



DOCUMENT DE RECHERCHE

EPEE

CENTRE D'ETUDE DES POLITIQUES ECONOMIQUES DE L'UNIVERSITE D'EVRY

**Confidence Region for Long Memory Based on
Inverting Bootstrap Tests: An Application
to Stock Market Indices**

Christian de PERETTI & Carole SIANI

08 - 02

Confidence Region for long memory based on Inverting Bootstrap Tests: an application to Stock Market Indices

Christian de Peretti *

Centre d'Étude des Politiques Économiques d'Évry (ÉPÉE)

Department of Economics

University of Evry Val d'Essonne (France)

Carole Siani

Laboratoire d'InfoRmatique en Image et Systèmes d'Information (LIRIS)

Department of computational Science

University of Claude Bernard Lyon 1 (France)

February 23, 2008

Abstract

In the context of long memory, the finite-sample distortion of statistic distributions is so large, that bootstrap confidence intervals (percentile and percentile- t) for the long memory parameter do not perform better than the corresponding asymptotic confidence interval. In this paper, we propose confidence intervals based on inverting bootstrap tests for the long memory parameter in the ARFIMA model. We show that classical confidence intervals have very poor performances, even the percentile- t interval, whereas confidence intervals based on inverting bootstrap tests have quite satisfactory performance. For this purpose, we use techniques for measuring **effectiveness** of confidence regions and for the graphical display of simulation evidence concerning the coverage and effectiveness of confidence regions in finite sample. **Monte Carlo results** on the confidence intervals for various situations are presented. These intervals are then applied on stock market indices.

Keywords: Long memory, confidence interval, inverting tests, bootstrap, graphical method.

J. E. L. Classification: C10, C13, C14, C15, C63.

*Correspondence to: Christian de Peretti. Address: Bâtiment Île-de-France - 4, Boulevard François Mitterrand - 91025 EVRY, FRANCE. Fax: +33 (0)1 69 47 70 50. Email: christian.deperetti@eco.univ-evry.fr.

1 Introduction

Many time series in diverse fields of application may exhibit long-memory or long-range dependence. This occurs when the autocorrelation function (a.c.f.) at large lags decays to zero at a slower rate than data following an ARMA(p,q) model and at a quicker rate than that of a unit root processes (see Mandelbrot and Wallis [1969], Mandelbrot [1971]). The class of fractionally integrated (FI) processes is characterised by hyperbolic decay rate, and it is often used to try to capture this strong data dependence. Granger and Joyeux [1980] and Hosking [1981] are two seminal articles on ARFIMA(p, d, q) processes.

Many methods for estimating and testing the long memory parameter d are described in Beran [1994]. However, a point estimator does not make sense if the uncertainty around the estimate is not taken into account. In the literature, the standard deviation of the estimators are provided. This standard deviation can be used to build an asymptotic Gaussian confidence interval. However, in the case of long memory, the distortion of the statistics due to the autocorrelation is so large that “asymptotic” means more than several thousands observations (see de Peretti and Marimoutou [2002]). Consequently, this implicit asymptotic confidence interval cannot necessarily be used in practice.

In this paper, we have first develop bootstrap confidence intervals (percentile and percentile- t) for the long memory parameter in the ARFIMA model. These intervals are based on Robinson [1995]’s estimator. However, in the context of long memory, the finite-sample distortion of statistic distributions is so large, that even bootstrap confidence intervals do not perform better than the corresponding asymptotic confidence interval. In a second step, we then propose a confidence interval based on inverting bootstrap tests following the methodology proposed by Davidson and MacKinnon [2001] and we adapte it to the estimate of the long memory parameter. Double bootstrap versions of the procedures are also discussed.

In order to present the results of the Monte Carlo study of the properties of confidence regions for long range dependence, coverage plots and coverage effectiveness curves are used. These methods are proposed by de Peretti [2004]. These graphical techniques are inspired by the graphical techniques of Davidson and MacKinnon [1998] in the case of tests. These techniques measure the effectiveness of confidence regions and display graphics of simulation evidence concerning the coverage and effectiveness of confidence regions in finite sample. These graphs convey much information, in a more easily assimilated form, than tables, *PP plots*, and *QQ plots* can do. It is often desirable to compare the effectiveness of alternative confidence regions, but this can be difficult to do if all the regions do not have the correct coverage probability. If the values of effectiveness criteria are plotted against (nominal) confidence level, the result will not be very useful, since a method can have a good effectiveness curve due to a coverage distortion and not because of a real effectiveness. Unfortunately, this is what is often implicitly done when region effectiveness is reported in a table. The graphical method we use for displaying effectiveness plots the effectiveness criterion against the coverage probability, *i.e.* the true confidence level; and then, the various methods can be compared. We use these graphical methods for investigating the finite-sample properties of Confidence Regions we propose.

Monte Carlo results on the confidence intervals for various situations are presented. Various confidence regions based on Robinson [1995]’s estimator are compared: the confidence region using the asymptotic distribution of the estimator, the percentile and the percentile- t confidence regions using the bootstrapped distribution, and the confidence region based on inverting bootstrapped Robinson [1995]’s test. We show that classical

confidence intervals have very poor performances, even the percentile- t interval, whereas confidence intervals based on inverting bootstrap tests have quite satisfactory performance.

These intervals are then applied to stock market index volatility.

The plan of the paper is as follows. In [section 2](#), classical and inverting tests confidence intervals are applied on the estimation of the LRD parameter in the ARFIMA models. In [section 3](#), a number of [Monte Carlo results](#) are presented on the various LRD confidence intervals to illustrate the use of coverage plots, and the use of coverage-effectiveness curves. In [section 4](#), the long memory confidence intervals are applied to stock market indices. [Section 5](#) concludes. In [section A](#), we present the graphs proposed by [de Peretti \[2004\]](#) for experiments dealing with confidence region coverage probability, as well as coverage-effectiveness curves.

2 Long range dependence confidence regions

This Section proposes various confidence regions for the long range dependence parameter. Various confidence regions based on [Robinson \[1995\]](#)'s estimator are proposed: there are the confidence region using the asymptotic distribution of the estimator ([subsection 2.2](#)), the percentile ([subsection 2.3](#)) and the percentile- t ([subsection 2.4](#)) confidence regions using the bootstrapped distribution, and the confidence region based on inverting bootstrapped [Robinson \[1995\]](#)'s test ([subsection 2.5](#)). Double bootstrap versions of the procedures will also be discussed. See [Table 1](#) for a listing of these methods.

Table 1: Listing of the confidence region methods

	Classical interval	Based on inverting tests
Asymptotic	Wald type confidence interval	Inverting asymptotic tests
Bootstrap	Percentile	
Bootstrap-t	Percentile- t with asymptotic variance	Inverting bootstrap tests
Double bootstrap-t	Percentile- t with bootstrap variance	Inverting double bootstrap tests

[Robinson \[1995\]](#)'s estimator is chosen rather than others, because, for example, Hurst and Lo methods have less satisfactory properties, and [Higuchi \[1988\]](#)'s estimator seems to be consistent only for $d \in [-0.5, 0.5]$ whereas [Robinson \[1995\]](#)'s estimator is consistent for $d \in [-0.5, 1]$.

Since [Robinson \[1995\]](#)'s statistic is not exactly pivotal, bootstrapping will not perform perfectly. However, [Robinson \[1995\]](#)'s statistic is asymptotically pivotal, which means that its distribution does not depend asymptotically on any nuisance parameters. In that case, bootstrapping should yield confidence regions that are accurate to higher order, in the sample size, than the confidence provided by asymptotic theory; see [Beran \[1988\]](#), [Horowitz \[1994\]](#), and [Davidson and MacKinnon \[1996b,a\]](#).

In our [Monte Carlo simulations](#), we deal with the univariate linear ARFIMA(p,d,q) model with Gaussian errors. Consequently, the error terms of the bootstrap samples in

the methods presented in this section are obtained from a Gaussian distribution rather than by resampling from the residuals, since the error terms are normally distributed. This case is deliberately chosen for the following reason. The finite sample distribution of the statistic can suffer from two types of departure from the asymptotic distribution:

1. The first type of departure comes from the error terms that can be not Gaussian. This departure due to non-Gaussian error terms is often quickly reduced thanks to the limit central theorem (in many bootstrap studies, parametric bootstrap works similarly than nonparametric bootstraps).
2. The second type of departure coming from the fact that the denominator of the studentised statistic is not independent from the numerator. Conversely, the second type of departure can be more persistent with respect to the sample size, especially in the case of long range dependence time series (see [de Peretti and Marimoutou \[2002\]](#) where parametric bootstrap and nonparametric bootstraps work similarly but both have size distortions).

Thus, in our [Monte Carlo experiments](#), we focus on the Gaussian case to show that asymptotic methods have serious problems that are not due to a misspecification of the error terms distribution but to the second type of departure. In this situation, the use of bootstrap combined with inverting tests is greatly advised.

In practice, macroeconomic series are often non-Gaussian and financial series are almost always strongly non Gaussian. The bootstrap procedure can be adapted to this kind of data by the use of nonparametric methods. In our case, when the replications of the series are generated, we just have to draw the bootstrap error terms in the empirical distribution of the residuals of the estimation of the series rather than in a Gaussian distribution (see [section 4](#) for an application of this nonparametric bootstrap, and [Davidson \[1998\]](#) and [subsection B.1](#) for examples and details of other nonparametric bootstrap distributions).

2.1 The model and the estimator

[Robinson \[1995\]](#)'s estimator pertains to the normal linear ARFIMA(p,d,q) model ¹:

$$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t \quad t \in \{1, \dots, T\}, \quad (1)$$

$$\{\varepsilon_t\} \sim i.i.d.N(0, \sigma^2), \quad (2)$$

where

- ϕ and θ are polynomials that have not necessarily all roots outside the unit circle,
- $\sigma^2 < \infty$,
- L is the lag operator,
- d is the differencing parameter and takes a real value.

In some circumstances, a long-memory process may be approximated by a fractionally integrated model; hence estimating for long-memory can be done by an estimation of d .

For applying [Robinson \[1995\]](#)'s procedure, a truncation point m has to be chosen and we choose the rule proposed by [de Peretti and Marimoutou \[2002\]](#):

$$m = \sqrt{T \times 2}.$$

¹ The ARFIMA model is presented in detail by [Granger and Joyeux \[1980\]](#) and [Hosking \[1981\]](#).

2.2 The classical asymptotic confidence region

The classical asymptotic confidence region for the long range dependence parameter d is obtained from [Robinson \[1995\]](#)'s estimator as follows:

$$[d_1, d_2] = \left[\hat{d} + \hat{\sigma}(\hat{d}) q_{\alpha/2}, \hat{d} + \hat{\sigma}(\hat{d}) q_{1-\alpha/2} \right],$$

where:

- \hat{d} is the estimate of d ,
- $\hat{\sigma}(\hat{d})$ is the asymptotic estimate of the standard deviation of \hat{d} , that is $0.5/\sqrt{m}$, (see [Robinson \[1995\]](#)),
- q_α is the α -quantile of the asymptotic distribution of the studentised \hat{d} statistic, that is the normal distribution, (see [Robinson \[1995\]](#)).

Since \hat{d} is asymptotically Gaussian only for $d_0 \in [\frac{1}{2}, \frac{3}{4})$ (see [Velasco \[1999\]](#)), the procedure is valid only for this range of values.

For using our graphical methods, we need to compute the confidence level $1 - \alpha_s$ such that d_0 , the true value for d , is included in the border of the confidence region ∂R , that is $\{d_1, d_2\}$ here. Consequently, we obtain:

$$1 - \alpha_s = F \left(\left(\frac{d_0 - \hat{d}}{\hat{\sigma}(\hat{d})} \right)^2 \right)$$

where F is the CDF of a χ^2 variable with one degree of freedom.

Dealing with the **effectiveness**, denoted $E(1 - \alpha)$ in [subsection A.4](#), we measure it by the confidence interval length expectation. It should be noted that the length of confidence intervals depends on the confidence level $1 - \alpha$. Consequently, the length expectation, corrected or not by the technique proposed in [subsection A.4](#), has to be computed for all the confidence levels. The length of the asymptotic confidence interval is:

$$e(1 - \alpha) = 2 q_{1-\alpha/2} \hat{\sigma}(\hat{d}).$$

For each replicated series, the length of the confidence interval is computed for each level $1 - \alpha$. For the whole **Monte Carlo experiment**, *i.e.* for all the replicated series, the expectations for each level are estimated by the sample mean of the lengths $e(1 - \alpha)$.

In the computational program, a set of values for the confidence level is defined:

$$\{0.001, 0.002, \dots, 0.009, 0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 0.9, 0.91, 0.92, \dots, 0.99, 0.991, 0.992, \dots, 0.999\}.$$

For each of these values, the vector $e(1 - \alpha)$ is computed (with matrix algebra, it is very straightforward).

2.3 The percentile confidence interval

For a general presentation of *percentile* and *percentile-t* methods, see [Davidson and MacKinnon \[1993\]](#), [Efron and Tibshirani \[1993\]](#), [Hall \[1992\]](#), [Hjorth \[1994\]](#), and [Shao and Tu \[1995\]](#).

Let $\{\hat{d}^b\}_{b=1}^B$ denotes the set of statistics simulated by bootstrap, where B is the number of bootstrap replications. The *percentile* confidence interval is given by

$$(b_1, b_2) = \left(\hat{d}^{[B\alpha/2]}, \hat{d}^{[B(1-\alpha)/2]} \right),$$

where $[\cdot]$ is the rank statistic of the integer part.

The level such that $d_0 \in \partial R$ is equal to:

$$1 - \alpha_s = 2 \min\{pv, 1 - pv\},$$

where

$$pv = \frac{1}{B} \sum_{b=1}^B I(\hat{d}^b \leq d_0).$$

Since \hat{d} is consistent only for $d_0 \in [-\frac{1}{2}, 1)$ (see [Velasco \[1999\]](#)), the procedure is valid only for this range of values.

More details about the bootstrap procedure have to be provided. We go through the following steps:

- First, the data must be estimated using a model as close as possible to the real data. This model will be used as a *data generating process* (DGP) for generating simulated samples, and then, the simulated replications \hat{d}^b of the statistic \hat{d} . We propose here to estimate an ARFIMA(1,d,1) model². For the parameter d , the value \hat{d} is chosen (estimated in a semi-parametric way here by Robinson's estimator). The remaining parameters are estimated by maximum likelihood conditionally to \hat{d} . We prefer this procedure rather than estimating all the parameters by an one step estimation by maximum likelihood, because we need the estimation of d in the ARFIMA(1,d,1) model in the same way as the estimation of d for the confidence interval. Otherwise, the replications \hat{d}^b can follow a probability law too different from the probability law that generates \hat{d} . Since the aim of the bootstrap procedure is to estimate the distribution of \hat{d} , the method can be biased.
- The second step of the bootstrap procedure is to generate B replications $\{x^b\}_{b=1}^B$ of the time series x . B has to be sufficiently large to avoid random effect of the bootstrap experiment: at least several hundreds, several thousands is desirable. It is also desirable that $(1 - \alpha)(B + 1)$ be an integer (see [Davidson and MacKinnon \[1993\]](#)). Each x^b is generated from a Gaussian ARFIMA(1,d,1) with the previously estimated parameters (see [section 4](#) for the non-Gaussian bootstrap). For each x^b , d is estimated by Robinson's procedure, providing \hat{d}^b , a bootstrap replication of \hat{d} . And then, the critical coverage is computed using the set of $\{\hat{d}^b\}_b$.

For easily calculating the **effectiveness**, here the length of the interval, we propose the following method. The length of the interval is given by

$$e(1 - \alpha) = \hat{d}^{[B(1-\alpha)/2]} - \hat{d}^{[B\alpha/2]}.$$

Rather than calculating $e(1 - \alpha)$ for a predetermined set of $1 - \alpha$, as in [subsection 2.2](#), we prefer calculating each possible different length since the number of possibilities is finite,

² A more sophisticated estimate can be done by estimating an ARFIMA(p,d,q) model and choosing p and q by a selecting criterion as AIC or BIC, but the large flexibility of the ARFIMA(1,d,1) model is sufficient for our illustration here.

because the bootstrap distribution is discrete. The corresponding levels are computed only later. Using matrix algebra, it is very straightforward: the vector of lengths is:

$$e = \hat{d}^*(\text{trunc}(B/2 + 1) : B) - \hat{d}^*(\text{ceil}(B/2) : 1),$$

where \hat{d}^* is the ordered vector of bootstrap replications. The function *trunc* gives the greatest integer smaller than the argument and the function *ceil* give the smallest integer greater than the argument. The associated confidence levels are the sequence:

$$\left\{ \frac{2b}{B+1} \right\}_b \text{ where } b \in \left\{ 0, \dots, \text{trunc} \left(\frac{B+1}{2} \right) - 1 \right\}.$$

2.4 percentile- t confidence interval

The *percentile- t* procedure is similar to the percentile procedure, but rather than using directly the estimator of d , the studentised form of this statistic is used.

$$\tau = \frac{\hat{d} - d}{\hat{\sigma}(\hat{d})},$$

$\hat{\sigma}(\hat{d})$ will be discussed latter. This leads a to statistic that is asymptotically pivotal, and this permits to obtain a higher rate of convergence for the method. τ is bootstrapped to obtain B replications, denoted τ^b , as follows:

$$\tau^b = \frac{\hat{d}^b - \hat{d}}{\hat{\sigma}(\hat{d}^b)}.$$

The percentile- t confidence interval is

$$\left[\hat{d} - \tau^{[1-\alpha/2]} \hat{\sigma}(\hat{d}), \hat{d} - \tau^{[\alpha/2]} \hat{\sigma}(\hat{d}) \right]$$

The coverage corresponding to $d_0 \in \partial R$ is equal to:

$$1 - \alpha_s = 1 - 2 \min\{pv, 1 - pv\},$$

where

$$pv = \frac{1}{B} \sum_{b=1}^B \mathbf{I}(\hat{\tau} \leq \tau_0),$$

and

$$\tau_0 = \frac{\hat{d} - \mathbf{d}_0}{\hat{\sigma}(\hat{d})}.$$

The length for the $(1 - \alpha)$ -confidence interval is:

$$(\tau^{[1-\alpha/2]} - \tau^{[\alpha/2]}) \hat{\sigma}(\hat{d}).$$

As for percentile procedure, the percentile- t procedure is valid only for $d_0 \in [-\frac{1}{2}, 1)$.

The DGP for replicating τ is determined in the same way as for the percentile as well as the lengths of the intervals.

$\hat{\sigma}(\hat{d})$, the estimated variance of \hat{d} , is computed in two different ways: the first one is the asymptotic estimate (see [subsection 2.2](#)), the second one is the bootstrap estimate. For

obtaining the bootstrap estimator of $\sigma(\hat{d})$, the time series are estimated and replicated B_2 times in the same way as previously. B_2 can be different from B . Since it is less important, it is generally taken smaller than B . For each replicated series, \hat{d} is computed leading to a set of $\{\hat{d}^b\}_b$ from which the standard error is computed. It should be noted that when the percentile- t method is used, and thus replicated series are generated for computing the test P value, the bootstrap estimator of $\sigma(\hat{d})$ must be applied on each bootstrap replication of the series for obtaining the studentised statistics. This leads to replications of replicated series. This method is often called *double bootstrap*.

A last remark about the percentile- t method: even if it has a better asymptotic convergence rate than the percentile method, in finite sample, the studentisation can produce a statistic that is farther from pivotal than the original statistic. This instability can be catastrophic, see among others [Li and Maddala \[1996\]](#), [Berkowitz and Kilian \[2000\]](#), [Davidson \[2000\]](#), and [Siani and Moatti \[2003\]](#). Thus, we have to check by [Monte Carlo experiments](#) that this method remains stable in finite sample.

2.5 Confidence region based on inverting tests

For a general presentation of confidence intervals based on inverting tests, see [Davidson and MacKinnon \[1993\]](#) Chapter 5, and for confidence intervals based on inverting bootstrap tests, see [Davidson and MacKinnon \[2001\]](#).

Let $T_{1-\alpha}(d')$ denote the result of a test for $H_0 : d = d'$ against $H_1 : d \neq d'$ at significance level α :

$$T_{1-\alpha}(d') = \begin{cases} 1 & \text{if } d = d' \text{ is retained} \\ 0 & \text{otherwise} \end{cases} .$$

The confidence region built by inverting tests for a confidence level of $1 - \alpha$ is defined as following: d' is in the region if and only if the test retains $d = d'$ for a significance level of $1 - \alpha$, *i.e.*

$$d' \in \hat{R}(1 - \alpha) \iff T_{1-\alpha}(d') = 1.$$

Let $p(d')$ denote the P value of the test for $d = d'$. We have

$$T_{1-\alpha}(d') = 1 \iff p(d') \geq \alpha.$$

In our case, the region is an interval. Consequently, it can be defined by both its limits, say d_{inf} and d_{sup} (corresponding respectively to the lower and the upper limit). These limits correspond to both the values of d' such that $p(d') = \alpha$ (see [Davidson and MacKinnon \[1993\]](#) Chapter 5). For computing the P value, see [subsubsection 2.5.1](#) and [subsubsection 2.5.2](#).

For using our graphical methods, the critical confidence level $1 - \alpha_s$ has to be computed. This level $1 - \alpha_s$ is such that $d_0 = d_{\text{inf}}^{(s)}$ or $d_0 = d_{\text{sup}}^{(s)}$, where (s) indicates that the values come from a simulated series. But it is known that $p(d_{\text{inf}}^{(s)}) = 1 - \alpha = p(d_{\text{sup}}^{(s)})$, thus

$$1 - \alpha_s = p(d_0),$$

where $p(d_0)$ is the P value of the underlying test at $d = d_0$. Fortunately, for confidence regions based on inverting tests, the critical confidence level is very easy to compute.

With regard to the [effectiveness](#) of this method, a problem arises since the test is not defined for all values for d . For computing the confidence interval limits, the values d' for d such that $p(d') = \alpha$ have to be found. Classically, $p(d)$ is a function that increases from

0 to 1 and then decreases from 1 to 0 when d goes from $-\infty$ to $+\infty$. In our situation, the method is only defined for $d_0 \in [-\frac{1}{2}, 1)$, thus, the function reaches its minimum values for $d_0 = -\frac{1}{2}$ and $d_0 = 1$ and the defined confidence levels go from 0 to a certain maximum value, say $1 - \alpha_{\max}$, that is lower than 1.

Since $p(d)$ is random, depending on the realisation of the observed time series, $1 - \alpha_{\max}$ will be random, going from 0 to 1. The consequence is that for all predefined confidence levels $1 - \alpha$, there is a positive probability that the confidence interval is not defined, and thus the expectation of the interval length is also undefined. It should be noted that this does not mean that the method does not work. In fact, it says rigorously that there is no sufficient information in the time series for allowing an estimation whereas the asymptotic confidence interval gives meaningless limits and the percentile(-t) methods give always numerical limits based on computational proportions that can have no sense. For illustrating all the same the **effectiveness** of confidence interval based on inverting tests, we detail results for two particular series in [section 4](#).

2.5.1 Inverting asymptotic tests

More precisely, if the asymptotic test based on an estimator of d (Robinson's one, or another estimator) is used, the P value can be built as following ³

$$p(d') = 1 - F_{d'}(\tau(d')),$$

where

$$\tau(d') = \left(\frac{\hat{d} - d'}{\hat{\sigma}(\hat{d})} \right)^2,$$

$F_{d'}$ is the asymptotic cdf of $\tau(d')$, and $\sigma(\hat{d})$ is the standard deviation of \hat{d} . Any explanations must be given for why F is indexed by d' since we can think that $\tau(d')$ follows simply a χ_1^2 . If $d' \in (-\frac{1}{2}, \frac{3}{4})$, the Robinson's estimator is Gaussian under the null that $d = d'$, and $F_{d'}$ is simply the cdf of a χ_1^2 random variable. In this case, the confidence interval built by inverting asymptotic tests is equal to the classical asymptotic confidence interval. However, when $1 - \alpha$ goes to 1, then d_{\inf} goes to $-\infty$ and d_{\sup} goes to $+\infty$, and we necessarily fall in the non Gaussian case, and F_d is not the cdf of a χ^2 .

It is not very useful to develop this method only for obtaining an extension of the classical asymptotic confidence interval (see [subsection 2.2](#)) for the case where d_0 is not included in $[-\frac{1}{2}, \frac{3}{4})$. Moreover, confidence regions based on inverting bootstrap tests (see [subsection 2.5.2](#)) will provide better approximation than with asymptotic tests. Consequently, this method is not presented.

2.5.2 Inverting bootstrap tests

We consider here bootstrap tests based on an estimator of d (Robinson's one, or another estimator). More precisely, for the **Monte Carlo experiments**, the parametric (single and double) bilateral bootstrap test based on Robinson's estimator is used: see its description in [subsection B.1](#) ⁴. Obviously, as for percentile methods, nonparametric versions of the

³The P value will be more complicated for bootstrap because it takes into account the asymmetry of the statistic distribution: see [subsection B.1.3](#).

⁴ Our procedure is inspired from the test of [de Peretti \[2003\]](#), but this last one is reduced to the case where the null hypothesis is "no long memory" ($d_0 = 0$), whereas our procedure extends to the case $d_0 \in [-\frac{1}{2}, 1)$. So, for more discussions about double and single bilateral bootstrap tests for long memory, see [de Peretti \[2003\]](#).

test can be used. In our experiments, nonparametric versions of the test are not used. We restrict our attention to Gaussian procedures for permitting to see the gain of the use of inverting tests confidence regions compared to not inverting tests, without noisy error due to nonparametric estimation. Nevertheless, the basic nonparametric test (only resampling) is applied in [section 4](#) and the details of this method and of others are in [subsection B.1](#). Both the single and the double bootstraps are discussed, since, even if the double bootstrap is asymptotically better than the single bootstrap, it has the disadvantage of having a computing time B_2 times larger than the single version, where B_2 is the number of bootstrap replications used for the computation of the standard deviation of the test statistic. Finally, the unilateral version of the test is not presented, since it has absolutely no advantage compared to the bilateral version.

2.5.3 Note on the LRD estimation of differentiated series

When the values for d such that $p(d) = 1 - \alpha$ are sought for building the confidence interval based on inverting tests, it can be frustrating to stop at -0.5 and 1 . A solution that appears naturally is to differentiate the series such that they are stationary for example. It can be done by successive augmented Dickey-Fuller (ADF) tests, differentiating the series at each step, until the series looks stationary. If the series is differentiated n times, the LRD estimate is then defined as equal to the LRD estimate of the differentiated series plus n : $\hat{d}(\Delta^n x) + n$. The problem is that the LRD estimate on the original series (when it can be computed) is in general unequal to $\hat{d}(\Delta^n x) + n$ because the bias coming from the estimation on the original series is in general different from the bias coming from the estimation on the differentiated series (even if both the estimators are consistent). This difference causes jumps into the P value function, leading to problematic resolution of the equation $p(d) = 1 - \alpha$. Consequently, we advise to not use differentiation for inverting tests without further improvement of this methodology in the future.

3 Monte Carlo experiments

[Coverage plots](#) and [coverage effectiveness](#) curves (see [section A](#)) are used to present the results of a study of the properties of the long range dependence confidence regions presented in the previous Section. These confidence regions are based on [Robinson \[1995\]](#)'s estimator: there are the confidence region using the asymptotic distribution of the estimator ([subsection 2.2](#)), the percentile ([subsection 2.3](#)) and the percentile- t ([subsection 2.4](#)) confidence regions using the bootstrapped distribution, and the confidence region based on inverting bootstrapped [Robinson \[1995\]](#)'s test ([subsection 2.5](#)). Double bootstrap versions of the procedures are not presented here because of time consumption.

All the following experiments deal with Gaussian ARFIMA(p,d,q) processes. Since the constant term is only a location parameter and does not influence the regions performances, it is set to zero. Similarly, σ , the standard deviation of the model, is only a scale parameter and does not affect at all the performances, is set to one. The test statistic depends then on the parameters (ϕ, d, θ) and T in [Equation 1](#). In a first step, we focus on the coefficient d and the sample size T , setting ϕ and θ to zero. $T = 2^n$ is used, where n is an integer. Each experiment is run with $S = 10,000$ replications of the time series, using the same random numbers for avoiding additional experimental errors (variance). For each replication in each experiment, the same random numbers for bootstrapping

are used for avoiding random effect in the result of the bootstrap method between the replications.

3.1 Case of ARFIMA(0,d,0) processes

We pick combinations of d and T in [Table 2](#) to investigate. B is the number of bootstrap

Table 2: choice of d and T

Case	d	T	B	B_2
1.1	-0.4	256	488	0
1.2	0	512	244	0
1.3	0.4	256	488	0
1.4	0.8	128	976	0

replications ⁵. It should be noted that the values for B chosen for these experiments are sufficient for illustrating the methods since in our case the distortions found in the results are not due to a gap of bootstrap replications (even if increasing B can improve a little the results) but to large distortions of the distribution of the test statistic depending on the parameter values. Since the bootstrap methods have to estimate these parameters, an error in the estimations lead to an error in the coverage. The plot of the P value function with respect to various parameters makes it clear (see [de Peretti \[2003\]](#)). However, for a real data study, there is not the same constraint than for Monte Carlo experiments, and we advise to use much more bootstrap replications than used in our experiments, since the user may want more precision for the coverage. B_2 is the number of bootstrap replications used in the estimation of the variance in the *double bootstrap*. It is taken equal to 0 for the moment (instead, the asymptotic estimation of the variance is used).

3.1.1 Coverage plots

The standard deviation of the [coverage plots](#) is equal to

$$\sqrt{\frac{c(1-\alpha)(1-c(1-\alpha))}{S}} \quad (3)$$

where $c(1-\alpha)$ is the coverage. Thus, when $1-\alpha$ goes from 0 to 1, and for $S = 10,000$, the standard deviation goes from 0 to 0.005.

Figure 1 presents the [coverage plots](#) of the asymptotic confidence interval, the [percentile](#) interval, the (single) [percentile-t](#) interval, and the interval based on [inverting](#) (single) bootstrap tests for the four cases of the parameters described in [Table 2](#). [Table 3](#) presents the same results than the third graph in [Figure 1](#) (*i.e.* the case 1.3 of the parameter in [Table 2](#)) but using a tabular presentation. Is this table pleasant to read? [Figure 1](#) shows clearly that the confidence interval based on [inverting](#) bootstrap tests is by far the best method on the basis of coverage accuracy criterion in all the situations. [Percentile](#) and [percentile-t](#) methods have large coverage distortions, that are as large as

⁵ B is chosen to meet the capacity of Gauss software to store a $T \times B$ matrix which is used in matrix algebra computation for these simulation experiments.

Figure 1: Coverage plots in the case of an ARFIMA(0,d,0) process

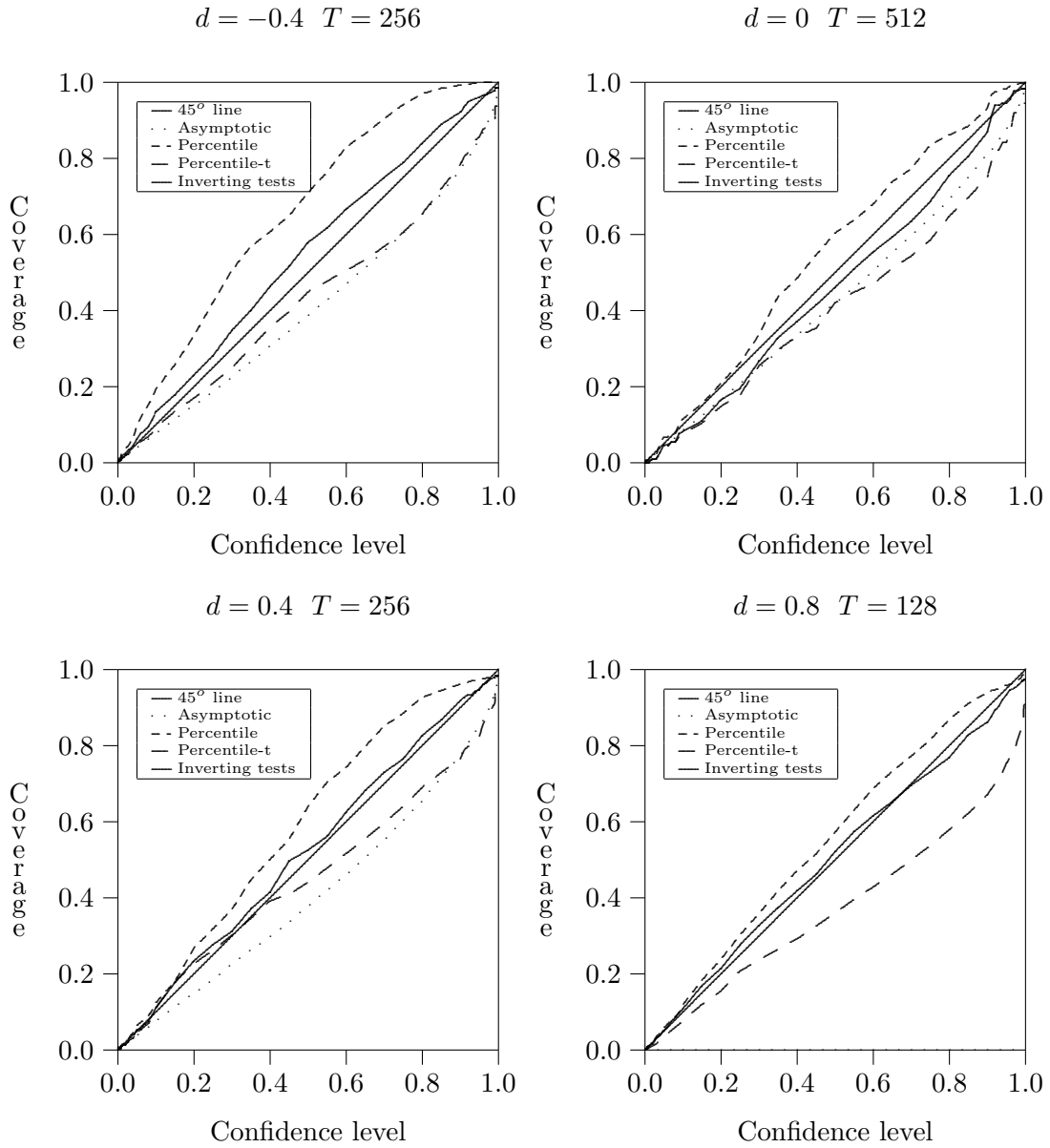


Table 3: Coverage in the case of an ARFIMA(0,d,0) process
 $d = 0.4$ $T = 256$

Method	Confidence levels			
	75%	90%	95%	99%
Asymptotic	0.60	0.77	0.85	0.94
Percentile	0.89	0.96	0.97	0.99
Percentile-t	0.64	0.77	0.82	0.91
Inverting tests	0.77	0.92	0.95	0.98

for the asymptotic confidence interval, except for the case $d = 0$ (*i.e.* i.i.d.) where the asymptotic method dominates them a little (there is no curve for the asymptotic method in the case of $d = 0.8$ since it is not applicable in this situation). It is not surprising that the **percentile** method does not perform for estimating the long memory parameter, since the **percentile** does not correct the bias in the estimates whereas it is known that the long memory estimators are generally very biased. What is more surprising is the unsatisfactory result for the **percentile-t** method that normally corrects the bias. The unsatisfactory result is probably due to the studentisation of the statistic that does not bring it closer to pivotal (closer to pivotal the statistic is, better the bootstrap methods will perform) since for the single bootstrap, only the asymptotic standard error of the statistic is used for studentising it, and for **Robinson [1995]**'s method, this asymptotic standard error is only a constant depending on the sample size (see **Robinson [1995]**). Double bootstrap should lead to more satisfactory results for the **percentile-t** methods, but also for confidence interval based on **inverting** double bootstrap tests, that will still dominate all the methods.

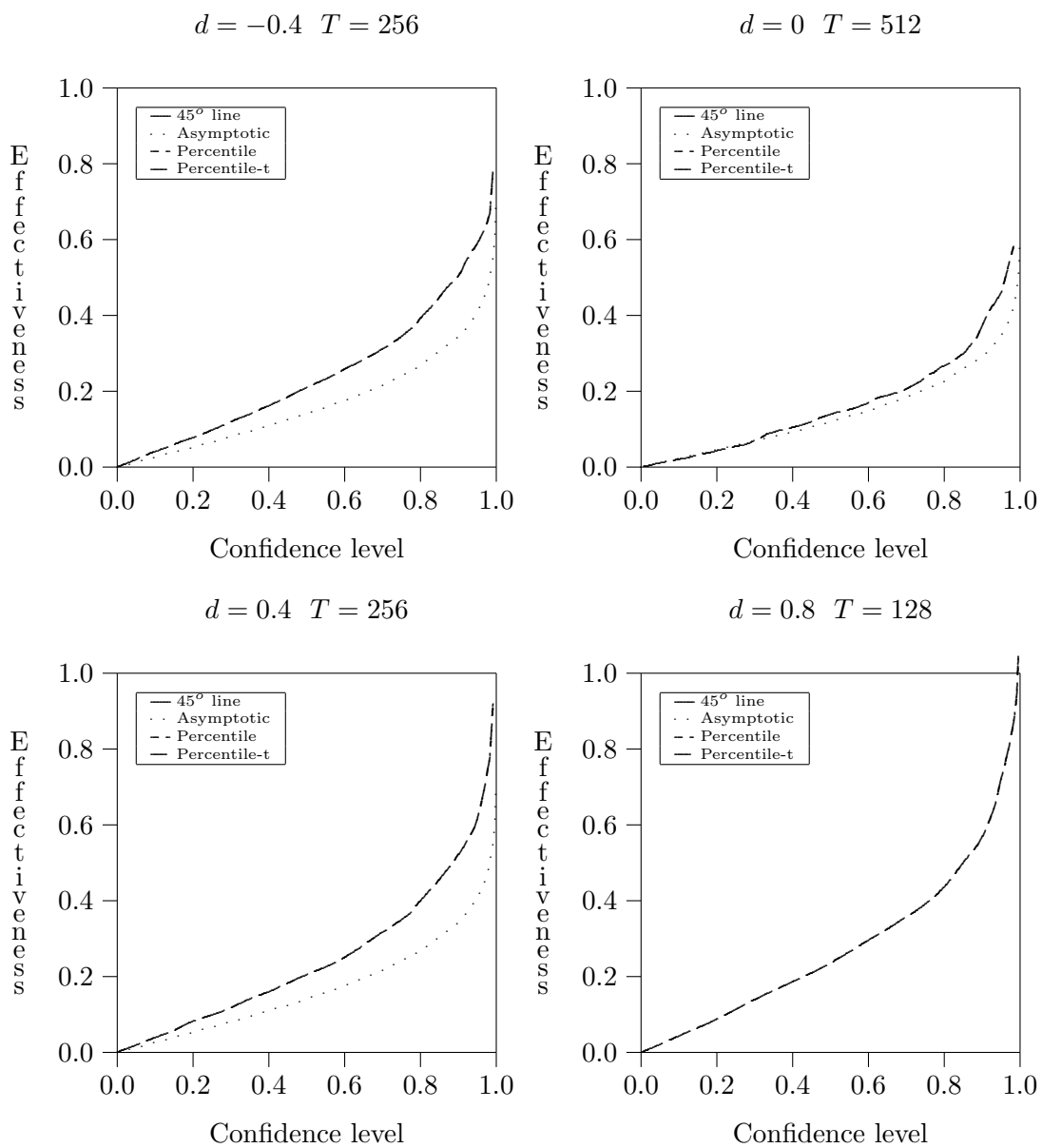
It should be noted that the specification of the data model is not necessarily well chosen by the bootstrap DGP. For example, in this paper, the bootstrap DGP is an ARFIMA(1,d,1) process whereas the true DGPs of the first set of simulations are ARFIMA(0,d,0) processes, and the true DGPs of the second set of simulations are ARFIMA(1,d,0) processes. Nevertheless, our Monte Carlo experiments are more realistic if the true DGP be used, since in practice the real DGP (the one of nature) is not known, and the orders of the ARFIMA(p,d,q) model have to be chosen *a priori* (the ARFIMA(1,d,1) model is often chosen) or estimated by any methods (AIC or BIC criteria or by inference tests).

Since the differences between the confidence regions are very clear with **coverage plots**, the use of **coverage discrepancy plots** is not necessary in this situation.

3.1.2 Coverage-effectiveness curves

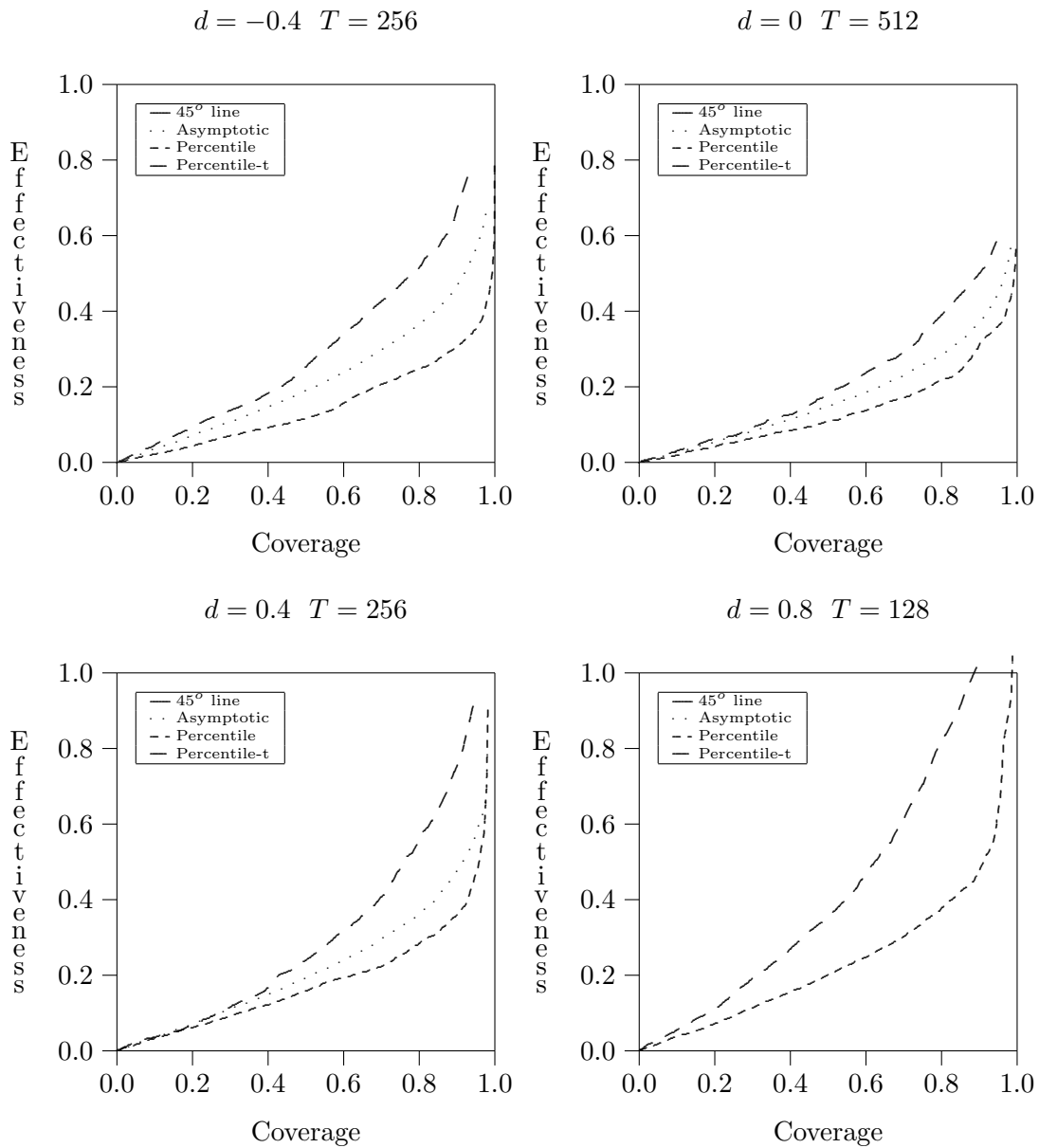
Figure 2 presents the confidence level-effectiveness curves of the asymptotic confidence interval, the **percentile** interval, the (single) **percentile-t** interval for the four cases of the parameters of the **Table 2**. The results for the interval based on **inverting** (single) bootstrap tests are not presented here because of the difficulty of estimating the length of the interval (see **subsection 2.5**). Several results will be provided later about this point. On the basis of confidence level-effectiveness curves, the **percentile** and the **percentile-t** seem to have the same effectiveness and to be dominated by the asymptotic method that has a lower average length.

Figure 2: Confidence level-effectiveness curves in the case of an ARFIMA(0,d,0) process



The effectiveness criterion is the average length here.

Figure 3: Coverage-effectiveness curves in the case of an ARFIMA(0,d,0) process



The effectiveness criterion is the average length here.

However, before concluding hastily, let us have a look to [Figure 3](#) that presents the [coverage-effectiveness](#) curves for the intervals obtained with the same methods. Following the results on [Figure 3](#), the [percentile](#) method has the most satisfactory “true” effectiveness. For these three methods, this information is not very interesting since these methods have too large coverage distortions to be used in practice. What the graphics in [Figure 3](#) can tell us is only that if the coverage distortions could be corrected, the [percentile](#) method will have the most satisfactory effectiveness compared to both other methods. Nevertheless, this kind of graphic is very useful for choosing among methods that have reasonable coverage distortions: it permits to make *arbitrage* between the coverage distortion and the true effectiveness for each methods and then to chose the most appropriate.

In addition, the standard errors of the lengths of the confidence intervals are plotted against their Coverage. This leads to what can be called *Coverage-‘Standard error of the length’ curves* that can be viewed as [coverage-effectiveness](#) curves for a secondary effectiveness criterion (see [Figure 4](#)). The same conclusion than previously holds. However, the graphics on [Figure 4](#) allow to see that the larger d is, the larger the standard error of the length. This is unsatisfactory since for large values of d , even if the average length of an interval is satisfactory, the probability to have a large length is great, and thus there is a big uncertainty on the value for the true value for d . This disadvantage cannot be viewed only on the basis of average length, but also on the basis of length standard error.

For the length of confidence interval based on [inverting](#) tests, see the analysis in [subsection 4.4](#).

3.2 Case of ARFIMA(1,d,0) processes

We choose combinations of d and T in [Table 4](#) for the investigation.

Table 4: choice of ϕ , d , and T

Case	ϕ	d	T	B
2.1	0.4	0	1024	122
2.2	0.8	0.4	1024	122

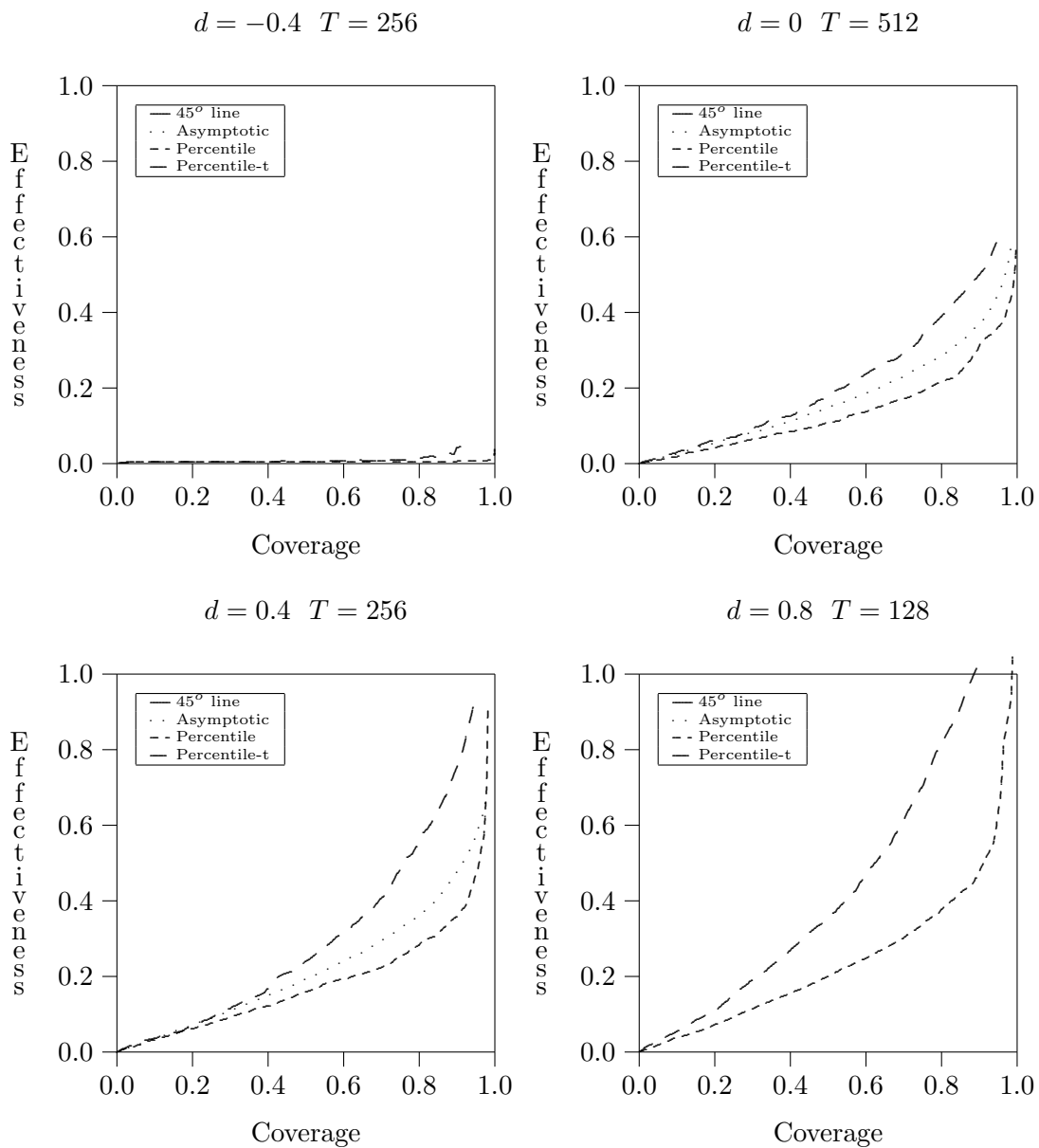
[Figure 5](#) presents the [coverage plots](#) of the asymptotic confidence interval, the [percentile](#) interval, the (single) [percentile-t](#) interval, and the interval based on [inverting](#) (single) bootstrap tests for the two cases of the parameters described in [Table 4](#)).

[Figure 6](#) presents the [coverage-effectiveness](#) curves for the three first methods. For $\phi = 0.4$ $d = 0$ $T = 1024$, the same conclusion than for the previous subsection holds. Conversely, the case $\phi = 0.8$ $d = 0.4$ $T = 1024$ shows catastrophic results which remind us that when the short memory parameters are too strong (close to unit root), the short memory property is confused with long memory property.

4 Application to financial time series

We propose to investigate the long memory properties of the following stock indices: ASX (all ordinaries index, Australia), Dow Jones Industrial 30, FTSE LATIBEX (all shares

Figure 4: Coverage-‘Standard error of the length’ curves in the case of an ARFIMA(0,d,0) process



The effectiveness criterion is the standard error of the length here.

Figure 5: Coverage plots in the case of an ARFIMA(1,d,0) process

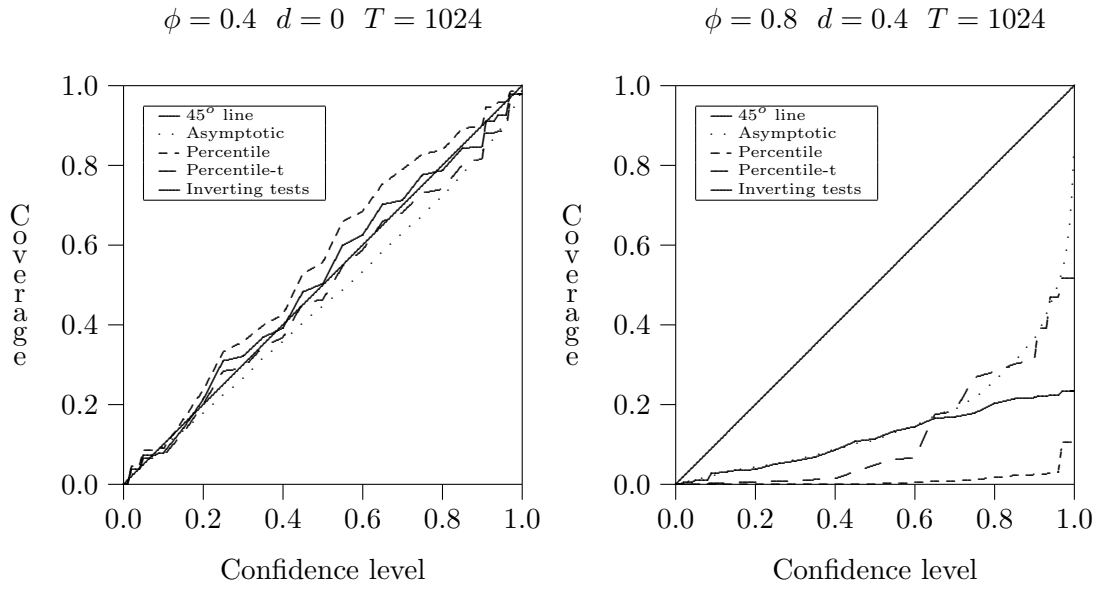
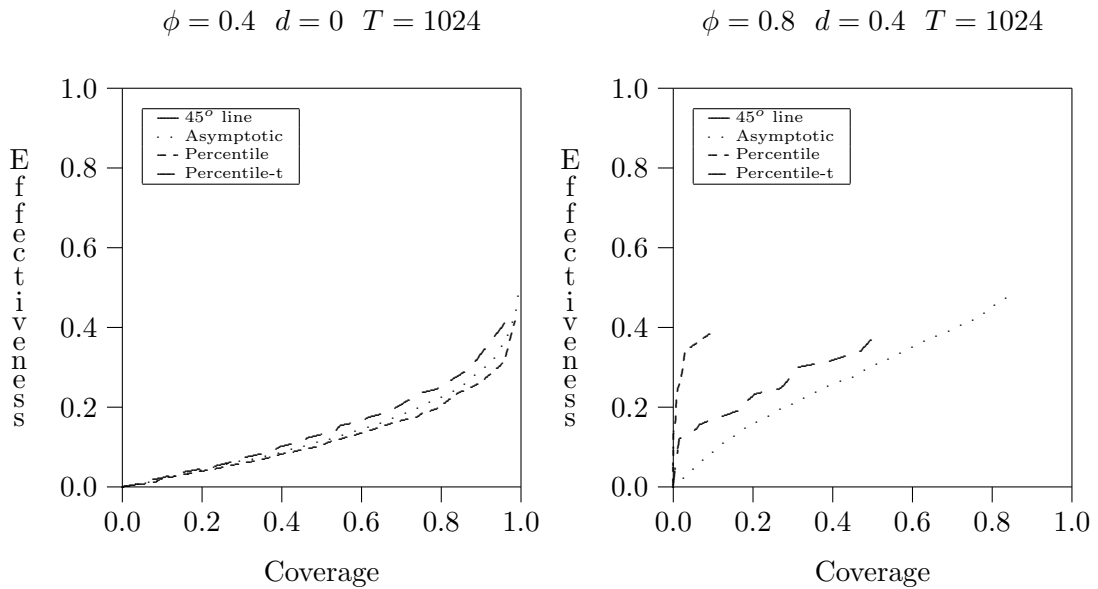


Figure 6: Coverage-effectiveness curves in the case of an ARFIMA(0,d,0) process



The effectiveness criterion is the average length here.

index), Hang Seng (Hong Kong), NASDAQ, NASDAQ Composite, NIKKEI 225 (stock average, Tokyo), and Singapore SE (all share) from 07/28/2006 to 07/27/2007. The number of observations is around 250 depending on the series.

4.1 The stock index levels

Figure 7 plots the stock indices index in level. It can be observed that the stock index series (denoted $(x_{i,t})_{t=1}^T$ for index $i = 1, \dots, N$) are clearly nonstationary. More rigorously, for studying the stationarity of the stock index levels, parametric and nonparametric bootstrapped versions of the augmented Dickey-Fuller (ADF) tests are used, with unilateral and bilateral bootstrap P values (see Appendix C). The number of bootstrap replications being 999. For selecting the number of augmentations, the residuals from the ADF regressions are tested for serial correlation using Ljung-Box and Box-Pierce tests (from 1 to 8 lags) until they look like white noise. All tests have been conducted by using our own programs via Gauss software. All the P values are close to one (they are not presented here) confirming that the series are nonstationary. Consequently, the returns of the series

$$r_{i,t} = \ln \left(\frac{x_{i,t}}{x_{i,t-1}} \right) \quad (4)$$

will be used in the following.

4.2 The stock index returns

For studying the stationarity of the stock index returns, the same bootstrapped ADF tests are used. All the P values are close to zero (they are not presented here) concluding that the series are clearly stationary.

Table 5 presents some descriptive statistics. For studying the normality hypothesis, three tests for normality are used. The first procedure tests for the nullity of the skewness. The second one tests whether the kurtosis is equal to 3. Finally, the last tests for both moments simultaneously: Jarque-Bera test (see [Bera and Jarque \[1980, 1981\]](#)). Since the critical values of Jarque-Bera test are not exact in finite sample due to the correlation between the estimated skewness and the estimated remind, and since the finite sample distribution for all the tests are not equal to their asymptotic distributions, we propose to use bootstrap versions of these tests (see Appendix D). Of course, it is a parametric bootstrap since the null hypothesis is the normality of the series, but in addition, we allow the Gaussian process under the null to be autocorrelated because if it is not taken into account, the autocorrelation amplifies the distortion of the test statistic distribution compared to the asymptotic distribution. Unilateral and bilateral bootstrap P values are also used (except for the asymptotic tests, since the asymptotic distributions are symmetric: both the P values are the same, and for Bera-Jarque test because it is unilateral by construction). The number of bootstrap replications is 999. There is a lag of the dependent variable in the regressor set. The number of lags is chosen by Ljung-Box and Box-Pierce tests using 1 to 8 lags until the residuals seem independent. Again, all the P values are close to zero, showing that the series are strongly non-Gaussian. Consequently, nonparametric methods, such as Robinson long memory estimator and nonparametric bootstrap confidence intervals, have to be used in the following.

The long memory feature is not search directly in the stock index returns since it is known that the financial returns do not exhibit long memory property, in accordance

Figure 7: Stock indices

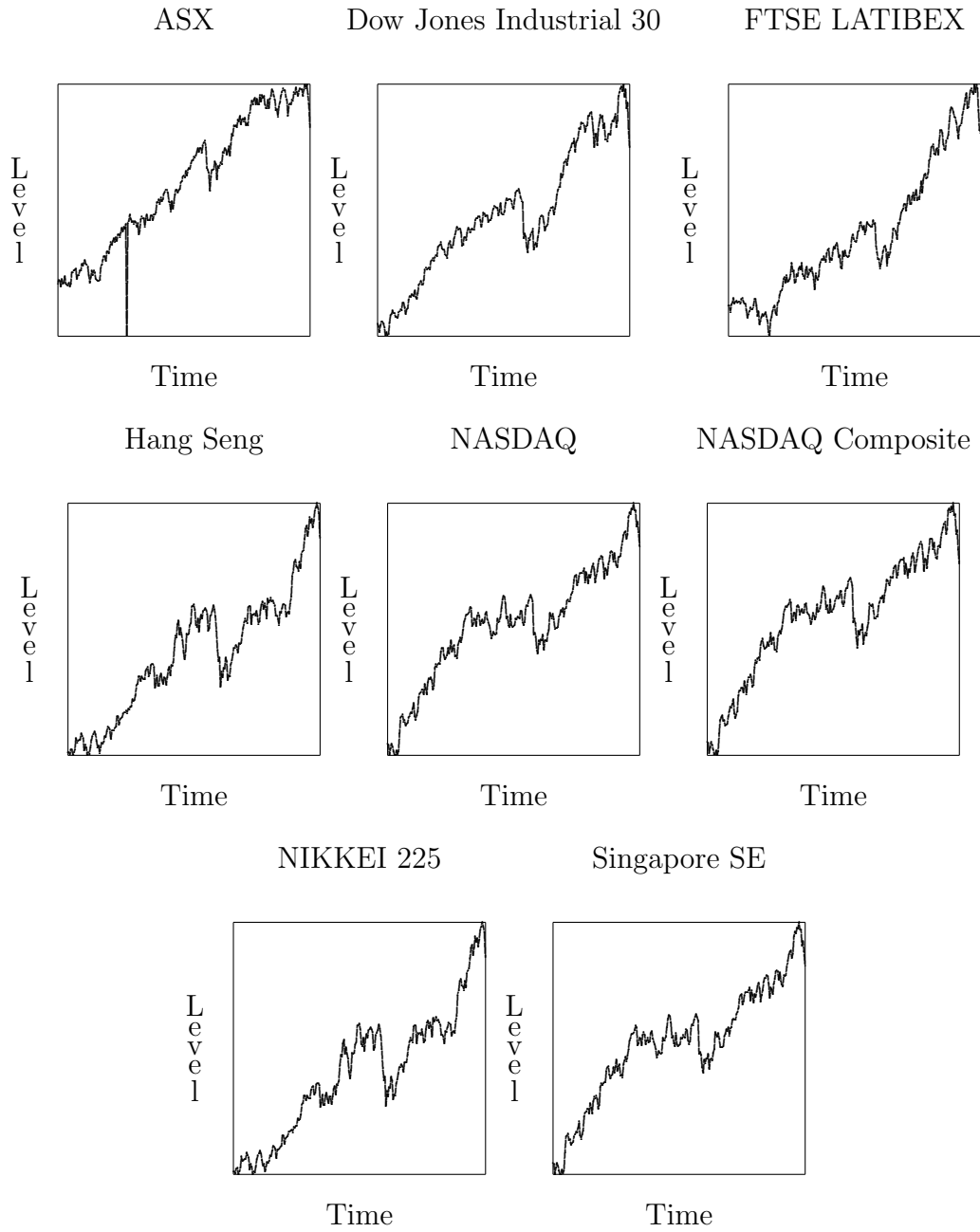


Table 5: Moments of the stock indices.

Stock index	Series	Mean	Standard deviation	Skewness	Kurtosis
ASX (Australia)	Returns	0.0009	0.0174	0.0043	80.2877
	Absolute returns	0.0074	0.0158	9.4691	100.2484
	Squared returns	0.0003	0.0027	11.0047	122.7773
Dow Jones Industrial 30	Returns	0.0007	0.0061	-1.1119	7.6627
	Absolute returns	0.0043	0.0044	2.3167	12.0106
	Squared returns	0.0000	0.0001	7.3523	75.4239
FTSE LATIBEX	Returns	0.0015	0.0152	-0.3068	3.5536
	Absolute returns	0.0120	0.0096	1.2406	4.5786
	Squared returns	0.0002	0.0004	3.0521	14.8703
Hang Seng (Hong Kong)	Returns	0.0012	0.0100	-0.5944	4.3368
	Absolute returns	0.0076	0.0066	1.4193	5.7393
	Squared returns	0.0001	0.0002	4.2412	29.9275
NASDAQ	Returns	0.0010	0.0088	-0.4611	4.9405
	Absolute returns	0.0067	0.0058	1.7211	7.9692
	Squared returns	0.0001	0.0002	5.9291	56.4770
NASDAQ Composite	Returns	0.0008	0.0082	-0.6492	5.1261
	Absolute returns	0.0062	0.0054	1.7628	8.3363
	Squared returns	0.0001	0.0001	6.2625	62.6264
NIKKEI 225 (stock average) (Tokyo)	Returns	0.0005	0.0092	-0.4976	4.0644
	Absolute returns	0.0069	0.0060	1.3689	5.3110
	Squared returns	0.0001	0.0001	3.7099	21.2146
Singapore SE	Returns	0.0018	0.0094	-1.0343	5.7145
	Absolute returns	0.0072	0.0063	1.8202	7.8889
	Squared returns	0.0001	0.0002	4.8071	32.6776

with financial market efficiency hypothesis. Consequently, the stock index volatility is explored.

4.3 The stock index volatility

The following measures of the volatility are used:

$$v_{i,t} = |r_{i,t}| \quad (5)$$

$$v_{i,t}^2 = r_{i,t}^2, \quad (6)$$

$$v_{i,t}^* = \ln(|r_{i,t}|). \quad (7)$$

The presence of long memory in the series is estimated using confidence intervals based on Robinson's test. The classical asymptotic confidence interval is used. In addition, nonparametric **percentile** and **percentile- t** are also used. The nonparametric bootstraps are based on the empirical distribution of the residuals: they resample the residuals. The number of bootstrap replications is $B = 3999$. (There are also versions of double bootstrap, but the consumption time is too large and are not.) Finally, the confidence interval based on **inverting** (single) nonparametric bootstrap tests is applied. The procedure is described in **subsection B.1**. Again, nonparametric methods are used (see **subsection B.1.2**). (A version of double bootstrap also exists and is described in **subsection B.2**. But the consumption time is too large and are not used here.) The number of bootstrap replications used for the inverted tests is $B = 399$. **Table 6** presents the confidence intervals for the long memory parameter in the stock market volatility series.

Non long memory is found in ASX, FTSE, NASDAQ, and NASDAQ composite. We know from **Monte Carlo experiments** that in the case of non long memory, the methods perform similarly. That can be observed in **Table 6**. Conversely, the presence of long memory feature is found in Dow Jones Industrial 30, Hang Seng, Nikkei 225, and Singapore SE. The value of the long memory parameter d seems to be between 0 and 0.5. It can be observed that the percentile confidence interval is systematically moved to the left with respect to the other methods, suggesting a left bias. It should be noted that the confidence interval based on inverting tests is systematically larger than the other confidence intervals. We know from **Monte Carlo experiments** that the confidence interval based on inverting tests is more reliable than the other methods. Consequently, the results on real data suggest that the uncertainty around the long memory estimate is not accounted for enough by the classical methods. The confidence interval based on inverting tests permit to account for this uncertainty, especially in the case of Singapore SE for which the long memory is large.

4.4 P value functions used to build confidence intervals based on inverting tests

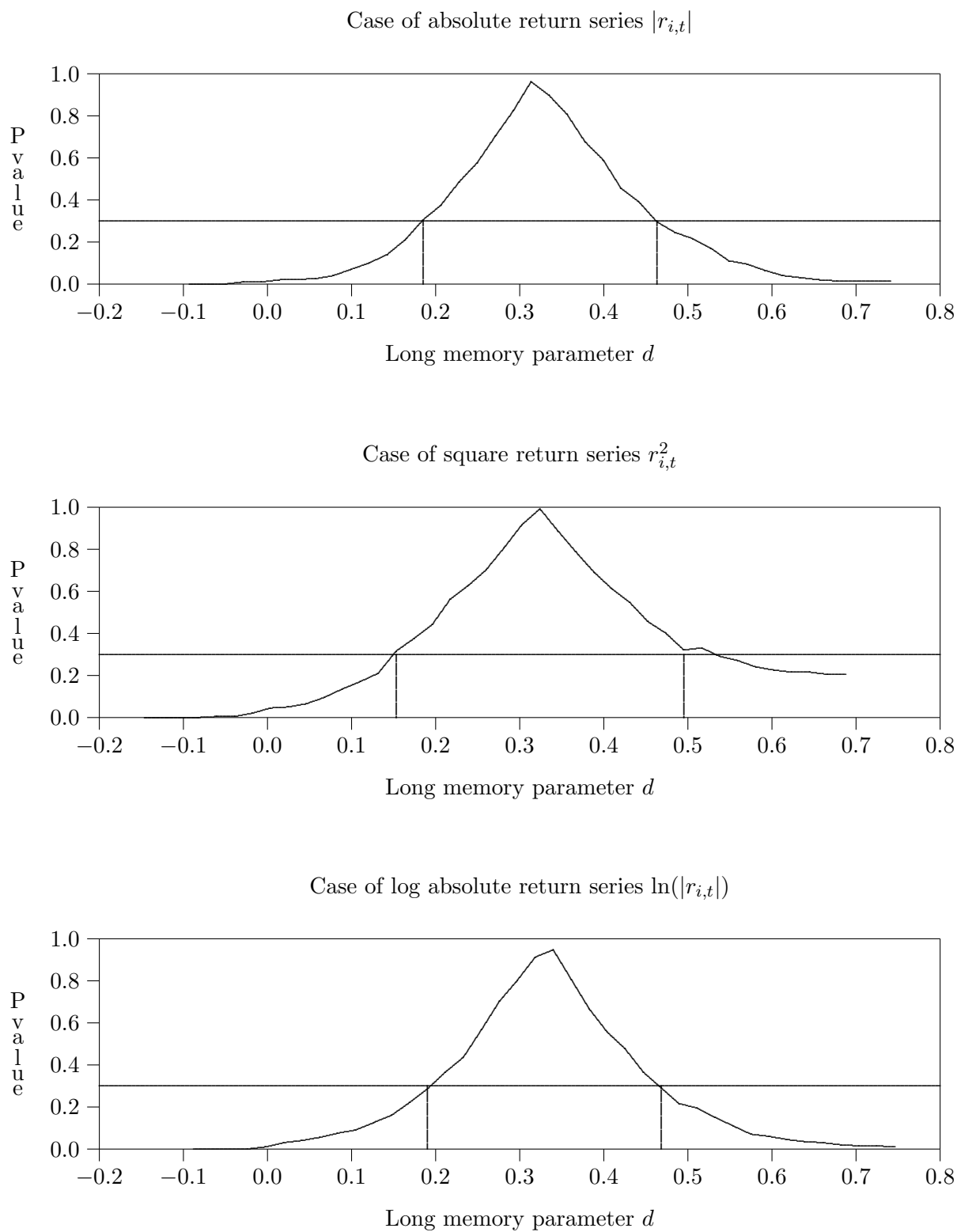
The limits of the confidence interval based on **inverting (single bootstrap) tests** presented in **Table 6** are determined by the values for d such that the test P value is equal to 30%. As an illustration, P value functions with respect to the values for d are presented in **Figure 8**, in the case of *Hang Seng (Hong Kong)*.

Table 6: Long memory 30%-confidence interval on the stock index volatility

Stock index	Confidence interval	$ r_{i,t} $		$r_{i,t}^2$		$\ln(r_{i,t})$	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
ASX	Asymptotic	-0.0032	0.0221	0.0615	0.0868	-0.0905	-0.0652
(Australia)	Percentile	-0.2090	0.0808	-0.1506	0.1422	-0.2962	-0.0124
	Percentile- t	-0.2039	0.0860	-0.1360	0.1568	-0.2853	-0.0015
	Inverting tests	-0.2005	0.0989	-0.1359	0.1635	-0.2956	0.0155
Dow Jones	Asymptotic	0.2765	0.3018	0.1971	0.2224	0.2563	0.2816
Industrial 30	Percentile	0.0462	0.3343	-0.0290	0.2571	-0.0843	0.2079
	Percentile- t	0.1020	0.3901	0.0204	0.3065	0.1880	0.4801
	Inverting tests	0.1005	0.4320	0.0211	0.3419	0.0803	0.4439
FTSE	Asymptotic	-0.0694	-0.0441	-0.1338	-0.1085	0.1366	0.1619
LATIBEX	Percentile	-0.2902	0.0016	-0.3354	-0.0465	-0.0917	0.1938
	Percentile- t	-0.2572	0.0347	-0.3377	-0.0488	-0.0373	0.2481
	Inverting tests	-0.2702	0.0695	-0.4167	-0.0500	-0.0394	0.2814
Hang Seng	Asymptotic	0.3826	0.4080	0.3291	0.3544	0.3877	0.4130
(Hong Kong)	Percentile	0.1224	0.4045	0.0865	0.3752	0.1263	0.4073
	Percentile- t	0.2441	0.5262	0.1664	0.4551	0.2514	0.5324
	Inverting tests	0.1853	0.4633	0.1532	0.4953	0.1903	0.4684
NASDAQ	Asymptotic	-0.0819	-0.0566	-0.0230	0.0023	-0.0629	-0.0376
	Percentile	-0.2916	-0.0151	-0.2390	0.0444	-0.2741	0.0190
	Percentile- t	-0.2654	0.0111	-0.2071	0.0763	-0.2616	0.0316
	Inverting tests	-0.2934	0.0212	-0.2148	0.0816	-0.2684	0.0338
NASDAQ	Asymptotic	0.0745	0.0998	0.0731	0.0984	-0.0210	0.0043
Composite	Percentile	-0.1409	0.1519	-0.1544	0.1382	-0.2325	0.0540
	Percentile- t	-0.1196	0.1731	-0.1087	0.1839	-0.2126	0.0739
	Inverting tests	-0.1229	0.1765	-0.1670	0.1751	-0.1929	0.1043
NIKKEI 225	Asymptotic	0.2338	0.2591	0.3111	0.3364	0.0717	0.0970
(stock average)	Percentile	0.0102	0.2958	0.0593	0.3586	-0.1468	0.1365
(Tokyo)	Percentile- t	0.0552	0.3407	0.1470	0.4463	-0.1098	0.1736
	Inverting tests	0.0365	0.3786	0.1352	0.5415	-0.1256	0.1738
Singapore	Asymptotic	0.4328	0.4581	0.4233	0.4486	0.1424	0.1677
SE	Percentile	0.1272	0.4141	0.1300	0.4219	-0.1549	0.1368
	Percentile- t	0.3347	0.6217	0.3080	0.5999	0.0313	0.3230
	Inverting tests	0.2996	0.9705	0.2687	(*)1.000	-0.0550	0.2230

(*) The tests retain the null hypothesis for every value of d up to $d = 1$. Robinson test is not valid for value greater than 1.

Figure 8: P value functions of bootstrap tests
in the case of *Hang Seng (Hong Kong)*



5 Conclusion

In the case of long memory processes, the statistics are far from pivotal, and classical confidence intervals suffer from large distribution distortions leading to weak performance, even when bootstrap techniques are used. In practice, bootstrap intervals (**percentile** and **percentile- t**) do not perform better than the asymptotic confidence interval.

To solve this problem, we provide a confidence interval based on **inverting bootstrap tests**. We used the bootstrap test proposed by **de Peretti [2003]**, that uses a bilateral p value accounting for asymmetric distributions, and thus performing better than classical tests since the long memory estimator distribution is asymmetric.

Monte Carlo experiments are carried out for obtaining information about the properties of confidence interval procedures in finite samples. However, the rich detail in the results they provide can be difficult to apprehend if they are presented in the usual tabular form. In this paper, we use graphical techniques provided by **de Peretti [2004]**, namely, **coverage plots**, **coverage discrepancy plots** (which may optionally be smoothed), and **coverage effectiveness** curves. These techniques can make the principal results of an experiment immediately obvious. All of these techniques are based on the construction of an estimated cumulative distribution function of the (true) coverage associated with some confidence regions and on effectiveness criteria that were discussed, without loss of computing time compared to classical tabular presentation.

These techniques are used for presenting the results of a number of experiments concerning long memory confidence intervals. We think that results, which are entirely presented in graphical form, are of interest and provide more information in a more easily assimilable fashion than a tabular presentation or *QQ plots* could possibly have done. The results show that **percentile** and **percentile- t** methods does not perform correctly at all, conversely to confidence interval based on **inverting** bootstrap tests that works quasi-perfectly compared to the confidence intervals obtained with the previous methods: the coverage is much closer to the confidence level.

Finally, the confidence intervals are applied to various measure of stock market volatility series. Half of the series do not exhibit long memory feature in their volatility. The other series exhibit long memory with long memory parameters greater than 0 and smaller than 0.5, suggesting stationary volatility series, except in the case of Singapore SE, which exhibit strong long memory.

Appendix

A Graphical methods

We use the Graphical Methods for Investigating the Finite-sample Properties of Confidence Regions proposed by [de Peretti \[2004\]](#). These methods are presented below.

A.1 The position of the problem

Let θ be the parameter vector of interest:

$$\theta \in \Theta \subset \mathbb{R}^k.$$

Let τ be the statistic used for constructing the confidence region:

$$\tau \equiv \tau_X(\theta),$$

where X is the observed finite sample. τ can also be a vector. The cumulative distribution function (CDF) of τ is denoted F_θ . The distribution of τ has to depend on θ for being able to estimate it.

Let R be a confidence region for θ with confidence level ⁶ $1 - \alpha$:

$$R \equiv R(\tau, \{F_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha).$$

However, there is an infinity of possibilities giving such a confidence region. Each of the possibilities corresponds to a different method. When the notation R is used in the following, we mean one of these methods.

If θ_0 is the true value of the parameter vector θ that generates the random sample X , then

$$\forall \theta_0 \in \Theta, \forall \alpha \in [0, 1], P(\theta_0 \in R(\tau(\theta_0), \{F_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)) = 1 - \alpha.$$

The graph of $\{(1 - \alpha, P(\theta_0 \in R)); \alpha \in [0, 1]\}$ is equal to the 45 degrees line. However, the family $\{F_\theta\}_{\theta \in \Theta}$ is not known in general.

Therefore, the family $\{F_\theta\}_{\theta \in \Theta}$ has to be estimated and we denote its estimation by $\{\hat{F}_\theta\}_{\theta \in \Theta}$. \hat{F}_θ can be the asymptotic limit of $\{F_\theta\}_{\theta \in \Theta}$ as $T \rightarrow \infty$, or it can be a distribution derived by bootstrapping, or it can also be some other approximations of F_θ (coming from a first order Taylor expansion, for instance). Let us denote

$$\hat{R} \equiv R(\tau(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha).$$

The difference between R and \hat{R} can also come from an approximation in the analytical calculus of R . Since \hat{F}_θ is not exact, we have in general

$$P(\theta_0 \in \hat{R}) \neq 1 - \alpha,$$

⁶The confidence level for R is the probability of observing the true value of the parameter vector in the random region R , according some distribution F . See [Davidson and MacKinnon \[1993\]](#), chapter 5, for more explanations.

but we wish \hat{R} such that

$$P(\theta_0 \in \hat{R}) \approx 1 - \alpha,$$

i.e. that the graph of

$$\left\{ (1 - \alpha, P(\theta_0 \in \hat{R})); \alpha \in [0, 1] \right\}$$

is near the 45 degrees line. $P(\theta_0 \in \hat{R})$ is the *coverage probability*, or just the *coverage*, of the random region \hat{R} . It is the true probability that the region will include, or cover, the true value of the parameter vector.

A.2 The Monte Carlo procedure

Consider a Monte Carlo experiment in which S realisations of the interest statistic $\tau(\theta)$ are generated using a data generating process (DGP) that is a special case of the model. We may denote these simulated values by $\tau_s(\theta)$, $s \in \{1, \dots, S\}$.

$P(\theta_0 \in R(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha))$ can be computed using the Monte Carlo experiment:

$$\hat{P}(\theta_0 \in R(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)) = \frac{1}{S} \sum_{s=1}^S \mathbb{I}(\theta_0 \in R(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha))$$

for S very large. $\mathbb{I}(\cdot)$ denoted an indicator function that takes the value 1 if its argument is true and 0 otherwise.

Set $1 - \alpha_s$ the value of $1 - \alpha$ such that

$$\theta_0 \in \partial R(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)$$

where ∂ represents the border of a set of values. $1 - \alpha_s$ can be called *critical coverage*. $1 - \alpha_s$ can be multiple, but it is easy to assume that it is unique by assuming the natural hypothesis that R (and \hat{R}) is increasing with respect to $1 - \alpha$ in the sense of inclusion. This hypothesis is obtained if R is optimised using the maximum likelihood principle, for instance, but not only.

It should be noted that

$$(1 - \alpha_s \leq 1 - \alpha) \iff (\theta_0 \in \hat{R}(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)),$$

thus,

$$\hat{P}(\theta_0 \in R(\tau_s(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)) = \frac{1}{S} \sum_{s=1}^S \mathbb{I}(1 - \alpha_s \in [0, 1 - \alpha]).$$

In practice, we just count the proportion of $1 - \alpha_s$ smaller or equal to $1 - \alpha$. In fact, it is the empirical cumulative distribution function of the random variable:

$$1 - \alpha_s \equiv 1 - \alpha_s(\tau_s, \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, R, \theta_0).$$

All the graphs we discuss are based on the empirical cumulative distribution function of the confidence level $1 - \alpha_s$ of $R(\tau_s, \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, \cdot)$ if the true value θ_0 of the parameter vector is just over ∂R (the border of R).

Let F_0 denotes the finite-sample distribution of $1 - \alpha_s$ such that the coverage probability is equal to the confidence level, *i.e.*

$$F_0(1 - \alpha) = P(\theta_0 \in R(\tau(\theta_0), \{F_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)) = 1 - \alpha.$$

Let F denote the generally unknown (true) finite-sample distribution of $1 - \alpha_s$, *i.e.*

$$\begin{aligned} F(1 - \alpha) &= P(1 - \alpha_s \leq 1 - \alpha), \\ &= P(\theta_0 \in R(\tau(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)), \end{aligned}$$

and \hat{F} , its Monte Carlo computation:

$$\begin{aligned} \hat{F}(1 - \alpha) &= \hat{P}(\theta_0 \in R(\tau(\theta_0), \{\hat{F}_\theta^{-1}\}_{\theta \in \Theta}, 1 - \alpha)), \\ &= \frac{1}{S} \sum_{s=1}^S \mathbb{I}(1 - \alpha_s \leq 1 - \alpha). \end{aligned}$$

At any point x_i in the $(0, 1)$ interval, it is defined by

$$\hat{F}(x_i) = \frac{1}{S} \sum_{s=1}^S \mathbb{I}(1 - \alpha_s \leq x_i).$$

See appendix A for the choice of the x_i 's.

A.3 Coverage Plots

The simplest graph that we will discuss is a plot of $\hat{F}(x_i)$ against x_i . We shall refer to such a graph as a *coverage plot* since it presents the (true) coverage probability against the (nominal) confidence level. If F is correct, *i.e.* $F = F_0$, each of the $1 - \alpha_s$ is distributed as an independent uniform random variable $U(0, 1)$. Therefore, when $\hat{F}(x_i)$ is plotted against x_i , the resulting graph should be close to the 45 degrees line.

A.4 Coverage-Effectiveness Curves

Coverage plots and Coverage discrepancy plots are very useful for dealing with coverage probability, but they are not useful at all for dealing with confidence region **effectiveness**. We will discuss graphical methods for comparing the effectiveness of competing regions using *coverage-effectiveness curves*. For an experiment (in which a given DGP is used), these curves can be constructed using the empirical cumulative distribution function of the critical coverage and the chosen effectiveness criterion. **Effectiveness criteria** will be discussed in the next subsection.

It is often desirable to compare the effectiveness of alternative confidence regions, but this can be difficult to do if all the regions do not have the correct coverage probability. If the values of an **effectiveness criterion** are plotted against (nominal) confidence level, the result will not be very useful, since claiming that a method is more satisfactory than another one on the basis of an **effectiveness criterion** has no sense if the methods suffer from different coverage distortions: for example, a criterion providing good results can be spurious due to a default of coverage. Unfortunately, this is what is often implicitly done when region effectiveness is reported in a table.

In order to plot effectiveness against the (true) coverage probability, an experiment has to be performed, preferably using the same sequence of random numbers for each region, to avoid experimental errors. Let the points on the approximate empirical cumulative distribution function be denoted $\hat{F}(x)$, and let the estimated effectiveness for a confidence level of $x = 1 - \alpha$ be denoted $\hat{E}(x)$. They have to be evaluated at a pre-chosen set of points $\{x_i\}_{i=1,\dots,N}$. As before, $F(x)$ is the probability of getting a critical coverage less than x . Similarly, $E(x)$ is the effectiveness for a confidence level of x . Tracing the locus of points $(F(x), E(x))$ as x varies from 0 to 1 thus generates a coverage-effectiveness curve on a correct coverage-adjusted basis. Plotting the points $(\hat{F}(x_i), \hat{E}(x_i))$, does exactly the same thing, except for experimental error due to the randomness of the Monte Carlo simulations (however, the experimental error converges to zero when the number of Monte Carlo replications goes to infinite). More precisely, it presents the effectiveness criterion against the coverage probability, *i.e.* the true confidence level; and then, the various methods can be compared.

Similarly to the coverage, calculating $E(x_i)$ for a set $\{x_i\}_i$ is done from the same simulated series, and consequently, it is not necessary to run an additional experiment for each confidence level x_i , and thus, there is no loss of computing time. Moreover, in practice, the calculation for a set $\{x_i\}_i$ is often straightforward by matrix algebra: for instance, for the asymptotic, percentile and percentile-t confidence intervals presented in [section 2](#), the computation is written with only two lines in the Gauss program, and it takes an almost nil computing time compared to the one for the bootstrap loop.

There is one practical problem with drawing coverage-effectiveness curves by plotting $\hat{E}(x_i)$ against $\hat{F}(x_i)$. For regions that under- or over-cover severely, there may be a region of the coverage-effectiveness curve that is left out by a choice of values of x_i . For solving this problem, a very large number of Monte Carlo replications should have to be chosen, but it is not necessarily possible in practice because of the computing time. Nevertheless, if a region under- or over-covers severely, it cannot be chosen for practical uses, and thus, it is not useful to compute its “true” effectiveness by coverage-effectiveness curves.

A.5 How to choose the effectiveness criterion?

The effectiveness criterion depends on the mathematical problematics under consideration, but also on the economic problematics. In this paper, we choose the expected length of the confidence interval as the primary criterion, and the standard deviation of the length of the confidence interval as a secondary criterion.

A.6 How to chose the “best” method ?

The coverage plots and the effectiveness curves are very useful for choosing among methods that have reasonable coverage distortions: they permit to make *arbitrage* between the coverage distortion and the true effectiveness for each methods and then to chose the most appropriate. There is no criterion that permits to select the “best” method from the combination of both the whole coverage and the whole effectiveness curves. However, the following rule can be used:

1. First, select the methods that have not their coverage plots too far from the 45° degree line. The coverage is the most important feature, since if the coverage error is too large, the method is false.

2. Among the methods selected in the first step, select the methods that have the best effectiveness curves. A large gain in the effectiveness can compensate the a small loss in the coverage. This largeness depends on the situation and on the decision-maker.

A.7 Details about the graphical representation

When S is large, storage space can be conserved by evaluating the EDF (1) only at N points $x_i, i = 1, \dots, N$, which should be chosen in advance so as to provide a reasonable snapshot of the $(0, 1)$ interval, or of that part of it which is of interest. As parsimonious way to choose the x_i is

$$x_i = 0.001, 0.002, \dots, 0.009, 0.01, 0.02, \dots, 0.09, 0.1, 0.15, \dots \\ \dots, 0.85, 0.9, 0.91 \dots, 0.98, 0.99, 0.991, \dots, 0.998, 0.999 \quad (N = 53) \quad (8)$$

There are extra points near 0 and 1 in order to ensure that we do not miss any unusual behaviour in the tails.

B Details on the LRD confidence intervals

B.1 Bilateral (single) bootstrap tests for long memory

For more details in the case of the null of no long memory, see [de Peretti \[2003\]](#). While the asymptotic tests are asymptotically valid, the tests based on the asymptotic distributions are not exact in finite samples, and so, it is natural to “bootstrap” them. For the conception of the bootstrap see [Efron \[1979\]](#), for its development, see [Davidson and MacKinnon \[1993\]](#), and for further analysis, see [Davidson and MacKinnon \[1996b,a\]](#).

B.1.1 The bootstrap procedure

The procedure is as follows:

1. Compute the test statistic (Hurst, Lo, Robinson, Higuchi, Jensen, or others), which will be denoted $\hat{\tau}$.
2. Estimate the ARFIMA(p,d,q) model by maximum likelihood under the null $H_0: d = d_0$, for obtaining the model parameters $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_\varepsilon^2)$ and the residuals $\hat{\varepsilon}$. For our simulations, $p = 1$ and $q = 1$ are chosen, but for estimating real data, we strongly advice users to determine p and q by information criterions or other efficient methods because an error in the choice of p and/or q generally yields to large bias in the estimation and the inference of d (see [Hauser et al. \[1999\]](#)).
3. Draw B sets of bootstrap error terms, ε^b , and use them to generate B bootstrap samples x^b . There are numerous ways to draw the error terms, four of which are described below. The elements of x^b should be generated from the equation

$$x_t = (1 - L)^{-d_0} \hat{\phi}(L)^{-1} \hat{\theta}(L) \varepsilon_t^b \quad t \in \{1, \dots, T\}. \quad (9)$$

In practice, we do as the GAUSS procedure for generating ARFIMA series: we calculate the MA(q) filter corresponding to $(1 - L)^{-d_0} \hat{\phi}(L)^{-1} \hat{\theta}(L)$ with a very large order q (T for example), and we apply it to ε^b for obtaining x_t . For that, of course, we have to generate ε^b with a larger sample size ($2 \times T$ for example).

4. For each bootstrap sample, compute the statistic (Hurst, Lo, Robinson, Higuchi, Jensen, or others), denoted τ^b , with x^b instead of x .
5. Then, compute the estimated bootstrap P value: see equation 11 or equations 12–13.

B.1.2 The error terms generating procedure

I examine four ways for generating the ε_t^b (see Davidson [1998]):

1. The parametric bootstrap, called b_0 : the ε_t^b are independent draws from the $N(0, \hat{\sigma}_\varepsilon^2)$ distribution.
2. The simplest nonparametric bootstrap, called b_1 : the ε_t^b are obtained by re-sampling with replacement from the vector of $\{\hat{\varepsilon}_t\}_{t=\hat{p}+1}^T$.
3. A slightly more complicated form of nonparametric bootstrap called b_2 : the ε^b are generated by re-sampling with replacement from the vector

$$\left\{ \sqrt{\frac{T}{T-2\hat{p}-1}} \left(\hat{\varepsilon}_t - \frac{1}{T-\hat{p}} \sum_{i=\hat{p}+1}^T \hat{\varepsilon}_i \right) \right\}_{t=\hat{p}+1}^T. \quad (10)$$

4. The most complicated nonparametric bootstrap, called b_3 : the ε^b are generated by re-sampling from the vector with typical element $\tilde{\varepsilon}_t$ constructed as follows:
 - let d_t be the t^{th} diagonal element of $P_{[(1-L)^{-d_0} \hat{\phi}(L)^{-1} \hat{\theta}(L)]}$, the matrix projecting onto the space spanned by $(1-L)^{-d_0} \hat{\phi}(L)^{-1} \hat{\theta}(L)$;
 - divide each element of $\hat{\varepsilon}$ by $\sqrt{1-d_t}$;
 - re-centre the resulting vector;
 - re-scale it so that it has variance $\hat{\sigma}_\varepsilon^2$.

This type of procedure is advocated by Weber [1984].

B.1.3 The choice of the bootstrap P value

By making a large number of drawings of bootstrap statistics τ^b , a bootstrap P value can be computed by the following formula:

$$\hat{p}_{bil}(\hat{\tau}^2) = \frac{1}{B} \sum_{b=1}^B I((\tau^b)^2 > \hat{\tau}^2), \quad (11)$$

see Davidson and MacKinnon [1993]. This formula corresponds to an unilateral test. This sort of formulae is often associated with symmetric bilateral tests. However, the size distortion is not necessarily symmetric. Thus, I prefer to use the following formula:

$$\hat{p}(\hat{\tau}) = 2 \min\{\hat{p}_{uni}(\hat{\tau}), 1 - \hat{p}_{uni}(\hat{\tau})\}, \quad (12)$$

where

$$\hat{p}_{uni}(\hat{\tau}) = \frac{1}{B} \sum_{b=1}^B I(\tau^b > \hat{\tau}), \quad (13)$$

that corresponds to a bilateral (asymmetric) test. Further considerations about this P value can be found in Chapter 5 of [Davidson and MacKinnon \[1993\]](#) dealing with confidence intervals.

B.2 Bilateral double bootstrap tests for long memory

The standard error of \hat{d} , say $\hat{\sigma}(\hat{d})$, used for studentising \hat{d} can be computed in two different ways: the first one is the asymptotic estimate (see subsection 2.2), the second one is the bootstrap estimate. For obtaining the bootstrap estimator of $\sigma(\hat{d})$, the time series are estimated and replicated B_2 times in the same way as previously: for each replicated series, \hat{d} is computed leading to a set of $\left\{ \hat{d}^b \right\}_b$ from which the standard error is computed. It should be noted that when this test is used, and thus replicated series are generated, the bootstrap estimator of $\sigma(\hat{d})$ must be applied on each replicated series for obtaining the test statistics leading to replications of replicated series. This method is often called *double bootstrap*.

C Bootstrapped ADF tests

The following variables have to be defined:

- B , the number of bootstrap replications,
- p , the number of augmentations in the ADF regressions.

B has to be chosen as large as possible, depending on the characteristics of the computer. The choice of p is more difficult. We recall that the ADF regressions are:

$$\begin{aligned} \Delta y_t &= \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p} + e_t, \\ \Delta y_t &= \text{constant} + \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p} + e_t, \\ \Delta y_t &= \text{constant} + \text{trend} + \alpha y_{t-1} + \beta_1 \Delta y_{t-1} + \dots + \beta_p \Delta y_{t-p} + e_t, \end{aligned}$$

where y_t is the time series, and e_t are the error terms, t goes from 1 to T . We propose the following procedure for choosing p : The procedure starts at $p = 0$. The residuals of each ADF regressions with p augmentations are tested for independence using both the Ljung-Box's and Box-Pierce's tests. The number of autocorrelation coefficients taken into account for the Ljung-Box's and Box-Pierce's tests go from 1 to 8. If the residuals are not independent, p is incremented by 1 until the residuals look independent.

The steps of the bootstrapped ADF test are the following:

1. The Student test statistics for α for each ADF regression are computed. Let the statistics be denoted t_α . At this step, the residuals can be kept to be tested for independence.
2. The bootstrap procedure needs a DGP for generating simulated samples under the null. This DGP is determined by estimating the model under the null using the data and the OLS procedure.
3. The bootstrap loop starts now. The simulated error terms, denoted e_t^b , are generated for a sample. There are four ways for generating the simulated error terms:

- (a) Parametric bootstrap: The simulated error terms are drawn from the normal distribution

$$e_t^b \sim N(0, s^2),$$

where s is the standard error of the error terms estimated from the ADF regression using the data.

- (b) Basic nonparametric bootstrap: The simulated error terms are drawn by ...
(c) Nonparametric bootstrap with corrected degree of freedom: since $E(\hat{e}_t^2) \neq E(e_t^2)$ where \hat{e}_t^2 are the residuals of the ADF regression, but $E(\hat{e}_t^2) =$

For our program, the parametric and the second nonparametric bootstrap are chosen.

4. The simulated time series under the null, denoted $(y_t^b)_t$, is generated recursively using both the following steps:

- (a) first, define Δy_t^b recursively:

$$\begin{aligned} \Delta y_t^b &= \hat{\beta}_1 \Delta y_{t-1}^b + \dots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \\ \Delta y_t^b &= \widehat{constant} + \hat{\beta}_1 \Delta y_{t-1}^b + \dots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \\ \Delta y_t^b &= \widehat{constant} + \widehat{trend} + \hat{\beta}_1 \Delta y_{t-1}^b + \dots + \hat{\beta}_p \Delta y_{t-p}^b + e_t^b, \end{aligned}$$

The p first values for y_t^b can be chosen equal to the p first values of y_t , interpreted as initial conditions. (Another way is to choose them randomly.)

- (b) second, compute y_t^b :

$$y_t^b = y_1 + \sum_{i=2}^t y_i^b.$$

y_1 is an initial condition.

5. The Student test statistics for α for each ADF regression are computed using the simulated series $(y_t^b)_t$. Let the statistics be denoted t_α^b .
6. The steps 3–5 are done again B times. A set of statistics t_α^b , $b = 1, \dots, B$, is then obtained for each the three ADF regressions, and for each both the parametric and nonparametric bootstraps (thus there are six statistics).
7. The bootstrap P value is finally computed depending on the test hypothesis:
(a) If the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha < 0$, the classical P value is

$$p_{\text{uni}} = \frac{1}{B} \sum_{b=1}^B I(t_\alpha^b \leq t_\alpha),$$

where I is the indicator function. This P value corresponds to an unilateral test.

- (b) If the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha \neq 0$, the classical bootstrap P value is

$$p_{\text{bil}_{\text{sym}}} = \frac{1}{B} \sum_{b=1}^B I(|t_{\alpha}^b| \geq |t_{\alpha}|).$$

This P value corresponds to a bilateral test.

- (c) In the case where the null hypothesis $H_0 : \alpha = 0$ is tested against the alternative hypothesis $H_1 : \alpha \neq 0$, we also propose the following bootstrap P value:

$$p_{\text{bil}_{\text{asym}}} = 2 \min\{p_{\text{uni}}, 1 - p_{\text{uni}}\}.$$

This P value also corresponds to a bilateral test, but it takes into account the asymmetry of the statistic distribution in addition. This P value can be found in [Davidson and MacKinnon \[1993, chapter 5\]](#), in the context of confidence regions.

In our program, the two last P values are used.

8. Finally, a significance level is chosen and compared to the P values. If a P value is lower to the significance level, H_1 is retained, otherwise H_0 is retained.

D Bootstrapped tests for normality

It is first assumed that the processes that will be tested are stationary and independent.

D.1 The test statistics

First, the estimated centred moments are presented:

$$\begin{aligned} \hat{\mu}_1(y) &= \frac{1}{T} \sum_{t=1}^T y_t, \\ \hat{\mu}_2^c(y) &= \frac{1}{T} \sum_{t=1}^T (y_t - \mu_1(y))^2, \\ \hat{\mu}_3^c(y) &= \frac{1}{T} \sum_{t=1}^T (y_t - \mu_1(y))^3, \\ \hat{\mu}_4^c(y) &= \frac{1}{T} \sum_{t=1}^T (y_t - \mu_1(y))^4, \end{aligned}$$

where T is the sample size, and y is the time series. The estimated skewness and estimated the kurtosis are defined as following:

$$\begin{aligned} \widehat{\text{sk}}(y) &= \frac{\hat{\mu}_3^c(y)}{(\hat{\mu}_2^c(y))^{3/2}}, \\ \widehat{\text{ku}}(y) &= \frac{\hat{\mu}_4^c(y)}{(\hat{\mu}_2^c(y))^2}. \end{aligned}$$

From the skewness and the kurtosis, three statistics can be built:

- a statistic based on the skewness:

$$t_{\text{sk}}(y) = \frac{\widehat{\text{ku}}(y)}{\sqrt{(6/T)}},$$

- a statistic based on the kurtosis:

$$t_{\text{ku}}(y) = \frac{\widehat{\text{ku}}(y) - 3}{\sqrt{(24/T)}},$$

- and a statistic based on both the the skewness and the kurtosis: the Jarque-Bera statistic (see [Bera and Jarque \[1980, 1981\]](#)):

$$t_{\text{jb}}(y) = (t_{\text{sk}}(y))^2 + (t_{\text{ku}}(y))^2.$$

D.2 The asymptotic tests

Under the null hypothesis H_0 : “ y is a Gaussian process”, the asymptotic theory establishes that:

$$t_{\text{sk}}(y) \sim N(0, 1), \tag{14}$$

$$t_{\text{ku}}(y) \sim N(0, 1), \tag{15}$$

$$t_{\text{jb}}(y) \sim \chi^2(2). \tag{16}$$

However, in finite sample, the two first statistics do not follow the normal distribution, even if the time series is Gaussian. Consequently, the third statistic do not follow the $\chi^2(2)$ distribution. This is a first reason for bootstrapping them. In addition, Jarque-Bera statistic follows asymptotically a $\chi^2(2)$ because of asymptotic independence of $t_{\text{sk}}(y)$ and $t_{\text{ku}}(y)$. However, in finite sample, these two components are dependent, leading to a second error in Jarque-Bera test. This is a second reason for bootstrapping Jarque-Bera statistic, since the bootstrap techniques are naturally able to take into account this dependence.

D.3 The bootstrap tests

Let B denotes the number of bootstrap replications. B has to be chosen as large as possible, depending on the characteristics of the computer. y_t is the time series, t goes from 1 to T . The steps of the bootstrapped test for normality are the following:

1. The test statistics defined in equations 14–16 are computed. Let $\tau(y)$ denote the chosen statistic.
2. The bootstrap procedure needs a DGP for generating simulated samples under the null that is simply $H_0 : y_t \sim \text{i.i.d.}N(\mu_1, (\mu_2^c)^2)$. This DGP is determined by estimating the model under the null using the data, *i.e.* by estimating μ_1 and μ_2^c . This estimation can be made simply by $\hat{\mu}_1(y)$ and $\hat{\mu}_2^c(y)$. Of course, it is a parametric bootstrap since the null hypothesis is the normality of the process.

3. The bootstrap loop starts now. The simulated time series under the null, denoted $(y_t^b)_t$, is generated by independently drawing from the normal distribution:

$$y_t^b \sim N(\hat{\mu}_1(y), (\hat{\mu}_2^c(y))^2),$$

It is necessarily a parametric bootstrap.

4. The three test statistics are computed using the simulated series $(y_t^b)_t$.
5. The steps 3–4 are done again B times. A set of bootstrap statistics is then obtained for each original statistic.
6. The bootstrap P value is finally computed: the classical bootstrap P value is

$$p_{\text{sym}} = \frac{1}{B} \sum_{b=1}^B I(|\tau(y^b)| \geq |\tau(y)|).$$

We also propose the following bootstrap P value:

$$p_{\text{asym}} = 2 \min\{p_{\text{uni}}, 1 - p_{\text{uni}}\},$$

where

$$p_{\text{uni}} = \frac{1}{B} \sum_{b=1}^B I(|\tau(y^b)| \leq |\tau(y)|).$$

This P value takes into account the asymmetry of the statistic distribution in addition ⁷ This P value can be found in [Davidson and MacKinnon \[1993, Chapter 5\]](#), in the context of confidence regions. This P value can be used only for $t_{\text{sk}}(y)$ and $t_{\text{ku}}(y)$, but not for $t_{\text{jb}}(y)$ because p_{uni} cannot be compute for $t_{\text{jb}}(y)$ since there are two underlying statistics. An extension of the asymmetric P value could be done for two dimensions, but in practice, it is very difficult.

7. Finally, a significance level is chosen and compared to the P values. If a P value is lower to the significance level, H_1 is retained, otherwise H_0 is retained.

D.4 Extension to linear dependence

We propose here an extension of the bootstrap tests to linearly dependent processes. For the asymptotic tests, the time series has to be independent, since the skewness and the kurtosis depend on the dependence parameters. However, if the linear dependence is specified in the bootstrap procedure, the bootstrap tests can apply easily to linearly dependent processes, if they are stationary. If the linear dependence is not taken into account, the autocorrelation amplifies the distortion of the test statistic distribution compared to the asymptotic distribution. Two steps change: the DGP estimation under the null, and the simulated series generation. We propose the following procedure:

- 2 The null is now $y_t \sim \text{Gaussian } AR(p)$. This DGP is determined by estimating the model under the null using the data and the OLS procedure. p is determined by the AIC and SIC criteria. Let the estimated DGP be denoted $\widehat{AR}(\hat{p})$. \hat{p} can be chosen by criteria as AIC and SIC, and also by Ljung-Box and Box-Pierce tests (using 1 to 8 lags for example) until the residuals seem independent.

⁷This bilateral bootstrap P value is not useful for the asymptotic tests, since the asymptotic distributions are symmetric: both the P values are the same, or unilateral.

- 3a The simulated error terms, denoted e_t^b , are generated for a sample. The simulated error terms are drawn from the normal distribution

$$e_t^b \sim N(0, s^2),$$

where s is the standard error of the error terms estimated from the AR(p) regression using the data.

- 3b The simulated time series under the null, denoted $(y_t^b)_t$, is generated recursively using both the Gaussian $\widehat{AR}(\hat{p})$ model.

D.5 Monte Carlo results

The Monte Carlo results, that are not presented here, suggest that the bootstrap tests are much more reliable than the asymptotic tests. The bootstrap tests are not perfect, and suffer from size distortion, but only for very small sample size.

References

- Anil K Bera and Carlos M. Jarque. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6 (3):255–259, 1980. [4.2](#), [D.1](#)
- Anil K Bera and Carlos M. Jarque. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte carlo evidence. *Economics Letters*, 7 (4):313–318, 1981. [4.2](#), [D.1](#)
- J. Beran. *Statistics for long-memory processes*. Chapman and Hall, 1994. London. [1](#)
- R. Beran. Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83(403):687–697, 1988. [2](#)
- Berkowitz and Kilian. Recent developments on bootstrapping time series. *Econometrics Reviews*, 19:49–54, 2000. [2.4](#)
- R. Davidson. Notes on the bootstrap. *GREQAM*, 1998. [2](#), [B.1.2](#)
- R. Davidson. Comments on ‘recent developments in bootstrapping time series’ by berkowitz and kilian. *Econometrics Reviews*, 19:49–54, 2000. [2.4](#)
- R. Davidson and J. MacKinnon. Graphical methods for investigating the size and the power of hypothesis tests. *The Manchester School*, 66:1–22, 1998. [1](#)
- R. Davidson and J. G. MacKinnon. *Estimation and inference in economics*. Oxford University Press, 1993. New York. [2.3](#), [2](#), [2.5](#), [6](#), [B.1](#), [B.1.3](#), [B.1.3](#), [7c](#), [7](#)
- R. Davidson and J. G. MacKinnon. The size distortion of bootstrap tests. *GREQAM Working Paper No 96A15 and Queen’s Institute for Economic Research Discussion Paper No 937*, 1996a. [2](#), [B.1](#)
- R. Davidson and J. G. MacKinnon. the power of bootstrap tests. *Queen’s University Institute for Economic Research, Discussion Paper 937*, 1996b. [2](#), [B.1](#)
- R. Davidson and J. G. MacKinnon. Improving the reliability of bootstrap confidence intervals. *Université de Montréal, Conference Resampling Methods in Econometrics*, 2001. [1](#), [2.5](#)
- C. de Peretti. Bilateral bootstrap tests for long memory: An application to the silver market. *Computational Economics*, page Forthcoming, 2003. [4](#), [5](#), [5](#), [B.1](#)
- C. de Peretti. Graphical methods for investigating the finite-sample properties of confidence regions: A gap in the literature? a new proposal. *Cass Business School working paper*, 2004. [1](#), [5](#), [A](#)
- C. de Peretti and V. Marimoutou. Are the long memory tests really effective? *GREQAM Working Paper*, No 02A14, 2002. [1](#), [2](#), [1](#)
- B. Efron. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979. [B.1](#)

- B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57, Chapman and Hall, 1993. London. 2.3
- C.W.J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional integration. *Journal of Time Series Analysis*, 1(1):15–29, 1980. 1, 1
- P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, 1992. New York. 2.3
- M. A. Hauser, B. M. Pötscher, and Reschenhofer E. Measuring persistence in aggregate output: Arma models, fractionally integrated arma models and nonparametric procedures. *Empirical Economics*, 24:243–269, 1999. 2
- T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica*, D 31:277–283, 1988. 2
- J.S.U. Hjorth. *Computer Intensive Statistical Methods*. Chapman and Hall, 1994. London. 2.3
- J. L. Horowitz. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, 61(2):395–411, 1994. 2
- J.R.M. Hosking. Fractional differencing. *Biometrika*, 68:165–176, 1981. 1, 1
- H. Li and G. S. Maddala. Bootstrapping time series models. *Econometric Reviews*, 15: 297–318, 1996. 2.4
- B. Mandelbrot. When can price be arbitrated efficiently? a limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, 53:225–236, 1971. 1
- B. Mandelbrot and J. R. Wallis. Some long-run properties of geophysical records. *Water Resource Research*, 5:321–330, 1969. 1
- P.M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23:1048–1072, 1995. 1, 2, 2, 2.1, 1, 2.2, 3, 3.1.1
- J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer-Verlag, 1995. New York. 2.3
- C. Siani and J. P. Moatti. The handling of uncertainty in economic evaluations of health care strategies. *Revue d’Epidémiologie et de Santé Publique (Frensh)*, 51:255–276, 2003. 2.4
- Carlos Velasco. Non-stationary log-periodogram regression. *Journal of Econometrics*, 91: 325–371, 1999. 2.2, 2.3
- N.C. Weber. On resampling techniques for regression models. *Statistics and Probability Letters*, 2:275–278, 1984. 4