventional GPH estimator \( \hat{d} \) and the skip-sampled estimator \( \hat{d}_{nk} \) defined in (4.2) can be analysed using the techniques developed in Hurvich et al. (1998) (henceforth HDB). In other words, letting \( \varepsilon_{nkj} = \log (I_{nk}(\lambda_{nj})/f(\lambda_{nj})) \) there exists a function \( f^* \) such that (reproducing the expression in HDB page 42)

\[
M_n^{1/2}(\hat{d}_{nk} - d) = -\frac{M_n}{2S_n} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \log f^*_{nj} - \frac{M_n}{2S_n} \frac{1}{M_n^{1/2}} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \varepsilon_{nkj}
\]

where \( S_n = \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n)^2 = O(M) \), and the first right-hand side term is \( o(1) \) on the conditions \( M \to \infty \) and \( (M \log M)/T \to 0 \). The case \( n = 1, k = 0 \) is the standard case, without skip-sampling, so that \( f^*_{nj} = f^*_j = g(\lambda_j) \), and \( \varepsilon_{nkj} = \varepsilon_j \), while the case \( n > 1 \) has

\[
\log f^*_{nj} = \log g\left( \frac{\lambda_j}{n} \right) - \log \left( \frac{\hat{H}(\lambda_{nj})}{\hat{H}(\lambda_{nj})} \right).
\]

Recalling that \( \hat{d} \) is \( M^{1/2} \)-consistent, and noting that \( H \) is twice-differentiable with respect to \( d \), expand \( \log \hat{H}(\lambda_{nj}) \) as

\[ \log \hat{H}(\lambda_{nj}) = \log H(\lambda_{nj}) + \frac{H(\lambda_{nj})'}{H(\lambda_{nj})} (\hat{d} - d) + O(M^{-1}). \]

Then, using Lemma 1 of HDB, and letting

\[ B_T(n, d) = \frac{1}{2S_n} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \frac{H(\lambda_{nj})'}{H(\lambda_{nj})} \]

we have

\[
M_n^{1/2}(\hat{d}_{nk} - d) = n^{-q/2} B_T(n, d) M^{1/2}(d - d) - \frac{M_n}{2S_n} \frac{1}{M_n^{1/2}} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \varepsilon_{nkj} + o(1).
\]
Be careful to note that the relevant properties of the $\varepsilon_{nkj}$ extend 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. These random variables represent continuous transformations of the discrete Fourier transforms of the data, which in the skip-sampled case are simply weighted sums of the original data points where the weights are zeros except for every $n$th observation and otherwise are the usual trigonometric functions, defined on the interval $[0, \pi/n]$. Since the regressors are the same for each $k$, we further find

$$M^{1/2}(\hat{d}_n - d) = B_T(n,d)M^{1/2}(\hat{d} - d) - \frac{n^{-q/2-1}M}{2S_n} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \sum_{k=0}^{n-1} \varepsilon_{nkj} + o(1). \quad (5.1)$$

In the appendix, we show the following

**Proposition 5.1** 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 $B(n,d)$.

It follows that apart from terms of small order, $M^{1/2}(\hat{d}_n - \hat{d})$ is asymptotically Gaussian with a mean of zero and finite variance. However, the distribution is not free of nuisance parameters since it depends on $n$ as well as on $B(n,d)$. These dependencies warrant the use of the bootstrap, as the most practical implementation of the test.

### 6 Properties under the alternative hypothesis

Testing the degree of persistence of time series is a problem that has attracted a degree of controversy, as has been 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 might be summarised as follows: cases of the null hypothesis constitute an open set in the parameter space, and leading cases of the alternative are contained in the closure of that set. 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 much more general parameterizations of the null, it is most transparent in the case where the "I(0)" property depends on the modulus of the maximal autoregressive root, and the null hypothesis is represented by the interval $[0,1)$. The present case is clearly similar, except that the null hypothesis, relating to the value of $d$, is the case of 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. For this reason, it is important to emphasize the context in which this type of test might be useful.

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 an estimator of zero. Since the estimators being compared are both consistent, albeit biased in finite samples, the test appears inconsistent. In the limit as the sample size tends to infinity, the null distribution simulated in the bootstrap test is converging on a case of the alternative (short memory) as the bias in both estimators of $d$ converges on zero. There is no reason to suppose that the the probability of exceeding the rejection criteria is always increasing in sample size.
However, suppose that the test is viewed as the second stage of a two-stage procedure, being performed only in the case of rejection in a significance test on \( d \). In the case of non-rejection in this first-stage test, the null hypothesis is to be regarded as rejected, and there is no need to proceed to the second stage. Since the log-periodogram estimators converge in probability to zero under the alternative, the combined test is consistent; that is to say, it will reject in the limit with probability \( 1 - \alpha \) where the first-stage significance level \( \alpha \) can be set as close to zero as desired, and should naturally be made a decreasing function of sample size. It is not feasible to compute an exact significance level for the composite test, although when the null hypothesis is true we have from the Bonferroni inequality that the probability of a rejection by the composite test exceeds that of rejection by the skip-sample test alone. But since the first-stage test is consistent so that the probability of proceeding to the second stage converges to 1 when the null is true, the combined test has the same size as the second-stage test asymptotically. Some evidence on this procedure is given in the following section.

An issue not so far addressed is the status, as cases of the null hypothesis, of nonstationary fractional processes having \( \frac{1}{2} \leq d \leq 1 \). 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 experiments, reported in the following section, report results for both stationary and nonstationary cases of the null, with similar results. As the observational equivalence issue raised in the Introduction would lead us to predict, autoregressively generated series with a root close to unity characteristically yield an estimated \( d \) in the non-stationary range. A unit root process is a case of the null hypothesis, noting that this case exhibits the invariance of the memory to skip-sampling characteristic of the fractional integration case. Considering a sequence of models with maximal autoregressive modulus ranging from unity down to zero, we expect (with fixed sample size) to find the nominal rejection probabilities initially increasing over this range. They may not vary monotonically over the range, but by interpreting non-rejection in the significance test on \( d \) as evidence for the alternative, we are able to discount the behaviour of the test in cases of \( d \) close to zero.

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 I(1) as \( T \) is extended with fixed parameters. As pointed out, the present test is not expected to have power in such cases. In essence a skip-sampled unit root process remains a unit root. However, there are also examples where the processes are "revealed" as I(0), and in particular these authors consider a simple independent process subject to Markov-switching. This is one of the case we consider in simulation experiments in the next section.

### 7 Monte Carlo Experiments.

Each of the tables in this section shows the results of experiments with four sample sizes (\( T \)), two choices of GPH bandwidth (\( M \), expressed as a fractional power of \( T \)), and three alternative skip rates (\( n \)). In each case, 5000 replications have been performed, and the bootstrap test is performed with 200 replications. The tables show the fifth percentiles of the distribution of the bootstrap \( p \)-values.

In this framework, the simulation evidence we present is falls into five categories, as shown in Tables 1-5. First, cases of the null hypothesis are generated as ARFIMA(1,\( d \),0) models, with the form

\[
(1 - \phi L)(1 - L)^d y_t = \varepsilon_t.
\]

Here and also in all the subsequent experiments reported, \( \varepsilon_t \sim N(0,1) \). Table 1 shows stationary
Taking spurious long memory, there is inevitably a trade-off of power for size in the choice of bandwidth. The width is chosen smaller. Since it is bias in the estimator that is also the basis for detecting source of such distortion is bias in the estimator of the GPH estimates of \( d \), which is of course smaller when the bandwidth is chosen smaller. Since it is bias in the estimator that is also the basis for detecting spurious long memory, there is inevitably a trade-off of power for size in the choice of bandwidth. Taking \( M \) in the vicinity of \( T^{0.7} \) might be the most advantageous choice from the point of view of power, since the biases and their potential difference are then exaggerated, and happily the size distortion appears moderate even in that case, unless the autoregressive component is large.

In Tables 3 and 4 we report cases of the alternative hypothesis. Note that these entries are the relative frequencies of rejection at the 5% level, not size-corrected powers which would be impossible to construct (see the remarks in Section 6). These tables also show the average values of the GPH estimates of \( d \) from each experiment, to put the results into context in terms of the findings in a conventional analysis of the data when the true \( d \) is zero.

Table 3 shows the results for the linear AR(1) case, where the model is

\[ y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \]
for three alternative values of $\phi_1$. Beyond this familiar class, our problem is to deal with the profusion of possible alternatives, and the cases we report are necessarily chosen rather arbitrarily, although sharing the characteristic that the values of $d$ obtained in log-periodogram regression are not too close either to zero or to unity.

- "Bilinear" is a model 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 $\phi_2 = \phi_3 = 0.3$.

- SETAR is the Self-Exciting Threshold AR case
  \[ y_t + (\alpha_{11} + \alpha_{12} y_{t-1}) G_t + (\alpha_{21} + \alpha_{22} y_{t-1})(1 - G_t) + \varepsilon_t, \]
  where $\alpha_{11} = 1$, $\alpha_{12} = 0.45$, $\alpha_{21} = 1$, $\alpha_{22} = 0.9$,
  \[ G_t = \frac{1}{1 - e^{-\gamma(y_{t-1} - y^*)}} \]
  with $\gamma = 10$ and $y^* = 1$.

- 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" is a model with Markov–switching intercepts. This model takes the form
  \[ y_t = \alpha(S_t) + \varepsilon_t, \]
  where $S_t = 1$ or $2$ with $P(S_t = 1|S_{t-1} = 2) = P(S_t = 2|S_{t-1} = 1) = 0.05$, and $\alpha(1) = 1$, $\alpha(2) = -1$.

Observe that all of these models under the alternative generate $I(0)$ series, in the sense that their memory decay is exponential.

Figure 1 shows realizations of 1000 observations of each of these test processes, together with a pure fractional process, to illustrate the different ways in which spurious long memory might arise. To an extent, the eye is often the best guide to the characteristic appearance of hyperbolic memory.

In Table 5 we report some experiments in which the test is treated as the optional second stage of a two-stage procedure. A significance test (a 1-tailed $t$-test computed with the bootstrap) is conducted on the full-sample log-periodogram estimator. If this test results in non-rejection at the 5% level, the second-stage test result is discarded and the procedure returns "rejection" (that is, the returned $p$-value is set to zero). The table shows the percentage of overall rejections when the second test is assigned a nominal 5% significance level. In the left-hand columns, the average estimated $d$ value appears as before. This table also includes the case of the alternative $\phi = 0.8$, which was not studied previously.

As noted above, this procedure necessarily yields a consistent test, although the rate of convergence is $M^{1/2}$, so large samples are needed for this fact to be decisive. The results show that the power versus size trade-off as a function of bandwidth is complicated. The rule that a broader bandwidth yields more rejections may be reversed, depending on the nature of the alternative. Comparing the cases $\phi = 0.8$ and $\phi = 0.88$ with the corresponding results in Table
Table 3: AR(1) Alternative

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$T$</th>
<th>$\hat{d}$</th>
<th>$M = [T^{0.55}]$</th>
<th>$\hat{d}$</th>
<th>$M = [T^{0.7}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n = 4$</td>
<td>$n = 8$</td>
<td>$n = 12$</td>
<td>$n = 4$</td>
</tr>
<tr>
<td>0.8</td>
<td>250</td>
<td>0.32</td>
<td>0.116</td>
<td>0.110</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.23</td>
<td>0.114</td>
<td>0.098</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.15</td>
<td>0.093</td>
<td>0.082</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.09</td>
<td>0.072</td>
<td>0.068</td>
<td>0.057</td>
</tr>
<tr>
<td>0.88</td>
<td>250</td>
<td>0.57</td>
<td>0.047</td>
<td>0.105</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.39</td>
<td>0.050</td>
<td>0.077</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.26</td>
<td>0.170</td>
<td>0.165</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.19</td>
<td>0.149</td>
<td>0.136</td>
<td>0.108</td>
</tr>
<tr>
<td>0.95</td>
<td>250</td>
<td>0.78</td>
<td>0.132</td>
<td>0.158</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.71</td>
<td>0.227</td>
<td>0.240</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.60</td>
<td>0.309</td>
<td>0.363</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.45</td>
<td>0.388</td>
<td>0.440</td>
<td>0.388</td>
</tr>
</tbody>
</table>

Table 4: Nonlinear Dynamic Models

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\hat{d}$</th>
<th>$M = [T^{0.55}]$</th>
<th>$\hat{d}$</th>
<th>$M = [T^{0.7}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 4$</td>
<td>$n = 8$</td>
<td>$n = 12$</td>
<td>$n = 4$</td>
</tr>
<tr>
<td>Bilinear</td>
<td>250</td>
<td>0.34</td>
<td>0.099</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.24</td>
<td>0.134</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.17</td>
<td>0.112</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.11</td>
<td>0.075</td>
<td>0.070</td>
</tr>
<tr>
<td>SETAR</td>
<td>250</td>
<td>0.58</td>
<td>0.144</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.45</td>
<td>0.195</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.34</td>
<td>0.214</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.24</td>
<td>0.187</td>
<td>0.187</td>
</tr>
<tr>
<td>ESTAR</td>
<td>250</td>
<td>0.47</td>
<td>0.140</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.38</td>
<td>0.163</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.31</td>
<td>0.167</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.18</td>
<td>0.137</td>
<td>0.134</td>
</tr>
<tr>
<td>Markov</td>
<td>250</td>
<td>0.45</td>
<td>0.143</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.38</td>
<td>0.207</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.30</td>
<td>0.232</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.21</td>
<td>0.162</td>
<td>0.138</td>
</tr>
</tbody>
</table>
The narrow band option looks more attractive here (assuming acceptable size characteristics) while the performances with the broader bandwidth are similar, thanks to the greater number of spurious first-stage rejections. In the case $\phi = 0.7$ with the broad bandwidth, the effect of the spurious rejections is dramatic. The skip-sampling test has little power in this region and rejection rates diminish as $T$ increases, while the rate of spurious rejection is too large to compensate. The overall rejection rate must eventually rise again, but evidently only in very large samples.

On the basis of these experiments, the best rule of thumb for effectively trading size and power appears to be to use the composite test in the context of narrow bandwidth estimation - perhaps, a compromise between the cases explicitly studied here – but stick to the broader bandwidth otherwise. There are however cases of the alternative where this rule clearly fails.

We note finally that there is is not a great deal to choose between the different rates of skip-sampling. However, $n = 8$ appears to offer a reasonable balance of advantages, independent of sample size.

### 8 Conclusion

In this paper we have investigated the potential 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 the skip-samples, poses a problem for the implementation of this test. A bias-corrected estimator permits the construction of a statistic that, although not asymptotically pivotal, allows the implementation of a bootstrap test using the sieve-autoregression method to model short-run dependence.

There is a size-power trade-off involved in the choice of bandwidth for the GPH estimation, and quite large samples prove necessary to yield a decent level of rejection under the alternative cases considered. This is the inevitable consequence of the use of a semiparametric method to construct the statistic, with correspondingly slow convergence to the asymptote. Nonetheless, the test may prove a useful addition to the arsenal of diagnostic procedures for long memory models, such as 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 $\gamma_k$ denote the $k$th autocovariance, defined by the well-known identity

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

For the skip-sampled data, the autocovariances are $\gamma_{nk}$ where

$$\gamma_{nk} = \int_0^{2\pi} \cos(nk\omega)f(\omega)d\omega = \sum_{j=0}^{n-1} \int_{2\pi j/n}^{2\pi(j+1)/n} \cos(nk\omega)f(\omega)d\omega = \sum_{j=0}^{n-1} \int_{2\pi j/n}^{2\pi} \cos(nk\omega)f(\omega + 2\pi j/n)d\omega = \int_0^{2\pi} \cos(k\lambda)f_n(\lambda)d\lambda.$$ 

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

$$f_n(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} f((\lambda + 2\pi j)/n).$$

Proof of Proposition 5.1

Letting

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

note first that

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

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 (4.1) 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 $Z_0 (T-p \times p)$ be the normalized data matrix whose columns are the vectors $z_j = (z_{p+1-j}, \ldots, z_{T-j})'$ for $j = 1, \ldots, p$. Also, let $Z_j$ for $j = 1, \ldots, p$ denote the matrix equal to $Z_0$ except that the $j$th column has been replaced by $z_0 = (z_{p+1}, \ldots, z_T)'$. Then, note that the coefficients $\hat{\phi}_j$ in the autoregression of order $p$ can be written using Cramer’s rule as

$$\hat{\phi}_j = \frac{|Z_j^T Z_j|}{|Z_0^T Z_0|}, \ j = 1, \ldots, p.$$
Let these elements define the \( p + 1 \times 1 \)-vector \( \hat{\psi} \) by also putting \( \hat{\psi}_0 = -1 \).

Now, let \( Q(\theta) \) \( (p + 1 \times p + 1) \) denote the Fourier matrix with elements \( q_{jk} = e^{i\theta(j-k)} \) for \( j, k = 0, \ldots, p \). Setting \( \theta_1 = (\lambda + 2\pi k)/n \) and \( \theta_2 = \lambda/n \), note that

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

where \( b \) is the \( p + 1 \)-vector having elements \( b_0 = -T^{-p} |Z_0' Z_0| \) and \( b_j = T^{-p} |Z_0' Z_j| \) for \( j = 1, \ldots, p \). In this notation we have

\[
\frac{d}{d\theta} \left( \frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} \right) = \frac{b' [Q(\theta_2) + Q'(\theta_2)] - b' Q(\theta_2)b' [Q(\theta_1) + Q'(\theta_1)]}{|b' Q(\theta_1)b|^2} \frac{db}{d\theta}
\]

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

Start with the elements of the \( Z_j \) matrices. Considering row \( t \), let \( m \) denote the generic \( Z_j \) matrix. Using the argument from Tanaka (1999), Section 3.1, the derivatives with respect to \( d \) can be written as

\[
\frac{dz_{t-m}}{dd} = \frac{d}{dd} (1 - L)^d x_{t-m}
\]

\[= \log(1 - L)(1 - L)^d x_{t-m}
\]

\[= - \sum_{k=1}^{\infty} k^{-1} z_{t-m-k}
\]

\[= z_{t-m}^d
\]

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

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

Since \( \{z_t\} \) is a weakly dependent process by hypothesis, the process \( z_t^d \) is covariance stationary. It follows directly that, for every finite \( p \), \( b^d \) converges in probability to a non-stochastic limit, depending on the autocovariances of \( \{z_t\} \). From the fact that \( b \) 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. Also, 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 \( b \) and \( b^d \) have sums converging to zero as \( p \) increases. \[\blacksquare\]
References


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.


